



**Universidade de Brasília
Faculdade de Tecnologia**

**Avaliando o impacto das modificações faciais
voluntárias em modelos de aprendizagem
profunda no reconhecimento facial**

Fernanda Vaz Borges Carneiro
Liz Carolina Jaber Costato

PROJETO FINAL DE CURSO
ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Brasília
2023

**Universidade de Brasília
Faculdade de Tecnologia**

**Avaliando o impacto das modificações faciais
voluntárias em modelos de aprendizagem
profunda no reconhecimento facial**

Fernanda Vaz Borges Carneiro
Liz Carolina Jaber Costato

Projeto Final de Curso submetido como requi-
sito parcial para obtenção do grau de Enge-
nheiro de Controle e Automação

Orientador: Prof. Dr. Flávio Barros Vidal

Brasília
2023

V393a Vaz Borges Carneiro, Fernanda.
Avaliando o impacto das modificações faciais voluntárias em modelos de aprendizagem profunda no reconhecimento facial / Fernanda Vaz Borges Carneiro; Liz Carolina Jaber Costato; orientador Flávio Barros Vidal. -- Brasília, 2023.
113 p.

Projeto Final de Curso (Engenharia de Controle e Automação)
-- Universidade de Brasília, 2023.

1. Reconhecimento Facial. 2. Aprendizagem Profunda. 3. Procedimento Estético. I. Jaber Costato, Liz Carolina. II. Barros Vidal, Flávio, orient. III. Título

**Universidade de Brasília
Faculdade de Tecnologia**

**Avaliando o impacto das modificações faciais
voluntárias em modelos de aprendizagem profunda no
reconhecimento facial**

Fernanda Vaz Borges Carneiro
Liz Carolina Jaber Costato

Projeto Final de Curso submetido como requi-
sito parcial para obtenção do grau de Enge-
nheiro de Controle e Automação

Trabalho aprovado. Brasília, 13 de Julho de 2023:

Prof. Dr. Flávio Barros Vidal, UnB/CIC
Orientador

Profa. Dra. Roberta Barbosa Oliveira,
UnB/IE/CIC
Examinador interno

Profa. Dra. Carla M. C. Cavalcante Koike,
UnB/IE/CIC
Examinador interno

Brasília
2023

Agradecimentos

Gostaria de agradecer primeiramente os meus pais, Eliene e Joaquim, pelo apoio incondicional às minhas escolhas acadêmicas e profissionais e por todo o esforço realizado para me proporcionar a melhor educação possível. Sou grata também ao nosso professor e orientador, Flávio Vidal, que esteve presente em todos os momentos do desenvolvimento desse trabalho demonstrando preocupação e nos instruindo da melhor maneira possível. Por fim, agradeço a todos os amigos que fiz durante a graduação e que me auxiliaram de alguma maneira a superar os desafios enfrentados no curso, em especial a minha amiga, dupla e co-autora desse trabalho, Liz Costato, pelo companheirismo e parceria nessa caminhada.

Fernanda Vaz Borges Carneiro

Gostaria de expressar meus sinceros agradecimentos a minha família e aos meus amigos, que são a minha base e o meu apoio, para tudo que preciso. Em especial agradeço aos meus pais, Alessandra e Rogério, que nunca mediram esforços para que eu tivesse uma vida feliz, equilibrada, repleta de oportunidades e com excelência educacional, sempre considerando e apoiando as minhas vontades e sonhos. Além disso, agradeço por serem verdadeiros exemplos de resiliência, dedicação e amor. Sou grata também ao meu excepcional irmão, Arthur, que sempre me ajudou em meus estudos e me incentivou com seu exemplo. Aprecio também os ensinamentos de todos os professores que contribuíram para a minha formação, permitindo-me construir a base de conhecimentos que possuo até o momento. Em especial, sou grata pelos esforços do nosso professor e orientador Flávio Vidal, que esteve presente em cada etapa desse trabalho, dedicando seu tempo para nos auxiliar da melhor maneira possível. Por fim, sou muito grata à Fernanda Vaz, que foi a minha dupla durante toda a graduação, por sua paciência, seu apoio e sua amizade nesses anos, e, em especial, ao longo desse projeto.

Liz Carolina Jaber Costato

*“Don’t let anyone rob you of your imagination,
your creativity, or your curiosity.
It’s your place in the world; it’s your life.
Go on and do all you can with it,
and make it the life you want to live.”
(Mae Jemison)*

Resumo

É notório o aumento do número de procedimentos estéticos realizados no rosto, em todo o mundo. Além disso, o uso de sistemas de reconhecimento facial é de grande importância em várias áreas. Sendo assim, este trabalho tem como objetivo avaliar o impacto de modificações faciais voluntárias em modelos estado da arte de aprendizagem profunda utilizados no reconhecimento facial. Para atingir este objetivo foi desenvolvida uma base de imagens composta por 90 (noventa) indivíduos que, em algum momento de sua vida, passaram por procedimentos estéticos e/ou cirúrgicos. Esse conjunto de imagens (C2FPW) é disponibilizado à comunidade acadêmica como um dos resultados do trabalho realizado através de links para imagens de domínio público em sites web. A metodologia proposta incluiu também outros três conjuntos de imagens disponíveis publicamente, que foram utilizados nas avaliações realizadas. Em um primeiro momento analisou exclusivamente a influência dos procedimentos estéticos e cirúrgicos nos modelos de aprendizagem profunda. Em seguida, utilizando quatro métodos diferentes, foi considerado o impacto dessas modificações faciais em conjunto com o envelhecimento. Enquanto as análises que consideram apenas os procedimentos estéticos e um dos métodos que incluíam o envelhecimento indicaram que os modelos respondem melhor na presença de procedimentos estéticos, outros dois métodos sugeriram que a maioria dos modelos se comporta estatisticamente igual com ou sem essas modificações faciais e o último sugeriu que o impacto é diferente para cada modelo analisado. Tendo em vista a quantidade de imagens consideradas e a atenuação do efeito da idade nos modelos que indicaram um comportamento estatisticamente igual com ou sem procedimentos estéticos, conclui-se que esta constatação é mais representativa do impacto das alterações faciais voluntárias no reconhecimento facial. Os resultados deste trabalho estimulam novas pesquisas nessa área desafiadora para algoritmos de reconhecimento facial.

Palavras-chave: Reconhecimento Facial. Aprendizagem Profunda. Procedimento Estético.

Abstract

The increase in the number of facial aesthetic procedures performed worldwide is notable. Additionally, the use of facial recognition systems is of great importance in various fields. Therefore, this study aims to evaluate the impact of voluntary facial modifications on state-of-the-art deep learning models used in facial recognition. To achieve this objective, an image base was developed, composed of 90 (ninety) individuals who underwent aesthetic and surgical procedures at some point in their lives. This set of images is made available to the academic community (C2FPW) as one of the results of the work carried out through links to public domain images on websites. The proposed methodology also included three other sets of publicly available images, which were used in the evaluations carried out. At first, it exclusively analyzed the influence of aesthetic and surgical procedures on deep learning models. Then, using four different methods, the impact of these facial changes in conjunction with aging was considered. While the analyses considering only the aesthetic procedures and one of the methods incorporating aging indicated that the models perform better in the presence of aesthetic procedures, other two methods suggested that most models behave statistically similar with or without these facial modifications. The last method suggested that the impact varies for each analyzed model. Bearing in mind the number of images considered and the attenuation of the effect of age in the models that statistically indicated a behavior that was the same with or without aesthetic procedures, it is concluded that this finding is more representative of the impact of voluntary facial changes on facial recognition. The results of this work stimulate further research in this challenging area for facial recognition algorithms.

Keywords: Facial Recognition. Deep Learning. Cosmetic Procedure.

Lista de ilustrações

Figura 1.1 – Exemplo 1 de indivíduo com significativa mudança de aparência devido a procedimentos estéticos.	19
Figura 1.2 – Exemplo 2 de indivíduo com significativa mudança de aparência devido a procedimentos estéticos.	19
Figura 2.3 – Exemplo de Redes Neurais Artificiais. Fonte: (IBM, s.d.).	23
Figura 2.4 – Arquitetura da LeNet-5 proposta por Lecun et al. (1998).	23
Figura 2.5 – Pipeline moderno de reconhecimento.	26
Figura 2.6 – Um esquemático da cascata de detecção. Adaptação da Figura 4 do artigo Viola e Jones (2001).	27
Figura 2.7 – Arquitetura do <i>Single Shot Multibox Detector</i> . Fonte: (LIU et al., 2016).	27
Figura 2.8 – Um esquemático da detecção por HoG seguido por SVM. Adaptação da Figura 1 do artigo (CORTES; VAPNIK, 1995).	29
Figura 2.9 – Taxa de erro e número de falso positivo por janela, para diferentes métodos de detecção de objetos. Fonte: (KING, D. E., 2015).	29
Figura 2.10–Pipeline da cascata do MTCNN. Fonte: Zhang et al. (2016).	30
Figura 2.11–Arquitetura das três etapas da cascata do MTCNN. Fonte: Zhang et al. (2016).	31
Figura 2.12–Saídas proporcionadas pelo RetinaFace. Fonte: Jiankang Deng, Guo, Yuxiang Zhou et al. (2019).	31
Figura 2.13–Visão geral do único estágio do RetinaFace. Os módulos de contexto de cada pirâmide são diferentes, e, para cada âncora em questão, é feito um cálculo de perda multitarefa. Fonte: Jiankang Deng, Guo, Yuxiang Zhou et al. (2019).	32
Figura 2.14–Exemplo de saída fornecida pelo BlazeFace em vermelho. Fonte: (BAZAREVSKY et al., 2019).	32
Figura 2.15–Arquitetura da <i>DeepFace</i> proposta em Taigman et al. (2014a).	34
Figura 2.16–Estrutura da rede convolucional da primeira versão <i>DeepID</i> proposta por Sun, Wang e Tang (2014)	35
Figura 2.17–Treinamento da <i>FaceNet</i> proposta em Schroff, Kalenichenko e Philbin (2015).	35
Figura 2.18–Aprendizado por <i>Triplet Loss</i> . Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015).	36
Figura 2.19–Arquitetura da CNN proposta em Parkhi, Vedaldi e Zisserman (2015).	37
Figura 2.20–Transformação 2D (afim) proposta pela <i>OpenFace</i> . Fonte: Baltrusaitis, Robinson e Morency (2016).	38

Figura 2.21–Estrutura da rede convolucional da <i>ArcFace</i> proposta por Jiankang Deng, Guo, Xue et al. (2019)	39
Figura 2.22–Esquemático da dinâmica de treinamento do SFace.	40
Figura 2.23–Exemplo de distribuição normal padrão (distribuição z).	45
Figura 4.24–Esquemático da metodologia seguida.	57
Figura 4.25–Exemplos de imagens da HDA, para cada cirurgia contida na base. As imagens foram parcialmente desfocadas, de forma a preservar as identidades dos indivíduos. Fonte: adaptação da Figura 2 de Rathgeb, Dogan, Stockhardt, Maria De Marsico et al. (2020)	58
Figura 4.26–Exemplo de imagens de um sujeito na FGNET. Fonte: Chang, Chu-Song Chen e Hung (2011)	59
Figura 4.27–Métodos propostos para a avaliação da influência de procedimentos estéticos no reconhecimento facial.	63
Figura 4.28–Exemplos de pares positivos e negativos na HDA. Fonte: adaptação da Figura 2 de Rathgeb, Dogan, Stockhardt, Maria De Marsico et al. (2020)	64
Figura 4.29–Exemplo de comparações realizadas no sujeito 002 da FGNET.	68
Figura 4.30–Procedimentos para a obtenção do valor final por indivíduo no Método 2, usando como exemplo o Sujeito 002 da FGNET.	70
Figura 4.31–Vetores formados pelos valores médios das 9 comparações de cada sujeito da base FGNET.	71
Figura 4.32–Vetor referente à FGNET, para o modelo VGG-Face, usando o resultado expresso na Tabela 5.25 para o sujeito 002.	72
Figura 5.33–Distribuição de indivíduos por sexo na C2FPW.	74
Figura 5.34–Distribuição de indivíduos por nacionalidade na C2FPW.	75
Figura 5.35–Distribuição de procedimentos cosméticos pelo mundo. Fonte: (MICHAS, 2023).	75
Figura 5.36–Distribuição de imagens por indivíduo na C2FPW.	76
Figura 5.37–Exemplo de imagens da C2FPW do indivíduo S014 com 43 imagens ao longo da vida.	76
Figura 5.38–Exemplo de imagens da C2FPW do indivíduo S025 com 16 imagens ao longo da vida.	77
Figura 5.39–Exemplo de detecção de face com modelo MTCNN do sujeito 002 da FGNET com 15 anos.	77
Figura 5.40–Imagens da FGNET não detectadas pelo modelo MTCNN.	78
Figura 5.41–Distribuição da quantidade de imagens por indivíduo nas bases FGNET e C2FPW.	84
Figura 5.42–Exemplo de imagens de uma pessoa contida na FGNET, indivíduo 002.	85
Figura 5.43–Dois exemplos de seleção de 10 imagens, por quartis, para o indivíduo 002 da FGNET.	85

Figura 5.44–Exemplo de seleção de fotos pelas extremidades, para o indivíduo 002 da FGNET.	86
Figura 5.45–Posição do <i>intervalo</i> para o indivíduo 002 da FGNET.	86
Figura 5.46–Distribuição de clusters nas duas bases C2FPW e FGNET.	87
Figura 5.47–Exemplo de resultados de comparações consecutivas do indivíduo 002 da base FGNET, utilizando o modelo de reconhecimento facial VGG-Face.	92

Lista de tabelas

Tabela 2.1 – Tabela comparativa da acurácia de reconhecimento dos diferentes modelos, quando testados na base de dados LFW (HUANG; RAMESH et al., 2007).	41
Tabela 4.2 – Tabela dos <i>thresholds</i> dos diferentes modelos do <i>DeepFace</i> para a distância por cossenos.	62
Tabela 4.3 – Matriz de Confusão.	65
Tabela 5.4 – Erros de detecção na base C2FPW.	77
Tabela 5.5 – Erros de detecção na base FGNET.	78
Tabela 5.6 – Sensibilidade calculada na HDA.	79
Tabela 5.7 – Especificidade calculada na HDA.	79
Tabela 5.8 – Acurácia calculada na HDA.	80
Tabela 5.9 – Resultados publicados de acurácia na LFW (HUANG; RAMESH et al., 2007).	80
Tabela 5.10–Sensibilidade calculada na LFW.	81
Tabela 5.11–Acurácia dos 9 modelos de reconhecimento na base FGNET.	82
Tabela 5.12–Sensibilidade e Especificidade dos 9 modelos de reconhecimento na base FGNET.	82
Tabela 5.13–Acurácia dos 9 modelos de reconhecimento na base C2FPW.	82
Tabela 5.14–Sensibilidade e Especificidade dos 9 modelos de reconhecimento na base C2FPW.	83
Tabela 5.15–Resultado dos modelos de reconhecimento facial para cada comparação do conjunto de imagens da Figura 5.43a.	88
Tabela 5.16–Resultado dos modelos de reconhecimento facial para cada comparação do conjunto de imagens da Figura 5.43a.	88
Tabela 5.17–Resultado final das 30 configurações de 10 imagens do indivíduo 002 da FGNET (Figura 5.42).	89
Tabela 5.18–Resultado final das 30 configurações de 10 imagens do indivíduo 002 da FGNET (Figura 5.42).	89
Tabela 5.19–Resultado do teste de normalidade Shapiro-Wilk nos 18 vetores obtidos no Método 2.	89
Tabela 5.20–Resultado das variâncias dos 18 vetores obtidos no Método 2.	90
Tabela 5.21–Resultado do teste de hipótese T-Teste nos pares de vetores apresentados em 5.19, cujo comportamento foi normal.	91
Tabela 5.22–Resultado do teste de hipótese <i>Wilcoxon Rank-Sum</i> nos pares de vetores apresentados na Tabela 5.19, cujo comportamento foi não-normal.	91

Tabela 5.23–Resultado do teste de hipótese nos pares de vetores apresentados nas Tabelas 5.21 e 5.22, cuja primeira hipótese foi rejeitada.	91
Tabela 5.24–Diferença entre comparações consecutivas do modelo VGG-Face, com as imagens da Figura 5.43a do indivíduo 002 da FGNET.	92
Tabela 5.25–Cálculo do valor final do indivíduo 002 da FGNET, como a soma das variações absolutas.	93
Tabela 5.26–Resultado do teste de normalidade Shapiro-Wilk nos 18 vetores obtidos no Método 3.	93
Tabela 5.27–Resultado das variâncias dos 18 vetores obtidos no Método 3.	94
Tabela 5.28–Resultado do teste de hipótese T-Teste nos pares de vetores apresentados em 5.26, cujo comportamento foi normal.	94
Tabela 5.29–Resultado do teste de hipótese <i>Wilcoxon Rank-Sum</i> nos pares de vetores apresentados em 5.26, cujo comportamento foi não-normal.	95
Tabela 5.30–Resultado do teste de hipótese nos pares de vetores apresentados em 5.28 e 5.29, cuja primeira hipótese foi rejeitada.	95
Tabela 5.31–Sensibilidade dos 9 modelos de reconhecimento facial, usando os resultados encontrados no Método 2	95

Lista de abreviaturas e siglas

C2FPW	<i>Clinical Cosmetic Facial Procedures in the Wild</i>	60
CALFW	<i>Cross-Age Labeled Faces in the Wild</i>	25
CFP-FP	<i>Celebrities in Frontal-Profile in the Wild</i>	25
CNN	<i>Convolutional Neural Network</i>	22
CPLFW	<i>Cross-Pose Labeled Faces in the Wild</i>	25
CPU	<i>Central Processing Unit</i>	36
HDA	<i>Hochschule Darmstadt Plastic Surgery Database</i>	56
HOG	<i>Histogram of Oriented Gradients</i>	28
K-NN	<i>K-nearest Neighbor</i>	41
LDA	<i>Linear Discriminant Analysis</i>	41
LFW	<i>Labeled Faces in the Wild</i>	25
LMNN	<i>Large Margin Nearest Neighbor</i>	35
MMOD	<i>Margin Object Direction</i>	28
MTCNN	<i>Multi-task Cascaded Convolutional Networks</i>	26
OACI	<i>Organização da Aviação Civil Internacional</i>	56
SSD	<i>Single Shot Multibox Detector</i>	26
SVM	<i>Support Vector Machine</i>	28
YTF	<i>YouTube Faces</i>	25

Sumário

1	Introdução	16
1.1	Contextualização	16
1.1.1	Reconhecimento Facial e o desafio de modificações faciais	16
1.1.2	Procedimentos Estéticos	17
1.2	Justificativa	18
1.3	Objetivos	20
1.3.1	Objetivo Geral	20
1.3.2	Objetivos Específicos	20
1.4	Organização do Trabalho	21
2	Fundamentação Teórica	22
2.1	Aprendizagem Profunda	22
2.1.1	Redes Neurais Convolucionais	22
2.2	Reconhecimento Facial	24
2.2.1	Principais <i>benchmarks</i> e conjuntos de dados de imagem	24
2.2.2	Fluxo de trabalho moderno de Reconhecimento Facial	26
2.2.3	Etapa de Detecção	26
2.2.4	Etapa de Alinhamento	33
2.2.5	Etapa de Representação	33
2.2.6	Etapa de Classificação	41
2.3	Análise Estatística	43
2.3.1	Amostragem da População	43
2.3.2	Distribuições Frequências	44
2.3.3	Teste de Hipótese	46
3	Trabalhos Relacionados	49
3.1	Reconhecimento Facial e Envelhecimento	49
3.2	Reconhecimento Facial e Procedimentos Estéticos	51
3.3	Elaboração do Conjunto de Dados para Reconhecimento Facial	54
4	Metodologia	57
4.1	Revisão bibliográfica	57
4.2	Elaboração e seleção de conjuntos de imagens faciais	57
4.2.1	Conjunto de dados de faces com cirurgias plásticas: HDA	57
4.2.2	Conjunto de dados com envelhecimento: FGNET	59
4.2.3	Desenvolvimento do conjunto de imagens proposto: C2FPW	60

4.2.4	Escolha de um <i>benchmark</i> para comparação dos resultados obtidos .	60
4.3	Definição de modelos de detecção e de reconhecimento facial	61
4.3.1	Avaliação dos modelos do <i>Deepface</i>	61
4.3.2	Escolha do modelo de detecção que será usado	61
4.3.3	Definição da métrica de distância que será usada	62
4.4	Avaliação da influência da idade e de procedimentos estéticos no reconheci- mento facial	62
4.4.1	Avaliando separadamente a influência de procedimentos estéticos .	62
4.4.2	Incluindo a influência do envelhecimento	66
4.5	Análise e discussão dos resultados obtidos	73
5	Resultados	74
5.1	Conjunto proposto de imagens ao longo da vida de pessoas as quais realizaram modificações faciais estéticas: C2FPW	74
5.2	Definição de modelos de detecção e de reconhecimento facial	76
5.2.1	Definição do modelo de detecção	76
5.3	Avaliação da influência da idade e de procedimentos estéticos	78
5.3.1	Método 1 - Acurácia dos modelos de reconhecimento	81
5.3.2	Método 2 - Comportamento médio de reconhecimento ao longo da vida	88
5.3.3	Método 3 - Variação absoluta de reconhecimento ao longo da vida .	92
5.3.4	Método 4 - Sensibilidade dos modelos de reconhecimento	95
6	Conclusões	97
6.1	Trabalhos Futuros	99
	Referências	100

1 Introdução

Este capítulo inicia com uma contextualização da questão dos procedimentos estéticos no âmbito do reconhecimento facial, abordando também a atual situação dessas modificações estéticas voluntárias em nível global. Em seguida, é fornecida uma justificativa para a realização da pesquisa, seguida por uma descrição clara dos objetivos pretendidos com este estudo, e, por fim, é delineada a organização dos próximos capítulos.

1.1 Contextualização

1.1.1 Reconhecimento Facial e o desafio de modificações faciais

O reconhecimento facial é uma aplicação da visão computacional que tem como objetivo identificar a identidade das pessoas presentes em uma determinada imagem. Sendo assim, trata-se de uma ferramenta de biometria não invasiva, adequada para uso em diversos contextos sociais (SINGH, S.; PRASAD, 2018).

Um dos trabalhos de maior relevância na história do reconhecimento facial foi realizado em 1996 por Ballantyne, Boyer e Hines (1996). Esse trabalho consistiu em reconhecer, dentre imagens de um livro de fotos, registros que correspondessem a pessoa de uma dada fotografia. O projeto foi chamado de “homem-máquina”, ou semi-automático, porque coordenadas de 20 determinadas características do rosto, como o centro das pupilas e as extremidades do contorno dos olhos, eram extraídas manualmente das imagens do livro, armazenadas computacionalmente, e então eram associadas à imagem e ao nome da pessoa. O programa comparava tais características com as coletadas da fotografia, calculava a distância entre a fotografia e cada imagem, e, por fim, retornava quais imagens eram as mais próximas da fotografia.

Essa primeira tentativa de reconhecimento facial semi-automático trouxe bons resultados e já apresentou características observadas em trabalhos atuais, como o uso de características da face e o cálculo de distâncias para concluir sobre a identidade do indivíduo na foto. Além disso, Ballantyne, Boyer e Hines (1996) apresentaram desafios que foram e têm sido enfrentados por quem estuda essa área. São eles: rotação e inclinação da cabeça, intensidade de iluminação, expressões faciais, envelhecimento, entre outros.

Em seguida, tentativas para obter melhores resultados foram se sucedendo (ADJABI et al., 2020). Os mesmos autores, Ballantyne, Boyer e Hines (1996), incluíram mais características da face, para serem utilizadas na identificação, proporcionando resultados ainda mais precisos. Já no trabalho de Turk e Pentland (1991) foi proposta uma abordagem de quase tempo real, facilmente implementada com redes neurais, que utiliza características

não necessariamente correspondentes à atributos específicos, como olhos, nariz e boca. Tais características da face, chamadas *eigenfaces*, são obtidas de um *set* de fotos de treinamento, e formam o espaço chamado de *face space*. A distância da imagem a esse espaço permite concluir se trata-se ou não de um rosto, e a distância entre duas imagens nesse espaço possibilita determinar se ambas pertencem ao mesmo indivíduo ou não.

Buscando instigar ainda mais o avanço do reconhecimento facial surgiram competições como o *Face Recognition Grand Challenge (FRGC)* (PHILLIPS; FLYNN et al., 2005) e programas como o *Face Recognition Technology (FERET)*, proposto por Phillips, Wechsler et al. (1998), que disponibiliza um banco de dados para o desenvolvimento e também para testes de algoritmos de reconhecimento facial. Além disso, grandes companhias como Google e Facebook começaram a criar e aprimorar seus próprios algoritmos para essa tarefa, como em Pike e K. Thompson (2019) e Taigman et al. (2014a). Com todo esse investimento, tecnologias de código aberto e também comerciais obtiveram resultados excelentes no reconhecimento facial, apesar de se tratar de uma tarefa complexa, e com diversas variáveis envolvidas. Em Taigman et al. (2014a), ainda em 2014, o algoritmo *DeepFace* alcançou performance próxima ao quão bem um humano reconhece faces.

Na última década, como será melhor apresentado no Capítulo 3, muitos trabalhos focaram na problemática de intervenções estéticas no reconhecimento facial. Inicialmente, resultados sugeriram que o desempenho dessa tarefa era significativamente afetado por procedimentos estéticos (SINGH, R.; VATSA; BHATT et al., 2010). Com o aprimoramento dos algoritmos de reconhecimento facial e com a elaboração de bases de imagem de qualidade, contendo fotos de sujeitos antes e depois de cirurgias plásticas no rosto, trabalhos recentes revelam que algoritmos atuais de aprendizado profundo não apresentam erros significativos ao tentar reconhecer faces de indivíduos que realizaram cirurgias plásticas (RATHGEB; DOGAN; STOCKHARDT; DE MARSICO et al., 2020).

1.1.2 Procedimentos Estéticos

São diversas as motivações para pessoas buscarem modificações estéticas. Segundo Haas, Champion e Secor (2008), tais motivos são uma combinação de fatores fisiológicos e emocionais, influências culturais e educacionais, como também devido a histórico de provocações e de baixa auto-estima. Ademais, além de somente o desejo de uma aparência física mais próxima de padrões, conforme Maisel et al. (2018), há outros estímulos para buscar cirurgias plásticas, como a procura por uma saúde física e psicológica melhor, bem como por um aumento na confiança e no conforto em situações sociais. Num geral, é constatado por Maisel et al. (2018) que a decisão é para satisfazer necessidades da própria pessoa que realizou o procedimento, e não de outros indivíduos.

Além disso, desde jovem, a aparência é muito relevante para a identidade e para a individualidade de uma pessoa (DAVISON; BIRCH, 2001). Ainda, de acordo com J. Kevin

Thompson et al. (1999), os ideais de beleza são dados e reforçados por colegas, pais e também pela mídia. Com isso em vista, e considerando a força de redes sociais no contexto atual, Walker et al. (2019) elaborou uma pesquisa evidenciando que ver com frequência fotos, em redes sociais, de outras mulheres as quais realizaram procedimentos estéticos aumenta o desejo de jovens mulheres por tais intervenções.

Com todas essas motivações e ainda com o aprimoramento e barateamento de cirurgias plásticas, estas tem se tornado ainda mais populares (SINGH, R.; VATSA; BHATT et al., 2010). Conforme Michas (2022b), procedimentos estéticos cirúrgicos aumentaram 73,54% de 1997 a 2021, enquanto a quantidade de procedimentos não cirúrgicos, como *fillers* e aplicação de toxinas, mais que dobrou no mesmo período nos Estados Unidos. Apenas nesse país, em 2021, foram realizados mais de 5 milhões de procedimentos estéticos não cirúrgicos e mais de 1 milhão cirúrgicos.

Para um panorama geral, segundo Rathgeb, Dantcheva e Busch (2019), as cinco intervenções mais comuns são: correção de sobrelance, correção das pálpebras, *facelift*, rinoplastia e correção dos ossos do rosto. Também segundo Michas (2022a), os países com maior número de procedimentos estéticos são os Estados Unidos, seguido pelo Brasil e pela Alemanha. Além disso, de acordo com a Sociedade Estadunidense de Cirurgias Plásticas (PLASTIC..., s.d.), em 2020 nos Estados Unidos, a maioria das cirurgias plásticas foi realizada por pessoas de 40 a 54 anos, e 92% dessas operações foram feitas por indivíduos do sexo feminino. Esses comportamentos também foram observados em anos anteriores.

1.2 Justificativa

O reconhecimento facial tem aplicações diversas na sociedade. Pode ser, por exemplo, uma ferramenta de controle de acesso não invasiva, silenciosa e ainda sem necessidade de contato com equipamentos compartilhados entre diversas pessoas, os quais podem transmitir doenças (MAHMOOD et al., 2017). Além disso, essa tecnologia pode auxiliar policiais na resolução de crimes (MITTAL; SINGH, R., 2022) e até mesmo atuar como facilitador para encontrar crianças desaparecidas (CAPLOVA et al., 2017).

De acordo com Li (2015), variações na face se referem a mudanças de condição de visualização, como iluminação e pose, ou a mudanças nas propriedades do rosto, como as ocasionadas pelo envelhecimento. Tais modificações podem influenciar significativamente e de forma negativa o desempenho de métodos de reconhecimento facial. Dado que cirurgias plásticas e outros procedimentos estéticos tem sido cada vez mais frequentes, e que o reconhecimento facial é uma ferramenta muito útil, pesquisadores têm estudado o resultado dessas intervenções no reconhecimento facial.

Há casos em que reconhecer pessoas pode ser desafiador, até mesmo para seres humanos, devido a alterações na face, como cirurgias plásticas e uso de toxinas injetáveis. Um

exemplo notável é a cantora norte-americana Kimberley Denise Jones, conhecida como Lil' Kim. As imagens da cantora com 23 e 43 anos, Figuras 1.1a e 1.1b, respectivamente, demonstram essa significativa diferença de aparência devido às modificações faciais voluntárias. Outro exemplo é o ator brasileiro Stenio Garcia. As Figuras 1.2a e 1.2b retratam diretamente o antes e o depois do procedimento *facelift* realizado por ele.



(a) Cantora Lil' Kim com 23 anos.
Fonte: (2NDTIDE, 2019)



(b) Cantora Lil' Kim com 43 anos.
Fonte: (HILL, 2017)

Figura 1.1 – Exemplo 1 de indivíduo com significativa mudança de aparência devido a procedimentos estéticos.



(a) Ator Stenio Garcia antes do *facelift*.
Fonte: (NT, 2023).



(b) Ator Stenio Garcia depois do *facelift*. Fonte: (NT, 2023).

Figura 1.2 – Exemplo 2 de indivíduo com significativa mudança de aparência devido a procedimentos estéticos.

Os trabalhos publicados sobre esse tópico divergiram, em parte, em suas conclusões. Enquanto estudos, como o de R. Singh, Vatsa, Bhatt et al. (2010), apontam para uma resposta

deteriorada do reconhecimento facial frente à cirurgias plásticas, a publicação de [Rathgeb, Dogan, Stockhardt, De Marsico et al. \(2020\)](#), sugeriu que algoritmos de aprendizagem profunda são mais robustos a tais modificações. Com os diferentes resultados encontrados na temática, faz-se interessante estudar como intervenções estéticas influenciam diferentes algoritmos de estado da arte, especificamente de aprendizagem profunda, e concluir se eles são menos influenciados por tais operações. Nesse sentido, o presente trabalho visa entender como nove modelos de reconhecimento facial, os quais utilizam aprendizagem profunda, são influenciados por modificações faciais estéticas considerando os procedimentos isoladamente e também de forma conjunta com o envelhecimento.

1.3 Objetivos

1.3.1 Objetivo Geral

O principal objetivo do presente trabalho é avaliar o desempenho de nove algoritmos de estado da arte de reconhecimento facial, quando testados em bases de imagens com e sem procedimentos estéticos, visando concluir acerca do impacto das modificações faciais nesses algoritmos. As alterações faciais consideradas para o presente trabalho podem ser procedimentos cirúrgicos ou não, desde que realizados de forma voluntária, para fins estéticos. Os modelos estudados no presente trabalho utilizam aprendizagem profunda, foram (ou ainda são) o estado da arte e possuem código aberto para uso.

1.3.2 Objetivos Específicos

A partir do objetivo principal, espera-se alcançar os seguintes objetivos específicos para o trabalho realizado, a saber:

- Realizar uma revisão bibliográfica de trabalhos que já abordaram a problemática de procedimentos estéticos no reconhecimento facial;
- Encontrar uma base de imagens publicamente disponível, com baixa probabilidade de conter sujeitos que realizaram operações no rosto, mas que apresente o indivíduo em diferentes idades. Dessa forma, essa base servirá de referência para estudar a influência apenas da idade no reconhecimento facial;
- Semelhantemente, deseja-se encontrar uma base de imagens de alta qualidade, a qual compare sujeitos antes e depois de procedimentos estéticos, para que a problemática do impacto dessas interferências sejam consideradas separadamente do envelhecimento no reconhecimento facial.
- Desenvolver uma base de imagens de indivíduos, os quais realizaram qualquer tipo de modificação facial, contendo imagens coletadas em diversas idades. Com essa base,

será possível avaliar simultaneamente a influência da idade e dos procedimentos no reconhecimento facial.

1.4 Organização do Trabalho

Esse trabalho é organizado nos seguintes capítulos: O Capítulo 2 exibe uma breve revisão teórica de conceitos fundamentais para a compreensão da metodologia proposta e para a interpretação dos resultados obtidos. Ainda, o Capítulo 3 apresenta trabalhos publicados no estado da arte, os quais abordam o impacto do envelhecimento e de procedimentos estéticos no reconhecimento facial. Além disso, estudos acerca da elaboração de banco de dados para o reconhecimento facial são resumidos nesse capítulo. O Capítulo 4 descreve a metodologia adotada neste trabalho, fornecendo detalhes sobre os procedimentos utilizados. Já o Capítulo 5 apresenta os resultados obtidos com essa metodologia e promove discussões acerca desses resultados. Por fim, o Capítulo 6 apresenta as principais conclusões do estudo e sugestões para trabalhos futuros.

2 Fundamentação Teórica

Nesse capítulo serão apresentados conceitos e fundamentações teóricas de tópicos que serão abordados durante o desenvolvimento do trabalho. Primeiramente há uma explicação acerca de aprendizagem profunda, redes neurais convolucionais e seus impactos nos trabalhos relacionados a imagens. Além disso, uma descrição de reconhecimento facial e suas características será realizada, bem como uma explicação sucinta dos modelos que serão utilizados. Por fim, de forma a fundamentar a análise dos métodos propostos, faz-se uma revisão de conceitos estatísticos.

2.1 Aprendizagem Profunda

O aprendizado de máquina é um campo da inteligência artificial que objetiva construir programas de computador capazes de aprimorar seu desempenho com base em experiências próprias (MITCHELL, 1997). Com o aumento da pesquisa nessa área, surgiu uma vertente de estudo conhecida como aprendizagem profunda, ou *deep learning* (SALAS; BARROS VIDAL; MARTINEZ-TRINIDAD, 2019). Essa área é responsável por desenvolver algoritmos de aprendizado que se assemelham à estrutura das redes neurais do cérebro humano.

