



**Universidade de Brasília
Faculdade de Tecnologia**

**Avaliando vieses de gênero e raça no
reconhecimento facial automático em modelos
de aprendizagem profunda**

Marina Carvalho Soares de Queiroz

**TRABALHO DE GRADUAÇÃO
ENGENHARIA DE CONTROLE E AUTOMAÇÃO**

Brasília
2022

**Universidade de Brasília
Faculdade de Tecnologia**

**Avaliando vieses de gênero e raça no
reconhecimento facial automático em modelos
de aprendizagem profunda**

Marina Carvalho Soares de Queiroz

Trabalho de Graduação submetido como re-
quisito parcial para obtenção do grau de Enge-
nheira de Controle e Automação.

Orientador: Prof. Dr. Flávio de Barros Vidal

Brasília
2022

C331a Carvalho Soares de Queiroz, Marina.
Avaliando vieses de gênero e raça no reconhecimento facial automático em modelos de aprendizagem profunda / Marina Carvalho Soares de Queiroz; orientador Flávio de Barros Vidal. -- Brasília, 2022.
74 p.

Trabalho de Graduação em Engenharia de Controle e Automação -- Universidade de Brasília, 2022.

1. Reconhecimento Facial. 2. Viés de gênero e raça. 3. Base de dados Faces. 4. Detector Facial. I. Vidal, Flávio de Barros, orient. II. Título

**Universidade de Brasília
Faculdade de Tecnologia**

Avaliando vieses de gênero e raça no reconhecimento facial automático em modelos de aprendizagem profunda

Marina Carvalho Soares de Queiroz

Trabalho de Graduação submetido como requisito parcial para obtenção do grau de Engenharia de Controle e Automação.

Trabalho aprovado. Brasília, 12 de Setembro de 2022:

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

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

Prof. Dr. Marcus Vinicius Lamar,
UnB/IE/CIC
Examinador interno

Brasília
2022

Resumo

Modelos de detecção e reconhecimento facial já são uma realidade nos dias atuais, porém existem ressalvas para sua utilização. Dentre estas ressalvas, surgem a baixa na acurácia do reconhecimento e identificação de faces em diferentes raças e gêneros, identificada como um possível viés no processo de reconhecimento facial automático. Com a intenção de demonstrar e procurar uma explicação para um possível existência destes vieses, esse trabalho propõe a realização do uso dos principais modelos e bibliotecas treinados na tarefa de reconhecimento facial, disponibilizados em (SERENGIL, S. I.; OZPINAR, 2020a). No uso dos modelos de detectores faciais (localizar a região da face na imagem) para identificar diversas raças em nos dois gêneros, masculino e feminino, o resultado teve pior desempenho para a raça indiana, o gênero feminino e no agrupamento de mulheres negras, indianas e homens indianos. Já no processo de reconhecimento facial entre duas imagens analisadas, em uma das faces é realizada uma cobertura em cada característica essencial da face: boca, olho direito, olho esquerdo e nariz. Neste processo foi feito para avaliar a influência de cada característica de modo a tentar identificar a possível relação entre alguma destas características e os agrupamentos com piores desempenhos demonstrados na detecção facial obtidos neste trabalho. Em todos os agrupamentos de raças e gêneros a cobertura do nariz resultou na redução da acurácia no reconhecimento facial, sendo um resultado geral, não podendo ser usado para analisar o viés em uma dada classe específica. Estudos feitos neste trabalho sobre como e quando foram treinados apontam como principal problema como bases de dados não-balanceadas e com pouca diversidade utilizadas para treinamento dos modelos, podem ser uma das principais fatores influenciadores no processo de redução da acurácia para estes grupamentos.

Palavras-chave: Reconhecimento Facial. Viés de gênero e raça. Base de dados Faces. Detec-tor Facial.

Abstract

Detection and facial recognition models are already a reality nowadays, but there are reservations about their use. Among these caveats is the low accuracy of recognition and identification of faces of different races and genders, identified as a possible bias in automatic facial recognition. To demonstrate and look for an explanation for the possible existence of these biases, this work proposes using the main models and libraries trained in the facial recognition task, available at (SERENGIL, S. I.; OZPINAR, 2020a). In the use of facial detector models (locate the region of the face in the image) to identify different races in both genders, male and female, the result had the worst performance for the Indian race, the female gender, and in the grouping of black, Indian women and Indian men. In the facial recognition process between two analyzed images, one of the faces is covered in each essential feature of the face: mouth, right eye, left eye, and nose. In this process, it was done to evaluate the influence of each characteristic to try to identify the possible relationship between any of these characteristics and the groups with the worst performances demonstrated in the face detection obtained in this work. In all race and gender groupings, nose covering resulted in reduced accuracy in facial recognition, being a general result and not being able to be used to analyze bias in a given class. Studies carried out in this work on how and when they were trained to indicate that the main problem is that unbalanced databases with little diversity used for training the models can be one of the main influencing factors in reducing the accuracy of these clusters.

Keywords: Face Recognition. Race and Gender Bias. Face Database. Face Detector.

Lista de ilustrações

Figura 1 – Exemplo de redes neurais profundas. Fonte: IBM (IBM, s.d.).	16
Figura 2 – Exemplo de uma rede convolucional simples	17
Figura 3 – Arquitetura dos estágios P-Net, R-Net e O-Net. Fonte: (XIANG; ZHU, 2017).	18
Figura 4 – Extração da imagem HOG. Fonte: (SULEIMAN et al., 2017).	19
Figura 5 – Hiperplano SVM.	19
Figura 6 – Pirâmide de recursos e módulos de contexto. Fonte: (DENG et al., 2020).	20
Figura 7 – Caixas delimitadoras detecção SSD. Fonte: (LIU et al., 2016).	21
Figura 8 – Modelo da estrutura Facenet. Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015).	22
Figura 9 – Modelo da estrutura Facenet. Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015).	23
Figura 10 – Arquitetura VGG-face. Fonte: (PARKHI; VEDALDI; ZISSERMAN, 2015).	24
Figura 11 – Modelo da estrutura do OpenFace. Fonte: (AMOS; LUDWICZUK; SATYA-NARAYANAN, 2016).	24
Figura 12 – Modelo da estrutura do Deepface. Fonte: (TAIGMAN et al., 2014).	25
Figura 13 – Fluxograma da metodologia proposta.	32
Figura 14 – Exemplo de imagens da base BFW.	34
Figura 15 – Comparações feitas para análise.	37

Lista de tabelas

Tabela 1 – Valores limites para distância de similaridade.	26
Tabela 2 – Distribuição de imagens da BFW. M - Mulheres, H - Homens.	33
Tabela 3 – Porcentagem de erro detectores com imagens originais e recortadas. . .	39
Tabela 4 – Matriz de confusão para melhor e pior detector.	39
Tabela 5 – Diferença quadrática entre os detectores.	40
Tabela 6 – Porcentagem de erro dos detectores com imagens originais e recortadas.	41
Tabela 7 – Matriz de confusão para o melhor e pior detector.	41
Tabela 8 – Diferença quadrática entre os detectores.	41
Tabela 9 – Porcentagem de erro dos detectores com imagens originais e recortadas.	42
Tabela 10 – Matriz de confusão para melhor e pior detector.	43
Tabela 11 – Diferença quadrática entre os detectores.	44
Tabela 12 – Acurácia do reconhecimento facial de acordo com as <i>features</i> cobertas para o reconhecimento facial cada raça.	45
Tabela 13 – Tabela verdade normalizada do pior e melhor resultado do reconheci- mento facial da raça asiática.	45
Tabela 14 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça asiática.	46
Tabela 15 – Tabela verdade normalizada do pior e melhor resultado do reconheci- mento facial da raça negra.	46
Tabela 16 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça negra.	46
Tabela 17 – Tabela verdade normalizada do pior e melhor resultado do reconheci- mento facial da raça indiana.	47
Tabela 18 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça indiana.	47
Tabela 19 – Tabela verdade normalizada do melhor e pior resultado do reconheci- mento facial da raça branca.	48
Tabela 20 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça branca.	48
Tabela 21 – Acurácia do reconhecimento facial de acordo com as <i>features</i> cobertas em uma das faces para o reconhecimento facial de gênero.	48
Tabela 22 – Tabela verdade normalizada do melhor e pior resultado do reconheci- mento facial entre faces do gênero masculino.	49
Tabela 23 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial do gênero masculino.	50
Tabela 24 – Tabela verdade do gênero feminino para imagens originais.	50

Tabela 25 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial do gênero feminino.	51
Tabela 26 – Acurácia do reconhecimento facial de acordo com as <i>features</i> cobertas para o reconhecimento facial entre faces de mulheres asiáticas.	51
Tabela 27 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres asiáticas.	52
Tabela 28 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de homens asiáticos.	52
Tabela 29 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres negras.	53
Tabela 30 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de homens negros.	53
Tabela 31 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres indianas.	54
Tabela 32 – Tabela verdade normalizada melhor e pior resultado do reconhecimento facial de homens indianos.	54
Tabela 33 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres brancas.	55
Tabela 34 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de homens brancos.	55
Tabela 35 – Acurácia do reconhecimento facial entre todas as raças e gêneros.	55
Tabela 36 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de todas as raças e gêneros.	56
Tabela 37 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial entre todas as raças.	57

Sumário

1	INTRODUÇÃO	12
1.1	Contextualização	12
1.2	Justificativa	13
1.3	Objetivos	14
1.3.1	Objetivo Geral	14
1.3.2	Objetivos Específicos	14
1.4	Organização do Trabalho	14
2	FUNDAMENTOS TEÓRICOS	15
2.1	Aprendizado profundo supervisionado em redes neurais	15
2.2	Detectores Faciais	17
2.2.1	Multi-task Cascaded Convolutional Networks - MTCNN	17
2.2.2	DLIB	18
2.2.3	Open Source Computer Vision Library - OpenCV	19
2.2.4	Retina Face	20
2.2.5	Single Shot MultiBox Detector - SSD	21
2.3	Modelos de reconhecimento facial	22
2.3.1	Facenet	22
2.3.2	Visual Geometry Group Face - VGG-face	23
2.3.3	OpenFace	24
2.3.4	DeepFace	25
2.4	Métricas de similaridade	26
2.4.1	Similaridade do Cosseno	26
2.4.2	Distância Euclideana	26
3	TRABALHOS RELACIONADOS	28
3.1	Reconhecimento Facial	28
3.2	Conjuntos de Dados	30
4	METODOLOGIA PROPOSTA	32
4.1	Levantamento bibliográfico	32
4.2	Levantamento e análise das principais bases de dados de imagens faciais	33
4.3	Estudo dos métodos de detecção, análise e reconhecimento facial	34
4.4	Avaliação dos resultados para cada métrica na análise facial	35

4.5	Investigar influência dos principais elementos faciais no processo de reconhecimento facial	36
4.6	Análise dos resultados obtidos	37
5	RESULTADOS	38
5.1	Desempenho dos detectores para identificação facial	38
5.1.1	Identificação da raça pela Face	38
5.1.2	Identificação de gênero pela face	40
5.1.3	Identificação de raça e gênero pela face	42
5.2	Influência dos principais elementos faciais no processo de reconhecimento facial	44
5.2.1	Reconhecimento facial dentro da mesma raça	44
5.2.1.1	Raça asiática	45
5.2.1.2	Raça negra	46
5.2.1.3	Raça indiana	46
5.2.1.4	Raça branca	47
5.2.2	Reconhecimento facial dentro do mesmo gênero	48
5.2.2.1	Gênero masculino	49
5.2.2.2	Gênero feminino	50
5.2.3	Reconhecimento facial dentro do mesmo gênero e mesma raça	51
5.2.3.1	Mulheres asiáticas	51
5.2.3.2	Homens asiáticos	52
5.2.3.3	Mulheres Negras	52
5.2.3.4	Homens Negros	53
5.2.3.5	Mulheres Indianas	53
5.2.3.6	Homens Indianos	54
5.2.3.7	Mulheres Brancas	54
5.2.3.8	Homens Brancos	55
5.2.4	Reconhecimento facial entre todas as raças e gêneros	55
5.3	Discussão dos Resultados	57
5.3.1	Identificação pela face	57
5.3.2	Reconhecimento facial	60
6	CONCLUSÃO	61
	REFERÊNCIAS	63

	ANEXOS	68
	ANEXO A – TABELAS NORMALIZADAS PARA FEATURES DE OLHOS E BOCA COBERTA NO RECONHECIMENTO FACIAL ENTRE RAÇAS	69
A.1	Raça asiática	69
A.2	Raça negra	69
A.3	Raça indiana	70
A.4	Raça branca	70
	ANEXO B – TABELAS NORMALIZADAS PARA FEATURES DE OLHOS E BOCA COBERTA NO RECONHECIMENTO FACIAL ENTRE GÊNEROS	71
B.1	Gênero masculino	71
B.2	Gênero feminino	71
	ANEXO C – TABELAS NORMALIZADAS PARA FEATURES DE OLHOS E BOCA COBERTA NO RECONHECIMENTO FACIAL MESMA RAÇA E GÊNERO	72
C.1	Mulheres asiáticas	72
C.2	Homens asiáticos	72
C.3	Mulheres Negras	73
C.4	Homens Negros	73
C.5	Mulheres Indianas	73
C.6	Homens Indianos	74
C.7	Mulheres Brancas	74
C.8	Homens Brancos	74

1 Introdução

1.1 Contextualização

A tecnologia de reconhecimento facial automático teve seu início na década de 60 com Woody Bledsoe ([BALLANTYNE; BOYER; HINES, 1996](#)). Inicialmente seus estudos focavam no reconhecimento de padrões, com o intuito de reconhecer letras. Com o desenvolvimento da tecnologia, ele viu uma oportunidade para introduzir faces ao invés de apenas letras. Assim desenvolveu um conjunto de dados para o treinamento de máquinas para os primeiros modelos de reconhecimento facial.

A base de dados utilizada vinha de livros policiais, dos quais se retirava um pequeno conjunto de imagens sendo que uma destas correspondia a face escolhida como parâmetro. O sucesso do resultado era medido pela razão entre a quantidade de respostas verdadeiras e o número de registros no banco de dados. A aprendizagem de máquina utilizada ainda necessitava de ajuda humana, uma pessoa extraía manualmente coordenadas de características faciais (*features*) essenciais, como olhos, boca e nariz. Essas características eram introduzidas no sistema computacional junto ao nome à quem pertenciam e usadas para o reconhecimento. Este método de aprendizagem é utilizado até hoje, porém hoje a extração destas características (*features*) também é feita pelo sistema computacional.

Devido a uma base de dados pequena, diferentes posições dos rostos, intensidade e ângulo da iluminação, expressão facial e idades, das que foram utilizadas para treinar o algoritmo, resultava numa alta taxa de erro. Alguns desses pontos, como o problema de ângulo, foram solucionados por um ajuste de normalização das distâncias calculadas de modo a sempre representar a imagem em uma orientação frontal ([KANADE, 1973](#)). Com o avanço da tecnologia, foi desenvolvida uma base de dados dentro da memória do computador. Agora poderia-se usar uma extensa quantidade de faces para treinamento dos algoritmos.

O interesse pelo método de reconhecimento facial ganhou visibilidade nos anos 90. Os esforços feitos por Woody Bledsoe não foram comercializados, pois foram financiados por uma empresa de inteligência artificial privada que manteve a pesquisa e resultados sob sigilo e só a partir de 1996 o aprendizado de máquina para o reconhecimento facial começou a ter viabilidade comercial ([PHILLIPS et al., 1998](#)). O governo norte americano criou o programa de Tecnologia de Reconhecimento Facial (FERET) ([PHILLIPS et al., 1998](#)) para utilizar como um modo de identificação e rastreamento de indivíduos onde o banco de dados utilizado para treinamento era feito em ambientes controlados. No século 21, se fez necessário o aumento de quantidade de faces em ambientes diversos, para que a precisão e acurácia dos sistemas melhorasse.

O aprendizado profundo de máquina foi demonstrado pela primeira vez por (MITCHELL, 1997) e nos anos seguintes aprimorado por diversos contribuintes. Seu uso para o reconhecimento facial foi de suma importância, visto que para um aumento significativo no banco de dados, seria inviável inserir manualmente cada *feature* e com o avanço do aprendizado profundo, o próprio algoritmo conseguiria extrair as *features* das faces.

Hoje existem diversos modelos de reconhecimento facial com diferentes camadas para o tratamento da imagem de entrada como, Facenet (SCHROFF; KALENICHENKO; PHILBIN, 2015), VGG-face (PARKHI; VEDALDI; ZISSERMAN, 2015), OpenFace (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) e outros. Com um maior número de faces nas bases de dados, hoje possui uma acurácia muito melhor que nos anos 90 e ainda está em processo para melhorias. Diversas faculdades e empresas investem nessa tecnologia para inúmeras utilizações (KAUR et al., 2020), como no desbloqueio de dispositivos móveis, investigações policiais, cadastramento de bancos digitais, filtros de redes sociais entre outros.

Sendo cada vez mais requerido na atualidade, há uma cobrança maior do desempenho ótimo de modelos de reconhecimento e detecção facial, onde um erro dessa aplicação pode causar um grande impacto ao indivíduo (CORBETT-DAVIES; GOEL, 2018). Esse tema tem sido muito abordado e questionado quanto sua capacidade de detectar e analisar corretamente faces de indivíduos de diversas raças e gêneros (BUOLAMWINI; GEBRU, 2018).

O trabalho realizado por (BUOLAMWINI; GEBRU, 2018) aponta a falta de uma base de dados com um balanceamento da diversidade de faces encontradas no mundo, entre raças e gêneros. Em sua conclusão acusa o algoritmo de reconhecimento facial de ser preconceituoso. O problema é que a carência de uma base de dados balanceada para treinamento dos modelos pode causar vieses, uma vez que quando se cobra do aprendizado de máquina algo que não lhe foi ensinado, não sendo algo intencional do algoritmo, sua classificação ou previsão pode não apresentar resultados satisfatórios para aquele novo conjunto de dados. Lembrando que a rede neural profunda ainda funciona como uma caixa preta onde os desenvolvedores não podem demonstrar exatamente o que ocorre entre os neurônios da aprendizagem e as analogias que são feitas por eles para chegar em determinado resultado.

1.2 Justificativa

Como será tratado no Capítulo 3, existem diversos estudos com a finalidade de tentar entender como e porque ocorre essa diferença de acurácia entre classes, tanto para detecção quanto para reconhecimento facial. Foram levantados pontos como uso de maquiagem, presença de cabelo e cor da pele. Nenhum deles resultou em uma explicação de porque ocorre o viés nos métodos de detecção e reconhecimento facial.

1.3 Objetivos

1.3.1 Objetivo Geral

Este trabalho tem como objetivo analisar os vieses presentes na detecção facial e uma tentativa de achar a *feature* que contribuem para o viés no reconhecimento facial de determinada classe. Assim como aprofundar a análise dos erros presentes na detecção e reconhecimento facial, realizando matrizes de confusão para entender onde o aprendizado de máquina comete maiores erros nas classes com vieses.

