UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Reduzindo viés em classificação de tons de pele em bases de dados de imagens

Luiz Augusto Vieira Manoel

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP
SÃO CARLOS

Luiz Augusto Vieira Manoel

Reduzindo viés em classificação de tons de pele em bases de dados de imagens

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Moacir Antonelli Ponti

USP – São Carlos
Novembro de 2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M266r

Manoel, Luiz Augusto Vieira
Reduzindo viés em classificação de tons de pele em bases de dados de imagens / Luiz Augusto Vieira Manoel; orientador Moacir Antonelli Ponti. -- São Carlos, 2022.
83 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2022.

1. Classificação de tom de pele. 2. Reconhecimento facial. 3. Classificação justa. 4. Processamento de imagens. 5. Diversidade de faces. I. Ponti, Moacir Antonelli, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Luiz Augusto Vieira Manoel

Reducing bias in skin tone classification in image databases

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Moacir Antonelli Ponti

USP – São Carlos
November 2022

Este trabalho é dedicado a todos que lutam contra a exploração e as opressões. Á memória de Vladimir Herzog, José Luís e Rosa Sundermann, Marielle Franco e Carlos Marighella.

Dedico também a todos cientistas pesquisadores que através de suas atividades de pesquisa, dentro e fora da Universidade, buscam servir aos interesses da maioria da população, a classe trabalhadora. Dedico àqueles que estão dentro da Universidade, lutando contra a privatização, contra a lógica produtivista, pela educação pública e pela construção de uma Universidade verdadeiramente popular. Dedico àqueles que não estão na Universidade, mas pesquisam incessantemente para contribuir com o avanço da ciência e da tecnologia no sentido da construção de uma sociedade verdadeiramente justa.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Meus agradecimentos são direcionados àqueles e àquelas que, de alguma forma, contribuíram com a execução dessa dissertação.

Primeiramente, agradeço ao professor Moacir Antonelli Ponti, meu orientador. Desde Julho de 2019, através de nossas reuniões semanais construímos juntos esse projeto de pesquisa. Escolhemos um tema novo para ambos e o desenvolvimento só foi possível através de muita colaboração, dedicação, orientação, paciência e ajuda do professor. Moacir é um grande cientista pesquisador, que através do seu vasto conhecimento e de sua intervenção na academia e na realidade e sua preocupação social, contribui e ainda contribuirá muito com o avanço da ciência em direção à uma sociedade verdadeiramente justa.

Meus colegas do grupo de pesquisa de Visualização, Processamento de Imagem e Computação Gráfica foram solícitos em me ajudar com problemas relacionados aos meus experimentos e ao acesso aos servidores da USP sempre que precisei. Obrigado à todos!

Sou grato aos trabalhadores da USP que através de suas atividades garantem a permanência de milhares de estudantes na universidade. Notadamente aos trabalhadores e trabalhadoras do restaurante universitário, que lutaram muito contra a privatização.

Agradeço também a minha família e notadamente aos meus pais, Liane e José, por nunca terem medido esforços para me proporcionar ensino de qualidade, saúde, acolhimento e todo tipo de auxílio necessário. Durante parte desse trabalho não recebi bolsa e a ajuda financeira deles foi essencial para que fosse concluído.

Não posso esquecer de todos meus amigos, amigas e camaradas que dividiram a vida comigo durante o período de realização desse trabalho, tornando mais fácil de ser vivida. Em especial a Beatriz França, Fernanda Neves, Gabriel Castro, Gabriel Colombo, Gabriel Luiz, Gabriel Zovaro, Gustavo Avellar, Jenifer Braz, Jorge Vilaça, Karen Zentil, Letícia Ribeiro, Matheus Manoel, Pedro Conde, Renan Meneghetti, Thiago Barssoti, Tiago Gimenez, Victoria Queiroz e Victor Durço.

E por fim, agradeço também aos comunistas do PCB, que unindo teoria e prática nas lutas diárias e na construção do Poder Popular, contribuíram muito para minha formação política e teórica.