Dentro desse campo, existem diversas subdivisões relacionadas às formas possíveis de aprendizado. Uma delas é o aprendizado supervisionado, que se baseia na utilização de exemplos, ou seja, dados previamente rotulados, para treinar o algoritmo. No caso desse tipo de modelo de aprendizagem, a qualidade e quantidade dos exemplos disponíveis para o algoritmo têm um impacto direto em sua acurácia, ou seja, em sua capacidade de modelar os dados de forma precisa (FAN; SHI, 2022).

2.1.1 Redes Neurais Convolucionais

Parte do fundamento da área de aprendizagem profunda está contido na definição das redes neuronais ou redes neurais artificiais (IBM, s.d.). Essas redes são caracterizadas por possuírem uma camada de entrada, uma de saída, e múltiplas camadas escondidas, como na Figura 2.3. Cada camada possui nós, ou neurônios, que foram propostos pela primeira vez em 1943 por McCulloch e Pitts (1943), e que realizam a soma ponderada das entradas e, através do limiar da função de ativação definida, eles decidem se o neurônio será ativado ou não. Esse processo ocorre para todos os nós da rede, e através da ativação ou não dos neurônios, eles realizam a passagem de informação pela rede, até a camada de saída. Um exemplo de uma rede neural profunda é apresentado na Figura 2.3.

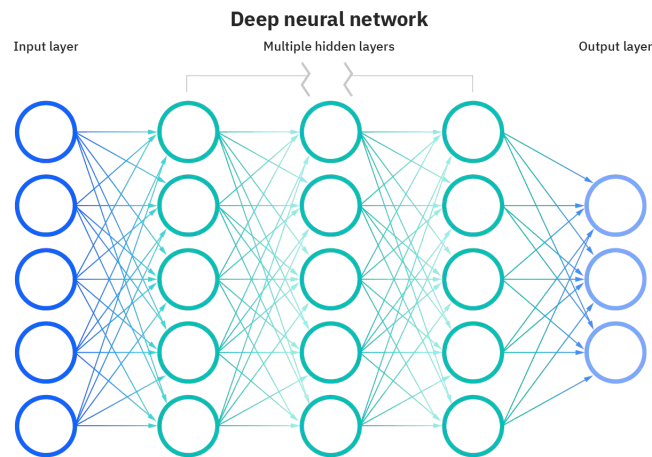


Figura 2.3 – Exemplo de Redes Neurais Artificiais. Fonte: (IBM, s.d.).

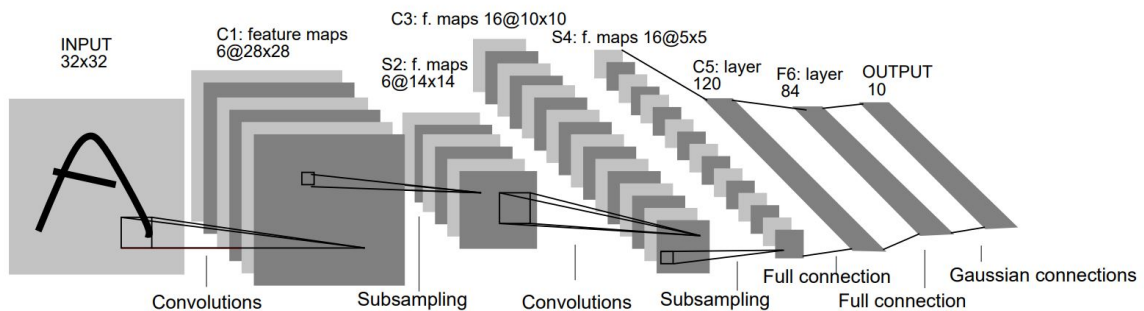


Figura 2.4 – Arquitetura da LeNet-5 proposta por Lecun et al. (1998).

As redes neurais convolucionais (CNN) como são encontradas hoje são inspiradas pelo conceito de "Neocognitron", idealizado em 1980 pelo Dr. Kunihiko Fukushima (FUKUSHIMA, 1980). A primeira CNN foi inicialmente proposta em 1998 por Lecun et al. (1998) para classificação de dígitos manuscritos. Eles utilizaram o banco de dados LeCun e Cortes (2010) e a CNN usada está representada pela Figura 2.4. Esses modelos são redes neurais especializadas que tentam se aproximar do funcionamento da visão humana, portanto, são muito utilizadas para tarefas de visão computacional. As CNNs conseguem identificar diferentes padrões em sua organização hierárquica de camadas ocultas com funções específicas. Com uma associação de camadas convolucionais, de *pooling* e camadas totalmente conectadas, essas redes são capazes de extrair de seus dados de entrada as informações que foram consideradas importantes no processo de treinamento e utilizar dessas informações para a tarefa para a qual foi programada (ALMEIDA; BARROS VIDAL, 2021).

Para ajustar os pesos dos nós das redes neurais, as CNNs passam por um treinamento, que geralmente necessita de grandes quantidades de dados previamente conhecidos e rotulados, além de um recurso computacional de alto nível. Como forma de reduzir os custos associados ao treinamento da rede, começou a se utilizar redes já pre-treinadas como ponto de partida dos pesos dos nós da rede no treinamento. Além disso, muitas redes construídas

para tarefas específicas, que são frequentemente utilizadas, ficam disponíveis para utilização do público, como é o caso de diversas redes de detecção e reconhecimento facial.

2.2 Reconhecimento Facial

Reconhecimento facial é uma das áreas de estudo que mais se desenvolveram nos últimos anos, e um dos grandes motivos dessa ascensão é a variedade de aplicações possíveis com essa tecnologia (ZHAO, W. et al., 2003). A tecnologia é responsável por realizar a identificação de um ou mais rostos no ambiente, podendo ser em foto, em vídeo, em tempo real ou em gravação, e extrair as características importantes que serão utilizadas para comparação e consequentemente para o reconhecimento.

Os primeiros sistemas de reconhecimento facial tinham como fluxo de trabalho (*pipeline*) a sequência: Detecção, Extração de Características e Classificação (ADJABI et al., 2020). A detecção aqui é o processo que o sistema realiza para encontrar rostos dentro do ambiente. A fase de extração de características é marcada pela análise do rosto identificado e extração do vetor de características dele, que será responsável por representá-lo. Por fim, a classificação consiste na comparação e verificação dos vetores de características a fim de encontrar outro rosto com mesmas características ou atestar a semelhança entre dois rostos, e assim realizar o reconhecimento.

2.2.1 Principais *benchmarks* e conjuntos de dados de imagem

No histórico do reconhecimento facial, já foram propostos e desenvolvidos diversos algoritmos, os quais podem ter abordagens 2D ou 3D, empregar ou não redes neurais ou mesmo utilizar diferentes características da face, para detectar uma identidade. Sendo assim, é necessário dispor de métricas e padrões que permitam comparar tais abordagens e definir quais possuem um desempenho mais interessante.

No contexto de detecção, reconhecimento e verificação facial, existem muitos conjuntos de dados de imagens, que são utilizados como *benchmarks*, de forma a comparar o desempenho dos algoritmos quando estes conjuntos de dados são usados como entrada. Neste caso, a porcentagem da base de dados que foi detectada, reconhecida ou verificada corretamente, a depender do propósito do algoritmo, é apresentada como um dos resultados do algoritmo naquele *benchmark*. Alguns dos mais conhecidos e utilizados são (DENG, J.; GUO; XUE et al., 2019):

- **Labeled Faces in the Wild (LFW)**: o LFW é um conjunto de dados de imagem público composto por mais de 13000 fotos, rotuladas com o nome da pessoa. Foi proposto visando ser uma ferramenta para auxiliar o desenvolvimento de pesquisas em verificação facial, e não para verificar o desempenho de algoritmos com propósito

comercial. Ainda, os criadores reforçam que a base de imagens é limitada em termos de variedade de idade, etnia e que há um desbalanço entre fotos de homens e mulheres (HUANG; RAMESH et al., 2007).

- **YouTube Faces (YTF)**: o conjunto de dados YTF foi proposto para o reconhecimento de faces em vídeos. Ele é composto por 3425 vídeos de 1595 pessoas. Os vídeos podem conter de 1 a 6 pessoas, e são rotulados com os nomes desses sujeitos. Os criadores ainda propõem alguns testes de *benchmark*, para medir a performance de reconhecimento de pares de vídeo que contém a mesma pessoa (WOLF; HASSNER; MAOZ, 2011).

Há também outros conjuntos de dados de imagens mais recentes, os quais focam em aspectos específicos e em desafios no reconhecimento facial, como possuir fotos em diferentes poses, níveis de iluminação e idade. Esse é o caso dos seguintes *benchmarks*:

- **Celebrities in Frontal-Profile in the Wild (CFP-FP)**: sabendo que a pose é um ponto que ainda gera muitos erros em algoritmos atuais de reconhecimento facial, o CFP-FP foi proposto, de forma a facilitar o progresso nesse aspecto. Trata-se de uma base de imagens aberta que contém fotos de 500 famosos de frente e de perfil (SENGUPTA et al., 2016).
- **Cross-Pose Labeled Faces in the Wild (CPLFW)**: semelhantemente ao CFP-FP, o CPLFW é uma base de imagens aberta que possui 3000 pares de fotos, sendo as fotos de um par pertencentes a uma mesma pessoa, em poses diferentes, de forma a possibilitar o estudo do impacto da pose no reconhecimento facial (ZHENG; DENG, W., 2018).
- **AgeDB**: já a AgeDB é uma base de imagens a qual aborda a questão do envelhecimento no reconhecimento facial. Esse conjunto de dados é composto por 16488 imagens, cada uma rotulada com a identidade, a idade e o sexo do indivíduo. A AgeDB é aberta para pesquisa, mas deve-se requisitar aos donos a senha de acesso, por email (MOSCHOGLOU et al., 2017).
- **Cross-Age Labeled Faces in the Wild (CALFW)**: ainda na problemática da idade, o conjunto de dados aberto CALFW reúne 3000 pares de fotos, em que há uma diferença de idade entre as duas fotos de uma mesma pessoa (ZHENG; DENG, W.; HU, 2017).

E conjuntos de dados de imagens os quais abordam diferentes desafios ao mesmo tempo, buscando demonstrar um desempenho geral do algoritmo, como:

- **VGG Face2**: VGG Face2 é uma base de 3,31 milhões de imagens de 9131 diferentes indivíduos. Tais fotos incluem variação de pose, idade, iluminação, etnia e profissão (CAO, Q. et al., 2018).

2.2.2 Fluxo de trabalho moderno de Reconhecimento Facial

Diferentemente do fluxo de trabalho (*pipeline*) inicial, o reconhecimento facial moderno possui uma etapa a mais: o Alinhamento. Segue-se então a ordem: Detecção, Alinhamento, Representação e Classificação [Taigman et al. \(2014a\)](#), como mostra a Figura 2.5. A seguir, cada uma dessas etapas será melhor explicada.

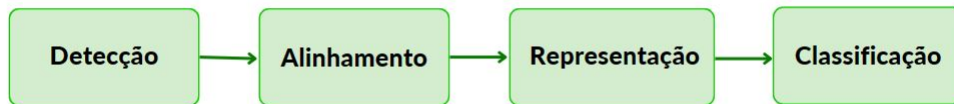


Figura 2.5 – Pipeline moderno de reconhecimento.

2.2.3 Etapa de Detecção

O primeiro passo para o reconhecimento facial é detectar um ou mais rostos na imagem. Existem muitos algoritmos para a detecção de contornos e de objetos e também algoritmos especializados para a detecção de faces. No presente trabalho, os algoritmos usados para essa etapa foram: *Haar Cascade* ([VIOLA; JONES, 2001](#)), *Single Shot Multibox Detector* (SSD) ([LIU et al., 2016](#)), dois modelos da biblioteca *Dlib* ([KING, D. E., 2009](#)), *Multi-task Cascaded Convolutional Networks* (MTCNN) ([ZHANG et al., 2016](#)), *RetinaFace* ([DENG, J.; GUO; ZHOU, Y. et al., 2019](#)) e *MediaPipe* ([PIKE; THOMPSON, K., 2019](#)). Uma breve descrição destes algoritmos será feita a seguir.

2.2.3.1 Haar Cascade

O *Haar F-based Cascade* foi proposto em 2001, por [Viola e Jones \(2001\)](#). Trata-se de um algoritmo de aprendizagem de máquina direcionado à detecção de objetos. [Viola e Jones \(2001\)](#) proporcionaram três grandes contribuições, as quais foram responsáveis pelos ótimos resultados apresentados. A primeira foi uma nova proposta para representação de uma imagem, chamada de "*Integral Image*". Tal representação permite avaliar *pixels* de uma forma computacionalmente mais barata do que representações de imagens por intensidades de cores, e facilitar a busca por contornos ou características relevantes na imagem ([OPENCV, s.d.](#)).

A segunda colaboração foi o uso de um algoritmo baseado no *AdaBoost*, do [Schapire \(1999\)](#), para determinar quais das características utilizadas são mais relevantes. De forma simplificada, a ideia é atribuir a cada uma das características um valor de limiar que permitira, baseado neste valor, distinguir se na imagem há ou não o objeto desejado, neste caso, um rosto. Ao final, são escolhidas as características que menos erraram em definir a existência da face ([OPENCV, s.d.](#)).

Por fim, a imagem é dividida em pequenas janelas de igual tamanho. Considerando apenas os valores de limiar de uma quantidade pequena de características, as janelas que possuírem menor probabilidade de conter uma face são descartadas. Ainda com essa ideia, como o nome do método indica, uma *cascata* é introduzida na Figura 2.6. Nos passos seguintes ao abandono das primeiras janelas, mais características são analisadas, e mais janelas são rejeitadas. Àquelas que chegarem ao último estágio da cascata são as faces detectadas pelo algoritmo. O procedimento é repetido com diferentes tamanhos de janelas (EMAMI; SUCIU, 2012). Com essa estratégia, mais processamento é gasto com trechos de interesse, enquanto trechos irrelevantes são logo descartados (OPENCV, s.d.).

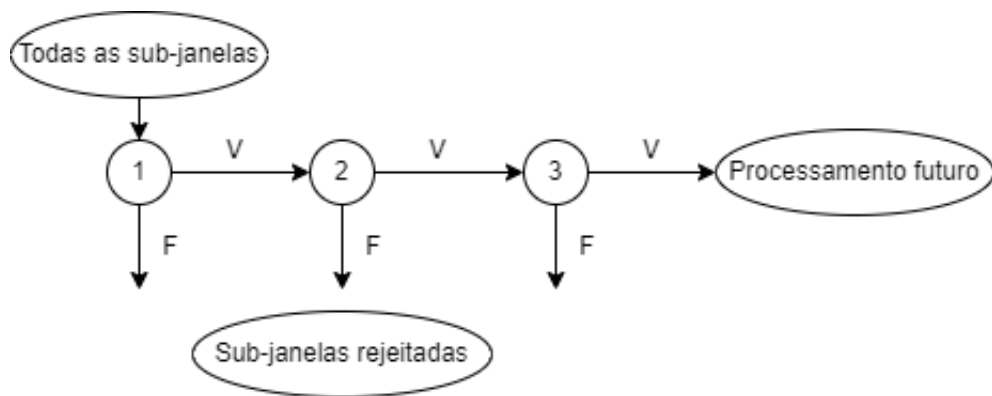


Figura 2.6 – Um esquemático da cascata de detecção. Adaptação da Figura 4 do artigo Viola e Jones (2001).

2.2.3.2 Single Shot Multibox Detector (SSD)

Este modelo foi inicialmente desenvolvido em um projeto na empresa Google, e, em seguida, foi continuado na *University of North Carolina at Chapel Hill*. Trata-se de uma única e rápida rede neural convolucional profunda, a qual produz um número fixo de saídas. Cada saída corresponde a uma caixa delimitada, que equivale a uma determinada região da imagem, e fornece uma conclusão se há ou não um objeto naquela caixa. Em seguida, um passo de *Non-Maximum Suppression* permite remover caixas com alta superposição, fornecendo ao final todos os objetos detectados. A Figura 2.7 explicita o modelo SSD.

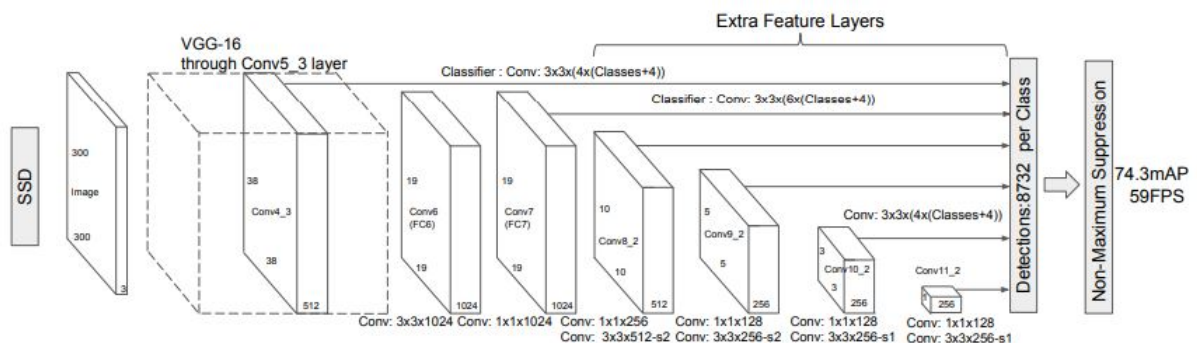


Figura 2.7 – Arquitetura do *Single Shot Multibox Detector*. Fonte: (LIU et al., 2016).

As primeiras camadas desta rede neural se baseiam em um modelo para classificação de imagens, mas removendo-se as etapas finais de classificação. No caso da imagem 2.7 e dos resultados apresentados por Liu et al. (2016), a arquitetura padrão utilizada como base foi a VGG-16, introduzido em Simonyan e Zisserman (2014). Já a escolha das caixas analisadas é semelhante ao *anchor boxes*, usado por Ren et al. (2015). Entretanto, são utilizados diversos mapas de característica com variadas resoluções, possibilitando discretizar, de forma eficiente, o espaço de possíveis formatos de caixa de saída.

2.2.3.3 Dlib

Dlib é uma biblioteca gratuita de uso geral, a qual contém algoritmos de aprendizado de máquina. Apesar de ser escrita em C++, a Dlib também provê uma interface em Python Davis E. King (2009). Para a detecção facial, a Dlib oferece duas opções: *Histogram of Oriented Gradients* (HOG) seguido por *Support Vector Machine* (SVM) e *Margin Object Direction* (MMOD). Enquanto o HOG é uma abordagem já mais antiga, para detecção de objetos num geral, o MMOD faz uso de *deep learning* (SERENGIL, 2018a).

- HoG seguido por SVM

Histogram of Oriented Gradients (HOG) é um descritor de características usado em visão computacional, para a detecção de objetos. A ideia foi proposta inicialmente por Sternberg (1983) e consiste em uma adaptação do método *Edge Orientation Histograms*. No HoG, entretanto, utiliza-se uma grande quantidade de células, as quais são pequenas regiões da imagem original, uniformemente espaçadas. Além disso, os gradientes de contrastes locais são normalizados, para que a performance computacional seja melhorada (DALAL; TRIGGS, 2005).

Para a detecção facial, as características chamadas de *Histogram of Oriented Gradient Features* são inicialmente extraídas da imagem, para, em seguida, serem inseridas em uma *Support Vector Machine* (SVM) (DALAL; TRIGGS, 2005). Uma SVM é um modelo de aprendizado de máquina que consegue determinar se dados são pertencentes à qual classe entre duas possíveis, baseado em seu aprendizado com amostras de treinamento. As duas classes no caso de detecção facial são se a imagem de entrada possui ou não uma pessoa. Esse método utiliza-se de teorias estatísticas e é um dos mais robustos métodos de predição (CORTES; VAPNIK, 1995). Na abordagem usada pela Dlib, o fluxo seguido para a detecção facial pode ser observado no fluxograma da Figura 2.8.

O gama apresentado na primeira etapa da Figura 2.8 diz respeito à Correção Gama, a qual é melhor explicada em Poynton (2003).

- MMOD

A maioria dos detectores de objetos utilizam apenas um subconjunto das janelas de uma imagem para aprenderem a detectar objetos. O método *Max-Margin Object*

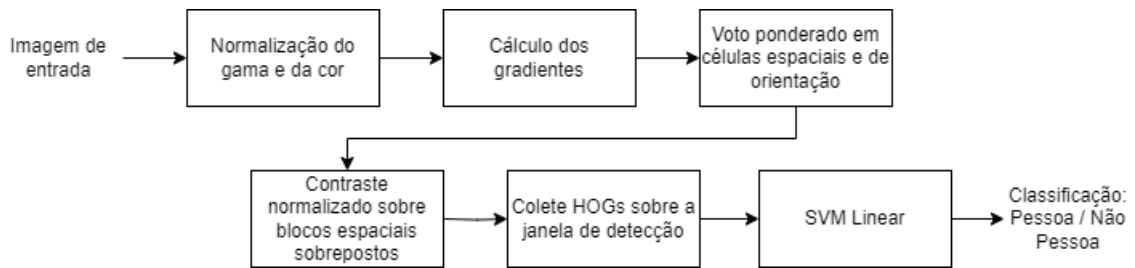


Figura 2.8 – Um esquemático da detecção por HoG seguido por SVM. Adaptação da Figura 1 do artigo (CORTES; VAPNIK, 1995).

Detection proposto por Davis E. King (2015) faz uso de todas as possíveis janelas em seu aprendizado, removendo aquelas que possuem mais de 50% da área coincidente com alguma das demais, e buscando ainda um tempo razoável de computação. Esse método pode ser utilizado para melhorar qualquer modelo de detecção que seja linear nos parâmetros de aprendizado, como é o caso do HoG, apresentado anteriormente.

Junto do MMOD são usadas apenas funções lineares de classificação do tipo $f(x, r) = \langle w, \phi(x, r) \rangle$, em que x é a imagem, r é a janela analisada e ϕ extrai as características da janela. É chamado de $F(x, y)$ a soma das pontuações $f(x, r)$ de todas as janelas da imagem x . O MMOD calcula um vetor de parâmetros, o qual é chamado "w" e cujo objetivo é obter o menor número possível de erros de detecção. Sendo assim, ele otimiza $f(x, r)$ para que o $F(x, y)$ do rótulo y correto seja maior que todos os $F(x, y)$ com y sendo a rotulação incorreta para x .

De forma a visualizar a otimização proporcionada pelo MMOD, a imagem 2.9 mostra que o MMOD usado no HOG (gráfico em rosa) possuiu menos falsos positivos e uma taxa de erro menor do que alguns modelos, incluindo o HoG sem MMOD (gráfico em amarelo na Figura 2.9). Esse resultado foi usando o conjunto de imagens INRIA Pedestrian (FARRAJOTA, 2017), o qual contém imagens positivas, com presença de pessoas, e negativas, sem nenhum indivíduo.

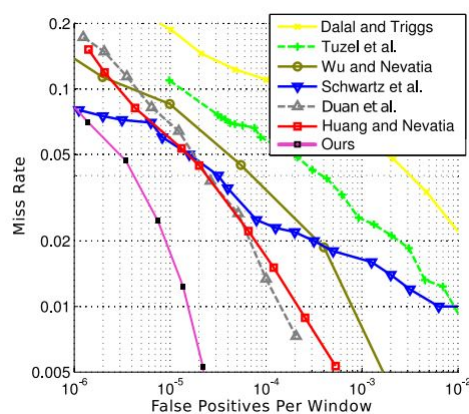


Figura 2.9 – Taxa de erro e número de falso positivo por janela, para diferentes métodos de detecção de objetos. Fonte: (KING, D. E., 2015).

2.2.3.4 Multi-task Cascaded Convolutional Networks (MTCNN)

Um dos pontos-chaves do MTCNN, proposto por [Zhang et al. \(2016\)](#), é levar em consideração a correlação existente entre as etapas de alinhamento e detecção de faces, no pipeline de reconhecimento facial. Sendo assim, a *Multi-task Cascaded Convolutional Network* consiste de uma Rede Neural Convolutiva (CNN) (Seção 2.1.1) de três estágios, como é exemplificado pelas imagens das Figuras 2.10 e 2.11, e realizam ambas as tarefas simultaneamente. Os três estágios são:

- *Proposal Network* (P-Net): uma CNN rasa, que produzirá como saída múltiplas janelas candidatas a conter um rosto.
- *Refine Network* (R-Net): com uma CNN um pouco mais complexa, esse estágio remove uma grande quantidade de janelas as quais não são de interesse.
- *Output Network* (O-Net): por fim, uma CNN mais profunda já entrega as faces detectadas com a posição de suas características principais especificadas.

Todos os estágios utilizam a técnica *Bounding Box Regression*, muito popular para refinar ou prever a localização de objetos em imagens ([LEE; KWAK; CHO, 2019](#)), e *Non-Maximum Suppression*, para remover candidatos com alta sobreposição.

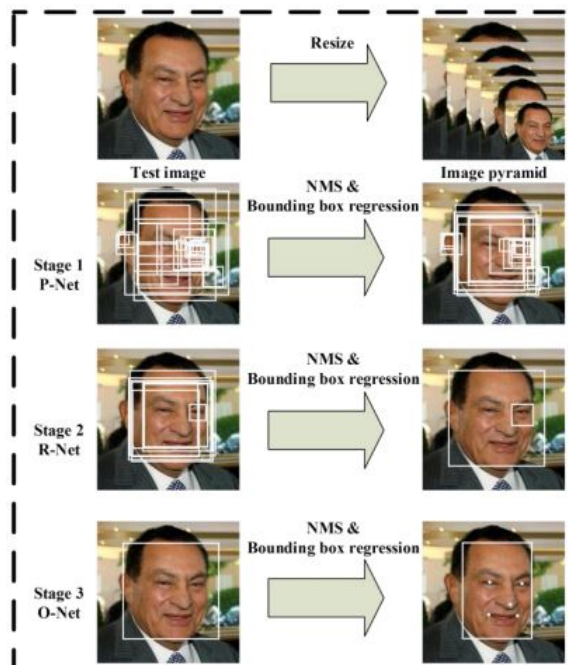


Figura 2.10 – Pipeline da cascata do MTCNN. Fonte: [Zhang et al. \(2016\)](#).

2.2.3.5 RetinaFace

Proposto por [Jiankang Deng, Guo, Yuxiang Zhou et al. \(2019\)](#), esse modelo não é apenas para detecção facial, mas sim, como apresentado pelo autor, para identificação facial.

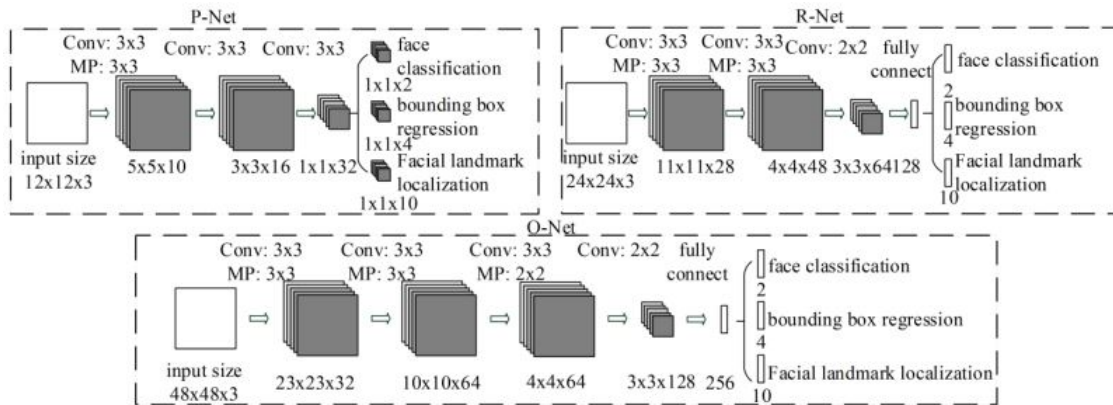


Figura 2.11 – Arquitetura das três etapas da cascata do MTCNN. Fonte: Zhang et al. (2016).

Essa tarefa conta também com a etapa de alinhamento e outras etapas seguintes, as quais podem ser úteis para algumas aplicações, como a análise pixel a pixel da face e também uma regressão 3D de densa correspondência. Seguindo a ideia de trabalhos anteriores bem sucedidos, Jiankang Deng, Guo, Yuxiang Zhou et al. (2019) utiliza uma rede convolucional de estágio único, a qual amostra densamente localizações e escalas de faces, em pirâmides de características, conforme apresentado na Figura 2.13.

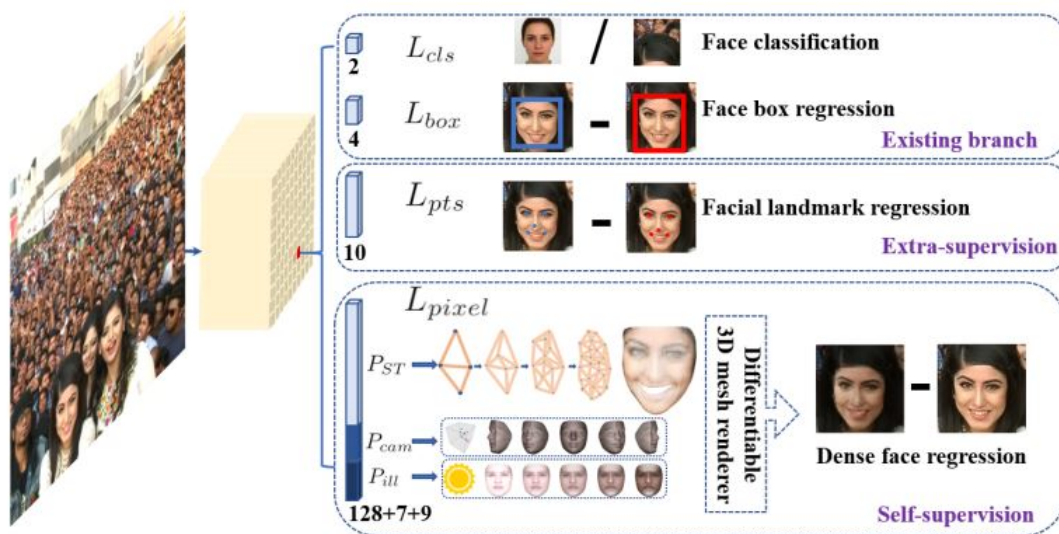


Figura 2.12 – Saídas proporcionadas pelo RetinaFace. Fonte: Jiankang Deng, Guo, Yuxiang Zhou et al. (2019).

Para o treinamento, foi feita uma anotação manual de cinco características faciais do WIDER FACE dataset (YANG et al., 2016), sendo elas o centro dos olhos, a ponta do nariz e as extremidades dos lábios. Já para a regressão 3D, é utilizada uma malha de decodificação auto-supervisionada, para prever a informação da face de forma 3D. Essas etapas proporcionaram um avanço considerável nos resultados do modelo. Ainda, por fazer uso de redes *backbone* leves, como MobileNet-0.25 (HOWARD et al., 2017), o RetinaFace pode ser utilizado em aplicações de tempo real, sem a exigência de um poder de processamento muito elevado.

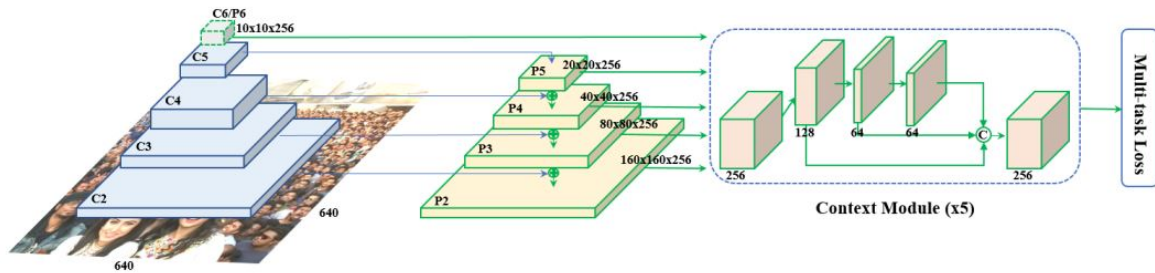


Figura 2.13 – Visão geral do único estágio do RetinaFace. Os módulos de contexto de cada pirâmide são diferentes, e, para cada âncora em questão, é feito um cálculo de perda multitarefa. Fonte: Jiankang Deng, Guo, Yuxiang Zhou et al. (2019).

2.2.3.6 MediaPipe

O MediaPipe é uma solução gratuita e de código aberto da empresa Google, que pode ser usada em diversas linguagens, como Python, Javascript e C++. Trata-se de um recurso para detecção de face ultra rápido, o qual permite reconhecer múltiplas faces simultaneamente. Suas principais vantagens são a facilidade de uso em diferentes plataformas e o fornecimento de uma rápida resposta, visando aplicações de tempo real, mesmo em hardwares comuns, como os de smartphones (PIKE; THOMPSON, K., 2019).

O MediaPipe é baseado no detector de face já existente, chamado BlazeFace, também da empresa Google (BAZAREVSKY et al., 2019). Inspirado no MobileNetV1/V2 (HOWARD et al., 2017) e utilizando ideias propostas por Liu et al. (2016), o BlazeFace pode servir como uma primeira etapa de aplicações de reconhecimento de faces, ao fornecer informações sobre a localização de seis pontos chave dos rostos, que são o centro dos dois olhos, a ponta do nariz, o centro da boca e a região do tragus das duas orelhas. Seguido deste passo, o pipeline da aplicação pode fornecer outros tratamentos desejados. A Figura 2.14 mostra em vermelho a saída do BlazeFace, enquanto em verde características refinadas são apresentadas por um modelo de estimativa de contorno de face mais complexo.

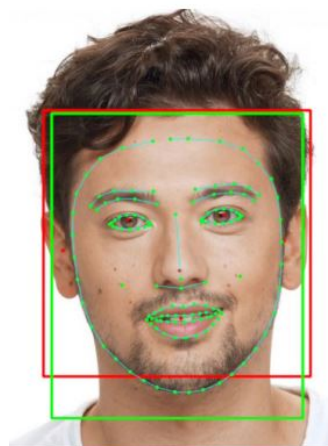


Figura 2.14 – Exemplo de saída fornecida pelo BlazeFace em vermelho. Fonte: (BAZAREVSKY et al., 2019).

Comparado ao MobileNetV2-SSD, modelo para aplicação semelhante ao MediaPipe, o desempenho deste é superior em precisão e em tempo de inferência. Além disso, o MediaPipe possui uma velocidade de inferência significativamente maior em GPUs de smartphones amplamente utilizados.

2.2.4 Etapa de Alinhamento

No pipeline moderno de reconhecimento facial, a adição da etapa de alinhamento da face foi fruto de estudos de aperfeiçoamento dos sistemas de reconhecimento. Alguns desses estudos, como o realizado por [Parkhi, Vedaldi e Zisserman \(2015\)](#), mostraram que o alinhamento da face após a etapa de detecção aumentava a acurácia de reconhecimento em até 1%. Nessa etapa, a imagem da face detectada é rotacionada de acordo à angulação entre os olhos, de forma que a linha que liga os olhos fique à 180° na imagem, ou seja, perpendicular a laterais da imagem. Nas grandes bibliotecas de reconhecimento facial, essa etapa de alinhamento está integrada à etapa de detecção e acontece de forma sequencial.

2.2.5 Etapa de Representação

Após as etapas de detecção e alinhamento, as imagens passam pela CNN para que sejam extraídas suas características principais antes de seguirem para a classificação final, essa etapa é a de representação. Como saída da CNN, é obtido um vetor de características que representa a imagem de entrada de forma compacta ([TAIGMAN et al., 2014a](#)), como na camada F7 da arquitetura do modelo *DeepFace* na [Figura 2.15](#).