1.3.2 Objetivos Específicos

Para realizar o trabalho serão determinados objetivos específicos. O primeiro será ter conhecimento sobre o assunto e como os problemas apontados na Subseção 1.3.1 já foram tratados em outros trabalhos.

Em seguida escolher a forma de como seria feita essa avaliação nos modelos, a base de dados a ser utilizada e um conhecimento mais profundo sobre os modelos e detectores existentes no reconhecimento facial.

Com os resultados em mãos será feita uma análise do desempenho dos modelos de detecção facial assim como os de reconhecimento facial. Por fim demonstrar os resultados realizando uma discussão sobre o que foi encontrado.

1.4 Organização do Trabalho

Este trabalho foi organizado com seis capítulos. O Capítulo 1 que conta com uma breve contextualização do assunto, os objetivos e organização do trabalho. Seguido do Capítulo 2 onde são apresentados de maneira breve como são feitas e treinadas os modelos e detectores utilizados para realização deste trabalho.

Trazendo informações e tentativas já feitas para encontrar os motivos de vieses existentes no reconhecimento facial, tem-se o Capítulo 3. No Capítulo 4 será demonstrado como foi feito este trabalho, desde da ideia de realização até os resultados.

No Capítulo 5 são apresentados os resultados e uma discussão sobre as informações encontradas. Por fim, as conclusões e sugestões para trabalhos futuros estão demonstradas no Capítulo 6

2 Fundamentos teóricos

Os métodos de reconhecimento facial de aprendizagem profunda em redes neurais que serão utilizados são automáticos e supervisionados. O algoritmo utilizado será o (SERENGIL, S. I.; OZPINAR, 2020a). Em Serengil *et al.* utiliza o *framework LightFace* (SERENGIL, S. I.; OZPINAR, 2020b) que integra 5 métodos de detecção, 4 modelos de reconhecimento facial e 3 métricas de distância para medir a similaridade das faces.

Os métodos de detecção são o *Multi-task Cascaded Convolutional Networks (MTCNN)* (XIANG; ZHU, 2017), *Open Source Computer Vision Library (OpenCV)* (BRADSKI; KAEHLER, 2008), *DLIB* (KING, 2009), *Retina Face* (DENG *et al.*, 2020) e *Single Shot MultiBox Detector (SSD)* (LIU *et al.*, 2016). Os modelos de reconhecimento são o *Facenet* (SCHROFF; KALENICHENKO; PHILBIN, 2015), *Visual Geometry Group Face (VGG-Face)* (PARKHI; VEDALDI; ZISSERMAN, 2015), *OpenFace* (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) e *DeepFace* (TAIGMAN *et al.*, 2014). As métricas de distância utilizadas para determinar a similaridade entre as faces são similaridade de cosseno (SINGHAL, 2001), distância euclidiana (DANIELSSON, 1980) e sua normalização (SHI; MALIK, 2000). Cada métrica possui um limite de resultado para identificação ou reconhecimento facial. Cada combinação de modelo de reconhecimento, detecção facial e distância, possui uma acurácia diferente.

Nesse capítulo será introduzido uma breve teoria sobre como funciona o aprendizado de máquina, redes neurais artificiais (RNA) e redes neurais convolucionais (CNN). Seguidos de explicações dos modelos de reconhecimento facial, métodos de detecção facial e métricas de similaridade utilizados no código do (SERENGIL, S. I.; OZPINAR, 2020a).

2.1 Aprendizado profundo supervisionado em redes neurais

Para uso da tecnologia de aprendizado de máquina, a preparação dos dados a serem utilizados no treinamento dos modelos é essencial. A coleta e análise destes, influenciará diretamente na acurácia e aprendizagem do computador. No aprendizado supervisionado o modelo aprende com resultados pré-definidos, os dados são compostos por pares de valores de entradas e saídas desejadas (MITCHELL, 1997). Deste modo o algoritmo mapeia as informações e cria uma regra geral para aprender que uma determinada entrada deve gerar uma saída específica.

Os primeiros passos para aprendizagem profunda foram dados por Alexey Grigorevich Ivakhnenko em 1965 (IVAKHNENKO; LAPA, 1973), e sua construção utiliza redes neurais. Uma rede neural é feita de diversos processadores simples, comparados a neurônios,

conectados para criar as redes, similar ao cérebro humano. Neste trabalho será utilizado o aprendizado profundo, que consiste em diversas camadas de redes, as camadas de entrada, múltiplas camadas escondidas e a saída (Figura 1). Esses processadores são neurônios matemáticos que são capazes de processar e compreender dados criando um método para cada entrada gerar a saída desejada.

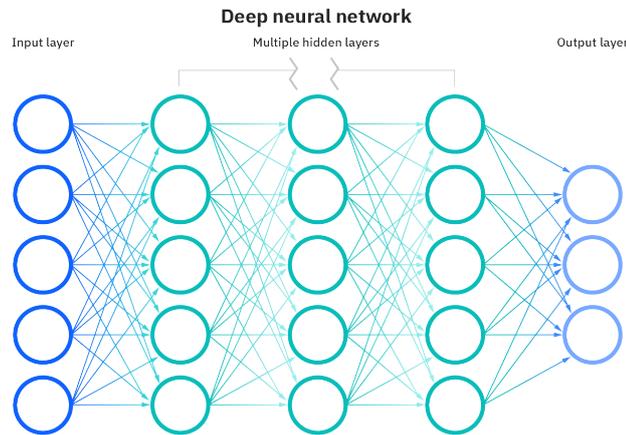


Figura 1 – Exemplo de redes neurais profundas. Fonte: IBM (IBM, s.d.).

Cada processador possui diferentes pesos e limites associados que direcionam a rede para acionar um neurônio ou não. Os pesos são atribuídos às *features* de entrada assim que são inseridos na rede. Esses ajudam o sistema a entender qual variável tem mais importância. Após essa atribuição, as entradas são multiplicadas pelos seus respectivos pesos e, em seguida, somadas. Elas percorrem o sistema sempre que a saída, a soma, ultrapassa o limite do nó. Sempre que isso ocorre ele envia os dados para a próxima camada.

Ao fazer a passagem de dados para a camada seguinte, esta usa as informações para executar uma determinada tarefa preditiva, ao final do treinamento da rede neural, ela compara o resultado obtido com o esperado para conseguir melhorar sua aprendizagem (MITCHELL, 1997). Assim, o computador aprende por meio do seu próprio processamento, mas para que isso ocorra de maneira eficiente, a quantidade e qualidade de dados para treinamentos deve ser alta. Além de que pode ser um processo demorado e exigir um computador mais avançado.

Para modelos de reconhecimento facial, são utilizadas as redes neurais convolucionais, idealizado nos anos 80 por Kuniyuki Fukushima (FUKUSHIMA, 1980). Esta rede é mais utilizada para o processamento de imagens. Seu funcionamento é similar a rede neural profunda, mas na rede convolucional, cada camada é responsável por extrair determinadas informações dos dados de entrada, e passá-las para a camada seguinte.

A rede convolucional, Figura 2, é composta de 3 camadas básicas, a de convolução, a de *pooling* e a totalmente conectada. A camada de convolução realiza o mapeamento

e extração de dados da imagem, esses dados são enviados para a camada de *pooling* que simplifica a informação para um único valor e o envia para a camada totalmente conectada, onde ocorre o processo de classificação das informações extraídas das camadas anteriores. A saída da última camada contém uma única linha com todas as informações.

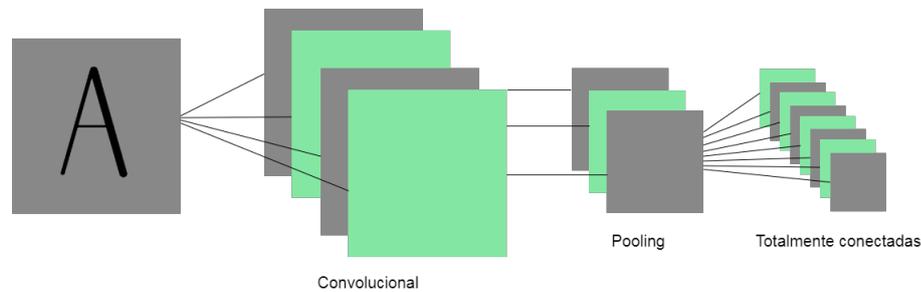


Figura 2 – Exemplo de uma rede convolucional simples

2.2 Detectores Faciais

2.2.1 Multi-task Cascaded Convolutional Networks - MTCNN

MTCNN (XIANG; ZHU, 2017) é um modelo idealizado em 2017, que propõe um novo *framework* usando dois métodos para a construção do modelo, as redes neurais convolucionais(RNC) em cascata e o aprendizado multitarefa, juntando o alinhamento e detecção de face. Essa combinação foi proposta afim de aprimorar a detecção rosto em ambientes sem restrições, imagens em inúmeras poses, diferentes iluminações e oclusões.

A cascata de redes neurais convolucionais utilizada é um pipeline onde há um redimensionamento da imagem, seguido dos estágios de RNCs (Figura 3). O primeiro, *Proposal Network (P-Net)* (REN et al., 2015) define uma rede de propostas à faces rapidamente sendo demarcadas por janela feitas em torno de fragmentos do rosto e dele completo. Essas são refinadas pela segunda RNC, *Refine Network (R-Net)* (GROUP, 2017) que remove as janelas que não contém fragmentos de faces e este processo é refinado pelo último estágio, *Output Network (O-Net)* (MAGHSOUDI et al., 2020), selecionando corretamente a face a ser analisada e aponta *features* importantes da face, como olhos, nariz e boca.

Para treinamento das três RNC's utilizadas, foram realizadas três tarefas, a classificação de faces, regressão da caixa delimitadora e a localização de *features* (XIANG; ZHU, 2017).

A primeira utiliza o parâmetro de erro *cross-entropy* para classificar a imagem em duas classes, uma contém possíveis partes da face e a outra com partes sem traços de face. A aprendizagem é um problema de regressão, é calculado com a distância euclidiana, a proximidade entre cada janela candidata a possuir uma face para a janela que realmente

a possui. Por último a localização de marcos faciais, que assim como no passo anterior é uma regressão que foi resolvida com o uso da distância euclidiana, no qual são definidos os cinco pontos principais da face: olho esquerdo, olho direito, cantos da boca e nariz. Eles são representados pelo seu centro e em forma de tuplas com as coordenadas dos pontos, eixo x e y.

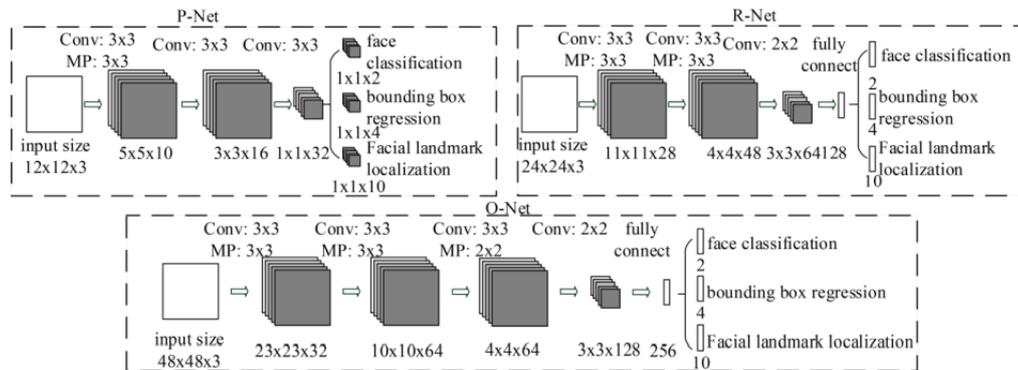


Figura 3 – Arquitetura dos estágios P-Net, R-Net e O-Net. Fonte: (XIANG; ZHU, 2017).

Os resultados do RNC em cascata proposto, superou métodos de última geração na detecção das principais *features* da face, incluindo modelos como *Face Detection Data Set and Benchmark Home*(FDDDB) (JAIN; LEARNED-MILLER, 2010) e *Wider Face* (YANG et al., 2016) e a *Annotated Facial Landmarks in the Wild*(AFLW) (MARTIN KOESTINGER PAUL WOHLHART; BISCHOF, 2011) para faces alinhadas, enquanto manteve o mesmo desempenho para análises de vídeos em tempo real.

2.2.2 DLIB

Proposto em 2015 por Davis King (KING, 2009), é um kit de ferramentas de código , que ainda é atualizado, com implementação de aprendizagem métrica profunda. Seu uso não se restringe a detecção facial, sendo útil em inúmeras áreas como aprendizagem de máquina, análise de texto, entre outras. Sua proposta de detector facial é detectar caixas que contenham a face, analisando as imagens por meio de vetores.

Esse modelo combina o Histograma de Gradientes Orientados(HOG) (DALAL; TRIGGS, 2005) com um classificador linear, a Máquina Vetorial de Suporte(SVM) (CORTES; VAPNIK, 1995), uma pirâmide de imagem e esquema de detecção de janela. O HOG descreve a imagem pela distribuição dos gradientes de intensidade que depois de extraído, os gradientes são colocados em um único vetor. O SVM é treinado com esse vetor e usado para problemas de duas classes e a cada entrada ele classifica de acordo com a probabilidade de ser de uma ou de outra. Entre elas, tem-se duas retas paralelas chamadas de hiperplano superior e inferior. Elas são determinadas pelos pontos mais próximos à cada uma, são os

chamados vetores suporte. Nas Figuras 4 e 5 pode ser observado como funciona o HOG e SVM respectivamente.

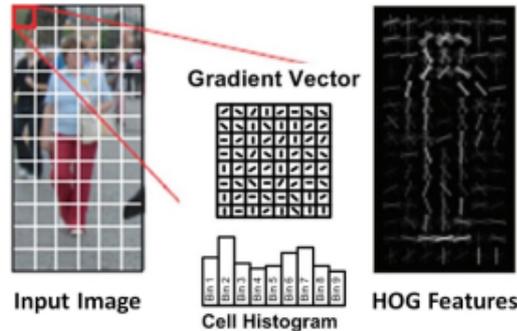


Figura 4 – Extração da imagem HOG. Fonte: (SULEIMAN et al., 2017).

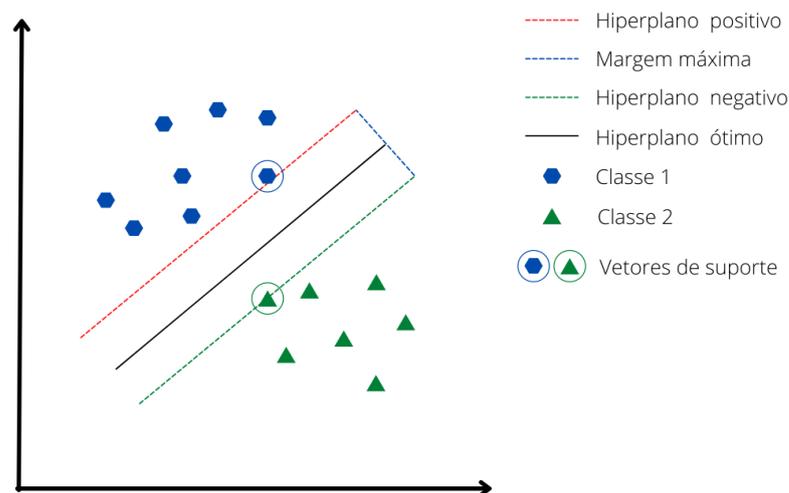


Figura 5 – Hiperplano SVM.

O método de *landmarks*, pontos chaves para rastrear faces, utilizado em (KING, 2009), contribuiu para um grande avanço na detecção de expressões faciais, sendo capaz de diferenciar sentimentos, quando visíveis no rosto, como: tristeza, alegria e raiva.

2.2.3 Open Source Computer Vision Library - OpenCV

Desenvolvido em 2008 pela Intel, OpenCV (BRADSKI; KAEHLER, 2008) é uma biblioteca de software de visão computacional e aprendizado de máquina de código aberto usado para extrair e processar dados significativos de imagens. Usado para imagens e vídeos, esses dados são usados para encontrar e reconhecer todo ou parte do objeto/humano/face, rastrear movimentos, determinar formas 2D e 3D de objetos a partir de uma ou mais imagens e associar dados de imagem a um significado categórico.

OpenCV é estruturado em quatro componentes principais, são eles: Visão computacional(CV), responsável pelo processamento de imagens, aprendizado de máquina(ML), que contém classificadores estatísticos e duas ferramentas de agrupamento, o *HighGUI* com rotinas para armazenar e carregar vídeos e imagens, e o *CXCore* com estrutura de dados e conteúdos básicos.

O processamento de imagens por CV é dado por uma entrada como imagem e a saída é sua interpretação como um todo. Dividido em três níveis, o tratamento das imagens, feito com redução de ruído, segmentação ou classificação e tarefas de cognição associadas com a visão humana.

Para detecção utiliza o método *Haar Cascade* proposto por (VIOLA; JONES, 2001). Ele percorre a imagem pixel a pixel analisando mudanças de intensidade de cores. Para isso, são utilizadas janelas em que o detector busca bordas na imagem para identificar onde está a face. Em cada etapa são excluídos características na foto que não contém nenhum traço de face, ao final é detectada a face.

2.2.4 Retina Face

Idealizado em 2020 por (DENG et al., 2020), é um detector de estágio único. Ele unifica a previsão da caixa de face, com localização de face por pixel em várias escalas de rostos em 2D e regressão de vértices 3D. Sua implementação contém apenas camadas convolucionais, Figura 6.

Utiliza uma pirâmide de redes convolucionais com contextos de módulos independentes. Essas são invariantes em escala, uma vez que ao haver uma mudança, outro nível é inserido na pirâmide. Isso permite com que a detecção aconteça independente de escalas ou posições.

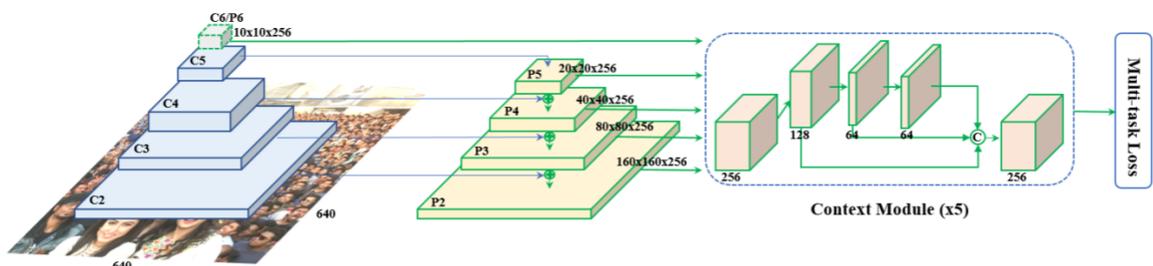


Figura 6 – Pirâmide de recursos e módulos de contexto. Fonte: (DENG et al., 2020).