"Que o futuro nos traga dias melhores e a capacidade de construir
a Universidade que está em nossos corações,
nas nossas mentes e nas nossas necessidades."
(Florestan Fernandes)

RESUMO

MANOEL, L. A. V. **Reduzindo viés em classificação de tons de pele em bases de dados de imagens**. 2022. 83 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Grandes conjuntos de dados de imagens de faces são frequentemente usados para treinar e implementar soluções de visão computacional para reconhecimento facial. Nesse contexto, a aplicação de tais modelos em diferentes populações levanta preocupações sobre a existência de exemplos suficientes e representativos para as diversas classes / grupos em termos de gênero, idade e cor da pele, entre outros atributos. O viés de seleção desses dados é inserido durante a coleta. É difícil encontrar bancos de dados que sejam anotados por cor da pele ou atributos étnico-raciais, o que também dificulta o estudo de viés de seleção nesse contexto em aprendizado de máquina. O objetivo deste projeto de pesquisa é propor e avaliar um método para detecção de tons de pele em imagens que tenham desempenho equilibrado para diferentes tipos de pele e que permita auditar bases de dados de forma a minimizar problemas com viés de seleção em modelos de reconhecimento facial em direção a uma classificação justa. O método proposto consiste em aplicar diferentes abordagens de processamento de imagens e algoritmos para rotulagem automática da cor da pele, selecionando as melhores abordagens para cada tipo de cor de pele (usando o sistema de classificação *Fitzpatrick Skin Type*) de acordo com o *F-score* obtido e aplicando-as em ordem de prioridade. Mostramos que o uso de uma única abordagem tende a direcionar os melhores resultados para faixas específicas de tons de pele, enquanto a combinação reduz o viés geral e melhora a classificação em diferentes tipos de pele. Aplicamos a proposta no banco de dados de faces *LFW* e no banco de dados dermatológico *Fitzpatrick17k* usando transformações gama, *CLAHE*, equalização de histogramas e filtros estatísticos de ordem não linear. Mostramos que um extrator de características com pesos pré treinados da *Facenet* usando o modelo de rede neural convolucional *ResNet50* como base tem pior desempenho na distinção de pessoas de pele escura e que é possível mitigar esse efeito através de técnicas de pré-processamento de imagens combinando abordagens que sejam melhores em cada faixa de tom de pele para obter um método de rotulação automática de grandes bancos de dados que se aproxime da rotulação manual. Por fim, disponibilizamos para futuros trabalhos, além da descrição do método, um destacamento da *LFW* com anotações manuais de tons de pele de 150 pessoas únicas e anotações de cor de pele para cada imagem da base *LFW* completa feitas a partir do método de classificação automática proposto.

Palavras-chave: Classificação de tom de pele, Reconhecimento facial, Classificação justa, Processamento de imagens, Diversidade de faces.

ABSTRACT

MANOEL, L. A. V. **Reducing bias in skin tone classification in image databases**. 2022. 83 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Large face datasets are often used to train and deploy Computer Vision solutions for face recognition. In this context, the application of such models in different populations raises concerns about the existence of sufficient and representative examples for the diverse classes/groups in terms of gender, age and skin color, among others attributes. The selection bias in such data may be inserted during the sample collection. It is difficult to find databases that are annotated by skin color or ethnic-racial attributes, which also makes it difficult to study selection bias in this context in machine learning. The objective of this work is to propose and evaluate an automatic classification method for skin tones in images with balanced performance across different skin types and that allows auditing databases in order to minimize problems with selection bias in facial recognition models towards a fair classification. The proposed method consists of applying different processing approaches and algorithms for automatic skin color labeling, selecting the best approaches for each skin color type (using the dermatologist approved Fitzpatrick Skin Type classification system) according to the F-score obtained and applying them in an order of priority. We show that using a single approach will bias the best results towards specific skin tone ranges, while combining it reduce the overall bias and improves classification across different skin types. We applied the proposal in the LFW faces database and in the Fitzpatrick17k dermatological database using gamma transformations, CLAHE, histogram equalization and non-linear order statistics filters. We showed that a feature extractor with pre-trained weights from *Facenet* using the convolutional neural network model *ResNet50* as a base has worse performance in distinguishing dark-skinned people and that it is possible to mitigate this effect through image pre-processing techniques combining approaches that are best across each skin tone range to achieve an auto-labeling approach of large databases that approximates manual labeling. Finally, we provide for future works, in addition to the description of the method, an LFW detachment with manual annotations of skin tones of 150 unique people and also skin color annotations for each image of the complete LFW base made from of the proposed automatic classification method.