Para conseguir distinguir quais as características importantes devem ser utilizadas no vetor de características da imagem, essas redes passaram por etapas de treinamento com milhares de imagens com as características desejadas conhecidas, de forma que conseguem generalizar bem para faces vindos de outros conjuntos de dados. Essa etapa de representação proposta por modelos de reconhecimento com aprendizagem profunda consegue condensar as informações importantes e necessárias para reconhecimento da face, que pode ser então comparada com maior facilidade e menor custo computacional a outros rostos ([ADJABI et al., 2020](#)). A seguir serão apresentados oito dos diversos modelos de reconhecimento facial existentes, e como é feita a representação em cada um deles.

2.2.5.1 Facebook DeepFace

Lançado em 2014 pelo Facebook, a *DeepFace* foi um marco no que diz respeito ao reconhecimento facial. A empresa tinha como objetivo principal identificar automaticamente usuários do Facebook em imagens publicadas na rede social. Seu lançamento foi marcado pela promessa de fechar quase toda a lacuna existente entre o processamento da máquina e o sistema visual humano ([TAIGMAN et al., 2014a](#)).

As contribuições realizadas por [Taigman et al. \(2014a\)](#) vão além do modelo entregue. Elas foram cruciais no estímulo de utilização de redes neurais na tarefa de reconhecimento facial na comunidade. Além disso, o *DeepFace* entrega o desenvolvimento de um método de aprendizado e de uma arquitetura de rede neural profunda, treinada em um banco de imagens que ultrapassa a casa dos milhões. É proposto um sistema de alinhamento facial (Figura 2.15) baseado em modelagem 3D da face.

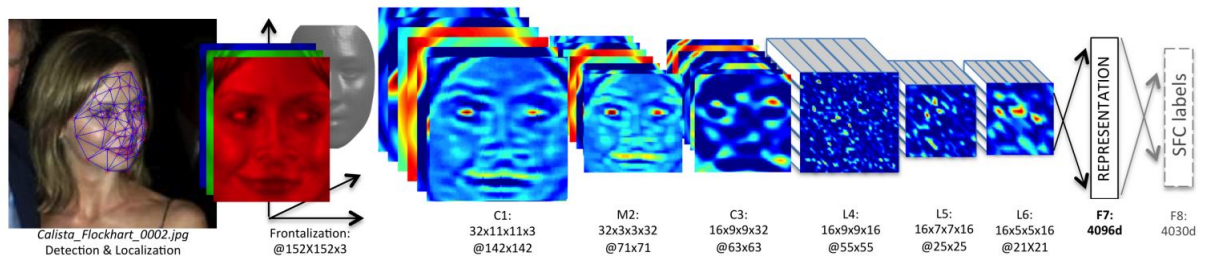


Figura 2.15 – Arquitetura da *DeepFace* proposta em [Taigman et al. \(2014a\)](#).

Com testes realizados nos *benchmarks* LFW e YTF, o *DeepFace* conseguiu entregar resultados significativos, com uma acurácia de 97% e 91.4%, respectivamente, em cada um dos conjuntos de dados. No YTF, esse resultado representou uma redução de mais de 50% do erro dos métodos mais renomados até o momento.

2.2.5.2 DeepID

No final de 2014, [Sun, Wang e Tang \(2014\)](#), através da Universidade de Hong Kong, publicaram a primeira versão do *DeepID*. A proposta realizada pelos autores foi de aprender características de identidade facial de alto nível, com modelos de aprendizagem profunda, por meio de classificação de imagens de treinamento, em uma dentre as 10.000 identidades consideradas.

As redes convolucionais utilizadas aprendem a classificar todos os rostos disponíveis para treinamento por suas identidades, utilizando a última camada oculta da rede como recursos, ou características, que são referidas como *DeepID*. As características faciais são extraídas pelas camadas da rede convolucional e o número vai reduzindo enquanto gradualmente características mais globais e de alto nível são formadas nas camadas finais. Ao final da rede, os resultados que são extraídos contém então muitas informações importantes de características fortes coletadas nas faces, como é mostrado na Figura 2.16, numa camada totalmente conectada de 160x1.

Essa abordagem, permite que os recursos de alto nível aprendidos tenham boa capacidade de generalização, pois não se ajustam excessivamente a um pequeno subconjunto de faces utilizadas em treinamento. Para atestar o funcionamento do algoritmo, a rede foi testada no LFW ([HUANG; RAMESH et al., 2007](#)) e conseguiu uma acurácia de reconhecimento de face de 97.45%. E por se tratar de uma abordagem que utiliza treinamento de

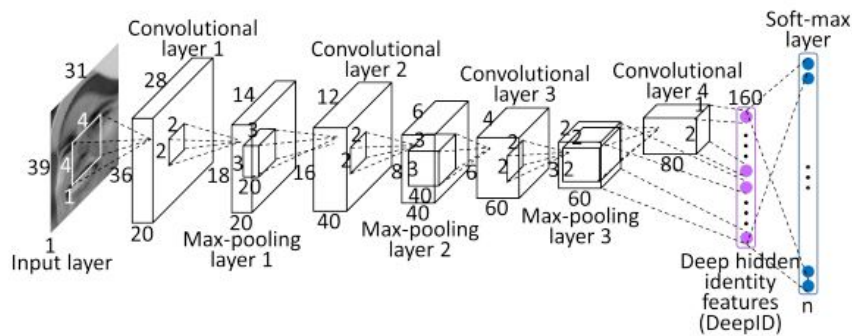


Figura 2.16 – Estrutura da rede convolucional da primeira versão *DeepID* proposta por Sun, Wang e Tang (2014)

classificação para extrair as características faciais principais, o aumento do banco de imagens de treinamento possibilita uma melhora na performance.

2.2.5.3 Google FaceNet

Uma empresa que está constantemente apresentando avanços, com novos modelos de reconhecimento facial é a Google. Em Schroff, Kalenichenko e Philbin (2015), três pesquisadores da Google apresentaram o FaceNet. Trata-se de um sistema de rede neural convolucional, cujo treinamento é mostrado na Figura 2.17, para mapear faces em um espaço Euclidiano compacto, no qual a distância corresponde a uma medida de similaridade entre rostos. A representação de uma face nesse espaço é dada por um vetor denominado *embedding*.

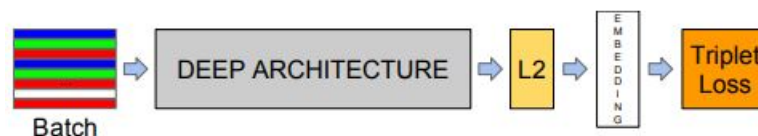


Figura 2.17 – Treinamento da *FaceNet* proposta em Schroff, Kalenichenko e Philbin (2015).

A ideia do FaceNet é treinar a CNN para gerar como saída diretamente *embeddings* de dimensão 128. O treinamento, por sua vez, utiliza-se de uma função de perda baseada em *triplets*, fundamentada em um algoritmo de aprendizado estatístico chamado *Large Margin Nearest Neighbor* (LMNN). Já os *triplets* consistem em um par de fotos de um mesmo indivíduo e uma terceira foto de um indivíduo diferente das demais. O treinamento com esses *triplets* visa separar o par positivo do negativo por uma margem de distância, como ilustrado na Figura 2.18. A escolha das fotos que compõe os *triplets* é fundamental para o treinamento e é bem explicada por Schroff, Kalenichenko e Philbin (2015).

Com essa abordagem, foi possível uma elevada eficiência de representação, utilizando um vetor de apenas 128 bytes para caracterizar uma face. Além disso, outra vantagem é que o *FaceNet* requer um alinhamento mínimo, dado por um corte justo ao redor da área

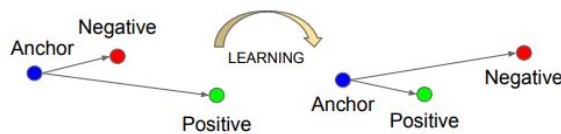


Figura 2.18 – Aprendizado por *Triplet Loss*. Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015).

do rosto. Ainda, com esse sistema, não apenas a tarefa de reconhecimento é possível, mas também a de verificação, dando como resposta se duas fotos pertencem a mesma pessoa, e a de agrupamento, podendo dizer quais das fotos de um conjunto pertencem a uma mesma pessoa. Na tarefa de reconhecimento, a acurácia obtida pelo *FaceNet* na base de imagens LFW foi de 99,63%, enquanto na YTF foi de 95,12%, superando os algoritmos de estado da arte do momento em que foi proposto.

Cabe ressaltar que há uma outra versão da *FaceNet* (SANDBERG, 2016) que se distingue da proposta por Schroff, Kalenichenko e Philbin (2015) basicamente no tamanho do vetor de *embedding* gerado. Há também influência de ideias do Visual Geometry Group (GROUP, s.d.). Nessa outra versão, ao invés de a saída ser um vetor de dimensão 128, essa dimensão é de 512 (SERENGIL, 2018b), e por isso é conhecido como *FaceNet512*.

Considerando as limitações ainda existentes, Schroff, Kalenichenko e Philbin (2015) propuseram como trabalhos futuros compreender os casos de erro obtidos, bem como reduzir o tamanho do modelo e a elevada demanda da CPU. Ainda, é interessante determinar estratégias para reduzir o longo período necessário para o treinamento. Foi sugerido, por exemplo, utilizar *batches* de treinamento de menor tamanho.

2.2.5.4 VGG-Face

Como justificativa para o nome, o *VGG-Face* foi desenvolvido pelo grupo *Visual Geometry Group*, da Universidade de Oxford (GROUP, s.d.). Ao avaliar arquiteturas propostas anteriormente, foi notado quais são as características importantes e quais os detalhes irrelevantes para obter bons resultados. Com essa análise, Parkhi, Vedaldi e Zisserman (2015) propuseram uma arquitetura "muito profunda", que compreende uma longa sequência de camadas convolucionais, consideravelmente mais simples que as de projetos anteriores, e ainda efetiva. A Figura 2.19 apresenta as principais características da profundidade do modelo da arquitetura da *VGG-Face*.

Para o treinamento, aprende-se, com *softmax log-loss*, um vetor de pontuações de dimensão 4096, que pode ser utilizado para identificação de faces. Em seguida, um treinamento utilizando *triplet loss* é realizado para aprender o descritor de *embedding*. Com isso, obtém-se uma camada de projeção da dimensão 4096 para 1024, dimensão escolhida para o descritor de *embedding* desse modelo.

Os resultados obtidos na LFW e na YTF são comparáveis aos de algoritmos de estado

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	-	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	-	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	-	3	-	64	-	-	64	-	128	-	-	128	-	256	-	256	-	-	256
num filts	-	64	-	64	-	-	128	-	128	-	-	256	-	256	-	256	-	-	512
stride	-	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	-	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1
layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmax
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	-	512	-	512	-	-	512	-	512	-	512	-	-	512	-	4096	-	4096	-
num filts	-	512	-	512	-	-	512	-	512	-	512	-	-	4096	-	4096	-	2622	-
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Figura 2.19 – Arquitetura da CNN proposta em Parkhi, Vedaldi e Zisserman (2015).

da arte, mesmo utilizando menos dados de treinamento e possuindo uma arquitetura mais simples. No LFW a acurácia de verificação do VGG-Face foi de 98,95%. Já no YTF, a acurácia de verificação foi de 91,6% sem o aprendizado do *embedding* e de 97,4%, quando essa etapa foi incluída.

Um estudo feito por Grm et al. (2017) busca comparar pontos fortes e fracos de quatro algoritmos de rede neural profunda, sendo eles AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), VGG-Face (GRM et al., 2017), GoogleLeNet (SZEGEDY et al., 2015) e SqueezeNet (IANDOLA et al., 2016). Nesse trabalho, em termos de desempenho médio de acurácia de verificação, o *VGG-Face* teve o melhor resultado. Entretanto, ele é o mais afetado por falta de dados na região periocular. Em termos de ruído, o *VGG-Face* se mostrou o mais robusto, mas também teve a pior performance quando houve variação no brilho da imagem.

2.2.5.5 OpenFace

Em 2015, a DeepFace (TAIGMAN et al., 2014a) e a FaceNet (SCHROFF; KALENICHENKO; PHILBIN, 2015) possuíam bastante destaque na área de reconhecimento facial, devido a suas elevadas acurácias, resultantes do treinamento de redes neurais com milhões de imagens de redes sociais. Com isso, Baltrusaitis, Robinson e Morency (2016) buscaram, com um algoritmo de código aberto, acompanhar o desenvolvimento dos sistemas privados de reconhecimento, com a publicação da *OpenFace*. A biblioteca tem então o propósito de utilização geral pela comunidade e promete explorar o reconhecimento para aplicações *mobile*, principalmente na área de reconhecimento facial em tempo real.

Com essa motivação por trás da criação do *OpenFace*, o foco dos autores foi na criação do modelo que resultasse em uma alta acurácia com um baixa necessidade de treinamento e de tempo de predição. Apesar de usar a arquitetura da rede neural da *FaceNet* como base, algumas mudanças são realizadas, como a entrada da rede, que agora é uma imagem menor e normalizada. Essa normalização é efetuada pelo *OpenFace* de forma a fazer com que os olhos e o nariz de todas as entradas da rede neural apareçam em locais semelhantes (Figura 2.20).

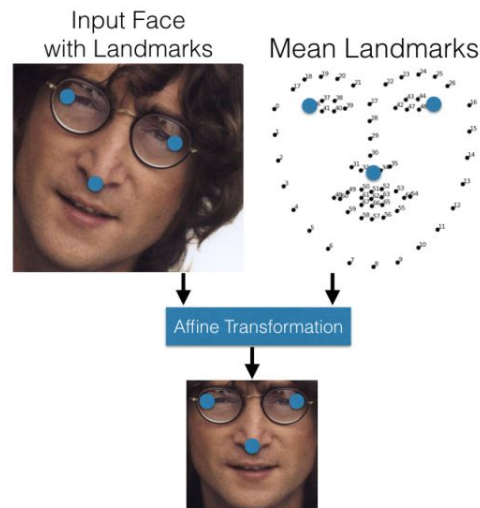


Figura 2.20 – Transformação 2D (afim) proposta pela *OpenFace*. Fonte: Baltrusaitis, Robinson e Morency (2016).

Para testagem do modelo, foi utilizado o *benchmark* LFW (HUANG; RAMESH et al., 2007). Os resultados obtidos se assemelham à técnicas de aprendizado bem estabelecidas no estado da arte, chegando à uma acurácia de 92,92%, próxima do reconhecimento facial humano (AMOS; LUDWICZUK; SATYANARAYANAN, 2016). Esse resultado é obtido com treinamento utilizando um conjunto de dados que chega a ser duas ordens de magnitude menor que o usado para o treinamento da *FaceNet*, por exemplo.

No entanto, estudos preliminares realizados por Fydanaki e Geradts (2018) mostram que o desempenho do modelo *OpenFace* é muito dependente da qualidade do conjunto de dados utilizado, não detectando faces quando qualquer ponto de referência estava ausente. Além disso, os resultados obtidos por Fydanaki e Geradts (2018) indicam que o tempo de execução do algoritmo estava mais correlacionado com a qualidade das imagens do que com o número de imagens. Com essa dependência de qualidade de imagem, os autores concluem a ineficiência de utilização do modelo *OpenFace* para aplicações forenses, visto que as fontes coletoras de imagens desses sistemas usualmente são de baixa qualidade.

2.2.5.6 Dlib

Como apresentado no tópico anterior, a Dlib é uma biblioteca gratuita de uso geral, a qual contém ferramentas de aprendizado de máquina (KING, D., 2015). Para o reconhecimento facial, o algoritmo da Dlib é inspirado no Modelo ResNet-34 (HE et al., 2015), que dispõe de redes residuais, visando facilidade de treinamento e otimização de resultados. Utilizando 29 camadas convolucionais, a rede neural desenvolvida produz vetores de saída, os quais descrevem a imagem de entrada no espaço de dimensão 128.

Ao ser testado no *benchmark* LFW (HUANG; RAMESH et al., 2007), em fevereiro de 2017, a versão do ResNet-34 da Dlib alcançou uma acurácia de 99,38%, fazendo este

algoritmo comparável a outros de estado da arte (KING, D., 2017). Apesar do expressivo resultado obtido nessa base, assim como outros modelos, a Dlib ainda possui limitações quando utilizada com imagens de baixa resolução, pouca iluminação e baixa profundidade de cor (YE et al., 2022). Por fim, a Dlib é uma biblioteca de código aberto e bem documentada, o que a faz uma opção interessante para aplicações de reconhecimento facial.

2.2.5.7 ArcFace

Ainda em 2015, num contexto em que muitos trabalhos propõe incluir margens na função de perda Softmax, visando maximizar a separação entre classes, pesquisadores do *Imperial College London*, em Jiankang Deng, Guo, Xue et al. (2019), introduziram à comunidade o *Additive Angular Margin Loss*, ou *ArcFace*. O objetivo de implementação do ArcFace é estabilizar o processo de treinamento e melhorar ainda mais o poder discriminativo de reconhecimento facial.

O Softmax tradicional não otimiza aproximações em uma classe e o distanciamento entre classes. Sendo assim, o algoritmo de reconhecimento é pouco robusto a variações, como pose e iluminação. Além disso, a distância entre classes é pequena, podendo levar à ambiguidade de interpretações (DENG, J.; GUO; XUE et al., 2019). A inclusão da margem angular do ArcFace visa mitigar esses problemas. Para isso, em um espaço normalizado de pesos e *embeddings*, é calculado o produto escalar entre o *embedding* de características e o peso de cada uma das N classes e 512 características utilizadas. Esses pesos são os centros da característica na hipersfera considerada. Utilizando a função arco-cosseno, calcula-se o ângulo entre o embedding atual e o centro alvo. Inclui-se, então, uma margem angular ao resultado obtido e atualiza-se o centro alvo (DENG, J.; GUO; XUE et al., 2019). O resumo dos passos adotados é apresentado na Figura 2.21.

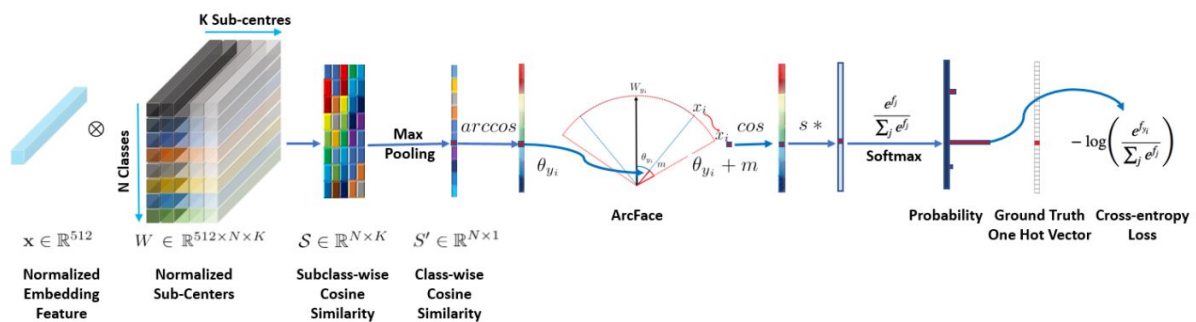


Figura 2.21 – Estrutura da rede convolucional da ArcFace proposta por Jiankang Deng, Guo, Xue et al. (2019).

Os testes realizados por Jiankang Deng, Guo, Xue et al. (2019) demonstram que o ArcFace otimiza a distinção entre classes, e ainda permite gerar, a partir do vetor de *embedding*, imagens de sujeitos presentes ou não no treinamento. Além disso, o ArcFace obteve uma acurácia de 99,81% no LFW (HUANG; RAMESH et al., 2007), comparável a

outros algoritmos de estado da arte. Entretanto, há ainda questões inerentes ao modelo proposto que dificultam a convergência e vão de encontro à redução da distância intraclasse (ZHAO, H. et al., 2020). Ademais, a facilidade de reconstruir faces a partir do vetor de *embedding* afeta a privacidade do indivíduo. Os pesquisadores de Jiankang Deng, Guo, Xue et al. (2019) pretendem, em trabalhos futuros, construir um modelo não inversível, ou seja, não seja possível reconstruir a face a partir do *embedding*.

2.2.5.8 SFace

O SFace foi apresentado em 2021 por alunos de PhD de BUPT (Beijing) e apoiado pela empresa Canon Information Technology (Beijing). O trabalho de Zhong et al. (2021) propõe uma forma de otimizar o objetivo do reconhecimento facial, o qual é minimizar a distância intraclasse, ou seja, entre fotos de um mesmo sujeito, e maximizar a distância interclasse, entre fotos de sujeitos distintos, considerando limitações que podem existir em bases de imagens de treinamento.

Para isso, Zhong et al. (2021) desenvolveram uma nova função de perda, *sigmoid-constrained hypersphere loss* (SFace), cujo nome pode ser traduzido, de forma literal, como "perda de hipersfera com restrição sigmoide". O objetivo dessa função é otimizar as distâncias intraclasse e interclasse apenas até um certo ponto, visando a melhorar a capacidade de generalização dos modelos.

O esquemático da Figura 2.22 visa facilitar a compreensão da proposta do SFace. W_{y_i} é a representação correta da amostra de treinamento x_i , e cuja rotulação é y_i . Já W_j é a representação de outra amostra, que não x_i . Sendo assim, tem-se que a velocidade de aproximação entre x_i e W_{y_i} diminui, quanto mais próximas esses pontos já estiverem, e a velocidade de afastamento entre x_i e W_j aumenta rapidamente quanto menor a distância entre eles. Essa lógica de mudança de velocidade é descrita por duas sigmóides. Já a direção da velocidade é sempre tangente à hipersfera. Tais características justificam o nome do método proposto.

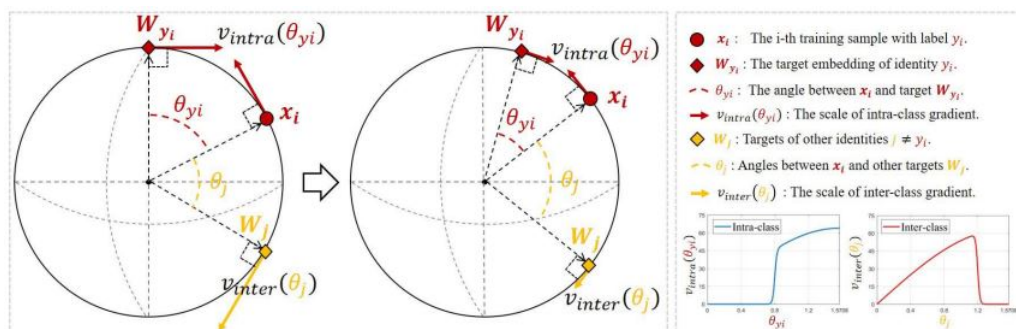


Figura 2.22 – Esquemático da dinâmica de treinamento do SFace.

O algoritmo apresentado visa otimizar as distâncias interclasse e intraclasse, apenas até certo ponto. Isso faz com que ruídos de rotulação sejam menos influentes no treina-

mento da rede neural, proporcionando maior robustez ao modelo. A melhor configuração de parâmetros testada resultou em um desempenho médio de 99,57% na LFW (HUANG; RAMESH et al., 2007). Foram também efetuados testes em outros *benchmarks*, como YTF (WOLF; HASSNER; MAOZ, 2011), CALFW Zheng, Weihong Deng e Hu (2017) e CPLFW (ZHENG; DENG, W., 2018), também apresentando resultados melhores que outros métodos comumente usados, como variações do Softmax.

Para facilitar a comparação, o resumo do desempenho dos modelos de reconhecimento apresentados neste capítulo, no conjunto de dados LFW (HUANG; RAMESH et al., 2007), está apresentado na Tabela 2.1.

Modelo de Reconhecimento	Acurácia (%)
Facebook DeepFace	97
Google FaceNet	99,63
VGG-Face	98,95
OpenFace	92,92
DeepID	97,45
Dlib	99,38
ArcFace	99,81
SFace	99,57

Tabela 2.1 – Tabela comparativa da acurácia de reconhecimento dos diferentes modelos, quando testados na base de dados LFW (HUANG; RAMESH et al., 2007).

2.2.6 Etapa de Classificação

Tendo já detectado, alinhado e representado a face, a etapa final é determinar a identidade daquele rosto. Nesse sentido, há duas abordagens possíveis. Pode-se comparar a face com uma base já conhecida de imagens, buscando determinar se o rosto em questão é compatível ou não com um dos sujeitos conhecidos da base. Esse problema é chamado de *Identificação Facial*, e o retorno será a identidade do sujeito, se esta puder ser identificada dentre as identidades conhecidas. A outra possibilidade é a chamada *Verificação Facial*. Nela, apenas duas imagens são comparadas, e o resultado é a conclusão se elas pertencem ou não a uma mesma pessoa.

Em ambos os casos, a etapa de classificação consiste em utilizar o vetor de características resultante da etapa de representação, para realizar a comparação entre duas imagens diferentes. Na Identificação Facial, esse processo é repetido várias vezes, até comparar e decidir sobre todas as possibilidades de sujeitos conhecidos. Para esses casos, é comum a utilização de classificadores como o *K-nearest Neighbor* (COVER; HART, 1967), o *Linear Discriminant Analysis* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e o *Support Vector Machine* (PHILLIPS, 1998).

Na Verificação Facial, a comparação é realizada calculando-se a distância entre os veto-

res de representação das duas imagens em questão, que resulta numa medida de similaridade entre eles. Pode-se dizer que a face detectada pertence a um determinado sujeito se sua representação é suficientemente semelhante à representação conhecida daquele sujeito. Essas distâncias entre vetores de representação da imagem podem ser calculadas de diferentes formas, e para cada uma haverá um valor que diz se o resultado é ou não suficientemente pequeno, para serem consideradas como faces da mesma pessoa. Esse valor é chamado de *threshold* ou, em português, *limiar*. Se a distância entre as duas representações é maior que o limiar, então as duas imagens pertencem a indivíduos distintos. Caso contrário, tem-se que as duas imagens são da mesma pessoa. Esse valor é frequentemente encontrado de forma empírica pelo programador do modelo, e depende de qual medida de distância está sendo usada.

2.2.6.1 Medidas de Similaridade

Há na literatura diferentes formas de calcular distância de vetores e consequentemente encontrar a medida de similaridade entre eles. Dentre os métodos mais conhecidos estão: similaridade cosseno, distância euclidiana e distância euclidiana normalizada, ou *L2-Norm*.

Considerando dois vetores de representação da face A e B , as distâncias podem ser calculadas por:

- Similaridade de Cosseno

A distância é medida entre dois vetores não nulos através do cálculo de seu produto interno.

$$\text{cossine}(A, B) = \frac{A^T \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2.1)$$

Onde A_i e B_i representam componentes do vetor A e B respectivamente, e n é o total de coordenadas em cada vetor (NGUYEN; BAI, 2010). O resultado obtido pode variar entre -1 e 1.

- Distância Euclideana

A distância euclideana tem seu cálculo baseado no Teorema de Pitágoras e no espaço Euclidiano (DANIELSSON, 1980). Essa distância corresponde ao comprimento do segmento de reta entre os vetores A e B e pode ser calculada com a seguinte equação:

$$\text{euclidean}(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (2.2)$$

- Distância Euclideana Normalizada (*L2-Norm*)

A diferença aqui é que é considerada o cálculo da norma do vetor, ou seja, sua distância

à origem. Para isso, um dos vetores da medida de distância euclidiana precisa ser numericamente igual à origem e para isso é realizada uma normalização dos vetores utilizados entre si.

2.3 Análise Estatística

Estatística é o ramo da ciência cujo foco é coletar, analisar, interpretar e apresentar dados de uma amostra que consigam representar uma população (DANGETI, 2017). Nesse campo, o termo *população* se refere a todos os dados que correspondam às características que estão sendo avaliadas ou que deseja-se analisar e/ou sobre as quais objetiva-se concluir, enquanto *amostra* representa uma parte selecionada da população que efetivamente será analisada. Essa divisão é necessária pois nem sempre uma análise da população inteira é conveniente e possível de ser realizada. Com análises da amostra através de ferramentas estatísticas, é possível inferir comportamentos da população como um todo, quando a amostra é representativa o suficiente.

Na estatística existem dois campos principais segundo Dangeti (2017):

- **Estatística Descritiva**, que é responsável por descrever os dados observados em formas que possibilitem um melhor entendimento. Entre as formas de representação estão: média, mediana, desvio padrão, frequência, percentagem, entre outros. Como o próprio nome diz, esse campo apenas sumariza os dados de forma a caracterizá-los de uma forma descritiva.
- **Estatística Inferencial**, que é caracterizada por métodos de quantificar propriedades de uma amostra de forma a representar toda a população. Portanto, os métodos desse campo da estatística são muito utilizados quando apenas uma amostra está disponível para os estudos e sendo necessária inferir as conclusões para a população toda.

Nos estudos realizados com bancos de imagens, os quais usualmente representam apenas uma pequena parcela de toda a população que possui as características estabelecidas na coleta de dados, a estatística é imprescindível na hora de concluir acerca de toda a população. Para os casos específicos de análise de modelos de aprendizagem de máquina, a aplicação de estatística está muito relacionada com a explicabilidade do comportamento desses modelos e na sua generalização em determinadas condições.

2.3.1 Amostragem da População

Na amostragem da população em que será realizada a análise, é preciso tomar alguns cuidados para que os dados finais consigam representar a população sem enviesar a

análise. Essa representatividade vem da necessidade de generalização dos resultados para a população, sendo preciso, portanto, que a amostra carregue as suas características gerais.

Existem dois tipos principais de amostragem os quais são muito utilizados, segundo McCombes, Shona (s.d.):

- **Amostragem Probabilística**, que envolve a seleção aleatória dos dados da população. Nesse tipo de amostragem, cada dado da população tem a mesma chance probabilística de ser selecionado como amostra. Dentro desse tipo existem alguns métodos específicos como: amostragem aleatória simples, amostragem sistemática, amostragem estratificada e amostragem por *clusters*.
- **Amostragem Não-Probabilística**, que é uma seleção não aleatória realizada de acordo aos critérios estabelecidos. Visto que a seleção é realizada com base em critérios, os dados da população não possuem a mesma chance de serem selecionados como amostra. Dentro desse tipo de amostragem estão alguns métodos conhecidos, como: amostragem por conveniência, amostragem por resposta voluntária, amostragem intencional e amostragem por quotas.

A decisão de como será realizada a amostragem da população é muito importante, e definirá o quão válidos são os resultados obtidos a partir da amostra coletada. Portanto, há um cuidado maior para reduzir os vieses de coleta e pesquisa, e aumentar a representatividade da questão estudada nos dados que serão utilizados.

2.3.2 Distribuições Frequenciais

Uma distribuição de frequência é responsável por dispor em tabela e/ou gráfico o número de vezes que cada valor possível de determinada variável ocorre na amostra de dados estudada. De forma semelhante, a distribuição probabilística ou distribuição de probabilidades é uma distribuição de frequência idealizada (TURNERY, SHAUN, s.d.) que visa descrever a probabilidade que cada valor tem de ocorrer na variável.

A distribuição de probabilidade é então uma descrição matemática das probabilidades que são utilizadas para descrever as populações de variáveis. Elas podem ser contínuas ou discretas, dependendo da variável cuja probabilidade ela está descrevendo. Quando se trata de uma variável do tipo contínua, é comum tratar a distribuição por meio da *função de densidade de probabilidade*, visto que a probabilidade de que a variável tenha qualquer valor específico, cujas possibilidades são infinitas, é tão pequena que considera-se como zero.

2.3.2.1 Distribuição Normal

Entre as diversas distribuições conhecidas, a *Distribuição Normal* ou *Distribuição de Gauss* é uma das mais importantes e mais encontradas no mundo real. Essa característica

é explicada pelo *Teorema do Limite Central*, que diz que toda soma de variáveis aleatórias independentes de média finita e variância limitada é aproximadamente normal, quando o número de amostras é suficientemente grande (LEHMANN; ROMANO; CASELLA, 2005).

A curva de distribuição normal tem um formato de sino idealmente simétrico e pode ser descrita pela sua média e desvio padrão através da equação 2.3. Dentre as características tem-se que o pico da curva acontece no valor de média e corresponde ao valor que mais tem probabilidade de ocorrência na população. Além disso, devido a sua simetria, os valores estatísticos de média, moda e mediana são iguais. Na Figura 2.23 é possível verificar uma distribuição normal padrão, ou *distribuição z*, onde a média é igual a zero e o desvio padrão é igual a um.

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad (2.3)$$

em que μ = média amostral, σ = desvio padrão amostral, σ^2 = variância, $f(\cdot)$ = função de probabilidade e x o valor da variável.

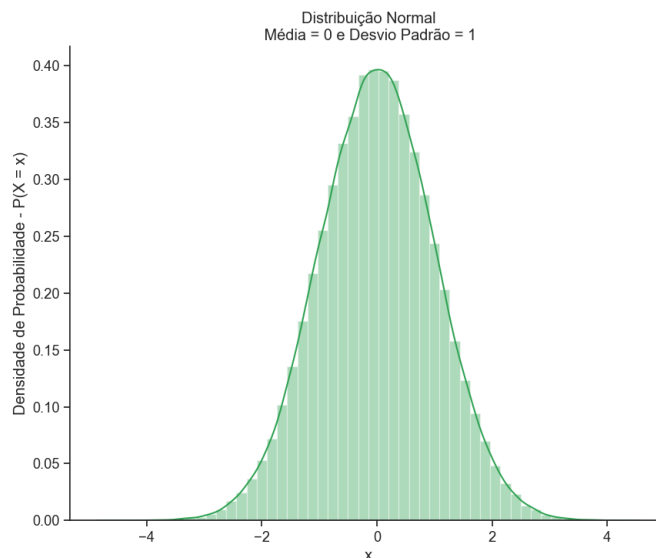


Figura 2.23 – Exemplo de distribuição normal padrão (distribuição z).

Outra característica muito importante encontrada em distribuições normais é que os valores equidistantes da média idealmente possuem mesma probabilidade de ocorrência e, quanto mais próximo da média, maior essa probabilidade. Como consequência da simetria também têm-se que valores menores que 1 desvio padrão do valor médio representam 68% de probabilidade de ocorrência, valores menores que 2 desvios padrão do valor médio representam 95% de probabilidade de ocorrência e por fim, valores menores que 3 desvios padrão do valor médio representam cerca de 99.7% de probabilidade de ocorrência.

2.3.3 Teste de Hipótese

Dentro desse campo, o teste estatístico de hipótese fornece uma metodologia objetiva de pesquisa que permite a verificação de hipóteses ou teoria estatística realizada acerca de dados amostrais. Segundo [Lehmann, Romano e Casella \(2005\)](#), este teste então busca decidir se a hipótese formulada está correta ou não, ou seja, aceita-se ou rejeita-se a hipótese. Quando aplicado o teste, é possível que haja erro na decisão e rejeite a hipótese quando ela é verdadeira ou a aceite quando ela é falsa, sendo erros de primeiro e segundo tipo, respectivamente. O controle de ambos os erros de uma só vez é difícil dada a sua natureza associativa, e usualmente decide-se pela limitação de ocorrência do erro do primeiro tipo através da utilização do *nível de significância* α ([BEVANS, REBECCA, s.d.](#)).

A hipótese nesse teste pode tomar os mais diversos caminhos, a depender de quais avaliações deseja-se fazer acerca das amostras coletadas. No teste haverá sempre duas hipóteses base: a hipótese nula H_0 e a hipótese alternativa H_1 . A hipótese nula corresponde usualmente a afirmação de que exista uma relação entre as variáveis que se está interessado, que deseja-se verificar. Já a hipótese alternativa é geralmente a hipótese oposta à nula que é tomada como verdadeira caso a hipótese nula seja rejeitada. A rejeição da hipótese nula leva, conseqüentemente, a aceitação a hipótese alternativa com um risco de erro de α .