São 5 níveis utilizados na pirâmide, quanto maior o nível, mais informação da imagem é aprendida pelo computador. O primeiro lida com os 5 *features* principais, olhos, nariz e boca. Em seguida pontua as 68 *landmarks*, como em DLIB (KING, 2009), são os traços de bordas da face, boca, nariz, olhos e sobrancelhas. Com essas informações monta no terceiro nível uma máscara na face; em seguida, os níveis 4 e 5 utilizam técnicas 3D para

redimensionar tridimensionalmente face com regressão de vértices. Com essa técnica se aprende todas as partes do rosto de vários meios diferentes.

Na conclusão do autor, esse modelo pode alcançar simultaneamente uma precisão na detecção facial, alinhamento de face 2D e reconstrução de face 3D com inferência eficiente de um único tiro.

2.2.5 Single Shot MultiBox Detector - SSD

O detector demonstrado em 2014 por (LIU et al., 2016) contém apenas camadas, conforme mostrado na Figura 7. A rede convolucional utilizada é a *feed-forward*, o modelo mais simples de redes neurais, onde as informações se movem em apenas uma direção, sem ciclos.

Este modelo funciona com uma coleção de tamanho fixo de caixas delimitadoras, um *grid* da imagem, e com elas gerar pontuações que indicam a presença de instâncias de objetos para cada classe dentro das caixas delimitadoras. Depois dessa etapa, ocorre uma supressão dos dados para a detecção final.

Para construção do modelo, foi utilizado como base a rede VGG-16 (SIMONYAN; ZISSERMAN, 2014) e a partir dela foram adicionadas camadas extras, camadas de *features* e com cada uma delas uma camada convolucional. As camadas de *features* guardam as feições semelhantes, agrupando-as. O modelo convolucional para prever detecções é diferente para cada camada de *features*.

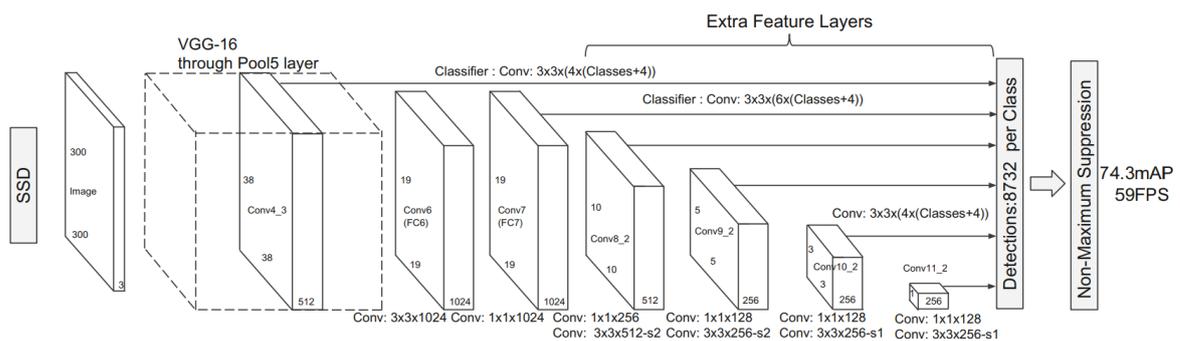


Figura 7 – Caixas delimitadoras detecção SSD. Fonte: (LIU et al., 2016).

A cada camada gerada por diferentes convoluções, seu tamanho diminui para que seja possível previsões de detecção em várias escalas. Conforme as camadas ficam mais profundas, o tamanho representado por cada instância fica maior. O diferencial das camadas do SSD é que cada uma produz um resultado para a detecção de objetos.

A construção é simples mas este detector tem um funcionamento mais simples mas ao mesmo tempo muito eficiente. O autor conclui que além de disso ele pode servir como base para outras redes, onde seriam adicionadas novas camadas. Seu modelo superou em

precisão o *Faster R-CNN* (GIRSHICK, 2015) e obteve boas velocidades na análise de faces em tempo real.

2.3 Modelos de reconhecimento facial

2.3.1 Facenet

Criado pela *Google* em 2015, Facenet (SCHROFF; KALENICHENKO; PHILBIN, 2015) é um sistema unificado de verificação, reconhecimento e *clustering* de faces. O *clustering* de faces é feito por meio se da aprendizagem não supervisionada, onde o treinamento consiste apenas de imagens, sem *labels*.

O funcionamento se baseia em incorporação euclidiana por imagem usando uma rede convolucional profunda e sua estrutura envolve uma camada de entrada *batch*, seguida da arquitetura profunda, normalização L2 resultando no *embedding* da face e a *triple loss* (SCHROFF; KALENICHENKO; PHILBIN, 2015), usada durante o treinamento como demonstrado na Figura 8.

A camada de *batch* tem como objetivo acelerar o treinamento de redes neurais profundas, padronizando as entradas dadas ao modelo para as demais camadas, reduzindo o erro de generalização. A saída desta camada vai para a arquitetura de redes profundas utilizada no modelo, Figura 8.



Figura 8 – Modelo da estrutura Facenet. Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015).

Após o modelo coletar o *embedding* de cada face, vetor de 128 números que representam as partes mais importantes da face, é possível calcular a distância entre os pontos da face e compará-los ao de outra imagem. Esse cálculo é feito com a distância euclidiana normalizada.

A última camada, com *triple loss*, maximiza essa distância para imagens que não tem a mesma identidade e minimiza para as que tem. Com este processo, o aprendizado de máquina é mais incentivado pelo acerto e mais penalizado pelo erro. Distâncias abaixo de um determinado limite resultam em resultado positivo de verificação. Este limite é determinado de acordo com o classificador, detector e métrica de similaridade utilizados.

layer	size-in	size-out	kernel	param	FLPS
conv1	220×220×3	110×110×64	7×7×3, 2	9K	115M
pool1	110×110×64	55×55×64	3×3×64, 2	0	
rnorm1	55×55×64	55×55×64		0	
conv2a	55×55×64	55×55×64	1×1×64, 1	4K	13M
conv2	55×55×64	55×55×192	3×3×64, 1	111K	335M
rnorm2	55×55×192	55×55×192		0	
pool2	55×55×192	28×28×192	3×3×192, 2	0	
conv3a	28×28×192	28×28×192	1×1×192, 1	37K	29M
conv3	28×28×192	28×28×384	3×3×192, 1	664K	521M
pool3	28×28×384	14×14×384	3×3×384, 2	0	
conv4a	14×14×384	14×14×384	1×1×384, 1	148K	29M
conv4	14×14×384	14×14×256	3×3×384, 1	885K	173M
conv5a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv5	14×14×256	14×14×256	3×3×256, 1	590K	116M
conv6a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv6	14×14×256	14×14×256	3×3×256, 1	590K	116M
pool4	14×14×256	7×7×256	3×3×256, 2	0	
concat	7×7×256	7×7×256		0	
fc1	7×7×256	1×32×128	maxout p=2	103M	103M
fc2	1×32×128	1×32×128	maxout p=2	34M	34M
fc7128	1×32×128	1×1×128		524K	0.5M
L2	1×1×128	1×1×128		0	
total				140M	1.6B

Figura 9 – Modelo da estrutura Facenet. Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015).

Este sistema foi testado com a base de dados *Labelled Faces in the Wild*(LFW) (HUANG et al., 2007) e foi o primeiro modelo que superou a acurácia de humanos utilizados nessa base, com um resultado de acurácia de 99.63%.

2.3.2 Visual Geometry Group Face - VGG-face

O VGG-face (PARKHI; VEDALDI; ZISSERMAN, 2015) foi desenvolvido pelo Departamento de Ciências da Engenharia da Universidade de Oxford em 2015. Sua proposta foi o reconhecimento facial tanto de uma única imagem quanto de um conjunto de faces rastreados em um vídeo.

Possui uma arquitetura baseada em (SIMONYAN; ZISSERMAN, 2015), Figura 10, onde é feito o estudo com redes convolucionais profundas, utilizando até 19 camadas para larga escala de classificação de imagem. As camadas convolucionais usam filtros de dimensão 3 e as camadas de pool usam um fator de 2 para realizar a subamostragem.

Sua implementação também conta com o aprendizado de *embedding* da face utilizando a função *triple loss*. Também faz uso do alinhamento de faces e a transformação de similaridade 2D, colocando a face em uma posição canônica.

(PARKHI; VEDALDI; ZISSERMAN, 2015) mostra um procedimento capaz de montar um conjunto de dados em grande escala, com pequeno ruído, processo utilizado para gerar o banco de dados utilizado no treinamento deste. Também mostra que a utilização de uma CNN profunda apenas com uma formação adequada, pode obter resultados comparáveis aos programas de última geração. Sua acurácia para base de dados *Labelled Faces in the Wild*(LFW) (HUANG et al., 2007) foi de 98.95%.

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 flts	-	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 flts	-	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 10 – Arquitetura VGG-face. Fonte: (PARKHI; VEDALDI; ZISSERMAN, 2015).

2.3.3 OpenFace

Apresentada por (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) em 2016, é um modelo de reconhecimento facial gratuito e de código aberto com construído utilizando redes neurais profundas. Foi baseado na estrutura do *FaceNet* (SCHROFF; KALENICHENKO; PHILBIN, 2015) e usa o detector DLIB (KING, 2009) pré-treinado para aumentar a precisão da detecção de faces do modelo. OpenFace foi desenvolvido com a intenção de ser um modelo utilizado em aplicações para celulares onde o reconhecimento facial em tempo real depende do contexto, Figura 11.

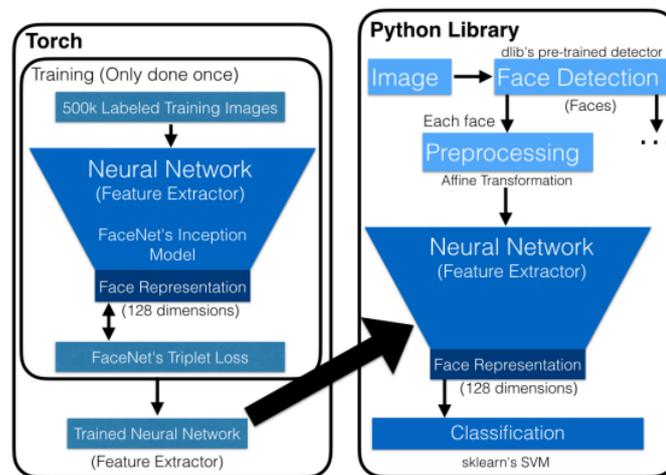


Figura 11 – Modelo da estrutura do OpenFace. Fonte: (AMOS; LUDWICZUK; SATYANARAYANAN, 2016).

Após a detecção da face e seus *landmarks*, o modelo faz um pré-processamento da imagem, a transformação *affine* no modelo da face 2D. Essa se baseia na rotação da face segundo as *landmarks* detectadas, fazendo com que os rostos sempre estejam na mesma posição para reconhecimento facial. Isso é feito porque a proposta do projeto é utilizar um banco de dados menor para treinamento. Em (SCHROFF; KALENICHENKO; PHILBIN, 2015) essa etapa não foi necessária pois o número de faces no banco de dados para treinamento foi suficientemente grande para suprir esse problema.

Com o pré-processamento feito, o reconhecimento facial é feito de modo similar ao FaceNet (SCHROFF; KALENICHENKO; PHILBIN, 2015), onde são extraídos *embedding* e aplicada a *triple loss* nas distâncias encontradas. e a classificação é feita utilizando o sistema linear SVM mostrado na Figura 5 da Secção 2.2.2 deste trabalho.

Foi utilizado a base de dados *Labelled Faces in the Wild*(LFW) (HUANG et al., 2007) para a comparação de acurácia com outros modelos e seu resultado foi de 92% para o reconhecimento facial.

2.3.4 DeepFace

Modelo desenvolvido por pesquisadores do *Facebook* em 2014, (TAIGMAN et al., 2014) apresenta um estrutura de rede neural profunda. Seus diferenciais estão no alinhamento e representação da imagem. Assim como em (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) também é feito uma transformação *affine*, porém neste é feito também com modelo de face 3D.

Ao fazer a detecção extraíndo os *embedding* da face, ocorre o recorte destacando o quadrado em que a face está inserida. Com isso é feita a detecção dos 67 pontos principais da face alinhada em 2D. Esses passos levam ao modelo 3D da face, e assim é feito o *affine* das faces.

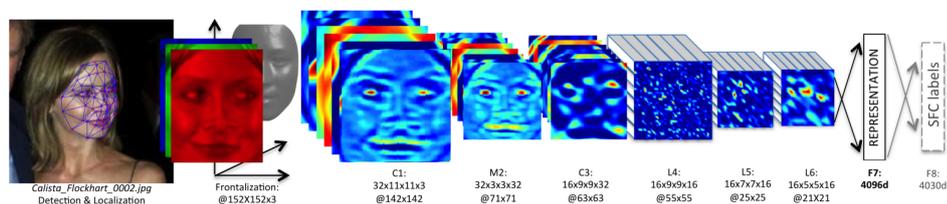


Figura 12 – Modelo da estrutura do Deepface. Fonte: (TAIGMAN et al., 2014).

Na Figura 12 apresenta o modelo apresentado por (TAIGMAN et al., 2014). Faz uso de convoluções (C1 e C3), pooling (M2), três camadas conectadas localmente (L4 a L6), que aplicam um filtro de bancos, onde toda *feature* mapeada da imagem aprende um diferente tipo de filtro, e duas camadas totalmente conectadas (F7 e F8), essas possuem cada unidade de saída conectada a todas as entradas, tendo acesso ao mapeamento de toda a face e assim ela consegue fazer correlações com todas as partes dos rosto.

Ao final do processo, é utilizada uma normalização nas *features* entre 0 e 1, para diminuir a sensibilidade do sistema a mudanças de iluminação.

Foi utilizado a base de dados *Labelled Faces in the Wild*(LFW) (HUANG et al., 2007) para a comparação de acurácia com outros modelos e seu resultado foi de 97%.

2.4 Métricas de similaridade

Para uma análise da imagem, o modelo de reconhecimento facial transforma a imagem em um vetor. Esse vetor é usado para comparar imagens, uma vez que se tem um vetor para cada. Para análise de similaridade das faces tratadas pelos métodos descritos acima, existem três métricas. Estas devem calcular a diferença da distância de valores em cada imagem, gerando assim um grau de similaridade. As métricas utilizadas no código do (SERENGIL, S. I.; OZPINAR, 2020a) são similaridade cosseno, distância euclideana e a euclideana normalizada (L2).

Para cada uma existe um limite, escolhido e calculado pelo programador por meio de estudos do desempenho do sistema, neste trabalho foram utilizados os valores descritos na Tabela 1, encontrados pelo autor do código (SERENGIL, S. I.; OZPINAR, 2020a). Esses limites são utilizados pelo modelo para classificar se as faces pertencem ou não a mesma pessoa. Esses limites também mudam de acordo com o modelo selecionado, conforme a tabela abaixo.

Modelo	Cosseno	Euclideana	L2
VGG-Face	0.40	0.60	0.86
Facenet	0.40	10	0.80
OpenFace	0.10	0.55	0.55
DeepFace	0.23	64	0.64

Tabela 1 – Valores limites para distância de similaridade.

2.4.1 Similaridade do Cosseno

Na similaridade de cossenos a análise é feita na distância angular entre X_a e X_b a partir da origem (SINGHAL, 2001). Este método é usado principalmente quando existem mudanças de dimensionalidade, uma vez que os ângulos são melhor preservados que as distâncias (Equação 2.1).

$$d(X_a, X_b) = 1 - \frac{(X_{ai}^T) \times X_{bi}}{\sqrt{\sum_{i=0}^n (X_{ai})^2} \times \sqrt{\sum_{i=0}^n (X_{bi})^2}} \quad (2.1)$$

onde n é o total de coordenadas dos pontos, X_{ai} e X_{bi} são os valores das coordenadas dos pontos X_a e X_b respectivamente. O resultado da equação pode variar entre 1 e -1, sendo 1 o caso em que os vetores apontarem para a mesma direção e sentido lugar e -1 quando apontarem para sentidos opostos.

2.4.2 Distância Euclideana

No método de similaridade euclideana (DANIELSSON, 1980), a distância media é feita entre os dois pontos X_a e X_b , definidos em um sistema de coordenadas, calculada a partir

do teorema de Pitágoras. Para uma imagem com n dimensões, o cálculo da distância entre as os pontos das matrizes da face, referentes a imagem base e a imagem a ser comparada, usando a forma euclideana é dada pela Equação 2.2.

$$d(X_a, X_b) = \sqrt{\sum_{i=0}^n (X_{ai} - X_{bi})^2} \quad (2.2)$$

onde n é o total de coordenadas dos pontos, X_{ai} e X_{bi} são os valores das coordenadas dos pontos X_a e X_b respectivamente. Também é usada a euclideana normalizada, que antes de realizar o cálculo da distância, primeiro normaliza as coordenadas, pegando o valor da matriz e dividindo pela sua norma, demonstrado na Equação 2.3. Esta foi uma combinação dos achados de (DANIELSSON, 1980) e (SHI; MALIK, 2000).

$$Matriz_{normalizado} = \frac{X}{\sqrt{\sum_{i=0}^n (X_i)^2}} \quad (2.3)$$

onde X é a matriz a ser normalizada e X_i os valores pertencentes a matriz.

O próximo capítulo será dos trabalhos relacionados ao tema tratado. Assuntos como erros encontrados na aprendizagem do reconhecimento facial assim como a importância de um banco de dados extenso e equilibrado para o treinamento do aprendizado de máquina, em especial para o reconhecimento facial.

3 Trabalhos Relacionados

Neste capítulo serão apresentados trabalhos que tiveram como objetivo detectar onde e porque ocorre vies no reconhecimento e identificação facial, propondo possíveis causas e soluções. A explicação do motivo de viés presente no reconhecimento facial ainda é incerto. Cada estudo relacionado a esse tema aumenta a compreensão do que acontece com o aprendizado de máquina e as possíveis causas da diferença de acurácia entre raças e gêneros.

Para tanto serão feitas 2 sessões: reconhecimento facial e banco de dados. A primeira trata de possíveis problemas levantados na aprendizagem do algoritmo e o segundo no desbalanceamento das base de dados utilizadas para treinamento.

3.1 Reconhecimento Facial

Em 2018, os ideais de igualdade de gêneros e tratamento igual entre diferentes raças começaram a surgir com mais força na sociedade. Incluindo problemas relacionados ao uso de reconhecimento e identificação facial. Essas questões deram início à busca para entender o que acontece no processo de reconhecimento e identificação facial, e porque determinadas classes possuem a acurácia melhor que outras.