Keywords: Skin tone classification, Face recognition, Fair classification, Image Processing, Representativeness.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

Tarefas tradicionalmente feitas por seres humanos têm sido progressivamente mais automatizadas por algoritmos computacionais, em particular envolvendo Inteligência Artificial (IA) (RUSSELL, 2010). A escolha por esses algoritmos é justificada pela velocidade na resposta e possibilidade de aplicação em larga escala e, assim, soluções que usam IA tem sido crescentemente adotadas por empresas e pelo poder público para realizar desde tarefas como detecção e reconhecimento de pessoas e objetos em cenas, até ajudar a determinar quem pode ou não ser contratado ou receber empréstimos (O'NEIL, 2016). Tais sistemas se apoiam largamente na disponibilidade de dados a partir dos quais é possível aprender os conceitos necessários a uma determinada tarefa. Esse aprendizado de conceitos por meio de exemplos é conhecido por Aprendizado de Máquina (MELLO; PONTI, 2018).

Sistemas de reconhecimento facial são um claro exemplo desse cenário. Assim como outros que usam tecnologias de Aprendizado de Máquina, precisam ser alimentados com bases de dados contendo milhares ou milhões de fotos de diversas pessoas. Essas imagens são fornecidas ao computador com anotações sobre o conceito a ser aprendido. Por exemplo: "essas 10 diferentes fotos são de uma mesma pessoa; essas outras 9 fotos são de outra". Isso faz com que os algoritmos aprendam quais são as características visuais presentes nas imagens que melhor diferenciam os rostos e assim sejam capaz de produzir representações matemáticas, comumente consistindo de uma série de números organizados em uma estrutura vetorial para cada um desses rostos. Essas representações também são chamadas de *embeddings* (que poderíamos traduzir por imersão porém o termo em Inglês é mais comum mesmo nos textos científicos em Português) (PONTI *et al.*, 2021). Esse processo de encontrar as melhores características que diferenciam as faces é a inteligência computacional envolvida no sistema (WECHSLER, 2009).

Produzir uma grande base de dados é difícil e custoso. As grandes bases para reconhecimento facial atualmente disponíveis publicamente são de pessoas famosas e figuras públicas, em sua maioria de grandes centros como os Estados Unidos e países europeus. Sabendo que o reco-

nhecimento feito por esses sistemas está intrinsecamente ligado com a base de dados com a qual eles são alimentados, o que acontece quando os exemplos presentes nessas bases não refletem a diversidade dos rostos da população sobre a qual esses sistemas serão aplicados? Essa pergunta se justifica pois estudos recentes mostram como muitas bases de dados não são suficientemente diversas levando em conta questões de gênero, idade, cor da pele e etnia (MERLER *et al.*, 2019). A aplicação de sistemas de reconhecimento treinados a partir de uma base que não reflete, por exemplo, os percentuais étnico-racial, de gênero e de idade da população alvo, pode fazer com que as representações matemáticas e o sistema de reconhecimento resultante privilegiem certos recortes de gênero, idade e étnico-raciais. Ainda, os métodos do estado-da-arte utilizados para essa aplicação (SCHROFF; KALENICHENKO; PHILBIN, 2015) podem reforçar esse viés ao otimizar seus modelos (YU *et al.*, 2018).

É ainda mais difícil avaliar problemas relacionados à falta de representatividade e ao viés de seleção (GORDON; DESJARDINS, 1995) no contexto do reconhecimento facial em um cenário em que grande parte das bases de dados disponíveis para a construção de sistemas desse tipo não trazem anotações sobre o tom de pele ou algum atributo étnico-racial que descreva seus elementos. O trabalho de Merler *et al.* (2019) deixa explicita essa lacuna a partir de um levantamento. Como podemos analisar se um sistema de reconhecimento facial treinado a partir de uma base de dados específica tem melhor performance em uma faixa de cor de pele se não temos a informação de qual é a faixa de cor de pele de cada amostra dessa base?

Pesquisas científicas nas áreas de diversidade de bases de dados e viés de seleção, podem conduzir um avanço considerável para uma classificação justa. Esse trabalho vem na direção de conduzir uma pesquisa na área de visão computacional que reduza o viés de seleção no contexto étnico-racial e forneça bases mais sólidas para o desenvolvimento de aplicações com critério mais justo.

1.1 Motivação

A sociedade em que habitamos vive uma contradição entre lucro, trabalho e avanço científico e tecnológico. Os que tem mais dinheiro tem mais poder e usam este poder em busca de mais dinheiro (MOROZOV; MARCONDES, 2018). Assim, também usam da mídia e outros aparelhos ideológicos (ALTHUSSER, 1985) para se elegerem como representantes do povo e garantir seus próprios interesses. O avanço científico e tecnológico só pode ser entendido dentro da história e das condições concretas da realidade. Hoje, em grande parte do mundo, inclusive no Brasil, está direcionado ao lucro e não a melhoria da qualidade de vida da maioria da população (MOROZOV; MARCONDES, 2018).