O teste de hipótese pode ser **paramétrico** ou **não-paramétrico** e a utilização de cada um está condicionado à quantidade de amostras com que se está trabalhando, bem como a distribuição da variável de estudo. Nos testes paramétricos alguns pressupostos são realizados, são eles:

- Amostras são independentes;
- Variâncias são homogêneas;
- Amostras seguem uma distribuição normal.

Já os testes não paramétricos correspondem àqueles em que não se pode presumir algum dos pressupostos citados. Nesse tipo de teste não são feitas grandes suposições sobre os dados, mas, por consequência, as inferências realizadas também não são tão fortes como as com os testes paramétricos, ou seja, não possuem o mesmo nível de confiança.

2.3.3.1 Teste de *Shapiro-Wilk*

O teste de *Shapiro-Wilk* foi apresentado por [Shapiro e Wilk \(1965\)](#) com o intuito de introduzir um procedimento de testagem de um conjunto de amostras quanto a distribuição normal. Portanto, a hipótese nula H_0 testada aqui é a de que os dados da amostra foram extraídos de uma distribuição normal. O teste retorna um valor p , que é utilizado para comparação com o nível de significância α estabelecido. Esse valor de α usualmente toma o

valor de 0.01 ou 0.05, a depender da confiança $(1-\alpha)$ desejada. Para que a hipótese nula seja aceita, o valor p deve ser maior que o α definido. Caso contrário, a hipótese H_0 é rejeitada.

$$\text{Hipóteses avaliadas} = \begin{cases} H_0 & : \text{Amostras tem distribuição normal} \\ H_1 & : \text{Amostras não tem distribuição normal} \end{cases} \quad (2.4)$$

2.3.3.2 Teste t de *Student*

Já o teste *t Student*, ou teste t, é um teste paramétrico proposto por [Student \(1908\)](#) em 1908 e é, até os dias atuais, um dos testes mais conhecidos e utilizados na estatística. O teste t pressupõe uma quantidade de amostras grande, considerada como tal se houver mais de 30 amostras, e tem como intuito verificar a diferença entre as médias de dois grupos de amostras. O teste t estima então a diferença verdadeira entre as médias dos dois grupos amostrais usando a razão da diferença das médias dos grupos sobre o erro padrão agrupado de ambos os grupos.

$$\text{Hipóteses avaliadas} = \begin{cases} H_0 & : \text{Média das duas amostras são estatisticamente iguais} \\ H_1 & : \text{Média das duas amostras são diferentes} \end{cases} \quad (2.5)$$

Existem três tipos de teste t básicos:

- Teste t de amostras independentes: que compara as médias de dois grupos de amostras diferentes e independentes;
- Teste t de amostras pareadas ou dependentes: que compara médias da mesma população;
- Teste t de uma amostra: que testa a média de um grupo amostral em relação a uma média determinada.

2.3.3.3 Teste de *Wilcoxon Rank Sum*

Muito utilizado como alternativa ao teste t independente, para quando alguma das pressuposições para utilização de teste paramétrico não é atendida, o teste de *Wilcoxon* é um dos testes não-paramétricos mais conhecidos, foi proposto por [Wilcoxon \(1945\)](#) e é também conhecido como *Wilcoxon signed-rank test*.

O teste segue a lógica de postos, ou *ranks*, onde os dados são ranqueados de acordo com a diferença entre as duas medidas pareadas de amostras independentes. Se as duas

amostras forem semelhantes, é esperado que exista pouca diferença entre os postos. Se houver uma diferença grande, haverá mais postos elevados.

$$\text{Hipóteses avaliadas} = \begin{cases} H_0 & : \text{Ambas amostras tem mesma distribuição} \\ H_1 & : \text{Amostras não possuem mesma distribuição} \end{cases} \quad (2.6)$$

3 Trabalhos Relacionados

Este capítulo apresenta uma revisão de trabalhos propostos na área de reconhecimento facial que buscaram entender e analisar as consequências de processos de envelhecimento e de cirurgias faciais nessa tarefa. Há uma divisão entre as duas subáreas de interesse, e busca-se entender o que há na literatura de concreto sobre os vieses associados a essas problemáticas e quais são as tendências esperadas para o resultado do presente trabalho. Ainda neste capítulo, serão abordados trabalhos específicos para a produção de *datasets*, voltados à pesquisa e ao desenvolvimento de modelos de reconhecimento facial.

3.1 Reconhecimento Facial e Envelhecimento

Desde o surgimento dos primeiros sistemas de reconhecimento facial, estudos sobre o impacto de fatores externos, como pose e iluminação, se tornaram recorrentes na comunidade (ZHAO, W. et al., 2003). Ainda no final do século vinte, os estudiosos começaram a perceber a importância de estudar também os impactos nesses sistemas de fatores internos e inerentes à natureza humana, como o envelhecimento.

No ano de 2004, o conjunto de imagens de envelhecimento *FGNET* foi publicado através do projeto *Face and Gesture Recognition Network* (FU et al., 2016). A *FGNET* reuniu 1002 imagens de 82 indivíduos, acompanhando-os desde à infância até idades de no máximo 69 anos. Esse foi um dos primeiros conjuntos de imagens de envelhecimento disponibilizado à comunidade de pesquisadores e ainda é muito utilizada em trabalhos atuais.

Em 2005, o trabalho de Ramanathan e Chellappa (2005) trouxe à comunidade um dos primeiros estudos de literatura na questão de envelhecimento facial. Trabalhos como o de Lanitis, Draganova e Christodoulou (2004) e o Lanitis, Taylor e Cootes (2002), pioneiros nessa área, foram citados no artigo com suas propostas de frameworks para simular efeitos de envelhecimento em imagens faciais.

Ainda no artigo de Ramanathan e Chellappa (2005), ele abordou a problemática da renovação de passaporte e o impacto da distância temporal entre o par de imagens na confiança da identificação do sujeito. Para realizar o estudo, eles criaram um conjunto de imagens retirado de passaportes com indivíduos entre 20 e 70 anos com espaçamento temporal variando de 1 à 10 anos entre pares de imagem. Além de disponibilizarem esse conjunto de imagens, que auxiliou muitos trabalhos posteriores na área, eles propuseram um classificador bayesiano de diferença de idade com minimização do impacto de pose e iluminação.

Um dos trabalhos que utilizaram o conjunto de imagens de passaporte para análise de

resultados foi realizado por [Ling et al. \(2007\)](#). Com o intuito de estudar o impacto de lacunas temporais em tarefas de reconhecimento facial, eles realizam a comparação de diferentes associações de descritores e classificadores de faces e propuseram um descritor utilizando *Gradient Orientation Pyramid*. De forma complementar no artigo, foi realizado um estudo empírico do efeito do processo de envelhecimento em algoritmos de reconhecimento.

Os resultados encontrados por [Ling et al. \(2007\)](#) mostraram uma saturação da dificuldade de reconhecimento quando a distância temporal entre pares de imagens ultrapassaram 4 anos, até uma distância de 10 anos, que é a máxima presente no conjunto de imagens utilizado pelos autores. Entretanto, a dificuldade associada ao processo de envelhecimento não ultrapassa a dificuldade associada aos fatores externos como pose e iluminação. Esse resultado qualitativo encontrado é crucial no estímulo de estudos quantitativos desse efeito ao longo dos anos.

Em 2006, [Ricanek e Tesafaye \(2006\)](#) publicaram o conjunto de imagens *MORPH* ([RICANEK, K., 2006](#)), que fornecia à comunidade um conjunto de imagens com distanciamento temporal que variava de alguns meses até mais de vinte anos. O intuito desse conjunto de imagens era fornecer aos pesquisadores material no qual eles pudessem aplicar seus estudos na área de envelhecimento, bem como utilizá-lo para estudo pelos próprios autores. Resultados encontrados pelos autores mostraram, de forma quantitativa, o impacto da progressão da idade nas taxas de reconhecimento, considerando gênero e raça dos indivíduos.

O trabalho realizado por [Yassin, Hoque e Deravi \(2013\)](#) utilizou do conjunto de imagens *MORPH* para estudar a sensibilidade de algoritmos de reconhecimento facial à idade. Eles realizaram um comparação de taxas de erros (*Equal Error Rates (EER)*) para 16 combinações diferentes de algoritmos e métricas de distância em um total de 100 indivíduos. Para permitir uma observação mais profunda dos impactos das idades nesses algoritmos, os indivíduos foram divididos em 5 faixas de idades diferentes. Ao final, eles chegaram a conclusão de que a taxa de erro foi maior para populações mais jovens e que os sistemas de detecção tem uma melhora significativa após a divisão entre grupos, permitindo que os algoritmos se adaptem à uma variação menor de idade.

Com a popularização das redes de aprendizagem profunda, pesquisadores começaram a estudar formas de utilizar das vantagens associadas as redes neurais convolucionais para tornar sistemas de reconhecimento facial mais robustos ao envelhecimento. Com a justificativa de impacto direto de intervalos temporais na segurança de sistemas biométricos de reconhecimento facial, [Hachim El Khiyari \(2016\)](#) trouxe à comunidade acadêmica um estudo acerca da utilização de *CNNs* para melhorá-los. A ideia proposta pelos autores é utilizar das *CNNs* para extrair características faciais que sejam mais robustas em relação à variação temporal.

Ao realizar o estudo, [Hachim El Khiyari \(2016\)](#) utilizou a arquitetura de rede *VGG-Face* como extrator de características e fizeram o experimento com diferentes métodos

de classificação. Realizando o experimento nas bases de imagens *FGNET* e *MORPH*, os autores compararam os resultados obtidos com a performance de outras abordagens de reconhecimento nas mesmas bases. A conclusão final é que o desempenho das redes como extratores de características é comparável aos métodos de última geração com resultados de até 90% de taxa de reconhecimento.

Em uma análise com mais algoritmos de aprendizagem profunda para extração de característica sendo testadas, os autores [Boussaad e Boucetta \(2022\)](#) utilizaram cinco redes neurais convolucionais pré-treinadas para seus estudos, sendo elas: *Alex-Net*, *GoogleNet*, *Inception V3*, *ResNet50* e *SqueezeNet*. As diferentes redes foram associadas com diferentes métodos de classificação, gerando ao todo 15 combinações diferentes que foram testadas no conjunto de imagens *FGNET*. Os resultados encontrados pelos autores são promissores, chegando a uma taxa de acurácia de 98.21% com a utilização da rede *AlexNet* juntamente com o classificador *SVM*. Eles chegam então a conclusão de que a utilização de redes pré-treinadas permite uma performance muito satisfatória no que diz respeito à busca de sistemas de reconhecimento facial invariante quanto ao envelhecimento.

3.2 Reconhecimento Facial e Procedimentos Estéticos

Como já comentado anteriormente, aspectos como pose, iluminação, expressão e idade são grandes desafios para o reconhecimento facial. Outro aspecto que tem sido fortemente explorado em trabalhos recentes é como procedimentos estéticos e cirurgias plásticas influenciam nessa tarefa ([MOEINI, A.; FAEZ; MOEINI, H., 2015](#)). Há também trabalhos como o de [Kohli, Yadav e Noore \(2015\)](#), o qual propõe um algoritmo especializado em reconhecer tais alterações estéticas. Excelentes resultados, como os apresentados por esse trabalho evidenciam como a presença de intervenções estéticas alteram a face, a ponto de poder-se distinguir pessoas que passaram por tais procedimentos.

Um dos trabalhos iniciais no estudo da interferência de procedimentos estéticos no reconhecimento facial, feito por [R. Singh, Vatsa e Noore \(2009a\)](#), apresentou conclusões as quais indicam que algoritmos de reconhecimento facial usados em sistemas comerciais da época, ou algoritmos que apresentaram bons resultados para cenários desafiadores, não foram capazes de mitigar variações causadas por procedimentos estéticos. Os algoritmos estudados utilizavam Principle Component Analysis (PCA) ([BELHUMEUR; HESPANHA; KRIEGMAN, 1997](#)), Fisher Discriminant Analysis (FDA) ([BELHUMEUR; HESPANHA; KRIEGMAN, 1997](#)), Geometric Features (GF) ([COX; GHOSN; YIANILOS, 1996](#)), Local Feature Analysis (LFA) ([PENEV; ATICK, 1996](#)), Local Binary Pattern (LBP) ([AHONEN; HADID; PIETIKAINEN, 2006](#)) e Neural Network Architecture based 2D Log Polar Gabor Transform (GNN) ([SINGH, R.; VATSA; NOORE, 2009b](#)) e, ao serem testados em uma base, gerada por [R. Singh, Vatsa e Noore \(2009a\)](#), com 506 indivíduos que realizaram cirurgias

plásticas, esses algoritmos obtiveram performances insatisfatórias ao compararem fotos antes e depois das operações.

Os mesmos autores, em (SINGH, R.; VATSA; BHATT et al., 2010) fizeram análises analíticas e experimentais para buscar identificar se tais modelos eram robustos às operações estéticas. Foi percebido que, em geral, os algoritmos reduziram a sua acurácia de 26 a 30% quando testados na base de imagens com cirurgia, em comparação com uma base sem operações. Esta base sem intervenções cirúrgicas foi gerada com 1800 fotos frontais, em expressão neutra, de cinco diferentes bases disponíveis publicamente. Esse estudo ainda possibilitou inferir que intervenções nas orelhas são as que menos afetam o reconhecimento facial, enquanto operações em regiões como nariz, bochecha, lábios, testa e pálpebras alteram significativamente os resultados dessa tarefa, por serem importantes na maioria das características locais utilizadas na etapa de representação do pipeline de reconhecimento facial.

Já em 2013, Bhatt et al. (2013) propôs um algoritmo para identificar sujeitos, a partir de fotos antes e depois de cirurgias plásticas. Tal algoritmo primeiro gera grânulos de face não disjuntos em três níveis de granularidade. Esses grânulos representam informações distintas da imagem, em diferentes tamanhos e resoluções. Em seguida, Bhatt et al. (2013) utiliza uma abordagem genética multiobjetiva para otimizar ao mesmo tempo a seleção do extrator de características e o peso dado a cada grânulo selecionado. Foram usados dois extractores de características diferentes, o Extended Uniform Circular Local Binary Pattern (EUCLBP) (BHATT et al., 2010) e o Scale Invariant Feature Transform (SIFT) (LOWEDAVID, 2004), entre os quais o algoritmo permite alternar, de forma que melhora a seleção de informações discriminatórias dos grânulos. O algoritmo foi testado nas duas bases usadas em R. Singh, Vatsa, Bhatt et al. (2010) e obteve resultados que superaram algoritmos de estado da arte na robustez a operações estéticas.

Em 2015, M. Marsico et al. (2015) comparou abordagens locais e globais de reconhecimento facial para abordar o problema de cirurgias plásticas nessa tarefa. Além disso, M. Marsico et al. (2015) propõe uma metodologia com o pressuposto de que um melhor desempenho ao comparar regiões menos modificadas por cirurgias pode contrabalancear variações drásticas em outras regiões e, assim, manter a qualidade do reconhecimento facial. Aplicando essa ideia junto a algoritmos já existentes, os autores obtiveram bons resultados. Alcançou-se uma taxa de reconhecimento de mais de 86%, ainda sem a necessidade de treinamento e sem precisar de imagens externas ou de mais imagens por sujeito.

No mesmo ano A. Moeini, Faez e H. Moeini (2015) propuseram uma nova abordagem para a extração de características da face, visando melhorar resultados de reconhecimento facial, quando há procedimentos estéticos e maquiagem. A ideia foi adicionar, além de características de textura facial, características faciais profundas, obtidas individualmente da reconstrução 3D da imagem, e as quais não possuem influência de cirurgias ou cosméticos.

Ao testar essa abordagem na base de dados gerada em [R. Singh, Vatsa, Bhatt et al. \(2010\)](#) e nas bases YouTube Makeup e Virtual Makeup [Dantcheva, C. Chen e Ross \(2012\)](#), foram obtidos resultados que superaram métodos de estado da arte, os quais consideravam apenas *texture images*.

Ainda em 2015, [Kohli, Yadav e Noore \(2015\)](#) apresentaram uma nova abordagem para diferenciar faces com e sem cirurgias. O algoritmo proposto aprende representações de cirurgias locais e globais na face, usando "Multiple Projective Dictionaries" e descritores binários de face compactos. Diferentemente de abordagens de aprendizado por dicionário tradicionais, [Kohli, Yadav e Noore \(2015\)](#) não utilizam as normas l_0 e l_1 , que são computacionalmente custosas. Já para verificar a identidade da pessoa, são utilizadas Regions of Interest (ROI), as quais são comumente alteradas por cirurgias, juntamente com pontuações de similaridade entre as características da face consideradas. Ao ser testado na base de imagens de [\(SINGH, R.; VATSA; BHATT et al., 2010\)](#), o algoritmo obteve um resultado de 97,96% de acurácia em detectar faces com e sem cirurgias.

Um novo algoritmo que visa mitigar interferências de cirurgias plásticas no reconhecimento facial foi proposto por [Aziz et al. \(2016\)](#). Nele, foram fundidos os resultados de dois algoritmos extratores de características. É primeiro aplicado o LBP ([AHONEN; HADID; PIETIKAINEN, 2006](#)) sobre um conjunto de pontos chave, denominados pelo autor como pontos silenciosos. Essas regiões foram definidas considerando-se que apenas algumas características, que possuem certo significado, serão úteis para o reconhecimento facial, em casos em que houve cirurgias. Além do LBP, utiliza-se o descritor de características global GIST ([OLIVA; TORRALBA, 2001](#)), e então, une-se os dois conjuntos de características obtidos. O algoritmo proposto, quando testado na base de [\(SINGH, R.; VATSA; BHATT et al., 2010\)](#), superou alguns sistemas de estado da arte e conseguiu uma acurácia de mais de 91% no reconhecimento facial após cirurgias. Além disso, teve-se uma performance estável para os diferentes procedimentos testados.

[Rathgeb, Dantcheva e Busch \(2019\)](#) faz uma rica pesquisa sobre como diferentes formas de embelezamento podem influenciar no reconhecimento facial. Nesta publicação, são diferenciados três tipos de embelezamento, sendo eles: Cirurgia Plástica Facial, em que mudanças são induzidas medicamente, visando corrigir características da face, por motivações de melhoria na aparência; Cosméticos Faciais, os quais proporcionam modificações não permanentes, como, por exemplo, maquiagens; e Retoque Facial, em que alterações são feitas no domínio digital, como com a inclusão de filtros prontos de aplicativos, ou por modificações específicas em ferramentas como Photoshop. Em seguida, é feita uma revisão de trabalhos que investigam a influência de cada um desses tipos de embelezamento no reconhecimento facial. Por fim, são discutidos os aspectos que continuam a ser desafiadores na pesquisa de impactos do embelezamento reconhecimento facial.

Em termos de embelezamentos por cirurgias plásticas, foi constatado que a estratégia

de alguns modelos de reconhecimento facial, que visam ser robustos a estes procedimentos, é considerar extrair apenas características as quais são menos afetadas por cirurgias, ou as considerar com mais peso que as demais. Além disso, há também propostas as quais fundem múltiplos algoritmos, visando aumentar a quantidade de informações extraídas das imagens, por complementação dos resultados individuais de cada algoritmos. Ambas as estratégias foram observadas nos trabalhos apresentados acima. Foi também concluído por [Rathgeb, Dantcheva e Busch \(2019\)](#) que é notável a melhora nos últimos anos da performance de algoritmos de reconhecimento facial, mesmo na presença de cirurgias plásticas.

Em um artigo mais recente, [Rathgeb, Dogan, Stockhardt, De Marsico et al. \(2020\)](#) expõe que muitos trabalhos sugerem que geralmente se tem uma queda de performance nos algoritmos de reconhecimento facial, quando se tem cirurgia plástica. Entretanto, [Rathgeb, Dogan, Stockhardt, De Marsico et al. \(2020\)](#) notou que a base de imagens usada pela grande maioria das pesquisas, a IIITD, de [\(SINGH, R.; VATSA; BHATT et al., 2010\)](#), é disponibilizada como uma coleção de links da web, que, por vezes já não estão mais disponíveis. Além disso, por meio de trabalhos de outros pesquisadores, soube que há imagens duplicadas e de baixa qualidade nessa base. Apesar de tais características serem parte de cenários da vida real, o uso dessa base para o estudo específico da influência de cirurgias plásticas pode levar a conclusões erradas.

Por esse motivo, [Rathgeb, Dogan, Stockhardt, De Marsico et al. \(2020\)](#) propõe uma nova base de imagens, com 1276 fotos de 638 sujeitos, os quais fizeram ao menos um procedimento estético. Todas as imagens seguem os prerrequisitos da Organização da Aviação Civil Internacional (OACI), para fotos aceitas em passaportes, de forma a garantir uma qualidade nas amostras da base. Testando-se o sistema comercial Cognitec FaceVACS e o algoritmo de código aberto ArcFace com a nova base de imagens proposta, foi constatado que a maioria dos procedimentos cirúrgicos estéticos mais comuns não causam erros significativos nesses sistemas de reconhecimento facial com aprendizado profundo. Ainda, para os dois sistemas testados, foi concluído que correções em ossos da face e nas sobrancelhas são as alterações mais impactantes para a acurácia do reconhecimento facial.

3.3 Elaboração do Conjunto de Dados para Reconhecimento Facial

Ao longo do desenvolvimento de pesquisas em reconhecimento facial, vários conjuntos de dados foram criados, para que fossem utilizados em testes de algoritmos e também como *benchmarks*, para comparação de resultados de diferentes sistemas. Os principais deles foram abordados na Seção de Fundamentação Teórica 2.2.1.

Sabe-se que, para implementações de aprendizagem profunda utilizando redes convolucionais, é necessário um grande conjunto de imagens de treinamento. Dessa forma, o

sistema consegue generalizar para dados encontrados na aplicação real, não pertencentes a esse conjunto, e proporcionar resultados aceitáveis. Além disso, quando trata-se de estudos na área de reconhecimento facial, é preciso testar a abordagem estudada em conjuntos de imagens com variados casos de teste, visando conhecer o desempenho do sistema em contextos gerais.

Alguns estudos se preocuparam não somente em testar os modelos que propuseram, mas também em construir novos conjuntos de imagens, para validar os requisitos propostos. Esse foi o caso do trabalho realizado por [Parkhi, Vedaldi e Zisserman \(2015\)](#), que buscou estudar o aprendizado para a tarefa de reconhecimento facial utilizando rede neural convolucional (CNN) e abordou também um passo-a-passo de criação de conjunto de imagens para treinamento em grande escala.

No artigo de [Parkhi, Vedaldi e Zisserman \(2015\)](#), a intenção é desenvolver um conjunto de dados significativamente grande, contendo fotos disponíveis na internet, e com trabalho minimizado de anotação manual. Para isso, a criação de conjunto de imagens é dividida em 5 estágios: *Listagem e filtragem inicial de candidatos*, os quais são figuras públicas ou celebridades, para evitar problemas de privacidade, *Coleta de imagens para cada identidade*, incluindo diferentes formas de pesquisa de fotos, *Filtragem automática*, removendo fotos indesejadas escolhidas automaticamente, utilizando um classificador, *Remoção de imagens duplicadas ou muito semelhantes*, também de forma automática, e *Filtragem manual final*.

Ainda na temática de gerar base de imagens a partir de fotos coletadas da internet, a publicação de [Erjin Zhou, Zhimin Cao e Yin \(2015\)](#) apresenta prós e contras dessa abordagem. Enquanto essa estratégia facilita a obtenção de grandes quantidades de dados para treinamento, que de fato melhoram o desempenho de algoritmos de reconhecimento facial, há o chamado *Long Tail Effect*, ou *efeito de cauda longa*, em que alguns indivíduos possuem muitas fotos, enquanto vários possuem poucas instâncias. Os testes do estudo de [Erjin Zhou, Zhimin Cao e Yin \(2015\)](#) indicam que incluir indivíduos com poucas imagens no treinamento atrapalha o desempenho do sistema, ao invés de melhorar seus resultados.

A primeira publicação que apresentou um conjunto de imagens publicamente disponível contendo retratos antes e depois de cirurgias plásticas faciais foi a de [R. Singh, Vatsa, Bhatt et al. \(2010\)](#). Foi proposto, neste artigo, um conjunto de 1800 imagens de 900 sujeitos. Para desenvolver a base de dados, [R. Singh, Vatsa, Bhatt et al. \(2010\)](#) inspiraram-se nos procedimentos propostos por [Kumar et al. \(2009\)](#), que publicou a base de imagens chamada de Pubfig, contendo fotografias obtidas da internet. As fotos então são de ambiente não controlado, embora uma filtragem manual tenha sido feita, para retirar imagens em que parte do rosto tenha sido omitido, além de selecionar fotos em posição frontal, com iluminação suficiente e expressão neutra.

Já em 2020, [Rathgeb, Dogan, Stockhardt, De Marsico et al. \(2020\)](#) propuseram uma nova base de imagens contendo fotos antes e depois das cinco mais frequentes cirurgias

plásticas, chamada de *Hochschule Darmstadt (HDA)*. Neste caso, também utilizou-se fotos publicadas na internet, entretanto, com restrições mais rígidas, conforme propostas pela Organização da Aviação Civil Internacional (OACI), de limitar à pose frontal, expressão neutra e distância inter-ocular mínima suficiente. Essa abordagem objetivou um conjunto de dados que abordasse a problemática das cirurgias plásticas no reconhecimento facial, minimizando interferências de outros possíveis desafios, como pose e expressão.

4 Metodologia

A metodologia proposta segue uma ideia sequencial, na qual as etapas são definidas conforme o fluxograma da Figura 4.24. A seguir, serão explicadas cada uma das etapas envolvidas, para atingir os objetivos pretendidos neste trabalho.

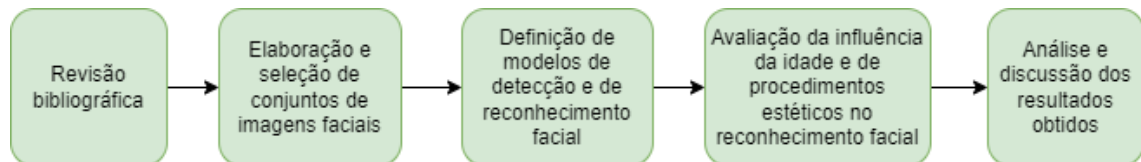


Figura 4.24 – Esquemático da metodologia seguida.

4.1 Revisão bibliográfica

O primeiro passo foi realizar a revisão de conceitos fundamentais relacionados à temática estudada, de forma que os procedimentos a serem seguidos pudessem ser melhor definidos, e os resultados coerentemente analisados. Os conteúdos revisitados estão explicados no Capítulo 2.

Ainda, visando a compreender as conclusões já publicadas sobre o reconhecimento facial e fatores que podem o influenciar, uma revisão de trabalhos semelhantes foi feita. Artigos e outros tipos de trabalhos científicos sobre o tema estudado foram sumarizados no Capítulo 3.

4.2 Elaboração e seleção de conjuntos de imagens faciais

Para realizar o estudo acerca do impacto de modificações estéticas, realizadas no rosto, no reconhecimento facial com aprendizagem profunda, se faz necessário entender também os impactos associados ao processo natural de envelhecimento. A associação ou não desses dois elementos ao desempenho dos algoritmos de reconhecimento facial precisa ser estudada de forma individual e também conjunta, para ambos os fatores, de forma que os resultados obtidos consigam exprimir a realidade.

4.2.1 Conjunto de dados de faces com cirurgias plásticas: HDA

Para entender o impacto direto das modificações faciais cirúrgicas no reconhecimento facial, procurou-se bancos de imagens que contivessem imagens do tipo "antes e

depois" de indivíduos que realizam esses procedimentos. Com imagens tiradas em ambiente controlado antes e após a cirurgia, é possível mitigar fatores externos, bem como o processo de envelhecimento, permitindo que o foco de análise dos resultados seja voltado unicamente para o processo de modificações estéticas.

O banco de dados *Hochschule Darmstadt* (RATHGEB; DOGAN; STOCKHARDT; MARSICO, M. D. et al., 2020) foi escolhido pois apresenta imagens tiradas antes e depois de cinco das mais populares cirurgias faciais: correção de sobrancelhas, correção palpebral, *facelift*, correção dos ossos faciais e correção de nariz. A base conta com 638 sujeitos ao todo distribuídos nas cirurgias citadas com um par de imagem por sujeito. Nas imagens da Figura 4.25 é possível verificar o exemplo de cada um dos tipos de cirurgia facial presentes no banco de imagens.

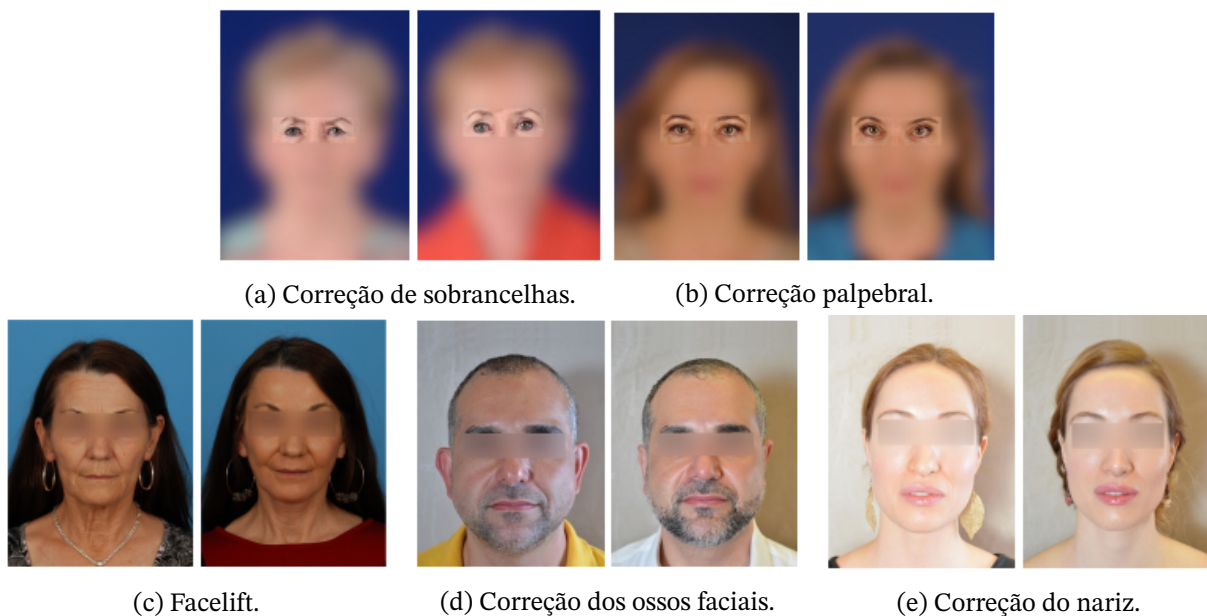


Figura 4.25 – Exemplos de imagens da HDA, para cada cirurgia contida na base. As imagens foram parcialmente desfocadas, de forma a preservar as identidades dos indivíduos. Fonte: adaptação da Figura 2 de Rathgeb, Dogan, Stockhardt, Maria De Marsico et al. (2020).

Cabe ressaltar que durante o processo de utilização do banco de imagens HDA, algumas alterações precisaram ser realizadas. São elas:

- Contatou-se o autor da base de imagens, e ele confirmou que os pares relativos ao sujeito 40 da pasta '*Facial Bones*' e ao sujeito 154 da pasta '*Nose*' são formados por pessoas diferentes, e que possivelmente serão removidos de uma futura atualização do conjunto de dados. Sendo assim, ambos os pares foram desconsiderados das análises deste trabalho.
- Visando a facilitar os códigos desenvolvidos para testes, os nomes dos arquivos '*a_a*', '*a_b*', '*b_a*' e '*b_a*' das pastas '*Facelift*' e '*FacialBones*' foram alterados, respectivamente,

para '103_a', '103_b', '104_a' e '104_b' e para '109_a', '109_b', '110_a' e '110_b', de forma a dar continuidade a numeração de cada pasta.

4.2.2 Conjunto de dados com envelhecimento: FGNET

Já com o intuito de compreender a influência das modificações faciais decorrentes somente da idade, buscou-se conjuntos de dados com fotos de diversos sujeitos, mas de forma que um mesmo indivíduo seja apresentado em diferentes idades, ao longo de vida. Além disso, deseja-se que a probabilidade de que esses sujeitos tenham feito alterações estéticas permanentes, cirúrgicas ou não, seja baixa. Dessa forma, entende-se que os resultados obtidos com esse banco de imagens podem ser associados diretamente ao envelhecimento.

Para esse caso, foram consideradas algumas opções. A primeira foi a amplamente conhecida base AgeDB ([MOSCHOLOU et al., 2017](#)). Entretanto, apesar de esta base ser aberta para pesquisa, ela é disponibilizada como uma pasta para *download* com acesso aos arquivos restrito por senha, que deve ser solicitada ao autor. Tentou-se contato com o proprietário, mas não foi obtida resposta.

Com isso, encontrou-se duas novas possibilidades as quais poderiam servir de referência para a questão do envelhecimento: a base MORPH ([RICANEK, K., 2006](#)) e a FGNET ([FU et al., 2016](#)). Enquanto a primeira possui um número elevado de sujeitos, atualmente com aproximadamente 70 mil em sua versão comercial, a versão gratuita da base apresenta poucas instâncias por indivíduo, que somam pouco mais de 13 mil. Já a segunda possui apenas 82 sujeitos, mas contém mais fotos por pessoa, com uma média de 12 imagens por sujeito. Dessa forma, optou-se por utilizar a FGNET nos testes relativos à influência da idade no reconhecimento facial, considerando-se que abranger um longo período da vida do indivíduo deve proporcionar resultados mais claros acerca do impacto do envelhecimento. É possível verificar na Figura 4.26 um exemplo de imagens ao longo da vida do sujeito 001 da base.



Figura 4.26 – Exemplo de imagens de um sujeito na FGNET. Fonte: [Chang, Chu-Song Chen e Hung \(2011\)](#).

Esse conjunto de imagens é disponibilizado com arquivos nomeados no formato *XXXAII*, em que *XXX* é o número atribuído aquele sujeito e *II* corresponde à idade do indivíduo naquela foto. Ainda, cabe ressaltar que a FGNET possui, para algumas pessoas, mais de uma foto em uma dada idade. Sendo assim, há imagens nomeadas como *XXXAIIa* e *XXXAIIb*, com o mesmo *XXX* e *II*, indicando que pertencem ao mesmo sujeito na mesma

idade. Devido à forma como serão utilizadas as fotos nos testes do presente estudo, para casos como esse, foram sempre mantidas as fotos *XXXAIIb* e excluídas as imagens *XXXAIIa*, mantendo-se uma única foto por idade do sujeito.

4.2.3 Desenvolvimento do conjunto de imagens proposto: C2FPW

Foram pesquisados bancos de imagens que contivessem indivíduos que sofreram modificações faciais e apresentassem quantidades significativas de imagens ao longo de toda sua vida, para que o processo de envelhecimento fosse analisado de forma conjunta às modificações estéticas. Entretanto, não encontrou-se um banco publicamente disponível que cumprisse com esses requisitos. Dessa forma, propôs-se elaborar um conjunto de dados com fotos disponibilizadas na internet, contendo retratos, em diferentes idades, de famosos e figuras públicas que realizaram intervenções no rosto. O nome escolhido para essa base é *Clinical Cosmetic Facial Procedures in the Wild* (C2FPW).