O autor de (BUOLAMWINI; GEBRU, 2018) utiliza diferentes modelos de reconhecimento facial para analisar a acurácia para indivíduos de pele clara e escura. Ele utiliza 3 modelos feitos por grandes empresas para analisar a acurácia para estes grupos.

Um dos problemas encontrados para realizar essa análise foi a falta de um banco de dados que tivesse um balanceamento de indivíduos. Sendo assim foi feito um banco de dados pelo autor utilizando imagens de representantes parlamentares e a determinação se eram faces pertencentes a pessoas de pele clara ou escura foi feita com auxílio da escala de cor *Fitzpatrick*. O estudo do desempenho dos modelos foi realizado em 4 grupos, homens de pele escura, homens de pele clara, mulheres de pele escura e mulheres de pele clara.

A conclusão e resultados de (BUOLAMWINI; GEBRU, 2018) apontaram que mulheres de pele escura possuem uma taxa de erro significativamente maior quando comparada aos demais grupos. No final aponta o erro sendo do algoritmo que de alguma forma faz isso de propósito, exigindo uma maior transparência. Porém como visto, o que acontece no aprendizado de máquina não é claro nem mesmo para quem os faz. Por isso utilizar um bom banco de dados para treinamento é tão necessário.

O trabalho apresentado por (MUTHUKUMAR et al., 2019) busca responder um dos problemas apontados por (BUOLAMWINI; GEBRU, 2018), a acurácia para mulheres de pele

escura ser significativamente menor que a de qualquer outro grupo. Para essa análise foram colocados dois pontos a serem estudados: se a cor do sujeito é um dos motivos e quais partes da face são mais preditores para um rosto feminino.

Para esclarecer a influência do tipo de pele, foi feita a utilização de métodos teóricos de cores, para que se pudesse estudar a influência das cores, como por exemplo, a mudança de fatores como a iluminação da imagem. Em análises feitas com as imagens escurecidas de mulheres brancas e iluminadas de mulheres negras, foi observado que em ambos os casos o score de acerto não aumentou significativamente, sendo menor que 0,1%. A conclusão de (MUTHUKUMAR et al., 2019) é que as classificações são relativamente estáveis à variações do tipo de pele.

Já para os erros de gênero, os suspeitos de maior influência para detecção de um rosto feminino foram o cabelo e o uso de maquiagem. O primeiro logo foi descartado, pois ao recortar as imagens excluindo os cabelos não houve melhoria significativa. Seguiu-se para o uso de técnicas de contraste *post-hoc* para *neural networks*, e com elas foi demonstrado que os pontos que mais contribuem para classificação feminina são olhos, bocas e maçãs do rosto. Esses resultados apontam que a maquiagem nos lábios e olhos são fortes preditores para um rosto feminino, o que indicam um outro problema sexista em relação ao aprendizado de máquina. Uma possível resolução é o tratamento das imagens no banco de dados, de modo que não seja visível a diferença de cor dessas partes do rosto, permitindo a máquina aprender sem levar em consideração aplicações de maquiagem.

Ao analisar motivos para que no grupo de mulheres de pele escura os resultados apresentam uma acurácia baixa, (MUTHUKUMAR et al., 2019) chegou a conclusão que a mudança de estruturas faciais entre mulheres europeias e afrodescendentes é o motivo pelo qual os resultados para o segundo grupo sejam muito inferiores. Isso porque os algoritmos de aprendizado de máquina são treinados utilizando imagens disponíveis ao público e muitas vezes pertencentes a celebridades, o que facilita a construção de *labels* para o treinamento. Esse problema pode ser solucionado aumentando o banco de dados oferecido ao ML para que possa aprender de maneira correta mais tipos de raças e rostos. Essa mudança pode ocasionar um bom aumento da acurácia dos algoritmos.

A baixa acurácia não acontece somente para mulheres negras, diferentes raças apresentam um desempenho inferior do que caucasianos. O trabalho de (NAGPAL et al., 2019) usa 4 modelos de redes neurais profunda pré-treinadas para fazer uma análise sobre o motivo dessa diferença. Em suas análises, foi encontrado um viés que é explicado pela aprendizagem da máquina, que é apontada como tendo feito uso de *own-race* e *own-age*.

Na utilização desses preditores, o que algoritmo aprende para uma raça ou idade, ele aplica para as demais, isso é, se para diferenciar mulheres e homens brancos ele aprende uma característica X, para diferenciar homens e mulheres negras ele aplicará essa mesma característica. O que também ocorre para crianças, onde o algoritmo aprendeu utilizando

faces de adultos, ele recria tenta aplicar as mesmas métricas aprendidas em faces de crianças.

Como dito antes, pelos bancos de dados de treinamento não serem balanceados e possuírem uma maior quantidade de faces caucasianas, tentar analisar diferentes raças e estruturas físicas da face, tendo maior conhecimento sobre apenas uma, pode gerar vieses e maiores erros no desempenho dos modelos.

A conclusão do autor em (NAGPAL et al., 2019) destaca a necessidade dos treinamentos serem feitos com uma base de dados maior e balanceada, para incorporar a maior variedade possível de imagens para aprendizado da máquina ser mais coerente com a realidade.

Contribuindo para os achados de (NAGPAL et al., 2019), (MERY; MORRIS, 2022) realizou estudos demonstrando quais partes são levadas mais em conta para o reconhecimento facial dos modelos. Para isso, (MERY; MORRIS, 2022) fez análises no desempenho de modelos criando o método de caixas pretas nas faces. Nele são cobertas algumas partes da face comparada e a cada cobertura realiza um novo score de comparação com a imagem base, através da repetição desse processo é determinado qual parte influencia mais a comparação de faces.

Em sua análise (MERY; MORRIS, 2022) utiliza um banco de dados com homens e mulheres, brancos e negros. Sua análise foi feita com duas faces que sabidamente pertenciam ao mesmo indivíduo, comparando quando ele apontava um maior grau de acuratividade de acordo com a cobertura da face. Sua conclusão mostra que o algoritmo sempre dá importância para os mesmos pontos, não havendo diferenciação para gênero ou raça.

Os trabalhos realizados por (NAGPAL et al., 2019), (MUTHUKUMAR et al., 2019) e (MERY; MORRIS, 2022) chegaram a resultados em que alguns dos problemas apontados vem da aprendizagem dos modelos e como isso influencia no modo de tratar faces de diferentes raças, gênero e idade. Todos contribuíram para que possam ser feitas melhorias significativas no reconhecimento facial. Saber as partes das faces que não são corretamente absorvidas pelo processo de aprendizagem do algoritmo, pode ser a chave para encontrar uma resolução para o viés.

3.2 Conjuntos de Dados

O conjunto de dados oferecido para o treinamento dos modelos de reconhecimento facial tem vital importância na sua aprendizagem. Quanto mais informação lhe for passada, mais o programa tem para julgar corretamente as faces. Os trabalhos (NAGPAL et al., 2019), (MUTHUKUMAR et al., 2019) e (MERY; MORRIS, 2022) apontam que os problemas relacionados aos vieses ocorre pela falta de conhecimento do algoritmo e a carência da falta de uma base de dados balanceados tanto para treinamentos quanto bases para testes,

é apontada no trabalho já citado realizado por (BUOLAMWINI; GEBRU, 2018). Nele é relatado a dificuldade de se encontrar e montar bases com esses requisitos.

Como visto em (MERLER et al., 2019), cada face é única. Logo para que haja igualdade nos resultados de reconhecimento facial, é necessário que haja também uma extensa base de dados com diversas faces, de diferentes raças e gêneros. (MERLER et al., 2019) também realiza estudos que demonstram que para uma melhor análise e reconhecimento facial, seria necessário um aprendizado muito mais profundo sobre cada face. Parâmetros como largura do crânio, extensão da face, simetria facial, seriam de grande ajuda para conhecer diferentes raças, vindas do mundo todo.

Um maior conhecimento de parâmetros faciais seria útil pois na utilização de algoritmos que contenham aprendizado de máquina, uma das maiores pré-análises devem acontecer no banco de dados, uma vez que devem ser analisados todos os tipos de informações passadas, além de garantir que elas sejam o suficiente para que o modelo tenha o desempenho esperado. A dificuldade está em encontrar uma grande quantidade de faces de diferentes lugares do mundo disponíveis para serem usadas com essa finalidade.

Não somente para treinamento de máquinas, mas um banco de dados com a quantidade e qualidade, são ideais para uma análise do desempenho geral dos modelos treinados. Uma quantidade de dados iguais para cada grupo é melhor para compreensão do que acontece com cada subgrupo em análises e para ensinar melhor o algoritmo quando usado para treinamento. Pensando nisso, (ROBINSON et al., 2021) criou a *Balanced Faces in the Wild* (BFW), uma base de dados onde todas as raças e gêneros tem o mesmo número de imagens. As faces utilizadas são de todos os ângulos, luzes e lugares.

4 Metodologia Proposta

A metodologia proposta para a realização deste trabalho segue a sequência demonstrada no fluxo de trabalho representado na Figura 13. São 6 etapas a serem realizadas, desde levantamentos bibliográficos até a análise dos resultados obtidos. A abordagem utilizada para a realização deste trabalho é a incremental, em que cada etapa é baseada na análise e conclusão da anterior. Neste capítulo serão descritas cada etapa realizada.

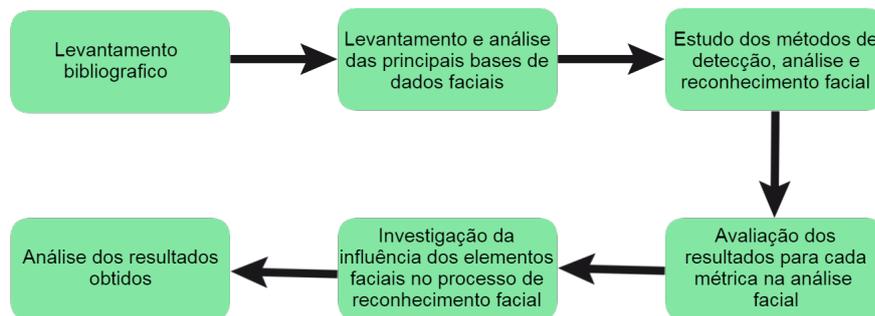


Figura 13 – Fluxograma da metodologia proposta.

4.1 Levantamento bibliográfico

A proposta deste trabalho é baseada nos achados citados por (BUOLAMWINI; GEBRU, 2018), que comprova uma diferença de desempenho para indivíduos com pele escura em especial para mulheres. A partir deste, foi utilizada a ferramenta de análise bibliométrica (EITAN et al., s.d.) para encontrar estudos relacionados que estudavam os vieses presentes na identificação e reconhecimento facial. Desta maneira, foram obtidas partes das bibliografias utilizadas para construir o Capítulo 3.

Com diversos papéis conectados referentes aos estudos sobre o impacto dos vieses do reconhecimento facial na sociedade, dois trabalhos tinham como objetivo analisar e responder algumas questões de como e porque ocorre a diferença de desempenho entre diferentes raças e gêneros. No primeiro, realizado em (MUTHUKUMAR et al., 2019), o autor busca responder exatamente o que ocorre para que o resultado seja o apresentado em (BUOLAMWINI; GEBRU, 2018) e no segundo, o autor de (MERLER et al., 2019), demonstra a importância de um extenso banco de dados, com diferentes raças e gêneros, para que em todas as diversidades se tenha uma acurácia aceitável e aponta outros parâmetros a serem observados para que o aprendizado de máquina possa ser mais eficiente.

Em busca de mais trabalhos relacionados foi utilizado o trabalho de (MERLER et al., 2019) no (EITAN et al., s.d.) e foi obtida uma conexão com o (NAGPAL et al., 2019). Este é um trabalho que tem como objetivo responder se e onde está o preconceito no código do

sistema de reconhecimento facial. Sua conclusão demonstrou que a máquina se concentra nas características analisadas em uma raça, a que ele tem maior informações, com variações entre diferentes raças. Acompanhando os trabalhos lançados com esse assunto, foi encontrado o estudo de (MERY; MORRIS, 2022), que reforça a essa conclusão. Sua análise mostra que o algoritmo usa as mesmas *features* como principais para verificar faces de diferentes raças.

Ao finalizar essa etapa, um dos pontos observados foi a necessidade da escolha de uma base em que houvesse uma variação considerável de gênero e raças, com uma quantidade suficiente de imagens, mesmo que para análise dos resultados de uma rede já treinada. Isso deu início a próxima etapa a ser realizada.

4.2 Levantamento e análise das principais bases de dados de imagens faciais

Para que os resultados deste estudo possam ser replicados, a base deverá ser de uso gratuito e de fácil acesso. Também será observada a quantidade de fotos disponíveis, assim como suas diversidades e informações. De início foram consideradas 3 bases para realização do trabalho. A *Large Scale Face Dataset*(UTKFace) (ZHIFEI; HAIRONG, 2017), *Labeled Faces in the Wild Home*(LFW) (HUANG et al., 2007) e *Balanced Faces in the Wild*(BFW) (ROBINSON et al., 2021). A primeira consiste em um banco de mais de 20 mil faces e cada imagem contém informações de idade, gênero e raças. A LFW (HUANG et al., 2007) possui por volta de 13 mil imagens de 5749 pessoas, apenas com o nome de cada indivíduo. A BFW assim como a UTK (ZHIFEI; HAIRONG, 2017), possui 20 mil faces em diferentes ambientes, porém apenas com informações de gênero e raças.

A base BFW (ROBINSON et al., 2021) possui um diferencial no balanceamento da quantidade de imagens por indivíduos e de número de indivíduos, igual para cada agrupamento proposto. Com quatro raças: asiáticos, negros, indianos e brancos. A base possui imagens com diferentes poses, expressões e iluminações. Devido ao seu balanceamento, foi a escolhida para realização deste trabalho. Sua distribuição de dados está demonstrada na Tabela 2.

	Asiáticos		Negros		Indianos		Brancos	
	M	H	M	H	M	H	M	H
Nº Faces	2500	2500	2500	2500	2500	2500	2500	2500
Nº Pessoas	100	100	100	100	100	100	100	100
Nº Faces/Pessoas	25	25	25	25	25	25	25	25

Tabela 2 – Distribuição de imagens da BFW. M - Mulheres, H - Homens.

Para que possa ser utilizada como parâmetro, será feito uma extração de dados desse banco de imagens. As informações retiradas serão sobre a raça, gênero, uma identificação para cada pessoa no arquivo, nome de cada imagem e o seu respectivo *path*. A informação



Figura 14 – Exemplo de imagens da base BFW.

de raças e gênero os nomes das pastas para seus respectivos. Dentro de cada pasta existem os 25 indivíduos, separados em pastas com seus id próprios.

Com os dados de cada imagem expostos facilmente, a análise para comparação de resultados será feita sem maiores dificuldades. Todos esses dados serão coletados por meio de códigos feitos na linguagem *python* e armazenados em um arquivo *json*. Essa extensão será utilizada pela facilidade de trabalhar dados contidos nela.

Essa etapa contribui para o desenvolvimento da etapa que envolvem a aplicação do código e avaliação dos resultados, Seções 4.3 e 4.4.

4.3 Estudo dos métodos de detecção, análise e reconhecimento facial

O trabalho feito por (BUOLAMWINI; GEBRU, 2018) fez uso de aplicações de reconhecimento facial de três grandes empresas, por isso o foco deste trabalho é fazer uma análise semelhante usando os modelos de reconhecimento facial *opensource* disponíveis. A aplicação encontrada com esses parâmetros foi o código feito por (SERENGIL, S. I.; OZPINAR, 2020a).

Sua estrutura foi preparada para executar diversos modelos, detectores e métricas, por isso se mostrou ideal para conseguir resultados satisfatórios. Em seu algoritmo ele conta com funções para análise, busca e verificação de faces. Neste trabalho serão utilizadas apenas a análise e verificação.

A função análise faz uso do classificador VGG-Face (PARKHI; VEDALDI; ZISSERMAN, 2015), com a opção dos 5 detectores: *Retina face* (DENG et al., 2020), MTCNN (XIANG;

ZHU, 2017), OPENCV (BRADSKI; KAEHLER, 2008), SSD e DLIB (KING, 2009). Esta função retorna as variáveis de idade, emoção, gênero e raça, mas apenas as duas últimas saídas fora observadas neste trabalho.

Na função de verificação, além da mesma variação possível entre os detectores da função citada acima, ela também possui a opção de 3 métricas de semelhanças e 8 classificadores. As métricas são: distância de cosseno, euclídeana e euclídeana normalizada. Para os classificadores foram escolhidos os 4 principais, o VGG-Face (PARKHI; VEDALDI; ZISSERMAN, 2015), *Facenet* (SCHROFF; KALENICHENKO; PHILBIN, 2015), *OpenFace* (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) e *DeepFace* (TAIGMAN et al., 2014).

Segundo análise feita pelo próprio (SERENGIL, S. I.; OZPINAR, 2020a), ao realizar comparações entre os classificadores e métricas disponíveis em seu código, buscou demonstrar qual combinação entre classificador e métrica possui a melhor acurácia. Esse estudo foi demonstrado em vídeo em (SERENGIL, S., s.d.).

O resultado desta etapa foi demonstrado no Capítulo 2, Seções 2.2 e 2.3 deste trabalho, sobre a fundamentação teórica estudada pra a confecção do mesmo, uma vez que para obter uma conclusão rica será necessário um conhecimento básico sobre como os modelos e detectores do código funcionam.

4.4 Avaliação dos resultados para cada métrica na análise facial

Nesta etapa será dado início aos testes nos detectores faciais apresentados até o momento, buscando avaliar e buscar a existência de vieses na classificação do indivíduo pela face. Com as imagens fornecidas pelo banco de dados BFW (ROBINSON et al., 2021), será utilizada a função de análise do código de (SERENGIL, S. I.; OZPINAR, 2020a). Aplicando os 5 métodos de detecção disponíveis, para que possa ser possível observar a diferença de desempenho entre eles. Nesta função o único classificador disponível é o VGG-Face (PARKHI; VEDALDI; ZISSERMAN, 2015).

Para comprovar o que já havia sido estudado no Capítulo 3 sobre a influência de cabelo nas fotos, os testes serão realizados com as imagens originais e recortadas. Para recortar as imagens será utilizado o detector MTCNN (XIANG; ZHU, 2017).

Será utilizado o banco de dados inteiro, passando uma vez pelo algoritmo. Com os resultados também será feita uma matriz de confusão dos resultados, sendo analisados separadamente por gênero e por raça, e por fim agrupando gênero e raça. Assim será possível detectar onde ocorre a maior quantidade de falsos positivos e qual a acurácia na análise de cada raça e gênero. Os resultados serão colocados em tabelas e analisados levando em consideração apenas a raças, apenas o gênero e no agrupamento raças e gênero. Os resultados

desta etapa estão demonstrados no Capítulo 5.