As tecnologias, também localizadas nesse mundo de injustiças, refletem os problemas sociais que vivemos. A Figura 1 evidencia um pequeno exemplo de viés na ferramenta de tradução da empresa *Google* que provoca injustiça com pessoas transgênero. O nome “Leo” é

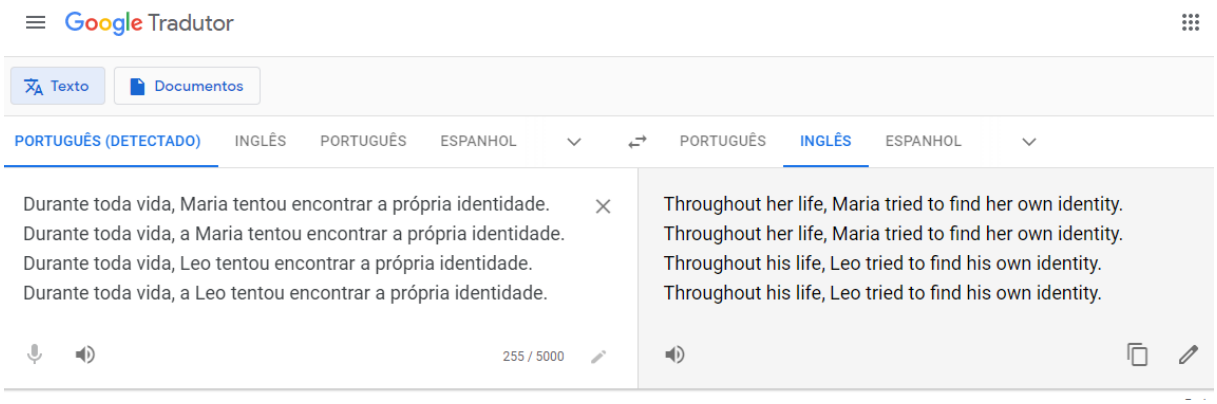

Figura 1 – Exemplo de viés de seleção no tradutor da *Google*

traduzido para a língua inglesa na ferramenta como uma pessoa do gênero masculino, mesmo quando acompanhada pelo pronome “a” na língua portuguesa, i.e. “A Leo”. Isso provavelmente se deve ao fato de que o tradutor é "treinado"a partir de exemplos que consideram que "Leo"só pode ser um homem e que o pronome "a"é um erro de português.

Segundo a pesquisadora Sherry Wolf, a LGBTfobia, assim como o machismo e o racismo, está historicamente ligada às formas de reprodução capitalista e serve para dividir a classe trabalhadora, nos colocando muitas vezes uns contra os outros quando na verdade deveríamos nos unir nas batalhas contra a injustiça (WOLF, 2009). Esses segmentos da classe são marginalizados e usados como trabalhadores a baixos custos para o Estado.

Outro exemplo de como as tecnologias refletem problemas sociais é o estudo recente de pesquisadores da USP que investigaram a relação entre reconhecimento facial e viés algorítmico no transporte público em grandes municípios brasileiros (BRANDÃO; OLIVEIRA, 2021). Fica claro que existem falhas e que a mera existência de protocolos não é suficiente para corrigir vieses algorítmicos de raça e/ou de gênero. No Capítulo 3 de revisão da literatura são apresentados mais exemplos onde a tecnologia reflete as injustiças sociais.

1.2 Objetivos

O objetivo deste projeto de pesquisa é propor e avaliar um método para detecção de tons de pele que permita auditar bases de dados de forma a minimizar problemas com viés de seleção em modelos de reconhecimento facial.

Em específico temos os seguintes objetivos:

- Obter abordagem de classificação automática de cor de pele que mais se aproxima de uma abordagem manual no contexto de imagens de face em grandes bases de dados para reconhecimento facial. Em particular, também queremos um método que funcione de maneira semelhante em diferentes tipos de pele;

- Entender se é possível mitigar o viés de seleção no reconhecimento facial através de métodos de processamento de imagens e classificação automática de cor de pele como pré-processamento para obtenção de *embeddings*;
- Analisar criticamente o desenvolvimento científico e tecnológico na área do aprendizado de máquina e mais especificamente do reconhecimento facial; e
- Fornecer bases para avaliar a representatividade étnico-racial em bases de dados utilizadas como *benchmark* na área no estudo de faces.