4.2.3.1 Pesquisa de famosos que realizaram procedimentos estéticos

Seguindo os passos iniciais propostos em [Parkhi, Vedaldi e Zisserman \(2015\)](#), de propor uma lista inicial de nomes de pessoas candidatas, e em seguida remover aquelas as quais não possuem um número suficiente de imagens distintas, primeiro produziu-se uma lista contendo 100 indivíduos. Estes são figuras públicas que admitiram ter realizado procedimentos estéticos faciais, cirúrgicos ou não, ou então famosos os quais um especialista na área julgou ter feito alguma modificação facial clínica.

4.2.3.2 Coleta de fotos publicamente disponíveis

Em seguida, pesquisando em sites que fornecem algum indício acerca da data da imagem, coletou-se fotos consideradas de qualidade aceitável, a critério humano, da maior quantidade de anos possível da vida dos sujeitos selecionados. Isso para que o envelhecimento também pudesse ser avaliado nessa base de imagens. Foram eleitos 90 dos 100 indivíduos escolhidos no passo anterior, sendo o critério de seleção a maior quantidade de fotos por sujeito possível.

4.2.4 Escolha de um *benchmark* para comparação dos resultados obtidos

Propôs-se escolher um conjunto de imagens amplamente utilizado como *benchmark*, visando a justificar o estudo realizado, bem como melhor interpretar os resultados obtidos. Como comumente publica-se a acurácia na LFW ([HUANG; RAMESH et al., 2007](#)), ao apresentar um novo modelo de reconhecimento facial, esse foi o *benchmark* escolhido. Além disso, outro fator que propiciou essa decisão foi o fato de essa base de imagens conter poucas fotos por sujeito, mas muitos sujeitos.

Considera-se ainda que a probabilidade de que os indivíduos que a compõe terem realizado modificações faciais voluntárias seja baixa. Sendo assim, ela permite uma comparação direta de seus resultados com os da base HDA, que contém procedimentos estéticos, e apenas duas fotos por sujeito.

Definida a escolha da LFW, serão considerados neste estudo, para os modelos de reconhecimento facial em análise, tanto os valores publicados de acurácia com esse *benchmark*, como os resultados de sensibilidade calculados no presente trabalho, conforme será explicado na Seção 4.4.1.2.

4.3 Definição de modelos de detecção e de reconhecimento facial

4.3.1 Avaliação dos modelos do *Deepface*

Serão analisados neste estudo os modelos já treinados do *Deepface* (SERENGIL; OZPINAR, 2020), os quais estão individualmente descritos na Seção 2.2.3: VGG-Face, FaceNet, FaceNet512, OpenFace, DeepFace, DeepID, ArcFace, Dlib e SFace. Sendo assim, foi necessário inicialmente verificar acerca do uso correto das funções disponibilizadas pela biblioteca, bem como se os parâmetros retornados por elas serão suficientes para as avaliações desejadas. Utilizou-se a versão 0.0.75 do *Deepface*, com o Python versão 3.8.10.

4.3.2 Escolha do modelo de detecção que será usado

O foco principal do presente trabalho é definir como modelos de reconhecimento facial são afetados por modificações estéticas. Entretanto, como foi explicado em seções anteriores, há uma etapa de detecção que antecede o reconhecimento. Sendo assim, é ideal que a etapa de detecção tenha sido realizada com sucesso, para a análise específica da etapa de reconhecimento facial. Dentre os modelos disponibilizados no pacote *Deepface* há 5 modelos de detecção: SSD, RetinaFace, OpenCV, MTCNN e Dlib.

Visto que o intuito do trabalho não é analisar o impacto desses detectores no reconhecimento facial, optou-se por selecionar o que possuísse a melhor performance na detecção, para que não fosse preciso estudar diferentes associações de modelos de detecção e de reconhecimento. O critério utilizado para definir qual o modelo selecionado é a porcentagem das fotos que foram corretamente detectadas por cada um na base desenvolvida C2FPW, bem como na base de imagens FGNET.

4.3.3 Definição da métrica de distância que será usada

Como explicado na Seção 2.2.6.1, a etapa de classificação dos reconhecedores faciais realizam um cálculo de similaridade entre os vetores de representação de duas faces para classificar as faces como precentoras ou não da mesma pessoa. A forma como essa similaridade é calculada é um dos parâmetros que o *DeepFace* possibilita escolha, e ele disponibiliza três métricas diferentes: distancia de cossenos, distância euclidiana e a *L2-Norm*.

Para os estudos aqui realizados, optou-se por seguir com a distância de cossenos como a métrica de distância, por ser uma das mais comumente utilizadas e ser a escolha *default* do *framework DeepFace*. Além disso, para os valores de *threshold*, ou limite de detecção, foram utilizados os valores também disponibilizados pelo *DeepFace*. Os valores para o cálculo de distância por cosseno podem ser encontrados na Tabela 4.2.

Modelo de Reconhecimento	Threshold
VGG-Face	0,40
FaceNet	0,40
FaceNet512	0,30
OpenFace	0,10
DeepFace	0,23
DeepID	0,015
ArcFace	0,68
Dlib	0,07
SFace	0,593

Tabela 4.2 – Tabela dos *thresholds* dos diferentes modelos do *DeepFace* para a distância por cossenos.

4.4 Avaliação da influência da idade e de procedimentos estéticos no reconhecimento facial

Para a avaliação da influência de procedimentos estéticos no reconhecimento facial de forma simultânea ao envelhecimento, serão implementadas diferentes metodologias de testes. Cada proposição possui uma finalidade específica e fará uso de dois conjuntos de imagens. Um resumo das análises que serão realizadas, bem como o intuito da proposta e as bases utilizadas em cada uma pode ser verificada na Figura 4.27.

4.4.1 Avaliando separadamente a influência de procedimentos estéticos

Para esse estudo, é preciso entender inicialmente o quantitativo direto da influência de procedimentos estéticos nos modelos de reconhecimento facial com aprendizagem profunda. Dessa forma, buscou-se primeiro entender esse impacto de forma isolada da questão do envelhecimento e, para isso, utilizou-se o banco de imagens *HDA*.

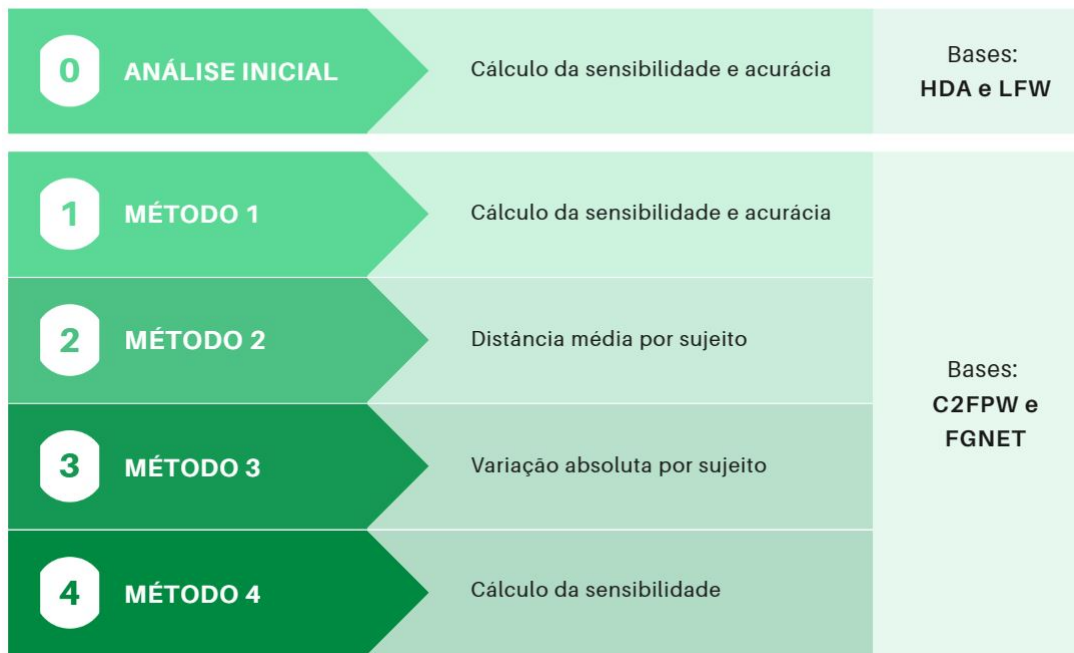


Figura 4.27 – Métodos propostos para a avaliação da influência de procedimentos estéticos no reconhecimento facial.

4.4.1.1 Obtenção da acurácia na HDA, para cada algoritmo de reconhecimento estudado

As imagens de antes e depois de indivíduos que realizaram cirurgias faciais famosas, presentes no banco de imagens *HDA*, foram utilizadas como um estimador direto do impacto das modificações nos sistemas de reconhecimento. Com esse banco de imagens, as funções de reconhecimento dos modelos do *DeepFace* puderam ser testadas e as distâncias entre as imagens dos mesmos indivíduos comparadas sem o impacto do envelhecimento, visto as fotos serem próximas temporalmente.

- Porcentagem de reconhecimento de pares positivos da HDA

Para essa análise serão avaliados todos os possíveis "pares positivos" da HDA, formados por fotos antes e depois das cirurgias, totalizando 637 pares. Considera-se como "pares positivos" aqueles compostos por fotos de um mesmo sujeito. Deseja-se, portanto, verificar quantas duplas de fotos foram corretamente analisados como pertencentes ao mesmo sujeito, pelos algoritmos estudados. É possível entender por esses resultados quais os modelos que possuem melhor performance, para a tarefa de reconhecimento, com presença de procedimentos estéticos.

- Porcentagem de reconhecimento de pares negativos da HDA

Em seguida, foram analisados "pares negativos" na HDA, que consistem em pares de imagens de pessoas distintas. A mesma quantidade de "pares positivos", avaliados no passo anterior, foi selecionada. Foram aleatoriamente escolhidos então 637 pares, compostos por quaisquer duas fotos da HDA, garantindo que nenhum par fosse formado pelas duas fotos do mesmo sujeito. Os 637 pares foram então avaliados pelos nove modelos de reconhecimento,

e os resultados foram armazenados para análise posterior.

Considerando as imagens apresentadas na Figura 4.25, pode-se gerar como exemplos de pares positivos e negativos na HDA os conjuntos das Figuras 4.28a e 4.28b, respectivamente.



Figura 4.28 – Exemplos de pares positivos e negativos na HDA. Fonte: adaptação da Figura 2 de Rathgeb, Dogan, Stockhardt, Maria De Marsico et al. (2020).

- Cálculo da acurácia na HDA

Por fim, com os resultados dos modelos tanto para os pares positivos, como para os negativos, a acurácia foi obtida, por meio da relação apresentada em Kulkarni, Chong e Batarseh (2020):

$$\text{Acurácia} = \frac{VN + VP}{VN + FP + FN + VP}, \quad (4.1)$$

em que,

"VN" refere-se a *Verdadeiro Negativo*, remetendo às imagens de pessoas diferentes que foram corretamente identificadas pelo modelo como tal. "VP" refere-se a *Verdadeiro Positivo*, representando os pares corretamente identificados pelo modelo como imagens da mesma pessoa. "FN" e "FP" são, respectivamente, *Falso Negativo* e *Falso Positivo*, correspondendo aos pares de fotos de pessoas diferentes ou da mesma pessoa, cuja resposta do modelo foi

contrária ao esperado. A Tabela 4.3 resume essas denominações e é conhecida como *Matriz de Confusão*.

		Resultado Esperado	
		Sujeitos diferentes	Mesmo sujeito
Resultado Obtido	Sujeitos diferentes	VN	FN
	Mesmo sujeito	FP	VP

Tabela 4.3 – Matriz de Confusão.

Com esses valores da matriz de confusão determinados, foi possível encontrar também algumas outras métricas relevantes à avaliação dos resultados. A sensibilidade, ou *recall*, estabelecida pela Equação 4.2 vai avaliar o quão bem o modelo consegue detectar os resultados positivos, ou seja, os pares positivos. Já a especificidade, abordada na Equação 4.3 vai ajudar a avaliação da capacidade do modelo em não reconhecer pares negativos da maneira esperada.

$$\text{Sensibilidade} = \frac{VP}{FN + VP}, \quad (4.2)$$

$$\text{Especificidade} = \frac{VN}{VN + FP}. \quad (4.3)$$

Com o objetivo de compreender o impacto dos procedimentos estéticos no reconhecimento facial, é coerente concentrar-se exclusivamente no desempenho dos algoritmos ao comparar imagens do mesmo sujeito, ou seja, avaliar a sensibilidade (Equação 4.2). O objetivo desse estudo é investigar se a presença dessas modificações dificulta ou facilita o reconhecimento de um determinado indivíduo. Portanto, não se faz necessário a análise de pares negativos de imagens ou concluir se os sujeitos tornam-se mais semelhantes ou diferentes na presença de alterações voluntárias. Isso seria escopo de um outro estudo. Portanto, a sensibilidade, a especificidade e a acurácia serão apresentadas nos resultados, mas, para obter conclusões, apenas a sensibilidade será considerada.

4.4.1.2 Cálculo da sensibilidade no benchmark LFW, para cada algoritmo de reconhecimento facial estudado

Juntamente com os resultados obtidos no banco de imagens *HDA*, para embasar a análise a ser realizada com a associação da influência da idade e de procedimentos estéticos, buscou-se outros resultados obtidos com os modelos de reconhecimento do *DeepFace* que conseguissem entregar resultados consideráveis, mesmo que de forma qualitativa acerca da questão do impacto dos procedimentos estéticos em outros bancos de imagens públicos.

Como comumente publica-se a acurácia na LFW (HUANG; RAMESH et al., 2007), ao apresentar um novo modelo de reconhecimento facial, considerou-se a acurácia nesse

benchmark, para referências nesse estudo. Entretanto, como explicado na Seção 4.4.1.1, para este trabalho é mais coerente a análise apenas da Sensibilidade. Portanto, decidiu-se calcular a sensibilidade dos nove modelos em análise no *benchmark* LFW. Com esse resultado, comparou-se o desempenho dos algoritmos na LFW e na HDA.

Para o cálculo da sensibilidade, a versão da LFW com imagens alinhadas por afinamento profundo (HUANG; MATTAR et al., 2012) foi obtida, e realizou-se o teste de reconhecimento com os nove modelos estudados, considerando um par por sujeito da base. Para sujeitos que possuem mais de duas imagens na LFW, selecionou-se aleatoriamente um par de imagens para o teste de reconhecimento. Já para sujeitos os quais apresentavam uma única foto no conjunto de imagens, por falta de dados, não fez-se nenhum teste de reconhecimento, visto que o intuito é trabalhar os pares positivos no cálculo da sensibilidade.

4.4.2 Incluindo a influência do envelhecimento

Inicialmente, foram estabelecidos os métodos a serem empregados no estudo, a fim de possibilitar a análise e interpretação de resultados diversos. Foram escolhidas quatro abordagens distintas, as quais serão apresentadas a seguir, e que foram denominadas Método 1, 2, 3 e 4. Esses procedimentos variam quanto ao número de sujeitos considerados e à quantidade e seleção das imagens. Além disso, definiu-se a forma de extrair conclusões a partir dos diferentes resultados obtidos.

4.4.2.1 Método 1

Foi proposto o cálculo da acurácia, da sensibilidade e da especificidade para os nove modelos de reconhecimento facial disponibilizados pelo framework em estudo *Deepface*, nas bases de dados C2FPW e FGNET. Dessa forma, poderá-se comparar o desempenho desses algoritmos em bases com e sem procedimentos estéticos, respectivamente, ambas contendo também o envelhecimento dos indivíduos. O procedimento para o cálculo dessas três métricas é melhor explicado a seguir.

Aleatoriamente, foram selecionados dois pares de imagens de cada sujeito da base. Espera-se que os modelos classifiquem esse conjunto como pares positivos, ou seja, que reconheçam as duas imagens do par como pertencentes à mesma pessoa. Ao todo, foram escolhidos 180 pares na base C2FPW e 164 pares na base FGNET, dado que possuem 90 e 82 sujeitos, respectivamente. O número de pares foi escolhido de forma a manter a mesma proporcionalidade de pares por sujeito de cada base.

Elegeu-se a mesma quantidade de pares em cada base, agora negativos, ou seja, constituídos de pessoas diferentes. Para formar esse conjunto, duas fotos foram selecionadas aleatoriamente do banco de imagens, certificando-se de que pertenciam a sujeitos distintos.

Após a seleção de um par, ele é removido da lista de imagens disponíveis e o processo

foi repetido até que o número desejado de pares fosse alcançado. Espera-se que os modelos de reconhecimento identifiquem corretamente esses pares como sendo compostos por imagens de pessoas diferentes.

Com os dois conjuntos estabelecidos, os 360 pares da C2FPW e os 328 pares da FGNET foram avaliados pelos nove modelos de reconhecimento facial. As conclusões obtidas pelos algoritmos foram então comparadas com as esperadas para cada par, e a acurácia foi calculada utilizando a Equação 4.1, a sensibilidade a Equação 4.2 e a especificidade a Equação 4.3.

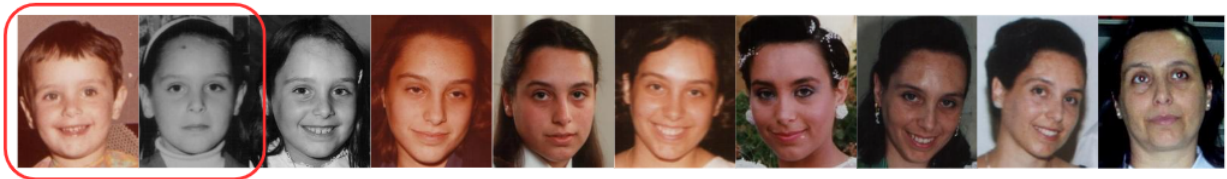
Assim como explicado ao final da Sessão 4.4.1.1, a sensibilidade, a especificidade e a acurácia serão apresentadas nos resultados, mas, para as conclusões deste método, apenas a sensibilidade será considerada.

Conforme será apresentado no Capítulo 5, observou-se uma grande queda da sensibilidade tanto na FGNET como na C2FPW, em comparação com resultados na LFW e na HDA. Entende-se que isso ocorreu devido à seleção de imagens aleatórias nas duas bases, para o cálculo da acurácia. Com isso, a chance de duas fotos com distância temporal significativa terem sido escolhidas é grande, favorecendo a influência do fator do envelhecimento no reconhecimento facial. Dessa forma, propôs-se três novos métodos, os quais utilizam mais fotos por indivíduo, buscando amenizar o impacto da idade.

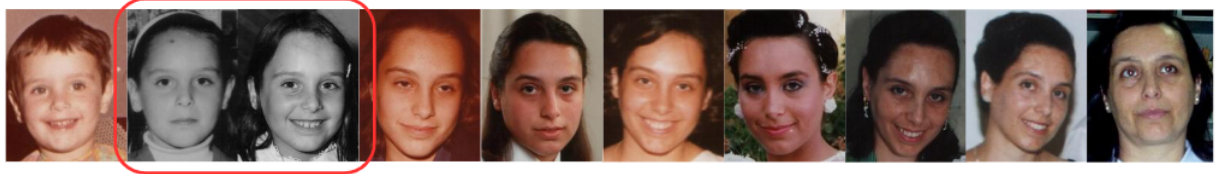
Nos Métodos 2, 3 e 4, almeja-se beneficiar da diversidade de fotos de um mesmo sujeito, em diferentes idades, que há nas bases C2FPW, a qual contém pessoas envelhecendo com procedimentos estéticos, e FGNET, que possui pessoas envelhecendo sem a probabilidade significativa de terem realizado tais modificações faciais. Dessa forma, o envelhecimento poderá ser considerado nas análises, e a diferença observada nos resultados das duas bases poderá ser atribuída aos procedimentos estéticos. Considerou-se que a probabilidade de um sujeito da base FGNET ter realizado modificações faciais voluntárias é a mesma de qualquer pessoa na população mundial tê-lo feito, enquanto na C2FPW 100% dos sujeitos se submeteram a tais alterações.

4.4.2.2 Seleção de fotos para os Métodos 2, 3 e 4

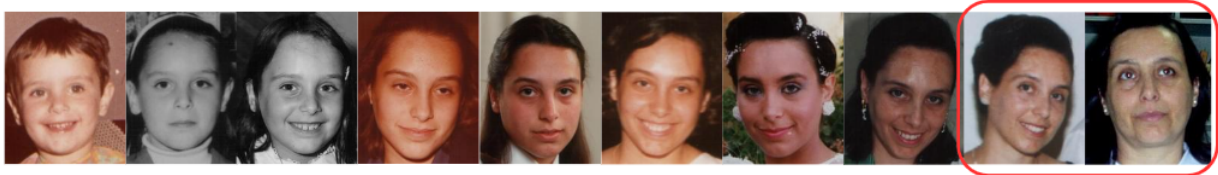
Desejando comparar o desempenho dos algoritmos entre a base C2FPW, que contém procedimentos estéticos, e a base FGNET, que não possui tais modificações faciais, é importante adotar um método semelhante para ambas as bases de imagens. Portanto, nos métodos 2, 3 e 4, optou-se por escolher a mesma quantidade de fotos por sujeito nesses conjuntos de imagens. Nesse sentido, considerando que a base FGNET possui menos fotos por sujeito, ela se tornou limitante para a definição desse valor, conforme será apresentado na Sessão 5.3.



(a) Comparação 1.



(b) Comparação 2.



(c) Última comparação, comparação 9.

Figura 4.29 – Exemplo de comparações realizadas no sujeito 002 da FGNET.

4.4.2.3 Método 2

Em resumo, foi proposto fazer uma análise do comportamento ao longo da vida de pessoas que realizaram ou não procedimentos estéticos em algum momento. O objetivo foi calcular a distância média entre as fotos de um mesmo sujeito, fornecida por cada um dos nove modelos de reconhecimento facial. A média foi calculada considerando 10 imagens do sujeito, em diferentes idades, selecionadas por quartis, conforme explicado na Seção 5.3, e a comparação foi feita entre pares de fotos em ordem cronológica. A seguir, esse método será melhor explicado.

- Comparação duas a duas

Após determinar as 10 fotos de cada sujeito que serão utilizadas nessa análise, conforme explicado na Seção 5.3, fez-se a comparação de duas a duas, em ordem cronológica, para cada modelo em estudo. O resultado, o qual indica uma distância entre as imagens, foi armazenado para todas as comparações.

A seguir, utilizando o exemplo de seleção de fotos para o sujeito 002 da FGNET, apresentado na Figura 5.43a, demonstra-se quais serão as duplas de fotos analisadas pelos modelos de reconhecimento facial, para esse conjunto de fotos. O primeiro par avaliado é o destacado na Figura 4.29a. Em seguida, compara-se as fotos evidenciadas na Figura 4.29b. A mesma lógica é seguida até o último par, apresentado na Figura 4.29c, resultando em 9 comparações entre as 10 imagens selecionadas.

- Resultados

A partir do passo anterior, são obtidos 9 resultados de distância entre pares de imagens por sujeito e por modelo de reconhecimento facial. Com base nisso, é calculada a distância média para cada par (indivíduo, modelo de reconhecimento) como sendo a média dos 9 valores obtidos das comparações, para cada modelo. É importante lembrar que o valor de distância fornecido pelo modelo de reconhecimento facial indica o quão semelhantes são as imagens. Se a distância for menor que um determinado limiar de similaridade, considera-se que as imagens pertencem ao mesmo indivíduo; caso contrário, são consideradas de indivíduos diferentes (ver Seção 2.2.6).

Para sujeitos que possuem mais de 10 fotos, os passos descritos até agora foram repetidos 30 vezes. Em cada execução, fez-se escolhas aleatórias das fotos, seguindo o procedimento explicado para seleção em quartis, resultando assim em 30 execuções com seleções de imagens diferentes. Em todos os testes, calculou-se a distância média de cada sujeito, como explicado anteriormente e exemplificado nas Tabelas 5.15 e 5.16 na Seção 5. Em seguida, a média dos resultados obtidos nas 30 execuções foi calculada para cada par (indivíduo, modelo de reconhecimento).

Em um exemplo prático desses cálculos, para o sujeito 002 da FGNET, o qual possui 18 imagens no total, considerou-se 30 seleções de 10 fotos, tais como as apresentadas nos exemplos de seleção (Figuras 5.43a e 5.43b). Em seguida, para cada modelo de reconhecimento, calculou-se a média dos resultados médios desse sujeito, obtidos nas 30 execuções realizadas, resultando em um valor final médio por par (sujeito, modelo de reconhecimento), como é exemplificado nas Tabelas 5.17 e 5.18 na Seção 5. A Figura 4.30 explica de forma visual esse procedimento.

Ao seguir os passos explicados tanto para a base C2FPW quanto para a base FGNET, é possível obter um valor médio para cada um dos 90 sujeitos da C2FPW e para os 73 indivíduos selecionados da FGNET, em relação a cada modelo de reconhecimento facial estudado. Isso resulta na criação de 18 vetores, com 2 vetores por modelo de reconhecimento, um para cada base de imagens. Esses vetores são compostos pelos valores médios finais de cada sujeito da respectiva base.

A Figura 4.31 ilustra o formato dos 9 vetores da base FGNET. Os resultados médios do sujeito 002, apresentados na Tabela 5.18, corresponderão ao segundo valor de cada vetor, uma vez que se referem ao segundo sujeito da lista. Os vetores relacionados aos sujeitos da C2FPW seguem o mesmo formato.

- Análise dos resultados

Objetivando concluir sobre a influência de procedimentos estéticos no reconhecimento facial ao longo da vida do indivíduo, propôs-se comparar o comportamento observado nos vetores obtidos no passo anterior. Para cada modelo de reconhecimento, compara-se o vetor correspondente na FGNET e na C2FPW.

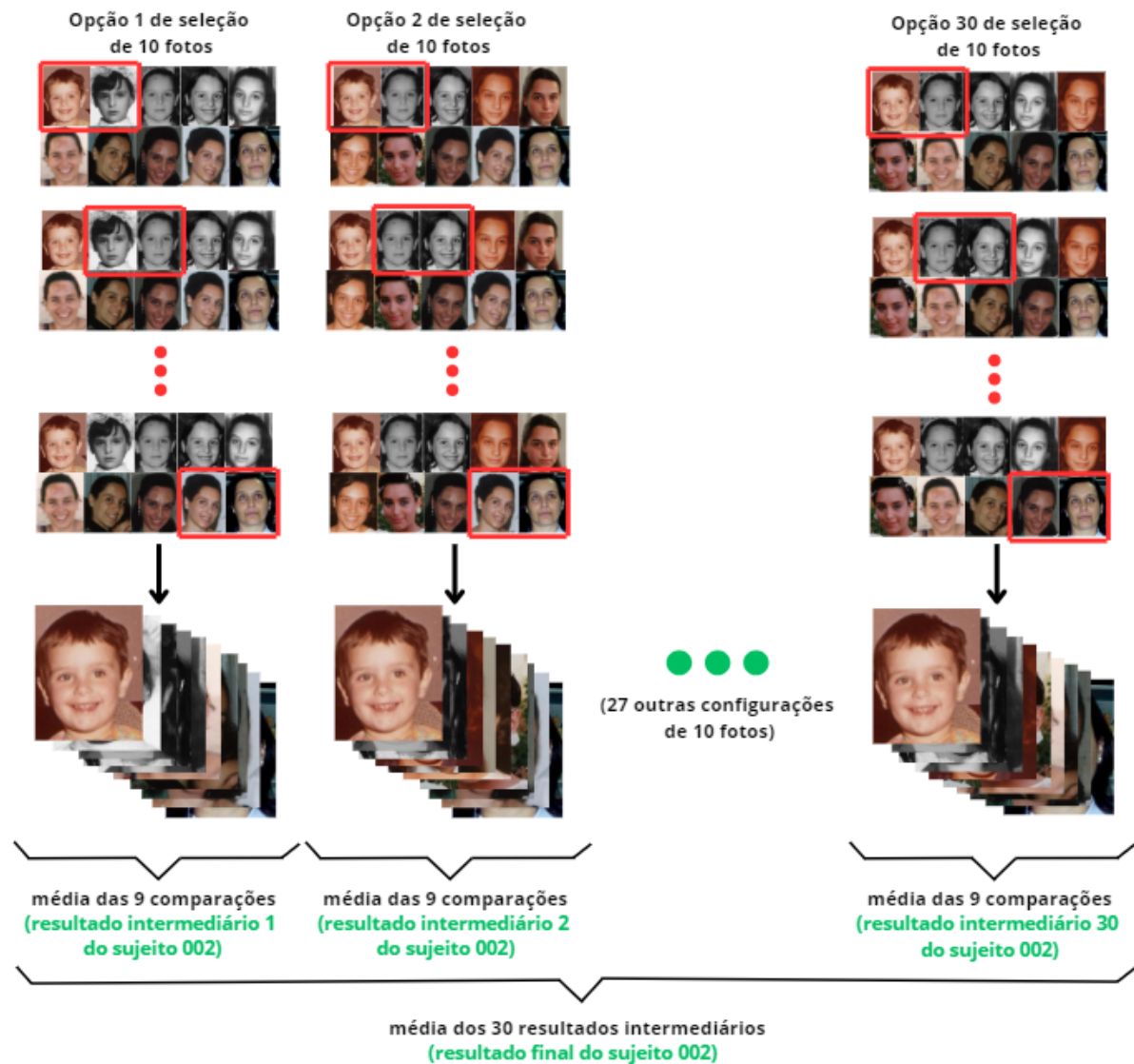


Figura 4.30 – Procedimentos para a obtenção do valor final por indivíduo no Método 2, usando como exemplo o Sujeito 002 da FGNET.

A fim de comparar os vetores, deve-se, primeiramente, constatar qual a distribuição dos valores que o formam. Por ser um teste amplamente usado, e fornecer uma medida efetiva acerca da normalidade de uma distribuição, o teste Shapiro-Wilk (Seção 2.3.3.1) foi realizado nos 18 vetores adquiridos.

Caso os vetores apresentem uma distribuição normal, é possível aplicar um teste de hipótese paramétrico. No entanto, se a distribuição não for normal, devem ser utilizados testes não paramétricos. O objetivo desses testes é validar ou refutar uma determinada hipótese. No presente estudo, a hipótese é que os dois vetores de um determinado modelo de reconhecimento facial, um em cada base de imagens estudada, sejam representativos da mesma população. Portanto, se a hipótese for aceita pelo teste, conclui-se que o comportamento dos sujeitos nas duas bases não possui diferença estatisticamente relevante, indicando que o modelo de reconhecimento é robusto em relação a procedimentos estéticos. Por outro lado, se a hipótese for rejeitada, significa que o modelo apresentou um comportamento

FGNET	
VGG-Face	→ [Média-S01, 0.22690, Média-S03, Média-S72, Média-S73]
Facenet	→ [Média-S01, 0.29747, Média-S03, Média-S72, Média-S73]
Facenet512	→ [Média-S01, 0.31151, Média-S03, Média-S72, Média-S73]
OpenFace	→ [Média-S01, 0.22255, Média-S03, Média-S72, Média-S73]
DeepFace	→ [Média-S01, 0.22913, Média-S03, Média-S72, Média-S73]
DeepID	→ [Média-S01, 0.04049, Média-S03, Média-S72, Média-S73]
ArcFace	→ [Média-S01, 0.46076, Média-S03, Média-S72, Média-S73]
Dlib	→ [Média-S01, 0.04233, Média-S03, Média-S72, Média-S73]
SFace	→ [Média-S01, 0.53467, Média-S03, Média-S72, Média-S73]

Figura 4.31 – Vetores formados pelos valores médios das 9 comparações de cada sujeito da base FGNET.

diferente nas duas bases, sugerindo que as modificações faciais voluntárias influenciaram no desempenho do algoritmo testado.

Como teste paramétrico, optou-se pelo teste t de [Student \(1908\)](#) (consulte a Seção 2.3.3.2). Já o teste de Wilcoxon, proposto por [Wilcoxon \(1945\)](#), foi selecionado como teste não paramétrico (consulte a Seção 2.3.3.3). Essa opção é amplamente utilizada como substituto do teste t quando não é possível aplicar este último, por exemplo, quando os dados não seguem uma distribuição normal. Vale ressaltar que ambos os testes requerem que os vetores testados tenham o mesmo tamanho. Como os vetores correspondentes à FGNET possuem 73 valores, enquanto os da C2FPW têm 90 elementos, considerando o número de sujeitos envolvidos, antes de realizar os testes de hipótese, são escolhidos aleatoriamente 73 dos 90 valores dos vetores da C2FPW. Assim, com os dois vetores de tamanho 73, é possível prosseguir com os testes.

Por fim, para os modelos que rejeitaram a hipótese nula de que possuem a mesma distribuição (com mesma média) nas duas bases, realizou-se um novo teste de hipótese com o objetivo de avaliar o desempenho do modelo em relação à presença ou ausência de procedimentos estéticos. Nesse caso, a hipótese nula testada foi de que a média da distribuição na base C2FPW é maior do que na base FGNET, indicando que a distância entre os pares de imagens analisados é maior na C2FPW, ou seja, que as imagens são mais diferentes quando há procedimentos estéticos. Considerando também o teste anterior que confirmou que há diferenças nas médias das distribuições, se a hipótese nula for rejeitada, conclui-se que a média é menor na base C2FPW e o modelo apresenta um desempenho melhor diante de procedimentos estéticos.

4.4.2.4 Método 3

Para quantificar a variação ao longo da vida de um sujeito, com e sem procedimentos estéticos, propôs-se calcular uma variação geral, por indivíduo, nas bases de dados FGNET e C2FPW. O cálculo é realizado utilizando os resultados obtidos no Método 2 (Seção 4.4.2.3),

de comparações de imagens duas a duas, em ordem cronológica, utilizando 10 fotos por sujeito. Aqui não são usados os 18 vetores gerados no Método 2, apenas as comparações duas a duas, como as exemplificadas nas tabelas 5.15 e 5.16. A partir desses resultados, calcula-se a diferença entre comparações consecutivas e, ao final, obtém-se a soma do valor absoluto de todas as diferenças.

Para esse método, os seguintes passos são seguidos:

- Passo 1: Obtenção dos resultados alcançados no Método 2 e cálculo da diferença entre comparações consecutivas;
- Passo 2: Os valores absolutos das diferenças de distâncias obtidos do sujeito são considerados, e o valor final do sujeito é calculado como a soma dessas variações absolutas. Esse procedimento é melhor entendido na Tabela 5.25 com os resultados referentes ao sujeito 002;
- Passo 3: Semelhantemente ao proposto no Método 2 (Seção 4.4.2.3), para cada um dos 9 modelos de reconhecimento facial, cria-se dois vetores, um relativo a cada base de imagens estudada: C2FPW e FGNET, totalizando 18 vetores. Os vetores são compostos pelo valor final de cada sujeito, naquela base, para aquele modelo de reconhecimento. A Figura 4.32 exemplifica esse processo para sujeito 002 e modelo VGG-Face;

VGG-Face → [ValorFinal-S01, ValorFinal-S02, ValorFinal-S03, ValorFinal-S72, ValorFinal-S73]
 VGG-Face → [ValorFinal-S01, 0,83127, ValorFinal-S03, ValorFinal-S72, ValorFinal-S73]

Figura 4.32 – Vetor referente à FGNET, para o modelo VGG-Face, usando o resultado expresso na Tabela 5.25 para o sujeito 002.