4.5 Investigar influência dos principais elementos faciais no processo de reconhecimento facial

Na segunda parte deste trabalho, será realizado um teste para verificar a acurácia no reconhecimento facial para diferentes raças, gêneros e agrupamentos. O modelo e métrica utilizada serão o *FaceNet* utilizando a métrica Euclideana. Em (SERENGIL, S., s.d.) é demonstrado que esta é a melhor combinação. O detector a ser utilizado sairá dos resultados obtidos na primeira etapa.

Nessa etapa a busca foi para detectar se algum elemento facial influenciava na acurácia para classes que possuem vieses. Para cada resultado também será feita uma tabela verdade para melhor visualização dos dados e desempenho do modelo.

Utilizando a função de verificação disponível no código (SERENGIL, S. I.; OZPINAR, 2020a), junto ao banco de dados BFW (ROBINSON et al., 2021). A comparação será realizada em três formas separadas, a primeira consiste em realizar as comparações no agrupamento de raças e gêneros. Seguida das comparações feitas dentro somente da mesma raça e por fim dentro somente do mesmo gênero. Ao final serão realizadas faces incluindo todo banco de dado disponível, onde por exemplo faces de mulheres brancas poderão ser comparadas com faces de homens indianos. Logo, as imagens passarão por quatro diferentes reconhecimentos faciais, com mudanças na escolha das faces.

Para comparações feitas dentro da mesma raças e gênero, serão selecionados 100 faces de uma base escolhida como principal e 500 faces usadas como teste. Para manter a mesma quantidade de comparações, ao comparar gêneros dentro da mesma raças, serão escolhidos como base e teste 220 faces femininas e 220 faces masculinas. Por fim, na comparação entre o mesmo gênero entre diferentes raças, serão escolhidas 100 faces para cada raça. A escolha de imagens será feita de forma aleatória.

O processo utilizado executará 5 análises do mesmo par de faces. O detector facial MTCNN (XIANG; ZHU, 2017) será utilizado para identificar, por meio de pares ordenados, as 4 principais *features* do rosto (olhos, nariz e boca). Essas coordenadas serão utilizadas para criar caixas coloridas para cobrir cada elemento facial em questão nas faces testes. O exemplo desta aplicação pode ser vista na Figura 15.

As caixas para cobrir os elementos faciais serão feitas com a biblioteca do OpenCV (BRADSKI; KAEHLER, 2008), que cria uma máscara para cada imagem, usando a adição entre duas imagens para que a caixa colorida cobrisse a componente de face correspondente. No primeiro processo de comparação serão realizados utilizando caixas nas cores pretas, azuis e vermelhas, afim de garantir que sua cor não influencie no resultado. Após análise

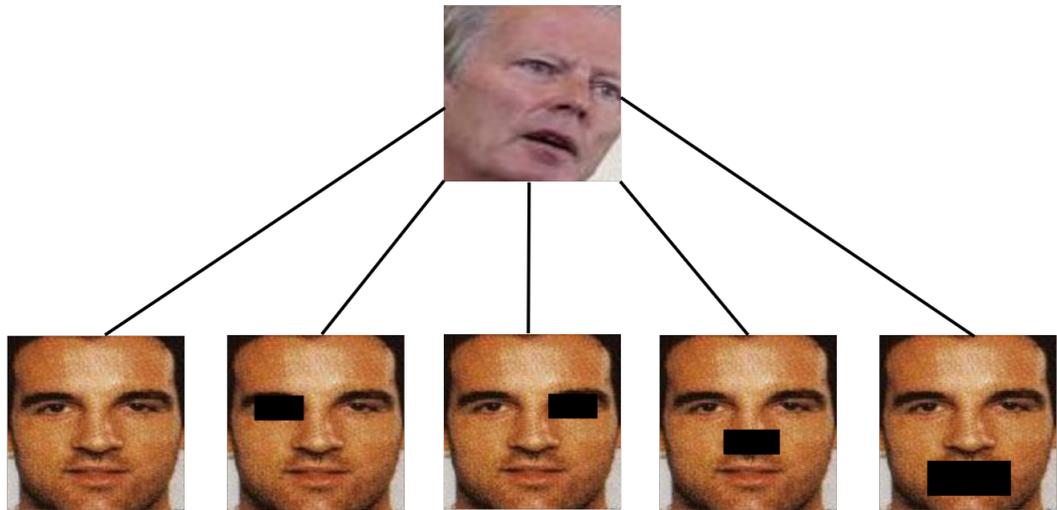


Figura 15 – Comparações feitas para análise.

da influência de cores, a diferenciação não causou vieses nas análises e a cor vermelha, por ser uma cor quente que difere da cor das faces utilizadas, foi utilizada para rodar as demais comparações.

No Capítulo 5 estão demonstrados os resultados e análises das comparações realizadas.

4.6 Análise dos resultados obtidos

Serão realizadas análises sobre as matrizes de confusão com os resultados obtidos. Haverá uma comparação os desempenhos dos diferentes detectores para a análise facial, obtendo o melhor e pior resultado para cada agrupamento.

As análises dos resultados serão feitas juntamente com o modelo de funcionamento do detector, métrica e classificador utilizado, buscando avaliar quais variações podem influenciar no desempenho de cada um e qual demonstrou melhores resultados. Para os resultados das comparações faciais, também será realizada uma análise de qual elemento facial demonstrou ter maior influência nos resultados e uma possível explicação do porque isso ocorre.

Os resultados desta etapa serão apresentados no Capítulo 5, juntamente com a apresentação da discussão e análise sobre a existência ou não de vieses no processo de reconhecimento facial.

5 Resultados

Neste capítulo serão apresentados os resultados dos testes realizados para identificação e reconhecimento facial. Ele será dividido em duas seções, na primeira será retratado o estudo sobre o desempenho dos detectores na identificação facial e na segunda serão apresentados os resultados do processo de reconhecimento facial, analisando a diferença de acurácia entre as faces originais e para cada máscara utilizada em uma das faces.

Serão apresentados os resultados gerais e em seguida o foco será entre o melhor e pior desempenho para cada estudo, tanto entre os detectores quanto entre as *features* cobertas no reconhecimento facial.

5.1 Desempenho dos detectores para identificação facial

As faces foram detectadas e classificadas uma vez pelo algoritmo em raça e gênero. As análises dos resultados foram feitas de três maneiras: na identificação somente entre raças, somente entre gêneros e agrupando ambos. Na identificação de raças poderá ser analisado qual delas, sendo homens ou mulheres, possui a melhor ou pior acurácia. No agrupamento de gênero demonstrará qual deles, independente de raça, ocorre um pior desempenho nos processos de identificação facial. Por fim, agrupamento raça e gênero será possível observar com maior clareza a acurácia na análise facial para cada combinação.

5.1.1 Identificação da raça pela Face

Com os resultados para identificação facial, foi feita uma matriz de confusão entre as raças originais e as raças classificadas pelo algoritmo. As raças disponíveis no algoritmo para detecção são seis: indianos, negros, asiáticos, brancos, latinos hispânicos e árabes. Já no banco de dados só existem amostras de quatro raças: indianos, negros, asiáticos e brancos. O algoritmo analisa a porcentagem presente de cada raça, entre as seis citadas, na face e a que possui a maior porcentagem de detecção é a raça que o algoritmo identifica para a face.

A Tabela 3 contém para cada raça, as taxas de erro de cada detector utilizado, com as imagens originais e recortadas. A diferença entre os desempenhos do recorte da imagem não se mostrou significativa, então as análises seguintes serão feitas em cima de imagens originais.

Os melhores resultados gerais foram feitos pelo detector DLIB (KING, 2009) com uma porcentagem de erro média de 24,08%. DLIB (KING, 2009) apresentou o melhor resultado

para identificação das raças negras e indianas com 19,26% e 58,15% de porcentagem de erro, respectivamente. A raça asiática foi melhor identificada pelo detector MTCNN (XIANG; ZHU, 2017), apresentando um erro de 7,26%. Para a raça branca a melhor identificação foi feita pelo detector SSD (LIU et al., 2016), com um erro de 8,96%.

Imagens originais					
	DLIB	MTCNN	Retina Face	SSD	OpenCV
Indianos	58,15%	72,13%	66,06%	65,09%	70,63%
Negros	19,26%	23,50%	23,82%	23,22%	19,84%
Asiáticos	8,94%	7,26%	9,87%	9,67%	11,02%
Branco	10,00%	14,36%	10,02%	8,96%	15,68%
Imagens recortadas					
	DLIB	MTCNN	Retina Face	SSD	OpenCV
Indianos	50,98%	73,86%	70,61%	70,43%	71,00%
Negros	22,32%	23,94%	22,52%	21,98%	22,26%
Asiáticos	8,10%	7,60%	9,43%	9,09%	9,99%
Branco	11,34%	15,16%	11,18%	11,08%	11,74%

Tabela 3 – Porcentagem de erro detectores com imagens originais e recortadas.

Os piores resultados gerais foram obtidos pelo detector MTCNN (XIANG; ZHU, 2017), com uma média de erro de 29,31%, apresentando os piores resultados na identificação da raça indiana e negra, com 72,13% e 23,50% de erro, respectivamente. Para as raças asiática e branca, a identificação que obteve piores resultados foi feita com o detector OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), com uma taxa de erro de 11,02% e 15,68%, respectivamente.

Utilizando a tabela de matriz de confusão do detector DLIB (KING, 2009) e do detector MTCNN, Tabela 4, é possível observar quais foram as raças que causaram maior confusão na identificação pela face.

Detector Dlib						
	Asiáticos	Negros	Indianos	Latinos	Árabes	Branco
Asiáticos	0,9106	0,0024	0,0008	0,0167	0,0046	0,0649
Negros	0,0676	0,8074	0,0178	0,0648	0,0080	0,0364
Indianos	0,0416	0,0274	0,4185	0,1890	0,2103	0,1133
Latinos	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Árabes	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Branco	0,0221	0,0039	0,0029	0,0315	0,0338	0,9059
Detector MTCNN						
	Asiáticos	Negros	Indianos	Latinos	Árabes	Branco
Asiáticos	0,9274	0,0018	0,0006	0,0163	0,0046	0,0492
Negros	0,1096	0,7650	0,0150	0,0708	0,0066	0,0330
Indianos	0,0712	0,0382	0,2787	0,2436	0,2137	0,1547
Latinos	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Árabes	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Branco	0,0596	0,0029	0,0018	0,0385	0,0325	0,8898

Tabela 4 – Matriz de confusão para melhor e pior detector.

Para asiáticos, cerca de 70% dos erros de identificação foram feitos na raça branca. Na raça negra foram duas raças que se obtiveram uma maior quantidade de faces identificadas como asiáticos e latinos, aproximadamente 35% para cada. Para indianos também foram duas raças que as mais identificadas erroneamente quando a face pertencia a indianos, as raças latina e árabe, com uma taxa de erro de aproximadamente 32% para cada. Por último a raça branca teve a maior porcentagem de erro na raça asiática, latinos e árabes, com uma taxa de 30% em cada raça.

Foi feito o cálculo da diferença ao quadrado das matrizes e a soma de todos os elementos, com as matrizes de confusão de todos detectores, duas a duas, a fim de demonstrar as diferenças entre os resultados dos detectores e analisar suas similaridades, Tabela 5. Um resultado próximo de zero representa que entre determinados detectores, os vieses presentes podem ser os mesmos.

	DLIB	MTCNN	Retina Face	SSD	OpenCV
DLIB	0,0000	0,0288	0,0104	0,0087	0,0214
MTCNN	0,0288	0,0000	0,0100	0,0136	0,0062
Retina Face	0,0104	0,0100	0,0000	0,0004	0,0106
SSD	0,0087	0,0136	0,0004	0,0000	0,0132
OpenCV	0,0214	0,0062	0,0106	0,0132	0,0000

Tabela 5 – Diferença quadrática entre os detectores.

As faces de negros e brancos foram todas detectadas e analisadas, totalizando 5000 faces para cada raça. Para asiáticos e indianos foram detectados e analisadas 4975 e 4927 faces respectivamente. A raça que obteve o melhor resultado no detector DLIB (KING, 2009) foram os asiáticos, seguidos dos brancos, negros e em último os indianos. Essa ordem se manteve em todos os detectores, com exceção do SSD (LIU et al., 2016) com imagens originais, onde o erro para raça branca foi menor do que para asiáticos.

5.1.2 Identificação de gênero pela face

Nesta subseção foi feita a matriz de confusão dos resultados analisando a identificação de gêneros. O algoritmo identifica as faces em dois: mulheres ou homens. Na Tabela 6 são apresentados as porcentagens de erros que cada detector obteve para cada gênero, usando imagens das faces originais e recortadas. Novamente a diferença no erro utilizando imagens originais e recortadas não foi significativo, então as demais análises serão feitas em cima dos resultados na identificação feita com imagens originais.

O detector SSD (LIU et al., 2016) apresentou a menor média de erro total com 7,47%, enquanto o OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) obteve a pior média, com 15,39% de erro.

Para a identificação de faces femininas, o detector SSD (LIU et al., 2016) apresentou a menor taxa de erro, com 12,15% e a pior taxa foi obtida com o detector OpenCV (AMOS;

LUDWICZUK; SATYANARAYANAN, 2016) com 28,72%. Para faces masculinas, o detector SSD (LIU et al., 2016) obteve a pior taxa de identificação, com 2,79% de erro enquanto o melhor detector foi o MTCNN (XIANG; ZHU, 2017) com 1,26% de erro.

Imagens originais					
	DLIB	MTCNN	Retina Face	SSD	OpenCV
Mulheres	24,51%	22,51%	17,24%	12,15%	28,72%
Homens	1,51%	1,36%	2,69%	2,79%	2,06%
Imagens recortadas					
	DLIB	MTCNN	Retina Face	SSD	OpenCV
Mulheres	23,60%	24,72%	19,57%	19,31%	22,51%
Homens	1,60%	1,26%	1,82%	1,78%	1,75%

Tabela 6 – Porcentagem de erro dos detectores com imagens originais e recortadas.

Na Tabela 7 são apresentados os resultados de matriz de confusão os detectores que tiveram o melhor e pior resultado. O OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) identificou mais faces de mulheres como faces de homens e menos faces de homens como faces de mulheres em relação ao detector SSD (LIU et al., 2016).

Detector SSD		
	Homens	Mulheres
Homens	0,97	0,03
Mulheres	0,12	0,88
Detector OpenCV		
	Homens	Mulheres
Homens	0,98	0,02
Mulheres	0,29	0,71

Tabela 7 – Matriz de confusão para o melhor e pior detector.

Novamente, foi feito o cálculo da diferença ao quadrado das matrizes e a soma de todos os elementos, com as matrizes de confusão de todos detectores, duas a duas, a fim de demonstrar as diferenças entre os resultados dos detectores e analisar suas similaridades, Tabela 8.

	DLIB	MTCNN	Retina Face	SSD	OpenCV
DLIB	0,0000	0,0008	0,0109	0,0309	0,0036
MTCNN	0,0008	0,0000	0,0060	0,0002	0,0078
Retina Face	0,0109	0,0060	0,0000	0,0051	0,0264
SSD	0,0309	0,0002	0,0051	0,0000	0,0550
OpenCV	0,0036	0,0078	0,0264	0,0550	0,0000

Tabela 8 – Diferença quadrática entre os detectores.

No total foram analisadas 9975 faces masculinas e 9927 femininas. Comparando com os números de faces não identificadas discriminados ao final da Seção 5.1.1, as faces não detectadas foram de homens asiáticos e mulheres indianas.

5.1.3 Identificação de raça e gênero pela face

Nesta subseção será feita a matriz de confusão do resultado na identificação de faces agrupando gênero e raça. O algoritmo classifica esse agrupamento em homens ou mulheres das raças asiática, negra, indiana, latina, árabes e brancos. As análises do algoritmo são feitas como citado nas subseções anteriores, a mudança ocorre na forma de realizar a matriz de confusão, onde foram agrupados os dois resultados.

Na Tabela 9 são apresentados as porcentagens de erros que os detectores apresentaram para cada agrupamento, usando imagens das faces originais e recortadas. Novamente a diferença no erro utilizando imagens originais e recortadas não foi significativo, então as demais análises serão feitas em cima dos resultados na identificação feita com imagens originais.

SSD (LIU et al., 2016) apresentou o melhor resultado geral, com um erro médio de 32,14%. OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) foi com o detector que obteve o pior desempenho, com uma taxa média de 40,06% de erro.

Para facilitar a visão dos resultados foram utilizadas as seguintes abreviações: H - homens, M - mulheres, A - asiáticos, N = negros, I - indianos, L - latinos, Ar - árabe B - brancos. As siglas contém duas letras, onde a primeira se refere ao gênero e a outra a raça.

		Imagens originais				
		DLIB	MTCNN	Retina Face	SSD	OpenCV
MA		24,84%	25,92%	20,24%	13,88%	31,08%
MN		66,48%	70,00%	63,84%	61,88%	68,60%
MI		65,02%	73,84%	68,04%	66,20%	78,90%
MB		20,76%	21,48%	15,16%	10,88%	26,00%
HN		5,68%	6,16%	6,04%	4,92%	5,56%
HA		15,64%	13,37%	20,12%	20,24%	18,42%
HI		60,48%	74,68%	66,96%	65,20%	70,24%
HB		14,80%	20,04%	14,72%	13,88%	21,68%
		Imagens recortadas				
		DLIB	MTCNN	Retina Face	SSD	OpenCV
MA		25,56%	29,08%	23,52%	23,20%	25,00%
MN		68,24%	71,00%	65,88%	65,80%	68,76%
MI		66,05%	75,94%	73,88%	69,24%	72,48%
MB		20,72%	24,16%	17,28%	16,88%	20,28%
HN		6,60%	6,80%	6,16%	5,52%	5,92%
HA		14,38%	13,45%	16,60%	16,16%	17,37%
HI		67,60%	76,24%	69,32%	73,64%	73,88%
HB		16,68%	20,68%	15,96%	15,84%	16,52%

Tabela 9 – Porcentagem de erro dos detectores com imagens originais e recortadas.

Faces de mulheres negras, asiáticas e brancas foram melhor identificadas com o detector SSD (LIU et al., 2016), com 61,88%, 13,88% e 10,88% respectivamente. Para mulheres indianas a menor taxa de erro ocorreu ao utilizar detector DLIB (KING, 2009), com 65,02% de erro. Os piores desempenhos de identificação para mulheres indianas, asiáticas e brancas,

ocorreram com uso do OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), com 78,90%, 31,08% e 26,00% respectivamente. O pior resultado para identificação de mulheres negras foi feito pelo detector MTCNN (XIANG; ZHU, 2017), com uma taxa de erro de 70%.