1.3 Hipótese

É possível obter um método de detecção automática de tons de pele que possua resultados objetivos similares para diferentes tons de pele, de forma a servir para auditar bases de dados com relação à recortes étnico-raciais.

1.4 Organização

No Capítulo 2 são descritos os conceitos fundamentais para o entendimento da dissertação. Buscamos fornecer as definições necessárias para que leitores que não são da área da computação consigam acompanhar a leitura do trabalho, sem deixar de lado o rigor científico nas conceituações.

Em seguida, é apresentada a revisão da literatura (Capítulo 3). Nesta parte são apresentados trabalhos que discutem o viés no contexto do aprendizado de máquina e o papel da tecnologia na sociedade. Através da revisão podemos enxergar a direção da literatura na área de pesquisa, levantando possíveis lacunas.

No Capítulo 4 são descritas as etapas principais da proposta, o detalhamento de cada etapa, as tecnologias e bases de dados utilizadas e as métricas de avaliação.

Em seguida, são apresentados todos os resultados (Capítulo 5) da investigação proposta através de gráficos, figuras, tabelas e textos explicativos.

Com os resultados apresentados, no Capítulo 6 é feita uma discussão que analisa cada um deles dentro do contexto da dissertação.

Por fim, o Capítulo 7 traça algumas das limitações desse trabalho de dissertação e também busca apontar caminhos para as próximas pesquisas na área de classificação justa no contexto do viés de seleção étnico-racial em sistemas de reconhecimento facial.

CAPÍTULO

2

CONCEITOS FUNDAMENTAIS

Neste Capítulo são apresentados os principais conceitos abordados na literatura relacionada e no texto deste trabalho, para que fique claro o significado atribuído para eles na redação e para que a compreensão da qualificação seja facilitada, considerando também leitores que não são estudiosos da área da computação. Nesse contexto, são desenvolvidos principalmente conceitos relativos ao Aprendizado de Máquina (AM), tipos de vieses encontrados no AM, Redes Nerais, algoritmos de classificação, agrupamento e projeção, processamento de imagens e sistema de cores.

2.1 Aprendizado de Máquina (AM)

O Aprendizado de Máquina é um conceito da área da Inteligência Artificial que trabalha com a ideia de que programas de computador podem aprender conceitos de forma automática através do acesso a um grande volume de dados. Estes dados representam experiências acumuladas, exemplos. A partir deles são compreendidas funções que criam ou detectam padrões e classificam novos exemplos com base no que a máquina aprendeu.

Seja X um subespaço de entrada, a partir dos quais são amostrados exemplos x_i, e Y um espaço de saída representado pelos valores a serem preditos a partir de X, por exemplo rótulos em um problema de classificação. Formalmente, um algoritmo que realiza Aprendizado de Máquina infere uma função $f : X \rightarrow Y$, que mapeia um exemplo de X em um valor do espaço Y. Após o aprendizado f pode ser considerado um modelo que é uma aproximação da distribuição de probabilidade conjunta $P(X,Y)$, que relaciona as variáveis X e Y (MELLO; PONTI, 2018).

No contexto do Aprendizado de Máquina, dados são valores atribuídos e sistematizados. Podem ser obtidos pela percepção através dos sentidos ou pela execução de um processo de medição ou captura. Em geral, é uma representação simbólica ou um atributo de uma entidade. Uma imagem da face de uma pessoa é um dado, interpretado por computadores como uma matriz

de valores (*pixels*) em uma escala de cores.

Bases de dados são coleções organizadas de dados que se relacionam entre si podendo gerar informações que podem ser estudadas. Grandes bases de dados são necessárias para uma máquina aprender conceitos através do uso da Inteligência Artificial.

Rótulos ou anotações são metadados, ou seja, dados que descrevem atributos de outros dados. Um texto que diz se uma imagem tem ou não um determinado objeto contido nela é um rótulo da imagem (ou metadado de um dado).

2.2 Viés em Aprendizado Supervisionado

Conforme apontado por Mello e Ponti (2018), viés em Aprendizado de Máquina, de forma geral pode ser definido pelo espaço de funções admissíveis de um algoritmo que realiza aprendizagem a partir dos dados. Estas funções tentam aproximar a distribuição de probabilidade conjunta $P(X,Y)$ e podem possuir características mais restritas (por exemplo, lineares como é o caso de Support Vector Classifiers), ou mais amplas, aproximando virtualmente qualquer tipo de função, como no caso das redes neurais Multilayer Perceptron. O viés também pode ser modificado por estratégias de treinamento como regularização.