- Passo 4: A análise do resultado é feita da mesma forma que no Método 2 (Seção 4.4.2.3), utilizando os testes de hipótese T-Test, quando este é possível, ou o Teste de Wilcoxon, caso contrário. Além disso, são selecionados aleatoriamente 73 dos valores dos vetores relativos à C2FPW, de forma que fiquem do mesmo tamanho dos vetores da FGNET e os testes possam ser aplicados. Por último, caso os testes indiquem um comportamento diferente dos modelos nas duas bases, um novo teste de hipótese é realizado, testando a hipótese nula de que a média da distribuição é maior na C2FPW.

4.4.2.5 Método 4

Almejando o cálculo de sensibilidade sem a influência tão significativa da idade, observada no Método 1 (Seção 4.4.2.1), utilizou-se as 9 comparações de cada sujeito, feitas no Método 2 (Seção 4.4.2.3), como sendo os pares positivos analisados. Com isso, a distância temporal entre as imagens é menos impactante.

Entretanto, sabendo que cada indivíduo da C2FPW realizou um ou mais procedimentos estéticos na face, é possível que somente uma das 9 comparações possua uma foto

antes e uma depois da modificação, e as demais comparações pares de imagens antes da mudança ou pares depois dela. Sendo assim, o cálculo da sensibilidade na C2FPW será um indicativo de uma base que possui sujeitos com alteração estética, mas nem todos os pares considerados contêm tais procedimentos. O mesmo foi feito na FGNET, de forma que o resultado em ambas as bases possam ser comparados.

4.5 Análise e discussão dos resultados obtidos

Obtidos os resultados dos 4 métodos apresentados na seção anterior, pode-se, então concluir sobre o impacto do envelhecimento bem como dos procedimentos estéticos nos 9 modelos de reconhecimento facial estudados. Essa avaliação será detalhada na próxima seção (Seção 5).

5 Resultados

Este capítulo apresenta os resultados obtidos com as implementações da metodologia proposta exposta no Capítulo 4. Os resultados apresentados foram organizados, todos de forma a demonstrar que os objetivos propostos na Seção 1.3 foram alcançados.

5.1 Conjunto proposto de imagens ao longo da vida de pessoas as quais realizaram modificações faciais estéticas: C2FPW

Seguindo os passos propostos na Seção 4.2.3, o conjunto de imagens C2FPW foi desenvolvido, sendo que estará disponível para domínio público para uso exclusivamente acadêmico-científico¹. A seguir, estatísticas desse conjunto são evidenciadas, juntamente com exemplos de amostras de imagens. Primeiramente, são evidenciadas as distribuições tanto de sexo como de nacionalidade na C2FPW, respectivamente nas Figuras 5.33 e 5.34.

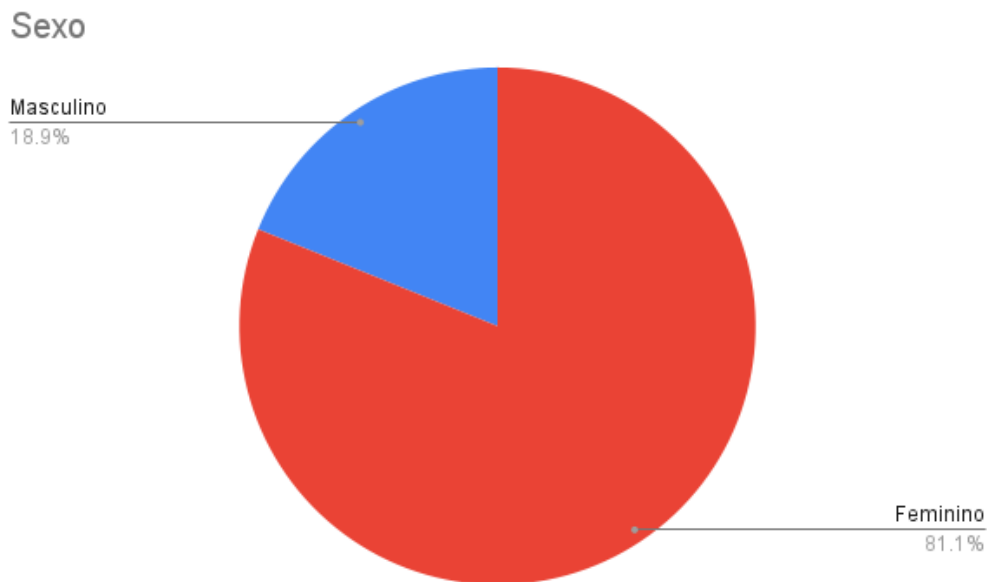


Figura 5.33 – Distribuição de indivíduos por sexo na C2FPW.

Segundo [Plastic Surgeons \(2021\)](#), de 2016 a 2020, observou-se uma constância de 92% dos procedimentos estéticos dos Estados Unidos sendo realizados por mulheres, enquanto 8% realizados por homens. Já de acordo com [Michas \(2023\)](#), em 2021, a distribuição de

¹ As imagens que compõe a base C2FPW foram disponibilizadas no sítio web <https://github.com/fbvidal/C2FPW-RecFac>.

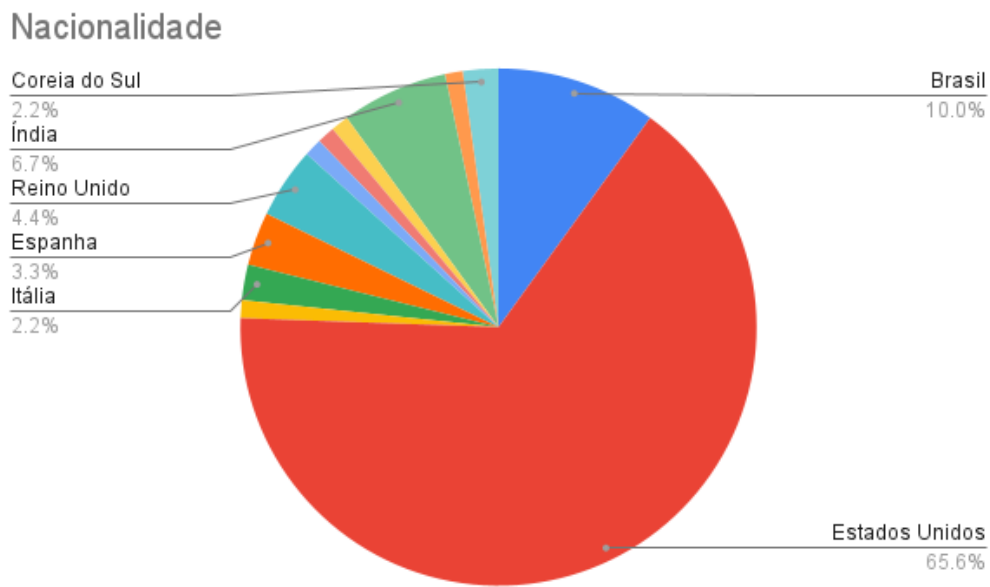


Figura 5.34 – Distribuição de indivíduos por nacionalidade na C2FPW.

procedimentos cosméticos pelo mundo é liderada pelos Estados Unidos e seguida pelo Brasil (Figura 5.35). Dessa forma, em ambos os critérios, sexo e nacionalidade, a C2FPW tem uma distribuição coerente com a população total que realiza tais alterações estéticas.

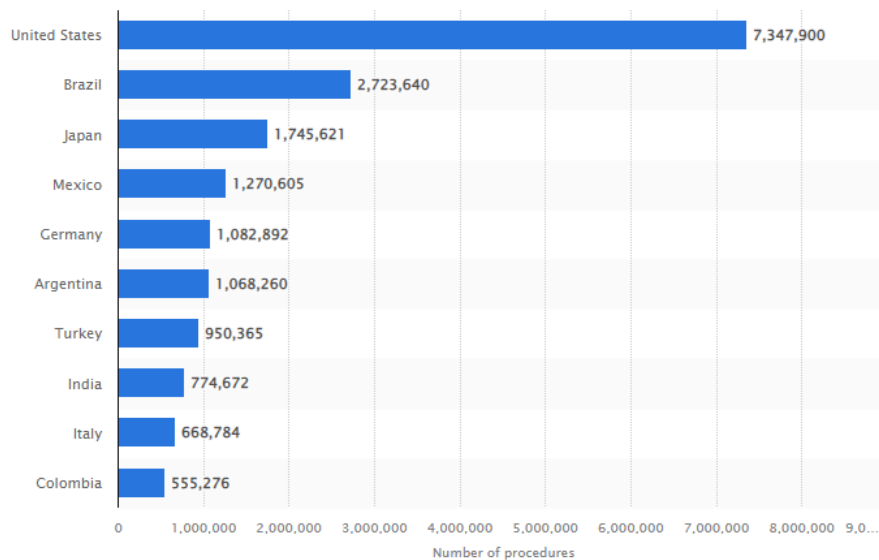


Figura 5.35 – Distribuição de procedimentos cosméticos pelo mundo. Fonte: (MICHAS, 2023).

Já a distribuição de imagens por indivíduo é bastante variada na C2FPW. O histograma na Figura 5.36 demonstra essa disposição. Especificamente, tem-se uma variação de 15 a 69 imagens por indivíduo, totalizando 3056 imagens de 90 indivíduos. As Figuras 5.37 e 5.38 exemplificam indivíduos com, respectivamente, 43 e 16 imagens ao longo de sua vida na C2FPW.

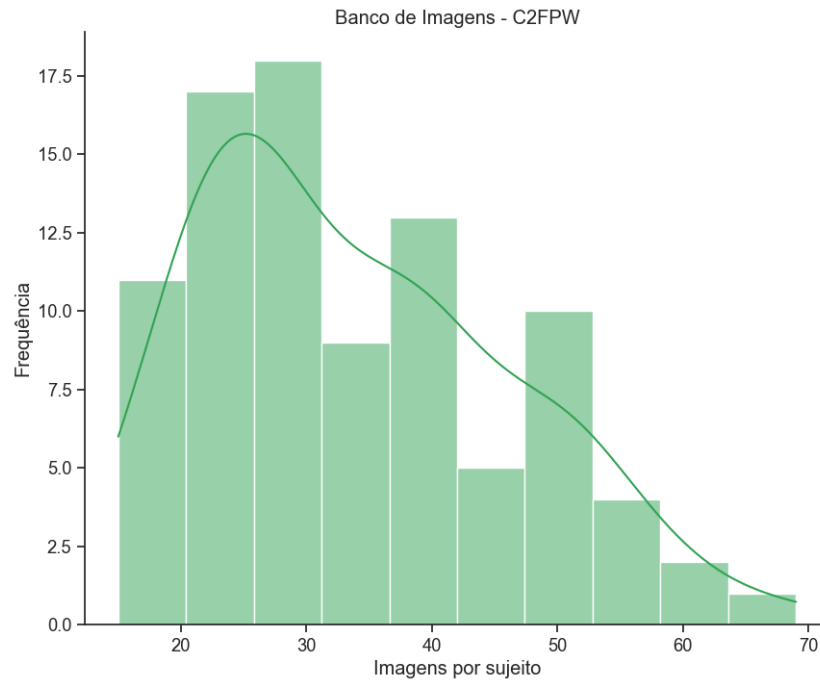


Figura 5.36 – Distribuição de imagens por indivíduo na C2FPW.



Figura 5.37 – Exemplo de imagens da C2FPW do indivíduo S014 com 43 imagens ao longo da vida.

5.2 Definição de modelos de detecção e de reconhecimento facial

5.2.1 Definição do modelo de detecção

Como exemplo de uma detecção realizada de forma adequada, tem-se a Figura 5.39, cujo processo de detecção foi realizado na imagem 002A15 do sujeito 002 da FGNET. É



Figura 5.38 – Exemplo de imagens da C2FPW do indivíduo S025 com 16 imagens ao longo da vida.

possível verificar que, além da detecção, o alinhamento é realizado na face, como evidenciado no pipeline de reconhecimento facial, explicado na Seção 2.

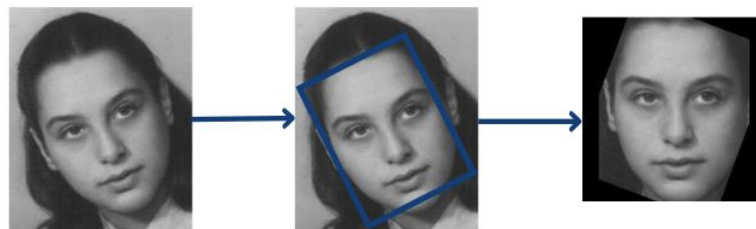


Figura 5.39 – Exemplo de detecção de face com modelo MTCNN do sujeito 002 da FGNET com 15 anos.

As Tabelas 5.4 e 5.5 apresentam, respectivamente, a quantidade total de falhas de detecção, considerando a base C2FPW e a base FGNET, para cada um dos cinco modelos oferecidos pelo *Deepface* (SERENGIL; OZPINAR, 2020). O total de falhas é apresentado em número absoluto e em porcentagem, considerando que as bases C2FPW e FGNET possuem ao todo 3056 e 1002 imagens, respectivamente.

Modelo de detecção	Total de falhas de detecção na C2FPW
Haar Cascade (OpenCV)	147 (4.81%)
Single Shot Multibox Detector (SSD)	94 (3.07%)
Dlib	26 (0.85%)
Multitask Cascaded Convolutional Networks (MTCNN)	2 (0.06%)
RetinaFace	23 (0.75%)

Tabela 5.4 – Erros de detecção na base C2FPW.

Dos modelos testados, o que menos apresentou erros de detecção foi o MTCNN (Seção 2.2.3.4), falhando em apenas 2 imagens (0.06%) do total de imagens no conjunto C2FPW e 5 imagens (0.49%) no conjunto da FGNET. Dessa forma, foi definido que este modelo de detecção será usado em todas as futuras etapas do estudo, e as fotos em que o rosto não foi detectado foram removidas dos conjuntos de imagem utilizados, visto que essa é uma etapa crucial no processo de reconhecimento.

Modelo de detecção	Total de falhas de detecção na FGNET
Haar Cascade (OpenCV)	81 (8.08%)
Single Shot Multibox Detector (SSD)	599 (59.8%)
Dlib	25 (2.49%)
Multitask Cascaded Convolutional Networks (MTCNN)	5 (0.49%)
RetinaFace	164 (16.37%)

Tabela 5.5 – Erros de detecção na base FGNET.

As 2 imagens que não foram reconhecidas pelo MTCNN no banco de imagem proposto, C2FPW, também não o foram por nenhum dos demais modelos de detecção facial. Dessa forma, elas foram removidas do conjunto e não serão disponibilizadas. Já para o banco de imagens FGNET, as imagens cujas faces não detectadas pelo MTCNN, modelo de detecção selecionado, foram apenas retiradas dos testes realizados. As 5 imagens retiradas da FGNET podem ser verificadas na Figura 5.40. Como é possível verificar, as faces não detectadas correspondem a imagens com baixa resolução de indivíduos com menos de 1 ano, em sua maioria.



Figura 5.40 – Imagens da FGNET não detectadas pelo modelo MTCNN.

Serão utilizadas para as futuras análises todos os nove modelos de reconhecimento facial disponibilizados pela versão utilizada do *Deepface*. São eles: VGG-Face, FaceNet, FaceNet512, OpenFace, DeepFace, DeepID, ArcFace, Dlib e SFace.

5.3 Avaliação da influência da idade e de procedimentos estéticos

O primeiro procedimento proposto foi com o objetivo de obter a influência direta de procedimentos estéticos no reconhecimento facial, sem o impacto conjunto da idade. Para isso, calculou-se a sensibilidade (Equação 4.2), a especificidade (Equação 4.3) e a acurácia (Equação 4.1) nas bases HDA e LFW, conforme explicado nas Seção 4.4.1.

Para o cálculo da porcentagem de reconhecimento de pares positivos, ou sensibilidade, na HDA, cada algoritmo de reconhecimento foi executado com todos os 637 pares positivos disponíveis na base, o cálculo da sensibilidade de cada modelo foi realizado utilizando a Equação 4.2. Os resultados podem ser verificados na Tabela 5.6.

Modelo	Verdadeiros Positivos	Falsos Negativos	Sensibilidade
VGG-Face	635	2	99,68%
FaceNet	541	96	84,93%
FaceNet512	621	16	97,49%
OpenFace	372	265	58,40%
DeepFace	623	14	97,80%
DeepID	475	162	74,57%
ArcFace	620	17	97,33%
Dlib	627	10	98,43%
SFace	584	53	91,68%

Tabela 5.6 – Sensibilidade calculada na HDA.

Já com os resultados do reconhecimento sob os 637 pares negativos aleatórios selecionados da base *HDA*, o cálculo da especificidade de cada modelo foi possível de ser realizado através da Equação 4.3. Os resultados dessa métrica na *HDA* podem ser verificados na Tabela 5.7.

Modelo	Verdadeiros Negativos	Falsos Positivos	Especificidade
VGG-Face	559	78	87,76%
FaceNet	636	1	99,84%
FaceNet512	637	0	100,00%
OpenFace	630	7	98,90%
DeepFace	290	347	45,53%
DeepID	484	153	75,98%
ArcFace	624	13	97,96%
Dlib	603	34	94,66%
SFace	630	7	98,90%

Tabela 5.7 – Especificidade calculada na HDA.

É possível verificar, através dos resultados de sensibilidade e especificidade dos modelos, as características específicas de cada um no que diz respeito ao reconhecimento testado. Há modelos que possuem um comportamento parecido, tanto para afirmação de reconhecimento de pares positivos como para negação de reconhecimentos dos pares negativos, como é o caso do modelo *ArcFace*, que possui ambas as métricas calculadas na casa dos 97%. Entretanto, há modelos que possuem um melhor comportamento quando se trata de reconhecimentos verdadeiros, como é o caso do *DeepFace*, cuja sensibilidade está na casa dos 97%, que é considerado um bom resultado, visto ser aproximadamente o desempenho humano nessa tarefa (TAIGMAN et al., 2014b), mas quando se trata de especificidade o percentual cai para 45%, indicando uma quantidade grande de falsos reconhecimentos pelo modelo. Em casos opostos, como é para o modelo *OpenFace*, em que a sensibilidade está próxima de 58% enquanto a especificidade está chegando aos 99%, os resultados indicam

que o modelo não consegue reconhecer de uma maneira satisfatória os pares positivos mas consegue um bom resultado na indicação de faces de diferentes pessoas.

Por fim, com os valores de VP, VN, FN e FP calculados, as acurácias de cada modelo foram estimadas através da Equação 4.1 e podem ser verificadas na Tabela 5.8. E o resultado indica que, para a base *HDA*, os modelos *OpenFace*, *DeepFace* e *DeepID* possuem comportamentos gerais menos satisfatórios que o restante, ficando na casa dos 70%.

Modelo	Acurácia
VGG-Face	93,72%
FaceNet	92,39%
FaceNet512	98,75%
OpenFace	78,65%
DeepFace	71,67%
DeepID	75,28%
ArcFace	97,65%
Dlib	96,55%
SFace	95,29%

Tabela 5.8 – Acurácia calculada na HDA.

Já para os resultados no *benchmark* da LFW, como foi apresentado no Capítulo 2 e pode ser verificado na Tabela 5.9, foram utilizados como base os resultados publicados de acurácia, dos modelos de reconhecimento que estão sendo analisados. Além disso, da mesma forma que foi calculado para a HDA, e conforme explicado na Seção 4.4.1.2, obteve-se a sensibilidade dos 9 modelos de reconhecimento na LFW. A Tabela 5.10 apresenta os resultados obtidos.

Modelo	Acurácia
VGG-Face	98,95%
FaceNet	99,63%
FaceNet512	99,60%
OpenFace	92,92%
DeepFace	97,00%
DeepID	97,45%
ArcFace	99,81%
Dlib	99,38%
SFace	99,57%

Tabela 5.9 – Resultados publicados de acurácia na LFW (HUANG; RAMESH et al., 2007).

Comparando-se os resultados calculados de sensibilidade na LFW (Tabela 5.10) com a sensibilidade na HDA (Tabela 5.6), entende-se que os procedimentos estéticos influenciam de forma positiva o desempenho dos nove modelos de reconhecimento facial em análise. Isso porque a HDA é composta por fotos antes e depois de procedimentos cirúrgicos faciais, enquanto a LFW possui indivíduos diversos, sobre os quais não se sabe acerca destas

Modelo	Verdadeiros Positivos	Falsos Negativos	Sensibilidade
VGG-Face	1578	98	94,15%
FaceNet	1160	516	69,21%
FaceNet512	871	805	51,97%
OpenFace	43	1633	2,56%
DeepFace	721	955	43,02%
DeepID	194	1482	11,57%
ArcFace	1480	196	88,30%
Dlib	1172	504	69,93%
SFace	958	718	57,16%

Tabela 5.10 – Sensibilidade calculada na LFW.

modificações. Assume-se, portanto, que a probabilidade de os indivíduos da LFW terem realizado procedimentos estéticos é pequena, enquanto na HDA é de 100%. Esse resultado vai de encontro ao esperado de forma intuitiva, de que os procedimentos estéticos reduziram o desempenho do reconhecimento facial.

Contudo, cabe ressaltar que ambas as bases possuem algumas diferenças relevantes e que devem ser consideradas. Enquanto a HDA possui apenas fotos em ambiente controlado, com boa iluminação e enquadramento, a LFW contém fotos em ambientes diversos, com diferentes iluminações e poses. Além disso, a HDA possui fotos muito próximas temporalmente, enquanto a LFW pode apresentar fotos de um mesmo indivíduo em momentos diferentes de sua vida. Há a possibilidade de essas constatações terem influenciado nos resultados observados. Portanto, faz-se interessante entender a influência de modificações voluntárias no reconhecimento facial estudando outros conjuntos de imagens, os quais possuam características semelhantes e contenham também o envelhecimento dos indivíduos.

Agora, na segunda abordagem seguida, com o intuito de estudar a influência do envelhecimento nos modelos de reconhecimento juntamente com os procedimentos estéticos, os métodos 1, 2, 3 e 4, explicitados nas Seções 4.4.2.1, 4.4.2.3, 4.4.2.4 e 4.4.2.5, respectivamente, foram executados.

5.3.1 Método 1 - Acurácia dos modelos de reconhecimento

Para o primeiro método proposto, a ideia é quantificar a acurácia nas bases C2FPW e FGNET, a fim de verificar se há diferenças significativas de desempenho nessas bases para os nove modelos de reconhecimento. Seguindo os procedimentos explicados na Seção 4.4.2.1, os resultados a seguir foram obtidos para o conjunto de imagens FGNET.

Com os mesmos resultados da Tabela 5.11, pode-se calcular mais dois parâmetros: a sensibilidade, também conhecida como Taxa de Verdadeiro Positivo, e a especificidade, também conhecida como Taxa de Verdadeiro Negativo (Tabela 5.12), através das Equações

Modelo	Teste com pares positivos		Teste com pares negativos		Acurácia
	Verdadeiros Positivo	Falsos Negativos	Verdadeiros Negativos	Falsos Positivos	
VGG-Face	95	69	143	21	72,56%
FaceNet	53	111	161	3	65,24%
FaceNet512	27	137	164	0	58,23%
OpenFace	3	161	164	0	50,91%
DeepFace	48	116	127	37	53,35%
DeepID	22	142	160	4	55,49%
ArcFace	109	55	154	10	80,18%
Dlib	100	64	143	21	74,09%
SFace	61	103	162	2	67,99%

Tabela 5.11 – Acurácia dos 9 modelos de reconhecimento na base FGNET.

4.2 e 4.3, respectivamente.

Modelo	Sensibilidade	Especificidade
VGG-Face	57,93%	87,19%
FaceNet	32,32%	98,17%
FaceNet512	16,46%	100%
OpenFace	1,83%	100%
DeepFace	29,27%	77,44%
DeepID	13,41%	97,56%
ArcFace	66,46%	93,90%
Dlib	60,97%	87,20%
SFace	37,20%	98,78%

Tabela 5.12 – Sensibilidade e Especificidade dos 9 modelos de reconhecimento na base FGNET.

A Tabela 5.13 apresenta os resultados obtidos para a C2FPW. Além da acurácia, também calculou-se a Sensibilidade e a Especificidade de cada modelo de reconhecimento facial nessa base (Tabela 5.14).

Modelo	Teste com pares positivos		Teste com pares negativos		Acurácia
	Verdadeiros Positivo	Falsos Negativos	Verdadeiros Negativos	Falsos Positivos	
VGG-Face	126	54	174	6	82,87%
FaceNet	94	86	180	0	76,11%
FaceNet512	57	123	180	0	65,83%
OpenFace	7	173	174	6	51,66%
DeepFace	78	102	131	49	58,06%
DeepID	26	154	166	14	53,33%
ArcFace	142	38	177	3	88,61%
Dlib	122	58	169	11	80,83%
SFace	91	89	178	2	74,72%

Tabela 5.13 – Acurácia dos 9 modelos de reconhecimento na base C2FPW.

Modelo	Sensibilidade	Especificidade
VGG-Face	70%	96,66%
FaceNet	52,22%	100%
FaceNet512	31,66%	100%
OpenFace	3,89%	96,66%
DeepFace	43,33%	72,78%
DeepID	14,44%	92,22%
ArcFace	78,89%	98,33%
Dlib	67,78%	93,89%
SFace	50,55%	98,89%

Tabela 5.14 – Sensibilidade e Especificidade dos 9 modelos de reconhecimento na base C2FPW.

Observa-se com os resultados apresentados nas Tabelas 5.12 e 5.14 que a sensibilidade dos nove modelos apresentados é consideravelmente baixa, em especial se comparadas com a acurácia esperada para esses modelos, como a publicada no *benchmark* LFW (Tabela 5.9) ou como a sensibilidade calculada também com esse conjunto de imagens (Tabela 5.10). Entende-se que a causa desses baixos resultados foi a escolha aleatória de imagens de cada indivíduo.

Como as bases FGNET e C2FPW possuem fotos dos indivíduos em diversas idades, é grande a probabilidade de duas fotos muito distantes temporalmente, com até mesmo décadas de diferença, terem sido consideradas como pares positivos. Dessa forma, o fator envelhecimento pode ter impactado o resultado mais do que os procedimentos estéticos, que são o escopo principal desse trabalho. Portanto, os próximos 3 métodos foram propostos, visando a reduzir a influência do envelhecimento, ou considerá-la de forma semelhante nas duas bases de imagem, e, assim, obter conclusões mais evidentes acerca do impacto de modificações faciais voluntárias no reconhecimento facial.

Para os Métodos 2, 3 e 4 propostos, estabeleceu-se a quantidade de imagens por indivíduo, bem como a forma de seleção das imagens de maneira que as conclusões na FGNET e na C2FPW pudessem ser comparadas. Primeiramente, a decisão acerca da quantidade de indivíduos utilizadas nestes métodos de cada base precisou ser tomada. Essa escolha foi realizada considerando dois critérios: ter um número suficiente de fotos por indivíduo, para representar o envelhecimento da pessoa, e manter uma parte significativa das bases de imagens, garantindo haver um número substancial de indivíduos diferentes na análise a ser realizada. Dessa forma, os histogramas da Figura 5.41 ilustram a distribuição da quantidade de fotos por indivíduo nas bases C2FPW e FGNET, após a remoção das imagens que não foram detectadas pelo modelo de detecção escolhido, conforme explicado na Seção 5.2.1.

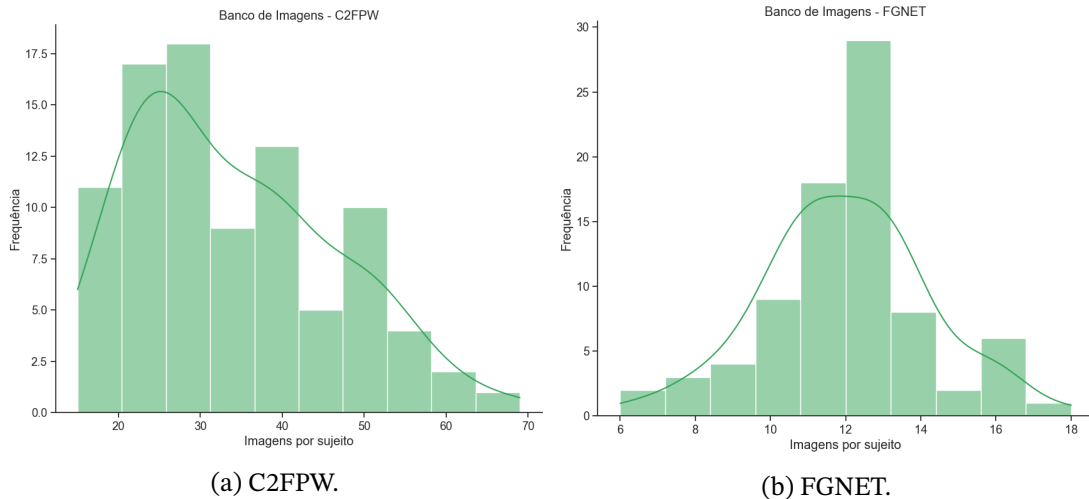


Figura 5.41 – Distribuição da quantidade de imagens por indivíduo nas bases FGNET e C2FPW.

Com o objetivo de preservar aproximadamente 90% da base FGNET e 100% da base C2FPW, optou-se por selecionar 10 fotos por indivíduo, considerando que essa quantidade é adequada para conseguir representar o envelhecimento de cada indivíduo e ainda utilizar, de forma proporcional, a maior parte dos bancos de imagens. Dessa forma, nos Métodos 2, 3 e 4, serão utilizados 90 indivíduos da C2FPW e 73 indivíduos da FGNET, incluindo todos aqueles que possuem 10 ou mais fotos disponíveis.

Decidida a quantidade de imagens, foi necessário determinar quais fotos serão utilizadas. Para garantir que o procedimento estético seja incluído na análise da base C2FPW e que um intervalo significativo de idade seja considerado na FGNET, a primeira e a última imagem de cada indivíduo são sempre selecionadas. Cabe lembrar que a data do procedimento estético não é conhecida na C2FPW. Sabe-se apenas que o indivíduo o realizou em algum momento.

Em seguida, para escolher as demais fotos, duas opções são consideradas: seleção pelas extremidades e por quartis. Ambas serão explicadas a seguir:

Seleção por quartis

Os quartis dividem o conjunto de fotos em quatro conjuntos igualmente distribuídos, contendo 25% das amostras, considerando a ordem dos dados (LOHNINGER, 2001). O conjunto de imagens do indivíduo é dividido em quartis, estando as fotos ordenadas cronologicamente e excluindo as duas imagens já escolhidas anteriormente (a primeira e a última de todos os indivíduos). A partir disso, duas fotos são aleatoriamente selecionadas de cada quartil.

A divisão em quartis foi escolhida de forma a distribuir as imagens ao longo da vida do indivíduo, e representar o envelhecimento. Já a escolha aleatória das fotos nos quartis foi feita visando possibilitar diferentes configurações de fotos analisadas em um mesmo indivíduo. Dessa forma, resultados utilizando diferentes conjuntos de 10 imagens podem ser

combinados, para representar melhor o comportamento daquele indivíduo.

A seguir são apresentados dois exemplos de seleção de 10 imagens de uma pessoa da FGNET. A Figura 5.42 exibe todas as fotos do indivíduo 002 dessa base, enquanto as Figuras 5.43a e 5.43b ilustram duas possibilidades de escolha das imagens desse indivíduo.



Figura 5.42 – Exemplo de imagens de uma pessoa contida na FGNET, indivíduo 002.



(a) Exemplo 1.



(b) Exemplo 2.

Figura 5.43 – Dois exemplos de seleção de 10 imagens, por quartis, para o indivíduo 002 da FGNET.

Seleção pelas extremidades

São selecionadas as 5 primeiras e as 5 últimas fotos de cada indivíduo. Aqui, portanto, só é possível uma configuração de fotos por indivíduo. A Figura 5.44 exemplifica a definição das fotos para o indivíduo 002 da FGNET.



(a) Todas as fotos do indivíduo 002 da FGNET.



(b) Fotos selecionadas pelas extremidades.

Figura 5.44 – Exemplo de seleção de fotos pelas extremidades, para o indivíduo 002 da FGNET.

Essa escolha visa distribuir melhor as imagens entre antes e depois do procedimento estético na base de dados C2FPW, considerando que a data exata da modificação facial não é conhecida. É assumido que, em termos de ordem cronológica, é mais provável que o procedimento estético esteja localizado no meio das imagens.

Ao aplicar esse mesmo método de escolha de fotos na base de dados FGNET, espera-se que o efeito da idade na distribuição das imagens seja semelhante ao observado na C2FPW, podendo-se, assim, separar o efeito do envelhecimento do efeito das mudanças estéticas no reconhecimento facial. Além disso, ao selecionar fotos consecutivas, tem-se uma pequena diferença de idade em 8 das 9 comparações, com apenas uma comparação apresentando uma variação de idade significativa.

Para facilitar a compreensão dos passos explicados a seguir, será chamada de *intervalo* a maior diferença de idade entre as 10 fotos selecionadas. Devido ao critério de seleção das fotos, o *intervalo* ocorrerá, em maioria, entre as imagens 5 e 6 escolhidas, ou seja, entre a última das 5 primeiras imagens e a primeira das 5 últimas imagens do indivíduo. A Figura 5.45 exemplifica onde ocorrerá o *intervalo* para o indivíduo 002 da FGNET.

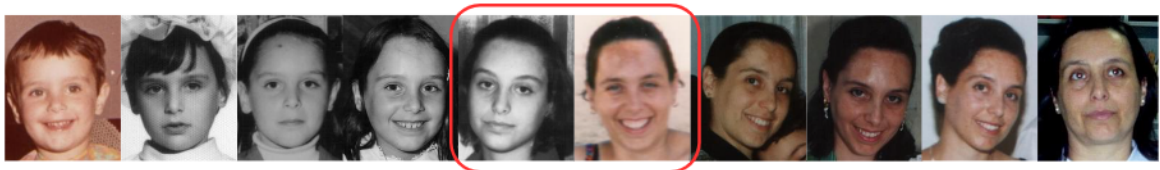


Figura 5.45 – Posição do *intervalo* para o indivíduo 002 da FGNET.

Dado que os indivíduos apresentam uma distribuição heterogênea de idades na base de dados, é esperado que o impacto do *intervalo* seja diferente nos casos em que um maior

período de vida está representado na base. Portanto, a análise comparativa entre C2FPW e FGNET será conduzida ao dividir os indivíduos em grupos (clusters), com base no intervalo, medido em anos, do *intervalo*.

Essa divisão foi realizada em faixas de 15 anos de intervalo, resultando ao final em uma separação de 4 clusters. A escolha de divisão de idades é realizada de forma a manter uma quantidade de indivíduos aproximadamente proporcional em cada cluster na C2FPW. Os *intervalos* definidos foram:

- **Cluster 1:** intervalo de 0 à 15 anos;
- **Cluster 2:** intervalo de 16 à 30 anos;
- **Cluster 3:** intervalo de 31 à 45 anos;
- **Cluster 4:** intervalo de 46 anos acima.

Entretanto fazendo-se a mesma divisão na FGNET, obtém-se conjuntos muito distintos, em termos de quantidade de indivíduos por cluster, o que pode ser observado na Figura 5.46. Sendo assim, não é possível comparar as duas bases, por cluster, com um número representativo de amostras por cluster. E, dado que a comparação das imagens que formam o *intervalo* propicia resultados muito distintos, a depender do tamanho dele, não é ideal obter conclusões misturando indivíduos de diferentes clusters. Dessa forma, optou-se por seguir com a **seleção por quartis** nos Métodos 2, 3 e 4.