As faces de homens indianos tiveram a menor taxa de erro na identificação com o detector DLIB (KING, 2009), com 60,48%. Homens asiáticos tiveram a melhor identificação com o detector MTCNN (XIANG; ZHU, 2017), com a taxa de erro igual a 13,37%. O detector SSD (LIU et al., 2016) obteve os melhores resultados na identificação para homens brancos e negros, com 13,88% e 4,92% de erro respectivamente. Para homens indianos e negros, os piores resultados da identificação ocorreram com o detector MTCNN (XIANG; ZHU, 2017), com 76,68% e 6,16% respectivamente. SSD (LIU et al., 2016), obteve o pior resultado na identificação de homens asiáticos, com 20,24%. OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) foi responsável pela maior taxa de erro na identificação de homens brancos, com 21,68%.

Na Tabela 10 estão as validações cruzadas do melhor, SSD (LIU et al., 2016), e pior detector, OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), para os agrupamentos dessa seção. Os padrões de falsos negativos para cada agrupamento se mantiveram em todos os detectores.

Detector SSD												
	HA	MA	HN	MN	HI	MI	HL	ML	HAr	MAr	HB	MB
HA	0,7956	0,0784	0,0032	0,0004	0,0012	0,0000	0,0271	0,0028	0,0081	0,0000	0,0739	0,0093
MA	0,0712	0,8612	0,0004	0,0004	0,0000	0,0004	0,0000	0,0120	0,0004	0,0044	0,0088	0,0408
HN	0,0204	0,0008	0,9508	0,0044	0,0028	0,0000	0,0056	0,0000	0,0040	0,0000	0,0108	0,0004
MN	0,0308	0,1184	0,1992	0,3812	0,0072	0,0280	0,0132	0,1324	0,0024	0,0152	0,0060	0,0660
HI	0,0448	0,0004	0,0524	0,0004	0,3480	0,0036	0,1788	0,0028	0,2240	0,0004	0,1412	0,0032
MI	0,0062	0,0346	0,0012	0,0025	0,0284	0,3181	0,0103	0,1784	0,0148	0,2666	0,0119	0,1269
HL	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ML	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
HAr	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MAr	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
HB	0,0283	0,0000	0,0082	0,0000	0,0037	0,0000	0,0446	0,0000	0,0669	0,0000	0,8409	0,0074
MB	0,0064	0,0128	0,0008	0,0012	0,0004	0,0016	0,0004	0,0096	0,0004	0,0116	0,0636	0,8912
Detector OpenCV												
	HA	MA	HN	MN	HI	MI	HL	ML	HAr	MAr	HB	MB
HA	0,8158	0,0529	0,0044	0,0004	0,0016	0,0004	0,0303	0,0016	0,0109	0,0004	0,0711	0,0101
MA	0,2216	0,6892	0,0004	0,0004	0,0000	0,0004	0,0036	0,0100	0,0012	0,0036	0,0252	0,0444
HN	0,0200	0,0008	0,9444	0,0032	0,0040	0,0000	0,0072	0,0000	0,0052	0,0000	0,0140	0,0012
MN	0,0480	0,0724	0,3424	0,3140	0,0104	0,0148	0,0284	0,0868	0,0036	0,0084	0,0220	0,0488
HI	0,0508	0,0008	0,0824	0,0004	0,2976	0,0036	0,1972	0,0016	0,2400	0,0004	0,1236	0,0016
MI	0,0161	0,0313	0,0095	0,0037	0,0750	0,2110	0,0688	0,2365	0,0334	0,1389	0,0429	0,1331
HL	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ML	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
HAr	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MAr	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
HB	0,0468	0,0000	0,0126	0,0000	0,0037	0,0000	0,0714	0,0000	0,0996	0,0000	0,7613	0,0045
MB	0,0132	0,0196	0,0040	0,0028	0,0004	0,0020	0,0152	0,0304	0,0024	0,0100	0,1600	0,7400

Tabela 10 – Matriz de confusão para melhor e pior detector.

Para asiáticos, as maiores taxas de erro ocorreram no outro gênero, mulheres asiáticas classificadas como homens e vice-versa, em seguida também houve erros significativos classificando asiáticos como brancos em seus respectivos gêneros. Homens negros a maioria dos falsos positivos foram para homens asiáticos. Mulheres negras demonstraram um alto valor de falso positivo em homens negros, seguidos também de mulheres asiáticas e latinas. Indianas foram mais erroneamente classificadas como latinos e árabes, em seus respectivos gêneros. Homens brancos obtiveram a maior taxa de falso positivo para homens árabes e mulheres brancas para homens brancos.

	DLIB	MTCNN	Retina Face	SSD	OpenCV
DLIB	0,0000	0,0602	0,0479	0,0903	0,0586
MTCNN	0,0602	0,0000	0,0589	0,1118	0,0303
Retina Face	0,0479	0,0589	0,0000	0,0142	0,0949
SSD	0,0903	0,1118	0,0142	0,0000	0,1682
OpenCV	0,0586	0,0303	0,0949	0,1682	0,0000

Tabela 11 – Diferença quadrática entre os detectores.

Novamente, foi feito o cálculo da diferença ao quadrado das matrizes e a soma de todos os elementos, com as matrizes de confusão de todos detectores, duas a duas, a fim de demonstrar as diferenças entre os resultados dos detectores e analisar suas similaridades, Tabela 11.

5.2 Influência dos principais elementos faciais no processo de reconhecimento facial

Nesta seção serão apresentados os resultados feitos de acordo com a Subseção 4.5. A ordem a ser apresentada será o reconhecimento facial feito dentro da mesma raça, seguidos do resultado dentro do mesmo gênero, depois os resultados dentro do agrupamento entre cada gênero e raça e por fim, o reconhecimento facial com todos os gêneros e raças disponíveis no banco de dados. Os principais resultados serão demonstrados em tabelas, os outros estão presentes por seção no Anexo A

5.2.1 Reconhecimento facial dentro da mesma raça

Os valores para acurácia no reconhecimento facial entre raças com ambas imagens sem máscara e uma das faces com as principais *features* cobertas estão na Tabela 12. Nas subseções abaixo os resultados serão descritos com maiores detalhes para cada raça observada, asiáticos, negros, indianos e brancos.

Para todas as raças, a maior diferença de acurácia foi entre faces sem máscara e quando em uma das faces a *feature* nariz foi coberto, logo esses dados serão demonstrados nas próximas seções. Os demais resultados podem ser vistos no Anexo A.

	Raça asiática	Raça negra	Raça indiana	Raça branca
Original	95,03%	90,40%	98,38%	98,57%
Olho direito	93,54%	88,44%	98,05%	98,47%
Olho esquerdo	93,47%	87,82%	97,8%	97,99%
Nariz	90,10%	83,76%	96,3%	96,72%
Boca	93,33%	89,45%	97,90%	98,21%

Tabela 12 – Acurácia do reconhecimento facial de acordo com as *features* cobertas para o reconhecimento facial cada raça.

Para analisar os falsos positivos, foram feitas as Tabelas 14, 16, 18, 20. Uma matriz de confusão nos resultados em que o algoritmo identificou a face sendo igual a outra, mesmo pertencendo a indivíduos diferentes. Nas linhas são representadas as raças da face usada como base de comparação e nas colunas as raças das face onde são inseridas as máscaras para cobrir as principais *features*.

5.2.1.1 Raça asiática

A melhor acurácia no reconhecimento facial entre asiáticos ocorreu ao usar faces sem máscara, com 95.030% . Ao cobrir uma das faces nas *features* dos olhos ou boca, o resultado da acurácia cai para 93% . E com o nariz coberto a acurácia fica em 90%.

Na Tabela 13 é demonstrada a tabela verdade para os resultados com maior significância. A maior diferença entre as tabelas verdades das imagens originais e olhos cobertos, é um aumento do falso positivo e diminuição do verdadeiro positivo. Entre as imagens originais, nariz e boca cobertos, há um aumento na quantidade de falsos positivos.

Imagens Originais		
	Positivo	Negativo
Positivo	0,47	0,53
Negativo	0,05	0,95
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,48	0,52
Negativo	0,10	0,90

Tabela 13 – Tabela verdade normalizada do pior e melhor resultado do reconhecimento facial da raça asiática.

Observando a Tabela 14, para ambas imagens, originais e com nariz coberto, a maior quantidade de erros é dentro do mesmo gênero, ou seja, mulheres confundidas com mulheres e homens com homens. Na matriz feita com os resultados do reconhecimento facial para o nariz coberto, aumenta o reconhecimento positivo de faces de mulheres asiáticas com faces de homens asiáticos.

Imagens Originais		
	Mulheres	Homens
Mulheres	0,73	0,27
Homens	0,40	0,60
Imagens <i>feature</i> nariz coberto		
	Mulheres	Homens
Mulheres	0,66	0,34
Homens	0,37	0,63

Tabela 14 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça asiática.

5.2.1.2 Raça negra

No reconhecimento facial entre faces negras, a maior acurácia foi com as imagens originais, com 90,40%. A cobertura das *features* dos olhos e boca, apresentaram um resultado em torno de 88% e novamente o nariz coberto apresentou o pior resultado, com 83,76%.

Imagens Originais		
	Positivo	Negativo
Positivo	0,46	0,54
Negativo	0,09	0,91
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,46	0,54
Negativo	0,16	0,87

Tabela 15 – Tabela verdade normalizada do pior e melhor resultado do reconhecimento facial da raça negra.

Observando a Tabela 16, para ambos os resultados, a maior quantidade de erros do gênero feminino são no gênero masculino, ou seja, mulheres confundidas com homens. A cobertura do nariz em uma das faces, não afetou, de maneira significativa, a distribuição de falsos positivos do reconhecimento facial.

Imagens Originais		
	Mulheres	Homens
Mulheres	0,41	0,59
Homens	0,21	0,79
Imagens <i>feature</i> nariz coberto		
	Mulheres	Homens
Mulheres	0,44	0,56
Homens	0,24	0,76

Tabela 16 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça negra.

5.2.1.3 Raça indiana

A melhor acurácia no reconhecimento facial entre faces indianas acontece ao usar imagens originais, com 98,38% . Ao cobrir as *features*, as acurácias ficaram similares a

original, sendo o nariz a menor delas, com 96,31%.

Na Tabela 17 é demonstrada a tabela verdade para os resultados com maior significância. O reconhecimento facial feito com imagens em que o nariz é coberto, tem um aumento de falsos positivos e o número de verdadeiro positivo diminui, ao comparar com os resultados obtidos com as imagens originais.

Imagens Originais		
	Positivo	Negativo
Positivo	0,47	0,53
Negativo	0,01	0,99
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,26	0,74
Negativo	0,03	0,97

Tabela 17 – Tabela verdade normalizada do pior e melhor resultado do reconhecimento facial da raça indiana.

Observando a Tabela 18, para ambos os resultados, a maior quantidade de erros é dentro do mesmo gênero, mulheres confundidas com mulheres e homens com homens. A cobertura do nariz em uma das faces não afetou, de maneira significativa, a distribuição de falsos positivos do reconhecimento facial.

Imagens Originais		
	Mulheres	Homens
Mulheres	0,81	0,19
Homens	0,33	0,67
Imagens <i>feature</i> nariz coberto		
	Mulheres	Homens
Mulheres	0,82	0,18
Homens	0,38	0,62

Tabela 18 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça indiana.

5.2.1.4 Raça branca

A melhor acurácia no reconhecimento facial entre faces brancas acontece ao usar imagens originais, com 98,57% . Ao cobrir as *features*, as acurácias ficaram similares a original, sendo o nariz a menor delas, com 96,72%.

Na Tabela 19 é demonstrada a tabela verdade para os resultados com maior significância. O reconhecimento facial feito com imagens em que o nariz é coberto, tem um aumento de falsos positivos e o número de verdadeiro positivo diminui, ao comparar com os resultados obtidos com as imagens originais.

Imagens Originais		
	Positivo	Negativo
Positivo	0,26	0,74
Negativo	0,01	0,99
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,21	0,79
Negativo	0,03	0,97

Tabela 19 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial da raça branca.

Observando a Tabela 20, para ambos os resultados, a maior quantidade de erros para as faces de mulheres brancas é equilibrado para faces de mulheres e de homens brancos. A cobertura do nariz em uma das faces não afetou, de maneira significativa, a distribuição de falsos positivos do reconhecimento facial dentro da raça branca.

Imagens Originais		
	Mulheres	Homens
Mulheres	0,44	0,56
Homens	0,30	0,70
Imagens <i>feature</i> nariz coberto		
	Mulheres	Homens
Mulheres	0,50	0,50
Homens	0,31	0,69

Tabela 20 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial da raça branca.

5.2.2 Reconhecimento facial dentro do mesmo gênero

A Tabela 21 traz os valores de acurácias do reconhecimento facial dos gêneros feminino e masculino com as imagens sem máscaras e as utilizando cobertura das principais *features* em uma das faces. Nas subseções abaixo, uma para cada gênero, serão detalhados os resultados mais significativos para análise.

	Masculino	Feminino
Original	96,72%	97,35%
Olho direito	95,82%	96,35%
Olho esquerdo	95,78%	96,38%
Nariz	93,47%	94,23%
Boca	96,01%	96,14%

Tabela 21 – Acurácia do reconhecimento facial de acordo com as *features* cobertas em uma das faces para o reconhecimento facial de gênero.

Para analisar os falsos positivos, foram feitas as Tabelas 23 e 25. Uma matriz de confusão nos resultados em que o algoritmo identificou a face sendo igual a outra, mesmo pertencendo a indivíduos diferentes. Nas linhas são representadas as raças da face usada

como base de comparação e nas colunas as raças das faces onde são inseridas as máscaras para cobrir as principais *features*.

Serão demonstradas nesta seção somente as *features* que demonstraram maior diferença de acurácia quando comparadas ao reconhecimento facial feito com imagens sem máscara. Os demais resultados podem ser vistos no Anexo B.

5.2.2.1 Gênero masculino

A melhor acurácia no reconhecimento facial entre faces masculinas acontece ao usar imagens sem máscara, com 96,72%. Ao cobrir as *features* em uma das faces utilizadas, as acurácias ficaram similares a original, sendo o nariz a menor delas, com 93,47%.

Na Tabela 22 é demonstrado a tabela verdade para os resultados com maior significância, com imagens sem máscara e utilizando a cobertura de nariz em uma das faces. As demais tabelas estão demonstradas no Anexo B. reconhecimento facial feito com imagens em que o nariz é coberto, tem num aumento de falsos positivos e o número de verdadeiro positivo diminui, ao comparar com os resultados obtidos com as imagens originais.

Imagens Originais		
	Positivo	Negativo
Positivo	0,45	0,55
Negativo	0,03	0,97
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,37	0,63
Negativo	0,06	0,97

Tabela 22 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial entre faces do gênero masculino.

Observando a Tabela 23, tanto para imagens sem máscara quanto com quando o nariz é coberto em uma das faces, para negros e asiáticos, o maior erro foi dentro da própria raça. Para o reconhecimento facial dos indianos e brancos, 45% e 41%, respectivamente, são confundidos com faces negras. Para imagens com o nariz coberto, as diferenças se apresentam no aumento de confusão de faces negras com faces brancas e dentro da mesma raça branca.

Imagens Originais				
	Asiáticos	Negros	Indianos	Branco
Asiáticos	0,50	0,30	0,14	0,06
Negros	0,22	0,62	0,12	0,04
Indianos	0,26	0,45	0,23	0,06
Branco	0,24	0,41	0,25	0,10
Imagens <i>feature</i> nariz coberto				
	Asiáticos	Negros	Indianos	Branco
Asiáticos	0,43	0,35	0,12	0,10
Negros	0,23	0,50	0,15	0,12
Indianos	0,22	0,40	0,22	0,16
Branco	0,22	0,35	0,22	0,21

Tabela 23 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial do gênero masculino.

5.2.2.2 Gênero feminino

No reconhecimento para faces do gênero feminino, o melhor resultado foi para comparações entre imagens originais da face, com 97,35%. Com uma redução pequena na acurácia, o pior resultado se deu ao cobrir o nariz em uma das faces na comparação, com 94,23%.

A Tabela 24 contém a tabela verdade para os resultados de maior significância para comparação, imagens originais e a *feature* nariz coberto em uma das faces utilizadas. O valor que mais se altera, em relação aos dois métodos de reconhecimento utilizados, são a quantidade de falsos positivos, que aumenta.

Imagens Originais		
	Positivo	Negativo
Positivo	0,34	0,66
Negativo	0,02	0,98
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,32	0,68
Negativo	0,06	0,94

Tabela 24 – Tabela verdade do gênero feminino para imagens originais.

Observando a Tabela 25, no reconhecimento facial para imagens originais dentro do mesmo gênero, as raças asiáticas, negras e indianas, obtiveram o maior erro de identificação dentro das próprias raças, já a raça branca foi mais confundida com faces asiáticas. Nas imagens com nariz coberto, em comparação as originais, as asiáticas continuaram com um erro maior dentro da própria raça, mas houve um leve aumento de confusão com a raça branca. Negras também foram mais confundidas com faces asiáticas e brancas.

Imagens Originais				
	Asiáticas	Negras	Indianas	Branças
Asiáticas	0,73	0,13	0,11	0,03
Negras	0,30	0,43	0,19	0,08
Indianas	0,32	0,17	0,44	0,07
Branças	0,43	0,25	0,18	0,14
Imagens <i>feature</i> nariz coberto				
	Asiáticas	Negras	Indianas	Branças
Asiáticas	0,68	0,13	0,11	0,07
Negras	0,35	0,34	0,19	0,12
Indianas	0,32	0,16	0,42	0,10
Branças	0,44	0,20	0,17	0,19

Tabela 25 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial do gênero feminino.

5.2.3 Reconhecimento facial dentro do mesmo gênero e mesma raça

A Tabela 26 apresenta a acurácia para cada agrupamento de raça e gênero, com a variante de cobertura de *features* e imagens sem aplicação da máscara. A seguir serão demonstrados com maiores detalhes dos resultados significantes observados, as demais tabelas estão disponíveis no Anexo C.

	MA	HA	MN	HN	MI	HI	MB	HB
Original	89,32%	91,05%	93,70%	83,11%	97,01%	95,49%	97,89%	97,13%
OD	87,98%	89,27%	92,86%	80,70%	97,04%	94,71%	97,85%	96,23%
OE	88,25%	89,47%	92,62%	80,05%	97,07%	94,23%	97,69%	96,07%
Nariz	81,68%	84,61%	88,52%	76,56%	94,26%	91,12%	96,28%	94,23%
Boca	85,74%	89,39%	93,35%	83,18%	96,56%	94,85%	97,79%	96,49%

Tabela 26 – Acurácia do reconhecimento facial de acordo com as *features* cobertas para o reconhecimento facial entre faces de mulheres asiáticas.

5.2.3.1 Mulheres asiáticas

O maior valor de acurácia para o reconhecimento facial no agrupamento de faces de mulheres asiáticas, foi ao utilizar imagens sem aplicação de cobertura de *features*, com 89,32%. Novamente o pior resultado se deu ao cobrir o nariz da face, com 81,68%. Comparando os resultados da Tabela 27, o valor que mais sofre alteração, aumentando a quantidade, é o número de falsos positivos.

Imagens Originais		
	Positivo	Negativo
Positivo	0,50	0,50
Negativo	0,10	0,90
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,49	0,51
Negativo	0,18	0,8201

Tabela 27 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres asiáticas.

5.2.3.2 Homens asiáticos

No reconhecimento facial entre homens asiáticos, a acurácia para imagens sem aplicação de máscara é a mais alta com 91,05%. O valor mais baixo novamente é para comparações onde uma das faces teve o nariz coberto, com uma acurácia de 84,61%. Comparando os resultados da Tabela 28, a diferença vista é o aumento de falsos positivos, quando feita a cobertura do nariz em uma das faces.

Imagens Originais		
	Positivo	Negativo
Positivo	0,55	0,45
Negativo	0,08	0,92
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,48	0,52
Negativo	0,15	0,85

Tabela 28 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de homens asiáticos.

5.2.3.3 Mulheres Negras

No reconhecimento facial entre mulheres negras, a acurácia para imagens originais foi de 93,70%. O valor mais baixo é para o reconhecimento feito com uma das faces com o nariz coberto, com uma acurácia de 88,51%. Observando a Tabela 29, a diferença entre os dois métodos utilizados para o reconhecimento facial, é no aumento de falsos positivos ao utilizar a cobertura do nariz em uma das faces.

Imagens Originais		
	Positivo	Negativo
Positivo	0,46	0,54
Negativo	0,06	0,94
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,36	0,64
Negativo	0,11	0,89

Tabela 29 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres negras.

5.2.3.4 Homens Negros

O reconhecimento facial entre homens negros obteve as melhores acurácias com imagens originais com 83,11%. A acurácia para faces com o nariz coberto teve o pior resultado com 76,56%. Observando a Tabela 30, tabela verdade, ao comparar os resultados, existe um aumento no valor de falsos positivos quando o reconhecimento é feito cobrindo o nariz de uma das faces.

Imagens Originais		
	Positivo	Negativo
Positivo	0,53	0,47
Negativo	0,17	0,83
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,42	0,57
Negativo	0,23	0,77

Tabela 30 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de homens negros.

5.2.3.5 Mulheres Indianas

Entre faces de mulheres indianas, o reconhecimento facial apresentou a melhor acurácia para imagens originais com 97,01%. O valor mais baixo é para reconhecimento com faces que tiveram o nariz coberto, com uma acurácia de 94,25%. Comparando os resultados da tabela verdade, Tabela 31, a diferença é no aumento de falsos positivos e falsos negativos, para os resultados quando a *feature* do nariz está coberta em uma das faces.

Imagens Originais		
	Positivo	Negativo
Positivo	0,41	0,59
Negativo	0,02	0,98
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,30	0,70
Negativo	0,05	0,95

Tabela 31 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres indianas.

5.2.3.6 Homens Indianos

Entre faces de homens indianos, o reconhecimento facial apresentou uma acurácia para imagens originais com 95,49%. O pior resultado foi para o reconhecimento com uma das faces teve o nariz coberto, com uma acurácia de 91,12%. A Tabela 32, tem os resultados da tabela verdade de ambos os reconhecimentos, a diferença é no aumento de falsos positivos e de falsos negativos quando é coberto o nariz em uma das faces.

Imagens Originais		
	Positivo	Negativo
Positivo	0,30	0,70
Negativo	0,04	0,96
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,22	0,78
Negativo	0,08	0,92

Tabela 32 – Tabela verdade normalizada melhor e pior resultado do reconhecimento facial de homens indianos.

5.2.3.7 Mulheres Brancas

Entre faces de mulheres brancas, o reconhecimento facial apresentou uma diferença baixa nas acurácias de cada método de cobertura de *feature*. Para imagens sem máscara em nenhuma *feature* a acurácia foi de 97,89%. O valor mais baixo de acurácia é para reconhecimento com uma das faces com o nariz coberto, num valor de 96,28%. Comparando os resultados da Tabela 33, tabela verdade, a diferença é no aumento de falsos positivos para o reconhecimento feito com uma das faces com o nariz coberto.

Imagens Originais		
	Positivo	Negativo
Positivo	0,29	0,71
Negativo	0,01	0,99
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,18	0,82
Negativo	0,03	0,97

Tabela 33 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de mulheres brancas.

5.2.3.8 Homens Brancos

O reconhecimento facial entre homens brancos apresentou, para faces originais a melhor acurácia, com 97,13%. O pior resultado é para o reconhecimento usando uma das faces com o nariz coberto, com uma acurácia de 94,23%. Na Tabela 34, tabela verdade, pode-se notar que a diferença é no aumento de falsos positivos quando usado uma das faces com a *feature* do nariz coberto.

Imagens Originais		
	Positivo	Negativo
Positivo	0,39	0,61
Negativo	0,02	0,98
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,25	0,75
Negativo	0,05	0,95

Tabela 34 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de homens brancos.

5.2.4 Reconhecimento facial entre todas as raças e gêneros

A Tabela 35 representa a acurácia para cada modo que foram realizados os reconhecimento faciais, com faces originais e com cada *feature* coberta em uma das faces. O desempenho para os métodos com a face original, uma face com olhos ou boca cobertos, foi similar, apresentando uma acurácia por volta de 97%, os resultados para essas *feature* estão demonstrados no Anexo ?? . O pior resultado foi para o reconhecimento com uma das faces com o nariz coberto, com um resultado de 94,74% de acurácia.

	Acurácia
Original	97,91%
Olho direito	97,05%
Olho esquerdo	97,07%
Nariz	94,74%
Boca	97,17%

Tabela 35 – Acurácia do reconhecimento facial entre todas as raças e gêneros.

Na Tabela 36, com a tabela verdade do melhor e pior resultado, é possível notar um aumento nos falsos positivos e falsos negativos no reconhecimento que usa uma das faces com o nariz coberto.

Imagens Originais		
	Positivo	Negativo
Positivo	0,38	0,62
Negativo	0,02	0,98
Imagens <i>feature</i> nariz coberto		
	Positivo	Negativo
Positivo	0,28	0,72
Negativo	0,05	0,95

Tabela 36 – Tabela verdade normalizada do melhor e pior resultado do reconhecimento facial de todas as raças e gêneros.

Para detectar onde ocorreram os maiores números de falsos positivos, foi feita a tabela de matriz de confusão para os métodos que apresentaram o pior e melhor desempenho, Tabela 37

No reconhecimento facial feito com a face original, para mulheres asiáticas, o algoritmo acusou um maior número de outras mulheres da mesma raça como mesma pessoa, mas também houve uma alta quantidade de confusão para serem homens asiáticos. Para faces de homens asiáticos, a maior confusão de falsos positivos ocorreu dentro do próprio agrupamento, seguido das faces de mulheres asiáticas, homens negros e homens indianos. Com a *feature* nariz coberto em uma das faces, os falsos positivos para o reconhecimento facial de mulheres asiáticas se estenderam também para faces de homens negros. Para faces de homens asiáticos, os falsos positivos continuaram nos mesmos grupos do que com imagens originais.

No agrupamento de faces de mulheres negras, o maior número de falsos positivos foi na própria raça, no gênero masculino, seguido de mulheres negras, homens e mulheres asiáticos. Já para homens negros a maior confusão se deu no próprio agrupamento, seguido de mulheres negras, homens e mulheres asiáticos. Com o nariz coberto em uma das faces, o comportamento de falsos positivos foi o mesmo para mulheres e homens negros.

Imagens Originais								
	MA	HA	MN	HN	MI	HI	MB	HB
MA	0,5275	0,2325	0,0175	0,0950	0,0050	0,0700	0,0525	0,0000
HA	0,1914	0,3730	0,0703	0,1758	0,0156	0,1133	0,0547	0,0059
MN	0,1635	0,1945	0,1979	0,2857	0,0241	0,0861	0,0361	0,0120
HN	0,1029	0,1648	0,1099	0,4276	0,0190	0,1209	0,0380	0,0170
MI	0,1915	0,2340	0,0426	0,1277	0,2021	0,1277	0,0745	0,0000
HI	0,1821	0,2299	0,0627	0,2299	0,0299	0,1881	0,0448	0,0328
MB	0,1522	0,3261	0,0000	0,1957	0,0000	0,1304	0,1957	0,0000
HB	0,1479	0,2296	0,0311	0,1829	0,0156	0,1984	0,0623	0,1323
Imagens <i>feature</i> nariz coberto								
	MA	HA	MN	HN	MI	HI	MB	HB
MA	0,3681	0,1986	0,0548	0,1468	0,0290	0,0962	0,0434	0,0631
HA	0,1661	0,2402	0,0959	0,2028	0,0351	0,1365	0,0374	0,0858
MN	0,1595	0,1459	0,2285	0,2311	0,0491	0,0891	0,0355	0,0613
HN	0,1145	0,1568	0,1401	0,3145	0,0370	0,1282	0,0295	0,0793
MI	0,1457	0,1424	0,1490	0,1457	0,1656	0,1026	0,0762	0,0728
HI	0,1660	0,1730	0,1026	0,2042	0,0563	0,1368	0,0523	0,1087
MB	0,1218	0,2179	0,0833	0,1795	0,0385	0,1795	0,0833	0,0962
HB	0,1391	0,1732	0,0735	0,2021	0,0407	0,1614	0,0472	0,1627

Tabela 37 – Matriz de confusão dentro dos falsos positivos no reconhecimento facial entre todas as raças.

O maior número de falsos positivos, ao utilizar as imagens originais para as faces de mulheres indianas, foram no próprio agrupamento e homens asiáticos, seguidos de mulheres asiáticas e homens negros. Para as faces de homens indianos, a maior quantidade de falsos positivos foi para homens negros e asiáticos, seguidos dos homens indianos e mulheres asiáticas. O padrão de confusão se manteve ao utilizar imagens com a *feature* nariz coberto em uma das faces.

Para faces de mulheres brancas, ao utilizar as imagens originais, faces de homens asiáticos foram mais confundidas como pertencendo a mesma pessoa, seguida de mulheres asiáticas e brancas, homens negros e indianos. As faces de homens brancos foi mais confundida com faces de homens asiáticos, seguido de mulheres asiáticas, homens negros, indianos e brancos. Novamente o padrão de confusão se manteve ao cobrir a *feature* do nariz em uma das faces.

5.3 Discussão dos Resultados

5.3.1 Identificação pela face

O modelo de reconhecimento facial utilizado nos resultados da Seção 5.1 foi o VGG-face (PARKHI; VEDALDI; ZISSERMAN, 2015), feito em 2015 e teve como base de treinamento com 2.6 milhões de imagens de 2622 pessoas retiradas da internet, pertencentes a celebridades. No mundo tem-se mais de 7 bilhões de pessoas, logo a quantidade de pessoas utilizadas para treinamento não foi nem de 1% da população mundial, o que é uma amostra

extremamente pequena. Menor ainda vinda de uma amostra de celebridades que não possui um balanceamento adequado de raça e gênero.

No ano de treinamento deste modelo as preocupações eram em se ter um modelo de reconhecimento facial, não se importando ainda com o desempenho para raças e gêneros especificamente. A quantidade de imagens pode ter ajudado ao aprendizado de máquina entender diferentes iluminações e posições da face, informações que já eram suficientes na época que foi feito o treinamento. Mas com o número restrito de pessoas, talvez não tenha sido possível analisar corretamente diferentes características faciais ao redor do mundo.

Para a identificação racial, os detectores DLIB (KING, 2009) e MTCNN (XIANG; ZHU, 2017) apresentaram o melhor e pior, com taxas de erro de 23,18% e 30,14% resultado respectivamente. O detector DLIB (KING, 2009) se destaca principalmente pela baixa porcentagem de erros para a raça indiana, se comparado ao resultados dos demais. Com as taxas de erros demonstradas na Tabela 3, o padrão de erro varia os melhores resultados entre asiáticos e brancos, seguido da raça negra, e por último, sempre com uma taxa de erro mais de 50%, os indianos.

DLIB (KING, 2009), feito em 2009 e atualizado em 2017, classifica as imagens fazendo uso da extração HOG (DALAL; TRIGGS, 2005) e hiperplano SVM, detectando caixas que contenham faces, como explicado na Subseção 2.2.2. O HOG, foi treinado em 2005 utilizando imagens do banco de dados INRIA, com 1805 imagens de pedestres em diversas posições e diferentes fundos nas imagens, retiradas de arquivos pessoais. O método de extração da face deste detector é diferenciado por ser dada em gradientes e não em *features* faciais.

O detector MTCNN (XIANG; ZHU, 2017) foi feito em 2017, e em seu treinamento foram usadas faces de dois bancos de dados. A WIDERFACE (YANG et al., 2015) para ter faces negativas, positivas e partes parciais da face. Ela foi extraída de um banco de imagens WIDER (XIONG et al., 2015), feito em 2015, com imagens em diversas situações e várias faces em uma única imagem, porém novamente sem preocupação com o balanceamento nas raças e gêneros apresentados nas faces. Com as informações das coordenadas das *features*, foi utilizado o banco de dados de CelebA (XIANG; ZHU, 2017), este conta com faces com diversas características, como cor de cabelo, bigode, óculos, nariz pontudo. No total são 10.177 pessoas, que mais uma vez é uma amostra pequena e novamente sem o cuidado de ter o balanceamento nas imagens.

O MTCNN (XIANG; ZHU, 2017) obteve taxa de erro consideravelmente maior para indianos se comparado com o DLIB (KING, 2009), as demais raças não apresentaram uma diferença acentuada.

Com a Tabela 5, a soma da diferença ao quadrado dos valores das matrizes de confusão de cada detector, tem-se o maior valor para os detectores DLIB (KING, 2009) e MTCNN (XIANG; ZHU, 2017), demonstrando que os vieses presentes em cada um, são os mais diferentes.

O detector SSD (LIU et al., 2016) é o que mais se assemelha ao DLIB (KING, 2009), com um resultado de 0.0087. Quanto ao detector que obteve pior desempenho, o OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) possui o menor valor de comparação, com 0.0062.

Na análise considerando apenas o gênero, observando a Tabela 6, os valores entre o melhor e pior detector, SSD (LIU et al., 2016) e OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), possuem a uma diferença da porcentagem de erro para o gênero feminino de mais de 15%, enquanto para o gênero masculino não chega a 2%. O melhor resultado feminino, 12,15%, é cerca de 10 vezes maior que o melhor masculino, 1,26%.

O detector SSD (LIU et al., 2016), demonstrado em 2015, foi treinado utilizando a base de imagens COCO (LIN et al., 2014), que contém não apenas faces, mas diversas categorias, incluindo 250000 pessoas com dados de pontos-chaves. Ele faz uso de um varrimento de janelas pela imagem para a detecção facial e analisa a presença de instâncias de objetos na imagem em diferentes escalas da imagem.

O detector OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), de 2016, foi treinado com uma mistura de dois grandes bancos de dados com rótulos, o CASIA-WebFace (YI et al., 2014) e FaceScrub (NG; WINKLER, 2014), totalizando 500 mil imagens. CASIA-WebFace (YI et al., 2014) tem 10.575 indivíduos e 494.414 imagens de faces, todos pertencentes a celebridades, devido a facilidade de achar informações sobre os indivíduos. FaceScrub (NG; WINKLER, 2014) com 141.130 faces de 695 indivíduos, novamente celebridades, porém se preocupou em manter o gênero balanceado, 50% das faces são identificadas com gênero masculino e as outras feminino. O OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) também faz uso de um varrimento de janelas pela imagem para a detecção facial, mas analisando as bordas das janelas para delimitar a face.

Com os resultados da Tabela 8, o detector que possui o viés mais próximo do SSD (LIU et al., 2016) foi o MTCNN (XIANG; ZHU, 2017), com um resultado de 0.0002, enquanto o OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) apresentou a maior diferença, com 0.0550.

No agrupamento de raça e gênero, novamente o melhor detector foi o SSD (LIU et al., 2016) e o com pior desempenho foi o OpenCV (AMOS; LUDWICZUK; SATYANARAYANAN, 2016). Observando a Tabela 9, nota-se um padrão de resultados para todos os detectores. Indianos de ambos os gêneros e mulheres negras estão entre os 3 piores resultados e apresentam sempre um erro maior que 60%. Todos os detectores também apresentam o melhor desempenho na análise de homens negros, onde a porcentagem mais alta de erro foi de 6.80%, com o detector MTCNN (XIANG; ZHU, 2017). Ambos os gêneros para asiáticos e brancos variam as taxas entre a casa dos 10% e 20%.

Com a Tabela 37 pode-se concluir que a dificuldade na classificação muitas vezes

está na diferenciação de gênero, dentro da raça corretamente escolhida, como é o caso para asiáticos, negros e brancos. Indianos, mostrados como a raça com pior análise, são os que possuem maior confusão na classificação, variando entre diversos agrupamentos, demonstrando baixo conhecimento sobre a raça indiana.

5.3.2 Reconhecimento facial

O reconhecimento facial nesse trabalho foi feito com o modelo Facenet (SCHROFF; KALENICHENKO; PHILBIN, 2015), treinado em 2015 com um banco de dados privado e com uso da distância euclideana normalizada para detecção. Obteve a melhor acurácia para o banco de dados LFW (HUANG et al., 2007), o que demonstra um maior conhecimento sobre faces. E foi o que apresentou os melhores resultados para o reconhecimento facial segundo teste feito pelo criador do código usado, (SERENGIL, S. I.; OZPINAR, 2020a), realizado em (SERENGIL, S., s.d.).

O intuito de realizar os testes cobrindo as *features* em uma das faces. era encontrar uma que causasse uma diferença de comportamento para o reconhecimento facial em indivíduos em que o viés existe, como mulheres negras e a raça indiana, como visto na identificação facial. Porém a cobertura no nariz resultou em uma diminuição da acurácia para todos os agrupamentos de raças e gêneros, não sendo um problema particular de alguma. Com os resultados demonstrados na Subseção 5.2.1, pode-se afirmar que há uma acurácia menor no reconhecimento facial feito entre pessoas da raça negra, ou seja, o modelo apresenta uma maior dificuldade de diferenciar pessoas negras uma das outras, o segundo pior resultado é para asiáticos, brancos e indianos apresentaram resultados similares. Na Subseção 5.2.2, não há uma diferença considerável entre as acurácias para os gêneros femininos e masculino. No reconhecimento feito dentro de cada agrupamento, na Subseção 5.2.3, quando feito entre homens negros, possui a menor acurácia entre todos, seguido de mulheres asiáticas, homens asiáticos e mulheres negras.

6 Conclusão

A proposta deste trabalho foi avaliar a influência de vieses de gênero e raça usando técnicas do reconhecimento facial. Na primeira parte avaliando o desempenho de diferentes detectores na identificação do indivíduo pela face e na segunda o desempenho para cada agrupamento no reconhecimento facial analisando a cobertura de *features* principais da face.