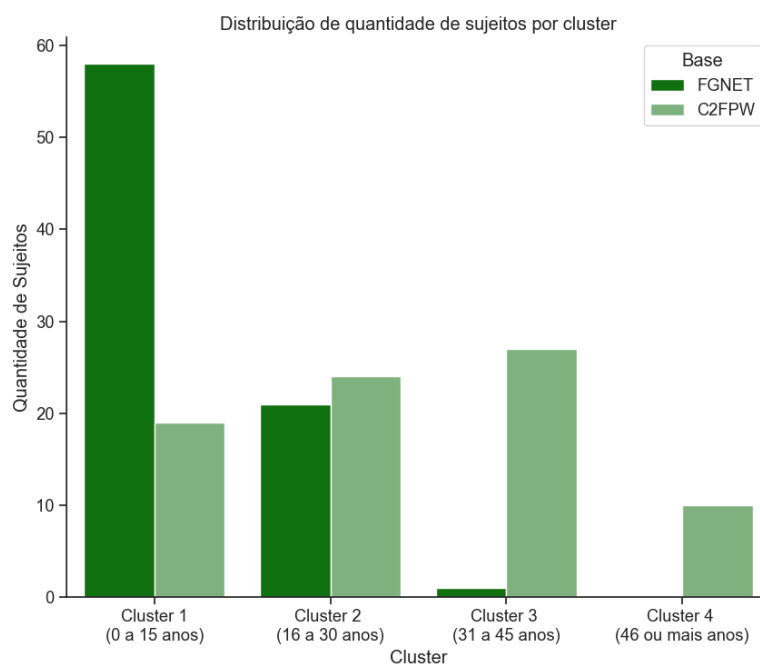


Figura 5.46 – Distribuição de clusters nas duas bases C2FPW e FGNET.

5.3.2 Método 2 - Comportamento médio de reconhecimento ao longo da vida

Como explicado na Seção 4.4.2.3, foram determinadas 10 imagens por indivíduo e a comparação foi realizada de duas a duas em ordem cronológica. As Tabelas 5.15 e 5.16 apresentam um resumo dos valores obtidos para o indivíduo 002 da base FGNET para um dos 30 testes realizados, utilizando a seleção de imagens apresentada na Figura 5.43a, bem como a média calculada para cada modelo de reconhecimento.

Comparação	Modelo de Reconhecimento				
	VGG-Face	Facenet	Facenet512	OpenFace	DeepFace
1	0,21267	0,27134	0,17657	0,25492	0,15288
2	0,13934	0,22670	0,23912	0,20862	0,13872
3	0,26354	0,21345	0,21309	0,18086	0,21649
4	0,16283	0,16731	0,20032	0,09308	0,16224
5	0,25865	0,24197	0,29166	0,26767	0,21645
6	0,38397	0,43274	0,29429	0,22885	0,17853
7	0,27991	0,44726	0,29027	0,23631	0,17565
8	0,11899	0,25986	0,13944	0,05648	0,11628
9	0,16590	0,39520	0,52491	0,38902	0,19201
Média	0,22064	0,29509	0,26330	0,21287	0,17214
Limiar de reconhecimento	0,4	0,4	0,3	0,1	0,23

Tabela 5.15 – Resultado dos modelos de reconhecimento facial para cada comparação do conjunto de imagens da Figura 5.43a.

Comparação	Modelo de Reconhecimento			
	DeepID	ArcFace	Dlib	SFace
1	0,00905	0,37289	0,04129	0,46952
2	0,00992	0,27457	0,02654	0,38934
3	0,01923	0,48406	0,03491	0,50226
4	0,04487	0,41990	0,03518	0,32769
5	0,05028	0,50138	0,04540	0,55775
6	0,04373	0,60581	0,05413	0,68145
7	0,05153	0,51200	0,02817	0,56493
8	0,02632	0,27127	0,02346	0,37771
9	0,01493	0,51327	0,06100	0,64972
Média	0,02998	0,43946	0,03890	0,50226
Limiar de reconhecimento	0,015	0,68	0,7	0,59328

Tabela 5.16 – Resultado dos modelos de reconhecimento facial para cada comparação do conjunto de imagens da Figura 5.43a.

Foram realizados ao todo 30 testes por indivíduo para cada conjunto de imagens (FGNET e C2FPW) de forma a conseguir um resultado mais significativo, visto que a seleção

de imagens por quartis acontece de forma aleatória, podendo haver enviesamento caso haja utilização de apenas uma seleção. Com isso, 30 resultados médios de comparações por indivíduo para cada modelo de reconhecimento foram obtidos. Com esses resultados por par (indivíduo, modelo de reconhecimento) foi encontrado o resultado final por indivíduo, através do cálculo de média dos testes. O resultado final do indivíduo 002 da FGNET, utilizado como exemplo aqui, está apresentado nas Tabelas 5.17 e 5.18.

Modelo de Reconhecimento					
	VGG-Face	Facenet	Facenet512	OpenFace	DeepFace
Média	0,22269	0,29747	0,31151	0,22913	0,22913

Tabela 5.17 – Resultado final das 30 configurações de 10 imagens do indivíduo 002 da FGNET (Figura 5.42).

Modelo de Reconhecimento				
	DeepID	ArcFace	Dlib	SFace
Média	0,04233	0,46076	0,04233	0,53467

Tabela 5.18 – Resultado final das 30 configurações de 10 imagens do indivíduo 002 da FGNET (Figura 5.42).

Foram obtidos ao todo então 18 vetores, como mencionados na Seção 4.4.2.3, e cujo formato foi ilustrado na Figura 4.31. Cada vetor corresponde a um par (modelo de reconhecimento facial, base de imagens) e contém as distâncias médias dos indivíduos da base em questão. Em seguida, foi inicialmente realizado um teste de normalidade nessas distribuições utilizando o teste de *Shapiro-Wilk* (Seção 2.3.3.1). Os resultados são apresentados na Tabela 5.19. Para esse teste, a hipótese nula (H_0) testada é que a média de resultados encontrados por indivíduo segue uma distribuição normal. O α , ou nível de significância, para esse teste foi definido como 0.05, significando uma probabilidade de 5% de rejeição da hipótese nula quando ela deveria ser aceita. Esse valor foi definido para o teste de *Shapiro-Wilk* pois é o valor comumente utilizado e citado como padrão na literatura (SEDGWICK, 2012).

Modelo	FGNET		C2FPW	
	Valor p	Resultado do teste Shapiro-Wilk	Valor p	Resultado do teste Shapiro-Wilk
VGG-Face	0.4395902	Normal	0.8090378	Normal
FaceNet	0.0144901	Não-Normal	0.7723836	Normal
FaceNet512	0.3426810	Normal	0.3091248	Normal
OpenFace	0.0032586	Não-Normal	0.0289974	Não-Normal
DeepFace	0.0194811	Não-Normal	0.4565085	Normal
DeepID	0.0097258	Não-Normal	3.5×10^{-12}	Não-Normal
ArcFace	0.1769404	Normal	0.4242640	Normal
Dlib	0.0184100	Não-Normal	0.6239643	Normal
SFace	0.5852066	Normal	0.1614409	Normal

Tabela 5.19 – Resultado do teste de normalidade Shapiro-Wilk nos 18 vetores obtidos no Método 2.

Após realizado o teste de normalidade nos vetores compostos pelos resultados de cada modelo, foram calculadas as variâncias amostrais de cada um deles. Objetiva-se verificar se a segunda suposição feita para testes paramétricos é satisfeita, de que as duas distribuições tem a mesma variância, e, assim, o teste pode ser realizado (Seção 2.3.3). As variâncias dos resultados para cada modelo pode ser verificado na Tabela 5.20.

Segundo [Bobbitt, Zach \(s.d.\)](#), se a razão entre a maior e a menor variância das duas distribuições em análise for menor que 4, então pode-se assumir que as variâncias são aproximadamente iguais e, assim, utilizar o teste de hipótese paramétrico, caso as outras suposições também sejam razoáveis.

Modelo	Variância na FGNET	Variância a C2FPW	Razão
VGG-Face	0.00271	0.00183	1.4809
FaceNet	0.01058	0.00315	3.3587
FaceNet512	0.00733	0.0032	2.2906
OpenFace	0.00619	0.00293	2.1126
DeepFace	0.00291	0.002	1.4550
DeepID	0.00022	0.00048	2.1818
ArcFace	0.00488	0.00589	1.2070
Dlib	0.00009	0.00006	1.5000
SFace	0.00596	0.00465	1.2817

Tabela 5.20 – Resultado das variâncias dos 18 vetores obtidos no Método 2.

Com os resultados obtidos (Tabela 5.20), foi possível verificar que todos os modelos tem então variâncias aproximadamente iguais para ambos os conjuntos de imagens e portanto, o teste paramétrico é passível de ser realizado nos modelos que também tiveram as suas distribuições atestadas como normal pelo teste anterior (Tabela 5.19).

Aplicou-se, então, o teste de hipótese t-Teste, para todos os modelos, cujos dois vetores apresentaram comportamento normal, que foram: VGG-Face, FaceNet512, ArcFace e SFace. Estabeleceu-se como hipótese nula (H_0) para o teste a de que ambos os vetores possuem mesma média, podendo ser representativos de uma mesma população, e escolheu-se um nível de significância α de 0.01. Dessa forma, aqueles que rejeitarem a hipótese demonstram que o comportamento nas duas bases é estatisticamente diferente, podendo-se, portanto, concluir em qual conjunto de imagens o algoritmo de reconhecimento facial teve melhor desempenho. Já para aqueles em que o teste não rejeitou a hipótese nula, tem-se que o modelo se comportou semelhantemente em ambas as bases, concluindo-se que o algoritmo possui um desempenho semelhante em presença ou não de procedimentos estéticos. A Tabela 5.21 apresenta os resultados obtidos com o T-Teste.

Dado que alguns modelos não atenderam o princípio básico estabelecido para o testes paramétricos, citados na Seção 2.3.3, de possuir uma distribuição normal em pelo menos uma das amostras de resultados a ser testado, aplicou-se o teste de hipótese *Wilcoxon Rank-Sum*. Esses modelos foram: Facenet, OpenFace, DeepFace, DeepID e Dlib. A hipótese

Modelo	Valor p	Resultado t-Teste
VGG-Face	0.0351565	Aceitou a Hipótese Nula
FaceNet512	0.1943596	Aceitou a Hipótese Nula
ArcFace	0.1400321	Aceitou a Hipótese Nula
SFace	0.0035606	Rejeitou a Hipótese Nula

Tabela 5.21 – Resultado do teste de hipótese T-Teste nos pares de vetores apresentados em 5.19, cujo comportamento foi normal.

nula foi rejeitada por esse teste apenas para o modelo OpenFace, considerando também um nível de significância α de 0.01 (Tabela 5.22). Dessa forma, entende-se que o modelo comportou-se diferentemente em ambas as bases de imagem, com e sem procedimentos estéticos.

Modelo	Valor p	Resultado Wilcoxon Rank-Sum
Facenet	0.0825970	Aceitou a Hipótese Nula
OpenFace	0.0024349	Rejeitou a Hipótese Nula
DeepFace	0.5453805	Aceitou a Hipótese Nula
DeepID	0.3005638	Aceitou a Hipótese Nula
Dlib	0.5401908	Aceitou a Hipótese Nula

Tabela 5.22 – Resultado do teste de hipótese Wilcoxon Rank-Sum nos pares de vetores apresentados na Tabela 5.19, cujo comportamento foi não-normal.

Para os modelos OpenFace e SFace, os quais rejeitaram a hipótese nula de mesma média nas distribuições na FGNET e C2FPW, testou-se uma nova hipótese nula (H_0), de que a média é maior na C2FPW do que na FGNET. Para o OpenFace utilizou-se o teste de Wilcoxon Rank-Sum, visto o comportamento não normal dos vetores, enquanto para o SFace o teste t-Teste foi empregado. Os resultados dos dois testes estão apresentados na Tabela 5.23.

Modelo	Teste	Valor p	Resultado
OpenFace	Wilcoxon Rank-Sum	0,000006	Rejeitou a Hipótese Nula
SFace	t-Teste	0.998219	Aceitou a Hipótese Nula

Tabela 5.23 – Resultado do teste de hipótese nos pares de vetores apresentados nas Tabelas 5.21 e 5.22, cuja primeira hipótese foi rejeitada.

Como o teste com a SFace aceitou a hipótese nula (Tabela 5.23), conclui-se, com o Método 2 (Seção 4.4.2.3), que o SFace se comporta melhor na FGNET do que na C2FPW. Dessa forma, procedimentos estéticos influenciam de forma negativa o desempenho desse modelo no reconhecimento facial. Já no teste com o modelo OpenFace, a hipótese de que a média de resultado é maior no C2FPW do que na FGNET foi recusada, indicando assim um desempenho melhor na base C2FPW, com procedimentos estéticos, visto que um resultado menor indica maiores proximidades entre as imagens comparadas.

5.3.3 Método 3 - Variação absoluta de reconhecimento ao longo da vida

Para o Método 3, foi utilizado os mesmos resultados encontrados no Método 2, entretanto, utilizando os valores resultantes das 9 comparações realizadas (Tabelas 5.15 e 5.16). Com esses resultados, os passos descritos na Seção 4.4.2.4 foram seguidos e inicialmente a diferença entre comparações consecutivas foi calculada para cada indivíduo. O resultado encontrado de variação de comparações para o indivíduo 002 pode ser visualizado na Tabela 5.24.

Comparação	Distâncias obtidas pelo VGG-Face	Varição
1	0,21267	-
2	0,13934	-0,07333
3	0,26354	+0,12420
4	0,16283	-0,10071
5	0,25865	+0,09582
6	0,38397	+0,12532
7	0,27991	-0,10406
8	0,11899	-0,16092
9	0,16590	+0,04691

Tabela 5.24 – Diferença entre comparações consecutivas do modelo VGG-Face, com as imagens da Figura 5.43a do indivíduo 002 da FGNET.

Graficamente observa-se as distâncias desse indivíduo na Figura 5.47, podendo-se perceber mais facilmente a diferença entre comparações consecutivas, calculada na Tabela 5.24. O eixo x da Figura 5.47 apresenta a idade do indivíduo na primeira foto considerada na comparação e o eixo y a resposta de distância entre as imagens, fornecida pelo modelo de reconhecimento.

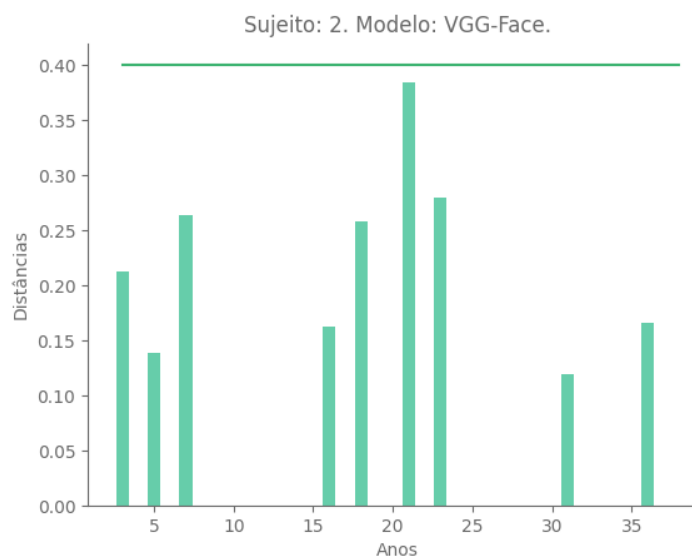


Figura 5.47 – Exemplo de resultados de comparações consecutivas do indivíduo 002 da base FGNET, utilizando o modelo de reconhecimento facial VGG-Face.

Com as diferenças de comparações dos indivíduos calculadas, a soma das variações absolutas do indivíduo foi estimada, resultado então em uma variação final que representa o quanto os resultados de distância de comparações daquele indivíduo variaram. Novamente, o resultado de um dos testes realizados com o indivíduo 002 da FGNET pode ser visualizado na Tabela 5.25.

Comparação	Distâncias obtidas pelo VGG-Face	Variação	Variação absoluta
1	0,21267	-	-
2	0,13934	-0,07333	0,07333
3	0,26354	+0,12420	0,12420
4	0,16283	-0,10071	0,10071
5	0,25865	+0,09582	0,09582
6	0,38397	+0,12532	0,12532
7	0,27991	-0,10406	0,10406
8	0,11899	-0,16092	0,16092
9	0,16590	+0,04691	0,04691
Resultado final do indivíduo 002:			0,83127

Tabela 5.25 – Cálculo do valor final do indivíduo 002 da FGNET, como a soma das variações absolutas.

Após obter a variação absoluta ao longo da vida dos indivíduos, foram gerados vetores para cada par (modelo de reconhecimento, base de imagens) para cada um dos 30 testes realizados. Esses vetores representam a variação absoluta de todos os indivíduos considerados em cada um dos conjuntos de imagens, resultando em 2 vetores por modelo de reconhecimento e um total de 18 vetores para os modelos e conjuntos de imagens analisados.

Inicialmente, os vetores foram submetidos também ao teste de normalidade de Shapiro-Wilk, e os resultados podem ser encontrados na Tabela 5.26. Novamente, o α , ou nível de significância, utilizado é de 0.05.

Modelo	FGNET		C2FPW	
	Valor p	Resultado do teste Shapiro-Wilk	Valor p	Resultado do teste Shapiro-Wilk
VGG-Face	0.1931522	Normal	0.1008004	Normal
FaceNet	0.0001907	Não-Normal	0.0370572	Não-Normal
FaceNet512	0.0352946	Não-Normal	0.1002461	Normal
OpenFace	0.0371815	Não-Normal	0.0000026	Não-Normal
DeepFace	0.0049356	Não-Normal	0.0030038	Não-Normal
DeepID	0.0006859	Não-Normal	7.6×10^{-13}	Não-Normal
ArcFace	0.0185867	Não-Normal	0.0053435	Não-Normal
Dlib	0.2071152	Normal	0.4955083	Normal
SFace	0.0680852	Normal	0.2359222	Normal

Tabela 5.26 – Resultado do teste de normalidade Shapiro-Wilk nos 18 vetores obtidos no Método 3.

Outra vez, as variâncias desses resultados por modelo e por base foram calculadas para verificar se o teste de hipótese paramétrico pode ser utilizado. Os resultados podem ser encontrados na Tabela 5.27 e verifica-se que apenas os resultados com a Facenet possuem variâncias significativamente diferentes, com razão maior que 4, nas bases.

Modelo	Variância na FGNET	Variância na C2FPW	Razão
VGG-Face	0.02919	0.01514	1.9280
FaceNet	0.32694	0.04523	7.2284
FaceNet512	0.10360	0.04143	2.2006
OpenFace	0.11449	0.05111	2.2401
DeepFace	0.03704	0.01466	2.5266
DeepID	0.00563	0.00972	1.7264
ArcFace	0.05830	0.03531	1.6609
Dlib	0.00178	0.00073	2.4383
SFace	0.11315	0.04647	2.4349

Tabela 5.27 – Resultado das variâncias dos 18 vetores obtidos no Método 3.

É possível verificar que os resultados dos modelos VGG-Face, Dlib e SFace são os que possuem uma distribuição normal em ambas as bases, C2FPW e FGNET, e nesses resultados, portanto, pôde ser aplicado o teste de hipótese paramétrico t-Teste. Os resultados do t-Teste nos vetores de variação absoluta estão expostos na Tabela 5.28, onde a Hipótese Nula (H_0) corresponde a hipótese de que ambos os resultados sendo testados possuem médias idênticas, e o nível de significância escolhido foi de $\alpha = 0.01$.

Modelo	Valor p	Resultado t-Teste
VGG-Face	0.0288317	Aceitou a Hipótese Nula
Dlib	0.6870852	Aceitou a Hipótese Nula
SFace	0.8957331	Aceitou a Hipótese Nula

Tabela 5.28 – Resultado do teste de hipótese T-Teste nos pares de vetores apresentados em 5.26, cujo comportamento foi normal.

Com os outros modelos, que possuíram algum comportamento não-normal em seus resultados em algum dos conjuntos de imagens, que foram: FaceNet, Facenet512, OpenFace, DeepFace, DeepID e ArcFace, aplicou-se o teste de *Wilcoxon Rank-Sum*. Novamente, a Hipótese Nula (H_0) sendo testada é a de que as distribuições dos resultados sendo testados para ambas as bases possuem uma distribuição semelhante, com nível de significância $\alpha = 0.01$. Os resultados desse teste podem ser verificados na Tabela 5.29.

É possível verificar que a hipótese de semelhança estatística entre os resultados de variação absoluta nas bases é aceita para quase todos os modelos, exceto o modelo FaceNet, que rejeita a hipótese.

Para o modelo FaceNet, que rejeitou a hipótese nula de mesma média nas distribuições na FGNET e C2FPW, testou-se uma nova hipótese, de que a média é maior na C2FPW

Modelo	Valor <i>p</i>	Resultado <i>Wilcoxon Rank-Sum</i>
FaceNet	0.0021941	Rejeitou a Hipótese Nula
FaceNet512	0.901879	Aceitou a Hipótese Nula
OpenFace	0.0191567	Aceitou a Hipótese Nula
DeepFace	0.5350259	Aceitou a Hipótese Nula
DeepID	0.2026731	Aceitou a Hipótese Nula
ArcFace	0.6625471	Aceitou a Hipótese Nula

Tabela 5.29 – Resultado do teste de hipótese *Wilcoxon Rank-Sum* nos pares de vetores apresentados em 5.26, cujo comportamento foi não-normal.

do que na FGNET. Para o esse modelo utilizou-se o teste de *Wilcoxon Rank-Sum*, visto o comportamento não normal dos vetores, e os resultados estão apresentados na Tabela 5.30.

Modelo	Teste	Valor <i>p</i>	Resultado
FaceNet	<i>Wilcoxon Rank-Sum</i>	0.001097	Rejeitou a Hipótese Nula

Tabela 5.30 – Resultado do teste de hipótese nos pares de vetores apresentados em 5.28 e 5.29, cuja primeira hipótese foi rejeitada.

Como o teste aceitou a hipótese nula (Tabela 5.30), conclui-se, com o Método 3 (Seção 4.4.2.4), que o FaceNet se comporta melhor na C2FPW do que na FGNET. Dessa forma, procedimentos estéticos influenciam de forma positiva o desempenho desses modelos no reconhecimento facial.

5.3.4 Método 4 - Sensibilidade dos modelos de reconhecimento

Como explicado na Seção 4.4.2.5, a sensibilidade (Equação 4.2), usando os resultados das 9 comparações por indivíduo do Método 2, para as bases FGNET e C2FPW, foi calculada e pode ser verificada na Tabela 5.31.

Modelo	FGNET	C2FPW
VGG-Face	89,85%	85,13%
Facenet	61,99%	70,62%
Facenet512	41,76%	49,79%
OpenFace	4,06%	7,47%
DeepFace	46,12%	49,12%
DeepID	14,91%	21,64%
ArcFace	93,71%	88,15%
Dlib	81,75%	78,29%
SFace	69,38%	60,79%

Tabela 5.31 – Sensibilidade dos 9 modelos de reconhecimento facial, usando os resultados encontrados no Método 2

A sensibilidade do modelo, que avalia o quão bem o modelo consegue reconhecer com sucessos os pares positivos, ou seja, de mesma pessoa, tem um comportamento parecido

nas duas bases. Há uma queda de sensibilidade na base FGNET para a base C2FPW, para os modelos VGG-Face, ArcFace, Dlib e Sface de aproximadamente 5,58%. Para os modelos Facenet, Facenet512, OpenFace, DeepFace e DeepID há um aumento médio de 5,96% nessa métrica.

6 Conclusões

A quantidade de procedimentos faciais voluntários, cirúrgicos ou não, tem crescido nas últimas décadas, e simultaneamente, há o aumento da utilização de sistemas de reconhecimento facial nos mais diversos campos. O presente trabalho propôs, portanto, a avaliação de nove modelos de aprendizagem profunda no reconhecimento facial, disponibilizados por [Serengil e Ozpinar \(2020\)](#), quanto ao impacto de modificações faciais realizadas voluntariamente nesses modelos.

Para isso, desenvolveu-se uma nova base de imagens, denominada C2FPW, cujas fotos foram obtidas de sites públicos da Internet, e que será disponibilizada de forma aberta após a apresentação deste trabalho. Ela possui, em média, 34 fotos em diferentes momentos da vida de 90 sujeitos que realizaram modificações estéticas voluntárias na face. O objetivo dessa base é possibilitar a avaliação da influência de procedimentos estéticos juntamente com o envelhecimento do sujeito no reconhecimento facial.

Numa motivação inicial para o estudo, foi realizada uma primeira avaliação do impacto dos procedimentos estéticos na sensibilidade dos nove modelos estudados. Avaliando-os na base HDA foi identificado um aumento da sensibilidade, em comparação com o cálculo realizado no *benchmark* amplamente utilizado LFW ([HUANG; MATTAR et al., 2012](#)). Dessa forma, entende-se que as modificações faciais voluntárias influenciam positivamente o desempenho dos algoritmos em análise. Entretanto, devido a diferenças como iluminação, pose e proximidade temporal das imagens que compõe a HDA e a LFW, propôs-se estudar o impacto dos procedimentos estéticos utilizando outros dois conjuntos de dados: FGNET e a base desenvolvida C2FPW.

Visto que os procedimentos estéticos faciais podem mudar significativamente regiões do rosto que usualmente os modelos de reconhecimento utilizam como informativos na hora de comparar duas faces, é intuitivo imaginar um decaimento no desempenho dos modelos quando aplicados na base C2FPW em comparação com a FGNET. Entretanto, a análise inicial com as bases HDA e LFW indicaram o comportamento contrário. Dessa forma, com essas duas bases, foram implementados 4 Métodos diferentes visando concluir acerca do impacto das modificações faciais voluntárias no reconhecimento facial.

O Método 1 indicou um desempenho inferior na FGNET para todos os 9 modelos em análise, em comparação com a C2FPW. Esse resultado indicaria que procedimentos estéticos facilitam o reconhecimento facial. Entretanto, como observou-se uma queda substancial da sensibilidade nessas duas bases, em comparação com as observadas na LFW e na HDA, assumiu-se que a escolha dos pares de imagens analisados foi ruim, favorecendo imagens com distância temporal elevada, podendo chegar a algumas décadas, e, portanto, o resultado

de desempenho seria mais associado ao envelhecimento do que de fato às modificações faciais, que são o escopo desse estudo. Dessa forma, os Métodos 2, 3 e 4 foram propostos.

Através dos testes de hipóteses no Método 2, foi verificado que as médias de resultados encontrados nos testes com ambas as bases, C2FPW e FGNET, para os modelos VGG-Face, Facenet, Facenet512, DeepFace, DeepID, ArcFace e Dlib, são estatisticamente iguais e portanto, possuem um comportamento parecido em relação à distância entre faces comparadas. Esse resultado sugere que tais modelos são igualmente robustos a modificações relativas ao envelhecimento quando com a presença simultânea ou não de modificações faciais voluntárias. Para os modelos cuja média não foi considerada estatisticamente igual nos testes, apenas o SFace apresentou uma média menor para a base sem procedimentos estéticos, demonstrando um decaimento de sua performance na base com procedimentos. Já o modelo OpenFace apresentou um comportamento divergente, apresentando uma melhora de performance na base com procedimentos estéticos, verificado por sua média de resultado maior na base sem procedimentos.

Com o Método 3, pode-se concluir que os modelos VGG-Face, Facenet512, OpenFace, DeepFace, DeepID, ArcFace, Dlib e SFace se comportaram sem diferença estatística nas bases FGNET e C2FPW. Dessa forma, esse Método indicou que esses modelos são robustos aos procedimentos estéticos. Já o FaceNet possuiu um comportamento melhor na C2FPW em comparação com a FGNET, ou seja, as modificações faciais voluntárias melhoraram o desempenho desse modelo de reconhecimento.

O Método 4 de fato apresentou resultados para a FGNET e a C2FPW mais próximos da LFW e HDA do que o Método 1. Dessa forma, entende-se que a escolha das imagens no Método 1 influenciou o resultado obtido, pela possibilidade de obter uma elevada distância temporal entre as fotos aleatoriamente selecionadas. Ainda, a conclusão obtida foi diferente do observado anteriormente. A sensibilidade foi maior na C2FPW do que na FGNET para 5 dos 9 modelos em análise e maior na FGNET para os 4 restantes, indicando que o impacto dos procedimentos estéticos no reconhecimento facial depende do modelo específico analisado.

Em suma, os diferentes Métodos empregados proporcionaram resultados variados. Enquanto as análises iniciais com a HDA e a LFW e o Método 1 com a C2FPW e FGNET indicaram que os modelos respondem melhor com a presença de procedimentos estéticos, os Métodos 2 e 3 sugerem que a maioria dos modelos se comporta de maneira estatisticamente igual com ou sem essas modificações faciais, e o Método 4 que essa conclusão depende do modelo específico em análise. Apenas um modelo em cada um dos Métodos 2 e 3 apresentaram melhora de desempenho frente a procedimentos estéticos, e um modelo no Método 2 obteve comportamento pior com essas modificações. Devido à quantidade de imagens consideradas e a atenuação do efeito da idade nos Métodos 2 e 3, bem como o uso de testes de hipótese para fundamentar as análises realizadas, considera-se que as conclusões deles são mais representativas do impacto de alterações faciais voluntárias no reconhecimento

facial do que as demais apresentadas.

Com base nos resultados encontrados, acredita-se que mais testes e análises são necessários para conseguir separar a influência de procedimentos estéticos do processo natural de envelhecimento e assim conseguir avaliar de forma quantitativa a influência dessas modificações nos modelos de reconhecimento. Os resultados deste trabalho estimulam assim novas pesquisas neste área importante, que está se estabelecendo como um grande desafio para algoritmos de reconhecimento facial.

6.1 Trabalhos Futuros

Como sugestão de trabalho futuro, propõe-se a criação de uma base de imagens semelhante a C2FPW, contendo imagens de sujeitos que realizaram modificações voluntárias faciais, porém, garantindo-se o conhecimento acerca da data do procedimento. Essa informação para os sujeitos avaliados na base permitiria uma separação da influência do envelhecimento da influência do procedimento nos modelos de reconhecimento, contribuindo ainda mais à análise desejada. Visto que a C2FPW é composta por figuras públicas, há ainda a possibilidade de contribuir com as informações das datas de procedimentos para os sujeitos presentes na base desenvolvida, como trabalho futuro.

Além disso, no contexto específico de crescimento do número de cirurgias de confirmação de gênero (*Gender confirming surgery*) realizadas por pessoas transgêneras, como é analisado por [Nolan, Kuhner e Dy \(2019\)](#), um estudo específico acerca do impacto dessas alterações no reconhecimento facial é uma área de grande interesse e pode ser objeto de estudos futuros. Com esse estudo seria possível identificar possíveis vieses de gênero associado à mudanças faciais voluntárias para os modelos de aprendizagem profunda no reconhecimento facial.

Por fim, outra possível abordagem para continuação desse estudo é aplicar transformações na base C2FPW as quais reduzem o impacto da idade, tais como a proposta por [R. Singh, Vatsa, Noore e S. Singh \(2007\)](#). Com isso, o impacto dos procedimentos estéticos deve ser evidenciado e conclusões mais precisas podem ser obtidas.

Referências

- 2NDTIDE. **2nd tide**. Out. 2019. Disponível em: <<https://2ndtide.tumblr.com/post/188706353902/lil-kim-1997>>. Citado na p. 19.
- ADJABI, I.; OUAHABI, A.; BENZAOUI, A.; TALEB-AHMED, A. Past, Present, and Future of Face Recognition: A Review. **Electronics**, v. 9, n. 8, 2020. ISSN 2079-9292. DOI: 10.3390/electronics9081188. Disponível em: <<https://www.mdpi.com/2079-9292/9/8/1188>>. Citado nas pp. 16, 24, 33.
- AHONEN, T.; HADID, A.; PIETIKAINEN, M. Face Description with Local Binary Patterns: Application to Face Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 28, n. 12, p. 2037–2041, 2006. DOI: 10.1109/TPAMI.2006.244. Citado nas pp. 51, 53.
- ALMEIDA, A. P. G. S. de; BARROS VIDAL, F. de. **Turning old models fashion again: Recycling classical CNN networks using the Lattice Transformation**. 2021. arXiv: 2109.13885 [cs.CV]. Citado na p. 23.
- AMOS, B.; LUDWICZUK, B.; SATYANARAYANAN, M. OpenFace: A general-purpose face recognition library with mobile applications. In. Citado na p. 38.
- AZIZ, A.; SAGAYAN, V.; MALIK, A.; ALI, A. Proposed face recognition system after plastic surgery. **IET Computer Vision**, v. 10, p. 342–348, ago. 2016. DOI: 10.1049/iet-cvi.2014.0263. Citado na p. 53.
- BALLANTYNE, M.; BOYER, R. S.; HINES, L. Woody Bledsoe: His Life and Legacy. **AI Magazine**, v. 17, n. 1, p. 7, mar. 1996. DOI: 10.1609/aimag.v17i1.1207. Disponível em: <<https://ojs.aaai.org/index.php/aimagazine/article/view/1207>>. Citado na p. 16.
- BALTRUSAITIS, T.; ROBINSON, P.; MORENCY, L.-P. OpenFace: An open source facial behavior analysis toolkit. In: WACV. IEEE Computer Society, 2016. P. 1–10. ISBN 978-1-5090-0641-0. Disponível em: <<http://dblp.uni-trier.de/db/conf/wacv/wacv2016.html#Baltrusaitis0M16>>. Citado nas pp. 37, 38.
- BAZAREVSKY, V.; KARTYNNIK, Y.; VAKUNOV, A.; RAVEENDRAN, K.; GRUNDMANN, M. **BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs**. arXiv, 2019. DOI: 10.48550/ARXIV.1907.05047. Disponível em: <<https://arxiv.org/abs/1907.05047>>. Citado na p. 32.

- BELHUMEUR, P.; HESPANHA, J.; KRIEGMAN, D. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 7, p. 711–720, 1997. DOI: [10.1109/34.598228](https://doi.org/10.1109/34.598228). Citado nas pp. 41, 51.
- BEVANS, REBECCA. **Hypothesis Testing | A Step-by-Step Guide with Easy Examples**. Disponível em: <https://www.scribbr.com/statistics/hypothesis-testing/> – acesso em 29 de Maio de 2023. Citado na p. 46.
- BHATT, H.; BHARADWAJ, S.; SINGH, R.; VATSA, M. Recognizing Surgically Altered Face Images Using Multiobjective Evolutionary Algorithm. **IEEE Transactions on Information Forensics and Security**, v. 8, n. 1, p. 89–100, 2013. DOI: [10.1109/TIFS.2012.2223684](https://doi.org/10.1109/TIFS.2012.2223684). Citado na p. 52.
- BHATT, H.; BHARADWAJ, S.; SINGH, R.; VATSA, M. On matching sketches with digital face images. In: p. 1–7. DOI: [10.1109/BTAS.2010.5634507](https://doi.org/10.1109/BTAS.2010.5634507). Citado na p. 52.
- BOBBITT, ZACH. **The Four Assumptions Made in a T-Test**. Disponível em: <https://www.statology.org/t-test-assumptions/> – acesso em 06 de Junho de 2023. Citado na p. 90.
- BOUSSAAD, L.; BOUCETTA, A. Deep-learning based descriptors in application to aging problem in face recognition. **Journal of King Saud University - Computer and Information Sciences**, v. 34, 6, Part A, p. 2975–2981, 2022. ISSN 1319-1578. DOI: <https://doi.org/10.1016/j.jksuci.2020.10.002>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1319157820304821>. Citado na p. 51.
- CAO, Q.; SHEN, L.; XIE, W.; PARKHI, O. M.; ZISSERMAN, A. VGGFace2: A dataset for recognising faces across pose and age. In: INTERNATIONAL Conference on Automatic Face and Gesture Recognition. 2018. Citado na p. 25.
- CAPLOVA, Z.; COMPASSI, V.; GIANCOLA, S.; GIBELLI, D. M.; OBERTOVÁ, Z.; POPPA, P.; SALA, R.; SFORZA, C.; CATTANEO, C. Recognition of children on age-different images: Facial morphology and age-stable features. **Science and Justice**, v. 57, n. 4, p. 250–256, 2017. ISSN 1355-0306. DOI: <https://doi.org/10.1016/j.scijus.2017.03.005>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1355030617300291>. Citado na p. 18.
- CHANG, K.-Y.; CHEN, C.-S.; HUNG, Y.-P. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: CVPR 2011. 2011. P. 585–592. DOI: [10.1109/CVPR.2011.5995437](https://doi.org/10.1109/CVPR.2011.5995437). Citado na p. 59.
- CORTES, C.; VAPNIK, V. Support Vector Networks. **Machine Learning**, v. 20, p. 273–297, 1995. Citado nas pp. 28, 29.

- COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, v. 13, n. 1, p. 21–27, 1967. DOI: [10.1109/TIT.1967.1053964](https://doi.org/10.1109/TIT.1967.1053964). Citado na p. 41.
- COX, I.; GHOSN, J.; YIANILOS, P. Feature-based face recognition using mixture-distance. In: PROCEEDINGS CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1996. P. 209–216. DOI: [10.1109/CVPR.1996.517076](https://doi.org/10.1109/CVPR.1996.517076). Citado na p. 51.
- DALAL, N.; TRIGGS, B. Histograms of Oriented Gradients for Human Detection. **Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on**, v. 1, p. 886–893, 2005. Disponível em: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1467360>. Citado na p. 28.
- DANGETI, P. **Statistics for Machine Learning: Techniques for Exploring Supervised, Unsupervised, and Reinforcement Learning Models with Python and R**. Packt Publishing, 2017. ISBN 1788295757. Citado na p. 43.
- DANIELSSON, P.-E. Euclidean distance mapping. **Computer Graphics and Image Processing**, v. 14, n. 3, p. 227–248, 1980. ISSN 0146-664X. DOI: [https://doi.org/10.1016/0146-664X\(80\)90054-4](https://doi.org/10.1016/0146-664X(80)90054-4). Disponível em: <https://www.sciencedirect.com/science/article/pii/0146664X80900544>>. Citado na p. 42.
- DANTCHEVA, A.; CHEN, C.; ROSS, A. Can facial cosmetics affect the matching accuracy of face recognition systems? In: p. 391–398. ISBN 978-1-4673-1384-1. DOI: [10.1109/BTAS.2012.6374605](https://doi.org/10.1109/BTAS.2012.6374605). Citado na p. 53.
- DAVISON, K. K.; BIRCH, L. L. Weight status, parent reaction, and self-concept in five-year-old girls. **Pediatrics**, v. 107, n. 1, p. 46–53, 2001. DOI: [10.1542/peds.107.1.46](https://doi.org/10.1542/peds.107.1.46). Citado na p. 17.
- DENG, J.; GUO, J.; XUE, N.; ZAFEIRIOU, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 4685–4694. DOI: [10.1109/CVPR.2019.00482](https://doi.org/10.1109/CVPR.2019.00482). Citado nas pp. 24, 39, 40.
- DENG, J.; GUO, J.; ZHOU, Y.; YU, J.; KOTSIA, I.; ZAFEIRIOU, S. **RetinaFace: Single-stage Dense Face Localisation in the Wild**. arXiv, 2019. DOI: [10.48550/ARXIV.1905.00641](https://doi.org/10.48550/ARXIV.1905.00641). Disponível em: <https://arxiv.org/abs/1905.00641>>. Citado nas pp. 26, 30–32.
- EMAMI, S.; SUCIU, V. Facial Recognition using OpenCV. **Journal of Mobile, Embedded and Distributed Systems**, v. 4, mar. 2012. Citado na p. 27.

- FAN, F. J.; SHI, Y. Effects of data quality and quantity on deep learning for protein-ligand binding affinity prediction. **Bioorganic and Medicinal Chemistry**, v. 72, p. 117003, 2022. ISSN 0968-0896. DOI: <https://doi.org/10.1016/j.bmc.2022.117003>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0968089622003960>>. Citado na p. 22.
- FARRAJOTA, M. **INRIA Pedestrian — dbcollection 0.2.6 documentation**. 2017. Disponível em: <https://dbcollection.readthedocs.io/en/latest/datasets/inria_ped.html>. Citado na p. 29.
- FU, Y.; HOSPEDALES, T. M.; XIANG, T.; XIONG, J.; GONG, S.; WANG, Y.; YAO, Y. Robust Subjective Visual Property Prediction from Crowdsourced Pairwise Labels. In: IEEE TPAMI. 2016. Citado nas pp. 49, 59.
- FUKUSHIMA, K. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. **Biological Cybernetics**, v. 36, p. 193–202, 1980. Citado na p. 23.
- FYDANAKI, A.; GERADTS, Z. J. M. H. Evaluating OpenFace: an open-source automatic facial comparison algorithm for forensics. **Forensic Sciences Research**, 2018. Disponível em: <<https://ivi.fnwi.uva.nl/isis/publications/2018/FydanakiFSR2018>>. Citado na p. 38.
- GRM, K.; Š TRUC, V.; ARTIGES, A.; CARON, M.; EKENEL, H. K. Strengths and weaknesses of deep learning models for face recognition against image degradations. **IET Biometrics**, Institution of Engineering e Technology (IET), v. 7, n. 1, p. 81–89, out. 2017. DOI: 10.1049/iet-bmt.2017.0083. Disponível em: <<https://doi.org/10.1049%5C%2Fiet-bmt.2017.0083>>. Citado na p. 37.
- GROUP, V. G. **Visual Geometry Group**. Acessado em 11 de janeiro de 2023. Disponível em: <<https://www.robots.ox.ac.uk/~vgg/>>. Citado na p. 36.
- HAAS, C. F.; CHAMPION, A.; SECOR, D. Motivating factors for seeking cosmetic surgery. **Plastic Surgical Nursing**, v. 28, n. 4, p. 177–182, 2008. DOI: 10.1097/psn.0b013e31818ea832. Citado na p. 17.
- HACHIM EL KHIYARI, H. W. Face Recognition across Time Lapse Using Convolutional Neural Networks. In: JOURNAL of Information Security, 2016. 2016. DOI: 10.4236/jis.2016.73010. Citado na p. 50.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. **Deep Residual Learning for Image Recognition**. arXiv, 2015. DOI: 10.48550/ARXIV.1512.03385. Disponível em: <<https://arxiv.org/abs/1512.03385>>. Citado na p. 38.

- HILL, T. **2017 Tribeca Film Festival - Screenings And Parties**. 2017. Disponível em: <https://www.gettyimages.pt/detail/fotografia-de-not%5C%C3%5CADcias/lil-kim-attends-the-cant-stop-wont-stop-fotografia-de-not%5C%C3%5CADcias/674272078?phrase=Lil%5C%27%5C%2BKim&adppopup=true>. Citado na p. 19.
- HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDRETTA, M.; ADAM, H. **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**. arXiv, 2017. DOI: [10.48550/ARXIV.1704.04861](https://doi.org/10.48550/ARXIV.1704.04861). Disponível em: <https://arxiv.org/abs/1704.04861>. Citado nas pp. 31, 32.
- HUANG, G. B.; MATTAR, M.; LEE, H.; LEARNED-MILLER, E. Learning to Align from Scratch. In: NIPS. 2012. Citado nas pp. 66, 97.
- HUANG, G. B.; RAMESH, M.; BERG, T.; LEARNED-MILLER, E. **Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments**. Out. 2007. Citado nas pp. 25, 34, 38, 39, 41, 60, 65, 80.
- LANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. **SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and less than 0.5MB model size**. arXiv, 2016. DOI: [10.48550/ARXIV.1602.07360](https://doi.org/10.48550/ARXIV.1602.07360). Disponível em: <https://arxiv.org/abs/1602.07360>. Citado na p. 37.
- IBM. **Redes Neurais**. Disponível em: <https://www.ibm.com/br-pt/cloud/learn/neural-networks> – acesso em 10 jan. 2023. Citado nas pp. 22, 23.
- KING, D. **dnn_{face,ecognition_e.cpp}**. 2017. Disponível em: http://dlib.net/dnn_face_recognition_ex.cpp.html. Citado na p. 39.
- KING, D. **Overview**. Out. 2015. Disponível em: <http://dlib.net/intro.html>. Citado na p. 38.
- KING, D. E. Dlib-ml: A Machine Learning Toolkit. **Journal of Machine Learning Research**, v. 10, p. 1755–1758, 2009. Citado nas pp. 26, 28.
- KING, D. E. **Max-Margin Object Detection**. arXiv, 2015. DOI: [10.48550/ARXIV.1502.00046](https://doi.org/10.48550/ARXIV.1502.00046). Disponível em: <https://arxiv.org/abs/1502.00046>. Citado na p. 29.
- KOHLI, N.; YADAV, D.; NOORE, A. Multiple Projective Dictionary Learning to Detect Plastic Surgery for Face Verification. **IEEE Access**, v. 3, p. 2572–2580, 2015. DOI: [10.1109/ACCESS.2015.2505243](https://doi.org/10.1109/ACCESS.2015.2505243). Citado nas pp. 51, 53.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In: PEREIRA, F.; BURGESS, C.; BOTTOU, L.; WEINBERGER, K. (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2012. v. 25. Disponível em: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Citado na p. 37.

- KULKARNI, A.; CHONG, D.; BATARSEH, F. A. 5 - Foundations of data imbalance and solutions for a data democracy. In: BATARSEH, F. A.; YANG, R. (Ed.). **Data Democracy**. Academic Press, 2020. P. 83–106. ISBN 978-0-12-818366-3. DOI: <https://doi.org/10.1016/B978-0-12-818366-3.00005-8>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780128183663000058>. Citado na p. 64.
- KUMAR, N.; BERG, A. C.; BELHUMEUR, P. N.; NAYAR, S. K. Attribute and simile classifiers for face verification. In: 2009 IEEE 12th International Conference on Computer Vision. 2009. P. 365–372. DOI: [10.1109/ICCV.2009.5459250](https://doi.org/10.1109/ICCV.2009.5459250). Citado na p. 55.
- LANITIS, A.; DRAGANOVA, C.; CHRISTODOULOU, C. Comparing different classifiers for automatic age estimation. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, v. 34, n. 1, p. 621–628, 2004. DOI: [10.1109/TSMCB.2003.817091](https://doi.org/10.1109/TSMCB.2003.817091). Citado na p. 49.
- LANITIS, A.; TAYLOR, C. J.; COOTES, T. Toward Automatic Simulation of Aging Effects on Face Images. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 24, p. 442–455, 2002. Citado na p. 49.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791). Citado na p. 23.
- LECUN, Y.; CORTES, C. MNIST handwritten digit database, 2010. Disponível em: <http://yann.lecun.com/exdb/mnist/>. Citado na p. 23.
- LEE, S.; KWAK, S.; CHO, M. **Universal Bounding Box Regression and Its Applications**. arXiv, 2019. DOI: [10.48550/ARXIV.1904.06805](https://doi.org/10.48550/ARXIV.1904.06805). Disponível em: <https://arxiv.org/abs/1904.06805>. Citado na p. 30.
- LEHMANN, E. L.; ROMANO, J. P.; CASELLA, G. **Testing statistical hypotheses**. Springer, 2005. v. 3. Citado nas pp. 45, 46.
- LI, S. Z. Face Variation. In: **Encyclopedia of Biometrics**. Edição: Anil K. Editor Jain. 2. ed.: Springer, 2015. P. 538–544. Citado na p. 18.
- LING, H.; SOATTO, S.; RAMANATHAN, N.; JACOBS, D. W. A Study of Face Recognition as People Age. In: 2007 IEEE 11th International Conference on Computer Vision. 2007. P. 1–8. DOI: [10.1109/ICCV.2007.4409069](https://doi.org/10.1109/ICCV.2007.4409069). Citado na p. 50.
- LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C.-Y.; BERG, A. C. SSD: Single Shot MultiBox Detector. In: **COMPUTER Vision – ECCV 2016**. Springer International Publishing, 2016. P. 21–37. DOI: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2). Citado nas pp. 26–28, 32.
- LOHNINGER, H. **Quartile**. 2001. Disponível em: <http://www.lohninger.com/tmdatanal.html>. Citado na p. 84.

- LOWEDAVID, G. Distinctive Image Features from Scale-Invariant Keypoints. **International Journal of Computer Vision**, 2004. Citado na p. 52.
- MAHMOOD, Z.; MUHAMMAD, N.; BIBI, N.; ALI, T. A review on state-of-the-art face recognition approaches. **Fractals**, v. 25, mar. 2017. DOI: [10.1142/S0218348X17500256](https://doi.org/10.1142/S0218348X17500256). Citado na p. 18.
- MAISEL, A.; WALDMAN, A.; FURLAN, K.; WEIL, A.; SACOTTE, K.; LAZAROFF, J. M.; LIN, K.; ARANZAZU, D.; AVRAM, M. M.; BELL, A.; CARTEE, T. V.; CAZZANIGA, A.; CHAPAS, A.; CRISPIN, M. K.; CROIX, J. A.; DIGIORGIO, C. M.; DOVER, J. S.; GOLDBERG, D. J.; GOLDMAN, M. P.; GREEN, J. B.; GRIFFIN, C. L.; HAIMOVIC, A. D.; HAUSAUER, A. K.; HERNANDEZ, S. L.; HSU, S.; IBRAHIM, O.; JONES, D. H.; KAUFMAN, J.; KILMER, S. L.; LEE, N. Y.; MCDANIEL, D. H.; SCHLESSINGER, J.; TANZI, E.; WEISS, E. T.; WEISS, R. A.; WU, D.; POON, E.; ALAM, M. Self-reported Patient Motivations for Seeking Cosmetic Procedures. **JAMA Dermatology**, v. 154, n. 10, p. 1167–1174, out. 2018. ISSN 2168-6068. DOI: [10.1001/jamadermatol.2018.2357](https://doi.org/10.1001/jamadermatol.2018.2357). eprint: https://jamanetwork.com/journals/jamadermatology/articlepdf/2696640/jamadermatology_maisel_2018_oi_180037.pdf. Disponível em: <https://doi.org/10.1001/jamadermatol.2018.2357>. Citado na p. 17.
- MARSICO, M.; NAPPI, M.; RICCIO, D.; WECHSLER, H. Robust face recognition after plastic surgery using region-based approaches. **Pattern Recognition**, v. 48, n. 4, p. 1261–1276, 2015. ISSN 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2014.10.004>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320314003884>. Citado na p. 52.
- MCCOMBES, SHONA. **Sampling Methods | Types, Techniques and Examples**. Disponível em: <https://www.scribbr.com/methodology/sampling-methods/> – acesso em 28 de Maio de 2023. Citado na p. 44.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na p. 22.
- MICHAS, F. **Cosmetic procedures number by top countries 2020**. Jul. 2022a. Disponível em: <https://www.statista.com/statistics/293356/leading-countries-by-total-number-of-cosmetic-procedures/>. Citado na p. 18.
- MICHAS, F. **Cosmetic procedures number by top countries 2021**. Mai. 2023. Disponível em: <https://www.statista.com/statistics/293356/leading-countries-by-total-number-of-cosmetic-procedures/>. Citado nas pp. 74, 75.

- MICHAS, F. **Surgical and nonsurgical cosmetic procedures U.S.** 2021. Mai. 2022b. Disponível em: <<https://www.statista.com/statistics/281211/us-surgical-and-nonsurgical-cosmetic-procedures/>>. Citado na p. 18.
- MITCHELL, T. M. **Machine learning, International Edition**. McGraw-Hill, 1997. (McGraw-Hill Series in Computer Science). ISBN 978-0-07-042807-2. Disponível em: <<https://www.worldcat.org/oclc/61321007>>. Citado na p. 22.
- MITTAL, N.; SINGH, R. 19 - Criminal identification system using face detection with artificial intelligence. In: TANWAR, S. (Ed.). **Blockchain Applications for Healthcare Informatics**. Academic Press, 2022. P. 421–430. ISBN 978-0-323-90615-9. DOI: <https://doi.org/10.1016/B978-0-323-90615-9.00001-3>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780323906159000013>>. Citado na p. 18.
- MOEINI, A.; FAEZ, K.; MOEINI, H. Face recognition across makeup and plastic surgery from real-world images. **Journal of Electronic Imaging**, v. 24, 053028, p. 053028, set. 2015. DOI: [10.1117/1.JEI.24.5.053028](https://doi.org/10.1117/1.JEI.24.5.053028). Citado nas pp. 51, 52.
- MOSCHOLOU, S.; PAPAIOANNOU, A.; SAGONAS, C.; DENG, J.; KOTSIA, I.; ZAFEIRIOU, S. Agedb: the first manually collected, in-the-wild age database. In: 3. PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. 2017. v. 2, p. 5. Citado nas pp. 25, 59.
- NGUYEN, H.; BAI, L. Cosine Similarity Metric Learning for Face Verification. In: v. 6493, p. 709–720. ISBN 978-3-642-19308-8. DOI: [10.1007/978-3-642-19309-5_55](https://doi.org/10.1007/978-3-642-19309-5_55). Citado na p. 42.
- NOLAN, I. T.; KUHNER, C. J.; DY, G. W. Demographic and temporal trends in transgender identities and gender confirming surgery. **Translational andrology and urology**, v. 8,3, 2019. DOI: [10.21037/tau.2019.04.09](https://doi.org/10.21037/tau.2019.04.09). Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6626314/>>. Citado na p. 99.
- NT. **Stênio Garcia rebate críticas e explica porque fez harmonização AOS 91 anos**. UOL, jun. 2023. Disponível em: <<https://natelinha.uol.com.br/famosos/2023/06/14/stenio-garcia-rebate-criticas-e-explica-porque-fez-harmonizacao-aos-91-anos-198385.php>>. Citado na p. 19.
- OLIVA, A.; TORRALBA, A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. **International Journal of Computer Vision**, v. 42, p. 145–175, 2001. Citado na p. 53.
- OPENCV. **Cascade Classifier**. Disponível em: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html – acesso em 15 jan. 2023. Citado nas pp. 26, 27.

- PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Deep Face Recognition. In: BRITISH Machine Vision Conference. 2015. Citado nas pp. 33, 36, 37, 55, 60.
- PENEV, P. S.; ATICK, J. J. Local feature analysis: a general statistical theory for object representation. **Network: Computation in Neural Systems**, Taylor e Francis, v. 7, n. 3, p. 477–500, 1996. DOI: [10.1088/0954-898X_7_3_002](https://doi.org/10.1088/0954-898X_7_3_002). eprint: https://doi.org/10.1088/0954-898X_7_3_002. Disponível em: https://doi.org/10.1088/0954-898X_7_3_002. Citado na p. 51.
- PHILLIPS, P. Support Vector Machines Applied to Face Recognition. In: KEARNS, M.; SOLLA, S.; COHN, D. (Ed.). **Advances in Neural Information Processing Systems**. MIT Press, 1998. v. 11. Disponível em: <https://proceedings.neurips.cc/paper/1998/file/a2cc63e065705fe938a4dda49092966f-Paper.pdf>. Citado na p. 41.
- PHILLIPS, P.; FLYNN, P.; SCRUGGS, T.; BOWYER, K.; CHANG, J.; HOFFMAN, K.; MARQUES, J.; MIN, J.; WOREK, W. Overview of the face recognition grand challenge. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005. v. 1, 947–954 vol. 1. DOI: [10.1109/CVPR.2005.268](https://doi.org/10.1109/CVPR.2005.268). Citado na p. 17.
- PHILLIPS, P.; WECHSLER, H.; HUANG, J.; RAUSS, P. J. The FERET database and evaluation procedure for face-recognition algorithms. **Image and Vision Computing**, v. 16, n. 5, p. 295–306, 1998. ISSN 0262-8856. DOI: [https://doi.org/10.1016/S0262-8856\(97\)00070-X](https://doi.org/10.1016/S0262-8856(97)00070-X). Disponível em: <https://www.sciencedirect.com/science/article/pii/S026288569700070X>. Citado na p. 17.
- PIKE, R.; THOMPSON, K. **GitHub - google/mediapipe: Cross-platform, customizable ML solutions for live and streaming media.** — [github.com](https://github.com/google/mediapipe). 2019. <https://github.com/google/mediapipe>. [Accessed 28-Jan-2023]. Citado nas pp. 17, 26, 32.
- PLASTIC SURGEONS, A. S. of. **Plastic surgery statistics**. 2021. Disponível em: <https://www.plasticsurgery.org/news/plastic-surgery-statistics>. Citado na p. 74.
- PLASTIC surgery statistics 2020. Disponível em: <https://www.plasticsurgery.org/news/plastic-surgery-statistics>. Citado na p. 18.
- POYNTON, C. **Digital Video and HD: Algorithms and Interfaces**. Elsevier Science, 2003. (Computer Graphics). ISBN 9781558607927. Disponível em: <https://books.google.com.br/books?id=ra1lcAwgvq4C>. Citado na p. 28.
- RAMANATHAN, N.; CHELLAPPA, R. Face verification across age progression. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005. v. 2, 462–469 vol. 2. DOI: [10.1109/CVPR.2005.153](https://doi.org/10.1109/CVPR.2005.153). Citado na p. 49.

- RATHGEB, C.; DANTCHEVA, A.; BUSCH, C. Impact and Detection of Facial Beautification in Face Recognition: An Overview. **IEEE Access**, PP, p. 1–1, out. 2019. DOI: [10.1109/ACCESS.2019.2948526](https://doi.org/10.1109/ACCESS.2019.2948526). Citado nas pp. 18, 53, 54.
- RATHGEB, C.; DOGAN, D.; STOCKHARDT, F.; DE MARSICO, M.; BUSCH, C. Plastic Surgery: An Obstacle for Deep Face Recognition? In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2020. P. 3510–3517. DOI: [10.1109/CVPRW50498.2020.00411](https://doi.org/10.1109/CVPRW50498.2020.00411). Citado nas pp. 17, 20, 54, 55.
- RATHGEB, C.; DOGAN, D.; STOCKHARDT, F.; MARSICO, M. D.; BUSCH, C. Plastic Surgery: An Obstacle for Deep Face Recognition? In: 15TH IEEE Computer Society Workshop on Biometrics (CVPRW). 2020. Citado nas pp. 58, 64.
- REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: CORTES, C.; LAWRENCE, N.; LEE, D.; SUGIYAMA, M.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2015. v. 28. Disponível em: <https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf>. Citado na p. 28.
- RICANEK, K.; TESAFAYE, T. MORPH: A longitudinal image database of normal adult age-progression. In: v. 2006, p. 341–345. ISBN 0-7695-2503-2. DOI: [10.1109/FGR.2006.78](https://doi.org/10.1109/FGR.2006.78). Citado na p. 50.
- RICANEK, K. **MORPH Academic Database**. Disponível em: https://uncw.edu/oic/tech/morph_academic.html – acesso em 01 fev. 2023. 2006. Citado nas pp. 50, 59.
- SALAS, J.; BARROS VIDAL, F. de; MARTINEZ-TRINIDAD, F. Deep Learning: Current State. **IEEE Latin America Transactions**, v. 17, n. 12, p. 1925–1945, 2019. DOI: [10.1109/TLA.2019.9011537](https://doi.org/10.1109/TLA.2019.9011537). Citado na p. 22.
- SANDBERG, D. **Daidsandberg/facenet: Face recognition using Tensorflow**. 2016. Disponível em: <https://github.com/daidsandberg/facenet>. Citado na p. 36.
- SCHAPIRE, R. A brief introduction to boosting. English (US). **IJCAI International Joint Conference on Artificial Intelligence**, v. 2, p. 1401–1406, dez. 1999. 16th International Joint Conference on Artificial Intelligence, IJCAI 1999 ; Conference date: 31-07-1999 Through 06-08-1999. ISSN 1045-0823. Citado na p. 26.
- SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. FaceNet: A unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, jun. 2015. DOI: [10.1109/cvpr.2015.7298682](https://doi.org/10.1109/cvpr.2015.7298682). Disponível em: <https://doi.org/10.1109%5C%2Fcvpr.2015.7298682>. Citado nas pp. 35–37.

- SEDGWICK, P. Confidence intervals: predicting uncertainty. **BMJ**, BMJ Publishing Group Ltd, v. 344, 2012. DOI: [10.1136/bmj.e3147](https://doi.org/10.1136/bmj.e3147). eprint: <https://www.bmj.com/content/344/bmj.e3147.full.pdf>. Disponível em: <https://www.bmj.com/content/344/bmj.e3147>>. Citado na p. 89.
- SENGUPTA, S.; CHEN, J.-C.; CASTILLO, C.; PATEL, V. M.; CHELLAPPA, R.; JACOBS, D. W. Frontal to profile face verification in the wild. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). 2016. P. 1–9. DOI: [10.1109/WACV.2016.7477558](https://doi.org/10.1109/WACV.2016.7477558). Citado na p. 25.
- SERENGIL, S.; OZPINAR, A. LightFace: A Hybrid Deep Face Recognition Framework. In: IEEE. 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). 2020. P. 23–27. DOI: [10.1109/ASYU50717.2020.9259802](https://doi.org/10.1109/ASYU50717.2020.9259802). Disponível em: <https://doi.org/10.1109/ASYU50717.2020.9259802>>. Citado nas pp. 61, 77, 97.
- SERENGIL, S. **Deep Face Recognition with VGG-Face in Keras**. 2018a. Online; accessed 2023-03-10. Disponível em: <https://sefiks.com/2018/08/06/deep-face-recognition-with-keras/>>. Citado na p. 28.
- SERENGIL, S. **Face recognition with FaceNet in Keras**. Jan. 2018b. Disponível em: <https://sefiks.com/2018/09/03/face-recognition-with-facenet-in-keras/>>. Citado na p. 36.
- SHAPIRO, S. S.; WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika**, [Oxford University Press, Biometrika Trust], v. 52, n. 3/4, p. 591–611, 1965. ISSN 00063444. Disponível em: <http://www.jstor.org/stable/2333709>>. Acesso em: 26 mai. 2023. Citado na p. 46.
- SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. arXiv, 2014. DOI: [10.48550/ARXIV.1409.1556](https://doi.org/10.48550/ARXIV.1409.1556). Disponível em: <https://arxiv.org/abs/1409.1556>>. Citado na p. 28.
- SINGH, R.; VATSA, M.; BHATT, H.; BHARADWAJ, S.; NOORE, A.; NOOREYEZDAN, S. Plastic Surgery: A New Dimension to Face Recognition. **Information Forensics and Security, IEEE Transactions on**, v. 5, p. 441–448, out. 2010. DOI: [10.1109/TIFS.2010.2054083](https://doi.org/10.1109/TIFS.2010.2054083). Citado nas pp. 17–19, 52–55.
- SINGH, R.; VATSA, M.; NOORE, A. Effect of plastic surgery on face recognition: A preliminary study. In: p. 72–77. DOI: [10.1109/CVPRW.2009.5204287](https://doi.org/10.1109/CVPRW.2009.5204287). Citado na p. 51.
- SINGH, R.; VATSA, M.; NOORE, A.; SINGH, S. Age Transformation for Improving Face Recognition Performance. In: GHOSH, A.; DE, R. K.; PAL, S. K. (Ed.). **Pattern Recognition and Machine Intelligence**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. P. 576–583. ISBN 978-3-540-77046-6. Citado na p. 99.

- SINGH, R.; VATSA, M.; NOORE, A. Face Recognition with Disguise and Single Gallery Images. **Image Vision Comput.**, Butterworth-Heinemann, USA, v. 27, n. 3, p. 245–257, fev. 2009b. ISSN 0262-8856. DOI: [10.1016/j.imavis.2007.06.010](https://doi.org/10.1016/j.imavis.2007.06.010). Disponível em: <https://doi.org/10.1016/j.imavis.2007.06.010>. Citado na p. 51.
- SINGH, S.; PRASAD, S. Techniques and Challenges of Face Recognition: A Critical Review. **Procedia Computer Science**, v. 143, p. 536–543, 2018. 8th International Conference on Advances in Computing Communications (ICACC-2018). ISSN 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2018.10.427>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050918321252>. Citado na p. 16.
- STERNBERG, S. R. (A. Method and apparatus for pattern recognition and detection. n. 4414685, nov. 1983. Disponível em: <https://www.freepatentsonline.com/4414685.html>. Citado na p. 28.
- STUDENT. The Probable Error of a Mean. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 6, n. 1, p. 1–25, 1908. ISSN 00063444. Disponível em: <http://www.jstor.org/stable/2331554>. Acesso em: 26 mai. 2023. Citado nas pp. 47, 71.
- SUN, Y.; WANG, X.; TANG, X. Deep Learning Face Representation from Predicting 10,000 Classes. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 1891–1898, set. 2014. DOI: [10.1109/CVPR.2014.244](https://doi.org/10.1109/CVPR.2014.244). Citado nas pp. 34, 35.
- SZEGEDY, C.; VANHOUCHE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. **Rethinking the Inception Architecture for Computer Vision**. arXiv, 2015. DOI: [10.48550/ARXIV.1512.00567](https://arxiv.org/abs/1512.00567). Disponível em: <https://arxiv.org/abs/1512.00567>. Citado na p. 37.
- TAIGMAN, Y.; YANG, M.; RANZATO, M.; WOLF, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014a. P. 1701–1708. DOI: [10.1109/CVPR.2014.220](https://doi.org/10.1109/CVPR.2014.220). Citado nas pp. 17, 26, 33, 34, 37.
- TAIGMAN, Y.; YANG, M.; RANZATO, M.; WOLF, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014b. P. 1701–1708. DOI: [10.1109/CVPR.2014.220](https://doi.org/10.1109/CVPR.2014.220). Citado na p. 79.
- THOMPSON, J. K.; HEINBERG, L. J.; ALTABE, M.; TANTLEFF-DUNN, S. Exacting beauty: Theory, assessment, and treatment of body image disturbance., 1999. DOI: [10.1037/10312-000](https://doi.org/10.1037/10312-000). Citado na p. 17.

- TURK, M.; PENTLAND, A. Eigenfaces for Recognition. **Journal of Cognitive Neuroscience**, v. 3, n. 1, p. 71–86, jan. 1991. ISSN 0898-929X. DOI: [10.1162/jocn.1991.3.1.71](https://doi.org/10.1162/jocn.1991.3.1.71). eprint: <https://direct.mit.edu/jocn/article-pdf/3/1/71/1932018/jocn.1991.3.1.71.pdf>. Disponível em: <https://doi.org/10.1162/jocn.1991.3.1.71>>. Citado na p. 16.
- TURNEY, SHAUN. **Probability Distribution | Formula, Types, and Examples**. Disponível em: <https://www.scribbr.com/statistics/probability-distributions/> – acesso em 28 de Maio de 2023. Citado na p. 44.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: PROCEEDINGS of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. 2001. v. 1, p. i–i. DOI: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517). Citado nas pp. 26, 27.
- WALKER, C. E.; KRUMHUBER, E. G.; DAYAN, S.; FURNHAM, A. Effects of social media use on desire for cosmetic surgery among young women. **Current Psychology**, v. 40, n. 7, p. 3355–3364, 2019. DOI: [10.1007/s12144-019-00282-1](https://doi.org/10.1007/s12144-019-00282-1). Citado na p. 18.
- WILCOXON, F. Individual Comparisons by Ranking Methods. **Biometrics Bulletin**, [International Biometric Society, Wiley], v. 1, n. 6, p. 80–83, 1945. ISSN 00994987. Disponível em: <http://www.jstor.org/stable/3001968>>. Acesso em: 26 mai. 2023. Citado nas pp. 47, 71.
- WOLF, L.; HASSNER, T.; MAOZ, I. Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011. 2011. P. 529–534. DOI: [10.1109/CVPR.2011.5995566](https://doi.org/10.1109/CVPR.2011.5995566). Citado nas pp. 25, 41.
- YANG, S.; LUO, P.; LOY, C. C.; TANG, X. WIDER FACE: A Face Detection Benchmark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. Citado na p. 31.
- YASSIN, D. H. P.; HOQUE, S.; DERAVID, F. Age Sensitivity of Face Recognition Algorithms. In: 2013 Fourth International Conference on Emerging Security Technologies. 2013. P. 12–15. DOI: [10.1109/EST.2013.8](https://doi.org/10.1109/EST.2013.8). Citado na p. 50.
- YE, R. Z.; SUBRAMANIAN, A.; DIEDRICH, D.; LINDROTH, H.; PICKERING, B.; HERASEVICH, V. Effects of Image Quality on the Accuracy Human Pose Estimation and Detection of Eye Lid Opening/Closing Using Openpose and DLib. **Journal of Imaging**, v. 8, n. 12, 2022. ISSN 2313-433X. DOI: [10.3390/jimaging8120330](https://doi.org/10.3390/jimaging8120330). Disponível em: <https://www.mdpi.com/2313-433X/8/12/330>>. Citado na p. 39.
- ZHANG, K.; ZHANG, Z.; LI, Z.; QIAO, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. **IEEE Signal Processing Letters**, v. 23, n. 10, p. 1499–1503, out. 2016. ISSN 1070-9908. DOI: [10.1109/LSP.2016.2603342](https://doi.org/10.1109/LSP.2016.2603342). Citado nas pp. 26, 30, 31.

-
- ZHAO, H.; SHI, Y.; TONG, X.; YING, X.; ZHA, H. Qamface: Quadratic Additive Angular Margin Loss For Face Recognition. In: 2020 IEEE International Conference on Image Processing (ICIP). 2020. P. 1901–1905. DOI: [10.1109/ICIP40778.2020.9191004](https://doi.org/10.1109/ICIP40778.2020.9191004). Citado na p. 40.
- ZHAO, W.; CHELLAPPA, R.; PHILLIPS, P. J.; ROSENFELD, A. Face Recognition: A Literature Survey. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 35, n. 4, p. 399–458, dez. 2003. ISSN 0360-0300. DOI: [10.1145/954339.954342](https://doi.org/10.1145/954339.954342). Disponível em: <https://doi.org/10.1145/954339.954342>. Citado nas pp. 24, 49.
- ZHENG, T.; DENG, W. **Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments**. Fev. 2018. Citado nas pp. 25, 41.
- ZHENG, T.; DENG, W.; HU, J. Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments. **CoRR**, abs/1708.08197, 2017. arXiv: [1708.08197](http://arxiv.org/abs/1708.08197). Disponível em: <http://arxiv.org/abs/1708.08197>. Citado nas pp. 25, 41.
- ZHONG, Y.; DENG, W.; HU, J.; ZHAO, D.; LI, X.; WEN, D. SFace: Sigmoid-Constrained Hypersphere Loss for Robust Face Recognition. **IEEE Transactions on Image Processing**, v. 30, p. 2587–2598, 2021. DOI: [10.1109/TIP.2020.3048632](https://doi.org/10.1109/TIP.2020.3048632). Citado na p. 40.
- ZHOU, E.; CAO, Z.; YIN, Q. **Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not?** arXiv, 2015. DOI: [10.48550/ARXIV.1501.04690](https://doi.org/10.48550/ARXIV.1501.04690). Disponível em: <https://arxiv.org/abs/1501.04690>. Citado na p. 55.