Na identificação do indivíduo foram detectado vieses em cada agrupamento feito. Para a raça indiana, a classificação obteve um índice de erro maior que 50% em todos os detectores. Para gêneros, mulheres tiveram o desempenho de todos os detectores cerca de 20 vezes pior que para homens. Entre o agrupamento de gênero e raça o viés foi observado para mulheres negras e indianas, homens indianos, todos tiveram uma taxa de erro maior que 50% em todos os detectores.

A classificação por faces é uma tarefa difícil de ser realizada até por seres humanos. Não foi encontrada uma acurácia comparativa entre uma visão humana e os classificadores utilizados, mas pode-se afirmar que ambas nunca serão feitas com uma taxa de erro de 0%. O que se espera é que a taxa de erro seja similar para todas as raças e gêneros.

No reconhecimento facial utilizando *features* cobertas, não foi detectada nenhuma que tivesse maior influência no erro do modelo nos agrupamentos apontados acima com presença de viés. Foi confirmado que o cabelo não influencia no reconhecimento, assim como apontado por (MUTHUKUMAR et al., 2019), mas não foi feita nenhuma conclusão quanto ao uso de maquiagem nas faces utilizadas. Também foi encontrado que a *feature* do nariz tem maior influência para todas as análises feitas, o que concorda com os achados de (MERY; MORRIS, 2022), onde é concluído que a parte central do rosto tem maior efeito na assertividade dos modelos.

Vale lembrar que como discutido no Capítulo 5, nenhum detector ou modelo utilizado foi treinado com um banco de dados balanceado em gênero e raça, o que faz com que eles não tenham total conhecimento sobre todas as raças e gêneros do mundo. Mesmo que óbvio, é necessário ressaltar nenhum código é feito com intenção de causar qualquer viés, o aprendizado de máquina pode inferir resultados de situações que não conhece com o que lhe é ensinado, porém podem não apresentar um desempenho ótimo.

Também é demonstrado nos resultados encontrados, onde ocorrem os erros de cada agrupamento, tanto da identificação quanto no reconhecimento facial, pelas tabelas de matriz de confusão apresentadas e discutidas em cada seção do Capítulo 5.

Dentro dos possíveis trabalhos futuros, primeiro pode ser realizado um treinamento

com uma base de dados desbalanceada para alguma raça, que não seja a branca, e detectar se ocorrem os mesmos problemas de vieses, quando se analisa uma face cuja raça não foi bem representada na base de treinamento. Dessa forma poderá ser feito um veredito o quanto o treinamento influencia nos resultados.

Relacionado ao primeiro, o segundo problema a ser resolvido para trabalhos futuros é o banco de dados utilizados nos treinamentos. Uma base de dados como a BFW (ROBINSON et al., 2021) ser rotulada devidamente com dados de raça, identidade e gênero para que seja possível utilizá-la de nos treinamentos ou usar a técnicas de transferência de aprendizado em modelos de reconhecimento facial. Assim se descarta o problema de um banco de dados desbalanceado, sendo possível uma análise mais correta de como o modelo se comporta para cada raça e gênero.

Importante a observação que a base de dados deve ser balanceada em relação a quantidade de representações para cada raça e gênero, logo o mesmo número de mulheres e homens por exemplo.

Um terceiro problema, se faz, na avaliação de bases de dados de pessoas em processo de redesignação de gênero. Estas bases são escassas e ainda poucas são disponíveis na literatura. Sendo um teste importante, por exemplo, avaliar se indivíduos que estão em processo final ou finalizado de redesignação, se estes encaixam nas categorias em que o viés foi percebido na realização deste trabalho (ex.: um homem negro que realiza o processo de redesignação para uma mulher negra). Por exemplo, com o uso do conjunto de dados *Hormone Replacement Therapy* (HRT) (SATOR et al., 2001), em que é apresentado imagens faciais no início e no final do tratamento de redesignação de gênero, no qual seria feita uma avaliação distinta nestes dois momentos (inicial e final) do tratamento. Observando e mensurando se as classes atribuídas pelos modelos de reconhecimento facial, irão indicar possíveis vieses para estes indivíduos.

Referências

- AMOS, B.; LUDWICZUK, B.; SATYANARAYANAN, M. **OpenFace: A general-purpose face recognition library with mobile applications**. 2016. Citado nas pp. 13, 15, 24, 25, 35, 39–43, 59.
- BALLANTYNE, M.; BOYER, R. S.; HINES, L. Woody bledsoe: His life and legacy. **AI magazine**, v. 17, n. 1, p. 7–7, 1996. Citado na p. 12.
- BRADSKI, G.; KAEHLER, A. **Learning OpenCV: Computer vision with the OpenCV library**. "O'Reilly Media, Inc.", 2008. Citado nas pp. 15, 19, 35, 36.
- BUOLAMWINI, J.; GEBRU, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: FAT. 2018. Citado nas pp. 13, 28, 31, 32, 34.
- CORBETT-DAVIES, S.; GOEL, S. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. **ArXiv**, abs/1808.00023, 2018. Citado na p. 13.
- CORTES, C.; VAPNIK, V. Support Vector Networks. **Machine Learning**, v. 20, p. 273–297, 1995. Citado na p. 18.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**, v. 1, 886–893 vol. 1, 2005. Citado nas pp. 18, 58.
- 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 nas pp. 15, 26, 27.
- DENG, J.; GUO, J.; VERVERAS, E.; KOTSIA, I.; ZAFEIRIOU, S. RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. In: CVPR. 2020. Citado nas pp. 15, 20, 34.
- EITAN, A. T.; SMOLYANSKY, E.; HARPAZ, I. K.; PERETS, S. **Connected Papers**. Disponível em: <<https://www.connectedpapers.com>>. Citado na p. 32.
- 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. 16.
- GIRSHICK, R. Fast R-CNN. **CoRR**, abs/1504.08083, 2015. Disponível em: <http://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf>. Citado na p. 22.
- GROUP, N. L. C. R-NET: Machine Reading Comprehension with Self-matching Networks. In: Disponível em: <<https://www.microsoft.com/en-us/research/publication/mcr/>>. Citado na p. 17.

- 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. 23, 25, 33, 60.
- IBM. **Redes neurais**. Disponível em: <<https://www.connectedpapers.com>>. Citado na p. 16.
- IVAKHNENKO, A.; LAPA, V. **Cybernetic Predicting Devices**. CCM Information Corporation, 1973. (Jprs report). Disponível em: <<https://books.google.com.br/books?id=FhwVNQAACAAJ>>. Citado na p. 15.
- JAIN, V.; LEARNED-MILLER, E. **FDDB: A Benchmark for Face Detection in Unconstrained Settings**. 2010. Citado na p. 18.
- KANADE, T. **Picture Processing System by Computer Complex and Recognition of Human Faces**. Nov. 1973. Citado na p. 12.
- KAUR, P.; KRISHAN, K.; SHARMA, S. K.; KANCHAN, T. Facial-recognition algorithms: A literature review. **Medicine, Science and the Law**, v. 60, n. 2, p. 131–139, 2020. PMID: 31964224. DOI: 10.1177/0025802419893168. eprint: <https://doi.org/10.1177/0025802419893168>. Disponível em: <<https://doi.org/10.1177/0025802419893168>>. Citado na p. 13.
- KING, D. E. Dlib-ml: A Machine Learning Toolkit. **Journal of Machine Learning Research**, v. 10, p. 1755–1758, 2009. Citado nas pp. 15, 18–20, 24, 35, 38–40, 42, 43, 58, 59.
- LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; BOURDEV, L.; GIRSHICK, R.; HAYS, J.; PERONA, P.; RAMANAN, D.; ZITNICK, C. L.; DOLLÁR, P. **Microsoft COCO: Common Objects in Context**. arXiv, 2014. DOI: 10.48550/ARXIV.1405.0312. Disponível em: <<https://arxiv.org/abs/1405.0312>>. Citado na p. 59.
- LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S. E.; FU, C.-Y.; BERG, A. C. SSD: Single Shot MultiBox Detector. In: ECCV. 2016. Citado nas pp. 15, 21, 39–43, 59.
- MAGHSOUDI, O. H.; GASTOUNIOTI, A.; PANTALONE, L.; DAVATZIKOS, C.; BAKAS, S.; KONTOS, D. O-Net: An Overall Convolutional Network for Segmentation Tasks. **Machine learning in medical imaging. MLMI**, v. 12436, p. 199–209, 2020. Citado na p. 17.
- MARTIN KOESTINGER PAUL WOHLHART, P. M. R.; BISCHOF, H. Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization. In: PROC. FIRST IEEE INTERNATIONAL WORKSHOP ON BENCHMARKING FACIAL IMAGE ANALYSIS TECHNOLOGIES. 2011. Citado na p. 18.
- MERLER, M.; RATHA, N. K.; FERIS, R. S.; SMITH, J. R. Diversity in Faces. **ArXiv**, abs/1901.10436, 2019. Citado nas pp. 31, 32.

- MERY, D.; MORRIS, B. On Black-Box Explanation for Face Verification. In: PROCEEDINGS of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Jan. 2022. P. 3418–3427. Citado nas pp. 30, 33, 61.
- MITCHELL, T. **Machine Learning**. McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673. Disponível em: <<https://books.google.com.br/books?id=EoYBngEACAAJ>>. Citado nas pp. 13, 15, 16.
- MUTHUKUMAR, V.; PEDAPATI, T.; RATHA, N.; SATTIGERI, P.; WU, C.-W.; KINGSBURY, B.; KUMAR, A.; THOMAS, S.; MOJSILOVIĆ, A.; VARSHNEY, K. R. Color-Theoretic Experiments to Understand Unequal Gender Classification Accuracy From Face Images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019. P. 2286–2295. DOI: 10.1109/CVPRW.2019.00282. Citado nas pp. 28–30, 32, 61.
- NAGPAL, S.; SINGH, M.; SINGH, R.; VATSA, M. **Deep Learning for Face Recognition: Pride or Prejudiced?** 2019. arXiv: 1904.01219 [cs.CV]. Citado nas pp. 29, 30, 32.
- NG, H.-W.; WINKLER, S. A data-driven approach to cleaning large face datasets. In: 2014 IEEE International Conference on Image Processing (ICIP). 2014. P. 343–347. DOI: 10.1109/ICIP.2014.7025068. Citado na p. 59.
- PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Deep Face Recognition. In: BRITISH Machine Vision Conference. 2015. Citado nas pp. 13, 15, 23, 24, 34, 35, 57.
- PHILLIPS, P. J.; WECHSLER, H.; HUANG, J.; RAUSS, P. J. The FERET database and evaluation procedure for face-recognition algorithms. **Image Vis. Comput.**, v. 16, n. 5, p. 295–306, 1998. Disponível em: <<http://dblp.uni-trier.de/db/journals/ivc/ivc16.html#PhillipsWHR98>>. Citado na p. 12.
- REN, S.; HE, K.; GIRSHICK, R.; SUN, J. **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks**. arXiv, 2015. DOI: 10.48550/ARXIV.1506.01497. Disponível em: <<https://arxiv.org/abs/1506.01497>>. Citado na p. 17.
- ROBINSON, J. P.; QIN, C.; HENON, Y.; TIMONER, S.; FU, Y. Balancing Biases and Preserving Privacy on Balanced Faces in the Wild. **arXiv preprint arXiv:2103.09118**, 2021. Citado nas pp. 31, 33, 35, 36, 62.
- SATOR, P. G.; SCHMIDT, J. B.; SATOR, M. O.; HUBER, J. C.; HÖNIGSMANN, H. The influence of hormone replacement therapy on skin ageing: a pilot study. **eng. Maturitas**, v. 39, n. 1, p. 43–55, jul. 2001. Place: Ireland. ISSN 0378-5122. DOI: 10.1016/s0378-5122(00)00225-5. Citado na p. 62.

- SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. FaceNet: A unified embedding for face recognition and clustering. **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 815–823, 2015. Citado nas pp. 13, 15, 22–25, 35, 60.
- SERENGIL, S. **The Best Face Recognition Model: FaceNet, VGG-Face, DeepFace, OpenFace**. Disponível em: <https://www.youtube.com/watch?v=i_M0wvhhLdI>. Citado nas pp. 35, 36, 60.
- SERENGIL, S. I.; OZPINAR, A. LightFace: A Hybrid Deep Face Recognition Framework. In: IEEE. 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). 2020a. P. 23–27. DOI: 10.1109/ASYU50717.2020.9259802. Disponível em: <<https://doi.org/10.1109/ASYU50717.2020.9259802>>. Citado nas pp. 4, 5, 15, 26, 34–36, 60.
- SERENGIL, S. I.; OZPINAR, A. LightFace: A Hybrid Deep Face Recognition Framework. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). 2020b. P. 1–5. DOI: 10.1109/ASYU50717.2020.9259802. Citado na p. 15.
- SHI, J.; MALIK, J. Normalized cuts and image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 8, p. 888–905, 2000. DOI: 10.1109/34.868688. Citado nas pp. 15, 27.
- SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **CoRR**, abs/1409.1556, 2014. Disponível em: <<http://arxiv.org/abs/1409.1556>>. Citado na p. 21.
- SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **CoRR**, abs/1409.1556, 2015. Citado na p. 23.
- SINGHAL, A. Modern Information Retrieval: A Brief Overview. **IEEE Data Eng. Bull.**, v. 24, n. 4, p. 35–43, 2001. Disponível em: <<http://dblp.uni-trier.de/db/journals/debu/debu24.html#Singhal01>>. Citado nas pp. 15, 26.
- SULEIMAN, A.; CHEN, Y.-h.; EMER, J. S.; SZE, V. Towards closing the energy gap between HOG and CNN features for embedded vision (Invited paper). **2017 IEEE International Symposium on Circuits and Systems (ISCAS)**, p. 1–4, 2017. Citado na p. 19.
- TAIGMAN, Y.; YANG, M.; RANZATO, M.; WOLF, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. **2014 IEEE Conference on Computer Vision and Pattern Recognition**, p. 1701–1708, 2014. Citado nas pp. 15, 25, 35.
- 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. Citado na p. 20.

-
- XIANG, J.; ZHU, G. Joint Face Detection and Facial Expression Recognition with MTCNN. **2017 4th International Conference on Information Science and Control Engineering (ICISCE)**, p. 424–427, 2017. Citado nas pp. [15](#), [17](#), [18](#), [34–36](#), [39](#), [41](#), [43](#), [58](#), [59](#).
- XIONG, Y.; ZHU, K.; LIN, D.; TANG, X. Recognize Complex Events from Static Images by Fusing Deep Channels. In: IEEE. COMPUTER Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. 2015. Citado na p. [58](#).
- YANG, S.; LUO, P.; LOY, C. C.; TANG, X. **WIDER FACE: A Face Detection Benchmark**. arXiv, 2015. DOI: [10.48550/ARXIV.1511.06523](https://arxiv.org/abs/1511.06523). Disponível em: <https://arxiv.org/abs/1511.06523>. Citado na p. [58](#).
- 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. [18](#).
- YI, D.; LEI, Z.; LIAO, S.; LI, S. Z. **Learning Face Representation from Scratch**. arXiv, 2014. DOI: [10.48550/ARXIV.1411.7923](https://arxiv.org/abs/1411.7923). Disponível em: <https://arxiv.org/abs/1411.7923>. Citado na p. [59](#).
- ZHIFEI, Z.; HAIRONG, Q. Age Progression/Regression by Conditional Adversarial Auto-encoder. In: IEEE. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. Citado na p. [33](#).

Anexos

ANEXO A – Tabelas normalizadas para features de olhos e boca coberta no reconhecimento facial entre raças

A.1 Raça asiática

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.4166	0.5833
Negativo	0.0617	0.9382
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.4074	0.5925
Negativo	0.0627	0.9372
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.4583	0.5416
Negativo	0.0640	0.9359

A.2 Raça negra

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.4663	0.5337
Negativo	0.1135	0.8864
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.4507	0.5492
Negativo	0.1197	0.8802
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.3834	0.6165
Negativo	0.1030	0.8969

A.3 Raça indiana

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.2472	0.7527
Negativo	0.0161	0.9838
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.3021	0.6978
Negativo	0.0184	0.9815
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.2527	0.7472
Negativo	0.0180	0.9819

A.4 Raça branca

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.1813	0.8186
Negativo	0.0161	0.9839
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.1987	0.8013
Negativo	0.0164	0.9836
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.1593	0.8406
Negativo	0.0141	0.9859

ANEXO B – Tabelas normalizadas para features de olhos e boca coberta no reconhecimento facial entre gêneros

B.1 Gênero masculino

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.3960	0.6039
Negativo	0.0403	0.9596
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.4059	0.5940
Negativo	0.0408	0.9591
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.3564	0.6435
Negativo	0.0383	0.9616

B.2 Gênero feminino

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.3243	0.6757
Negativo	0.0347	0.9653
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.2882	0.7118
Negativo	0.0343	0.9657
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.3063	0.6937
Negativo	0.0367	0.9633

ANEXO C – Tabelas normalizadas para features de olhos e boca coberta no reconhecimento facial mesma raça e gênero

C.1 Mulheres asiáticas

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.4277	0.5722
Negativo	0.1155	0.8844
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.4415	0.5584
Negativo	0.1129	0.8870
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.4514	0.5485
Negativo	0.1384	0.8615

C.2 Homens asiáticos

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.4940	0.5059
Negativo	0.1004	0.8995
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.4822	0.5177
Negativo	0.1001	0.8998
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.4249	0.5750
Negativo	0.1022	0.8977

C.3 Mulheres Negras

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.4231	0.5768
Negativo	0.0662	0.9337
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.4211	0.5789
Negativo	0.0686	0.9313
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.3652	0.6347
Negativo	0.0607	0.9392

C.4 Homens Negros

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.507	0.4929
Negativo	0.1911	0.8099
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.5020	0.4919
Negativo	0.1965	0.8034
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.4349	0.565
Negativo	0.1655	0.8344

C.5 Mulheres Indianas

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.3611	0.6388
Negativo	0.0236	0.9764
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.3590	0.6409
Negativo	0.0234	0.9765
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.2839	0.7160
Negativo	0.0278	0.9721

C.6 Homens Indianos

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.2773	0.7226
Negativo	0.0464	0.9535
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.2815	0.7184
Negativo	0.0513	0.9486
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.2226	0.7773
Negativo	0.0445	0.9554

C.7 Mulheres Brancas

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.2460	0.7540
Negativo	0.0140	0.9860
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.2244	0.7756
Negativo	0.0153	0.9847
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.1712	0.8288
Negativo	0.0138	0.9862

C.8 Homens Brancos

Imagens feature olho direito coberto		
	Positivo	Negativo
Positivo	0.3286	0.6714
Negativo	0.0313	0.9687
Imagens feature olho esquerdo coberto		
	Positivo	Negativo
Positivo	0.3206	0.6794
Negativo	0.0328	0.9672
Imagens feature boca coberta		
	Positivo	Negativo
Positivo	0.2965	0.7035
Negativo	0.0283	0.9717