Além do viés definido pelo espaço de funções admissíveis por um determinado algoritmo de aprendizado, há ainda um viés que pode advir da base de dados utilizada para treinamento. Considerando o espaço de variáveis de entrada X, comumente temos acesso a uma amostra pequena do espaço completo, e.g. $X' \subset X$. Muitos algoritmos assumem que os dados amostrados serão independentes e identicamente distribuídos (i.i.d), ou seja, que cada exemplo na população tem a mesma chance de ser observado. Assim, a estratégia de amostragem deve garantir essa premissa do contrário o aprendizado terá viés definido por X'.

Formalmente, assuma que amostramos um par (x_i, y_i) que será utilizado no treinamento do modelo de aprendizado com o fim de estimar $P(X,Y)$. Para que a amostra seja independente, o par i não pode afetar a probabilidade de amostrar um novo (x_{i+1}, y_{i+1}). Para que seja identicamente distribuída, é preciso garantir que os pares amostrados possam vir de qualquer subespaço da variável X. Nesse trabalho estamos interessados em particular no viés advindo da amostra X' selecionada para treinar um modelo de aprendizado.

Viés de seleção no Aprendizado de Máquina ocorre quando uma base de dados não reflete em termos de representatividade a porcentagem de determinados grupos presentes no conjunto de dados onde os resultados do aprendizado serão testados ou aplicados (GORDON; DESJARDINS, 1995). No domínio do reconhecimento facial, pode acontecer quando a base de dados usada para o treino da máquina apresenta uma porcentagem mais baixa de pessoas negras do que a população onde a máquina irá aplicar seu novo conhecimento. Esse tipo de viés é facilmente identificado em sistemas de reconhecimento facial.

No contexto do viés de seleção, estudaremos a questão da falta de representatividade. Para isso, pretendemos aplicar uma análise interseccional. que é o estudo da sobreposição de duas ou mais características ou identidades sociais oprimidas na sociedade (MERLER *et al.*, 2019). Segundo Crenshaw (1989) uma análise desse tipo é capaz de identificar como uma injustiça se expressa em uma base multidimensional. Em bases de dados para reconhecimento facial podemos estudar o viés provocado pela ausência de exemplos suficientes de mulheres negras no conjunto de treino, por exemplo.

2.3 Redes neurais artificiais

Na ciência de computação, redes neurais artificiais são modelos matemáticos inspirados na estrutura biológica do sistema nervoso de animais. São compostas por várias unidades de processamento que se ligam através de uma comunicação com um peso associado. Estes modelos são usados para o Aprendizado de Máquina, uma vez que a maioria destas redes são estruturadas de forma que os pesos de suas conexões são ajustados de acordo com os padrões experienciados.

Para Dawson e Wilby (1998), uma rede neural pode ser formalmente descrita pela por uma função f que é uma composição de outras funções g_i que por sua vez podem ser decompostas em outras funções e assim sucessivamente. Pode ser representada por uma estrutura em rede, com linhas indicando dependências entre funções (neurônios), assim como exposto na Figura 2.

Redes neurais profundas são redes neurais artificiais que possuem múltiplas camadas de unidades de processamento (ou nerônios) entre a entrada (dados) e a saída. As camadas entre entrada e saída são chamadas de ocultas. Os dados de entrada são transformados sucessivamente, com a saída da camada anterior fornecendo entrada para a próxima camada. A saída da última camada pode ser comparada com anotações, em inglês também chamado de *ground truth* (quando há essa informação temos aprendizado supervisionado), e os pesos da conexões reajustados a partir de uma função objetivo computada em função dos dados e/ou dessas anotações (PONTI *et al.*, 2017).

2.3.1 Embedding

Um *embedding* é uma representação vetorial de um dado de alta dimensionalidade. No contexto do Aprendizado de Máquina, é obtido por meio do aprendizado das principais características de um objeto não estruturado. Considerando imagens por exemplo, podemos extrair *embeddings* de dimensões menores que a matriz bruta de pixels através de uma rede neural profunda. Neste caso, a camada de entrada é alimentada com todos os *pixels*, as conexões representam os pesos, as camadas intermediárias são unidades de processamento e as saídas das camadas intermediárias podem ser utilizadas como representações da entrada, as quais chamamos *embedding* (PONTI *et al.*, 2017).

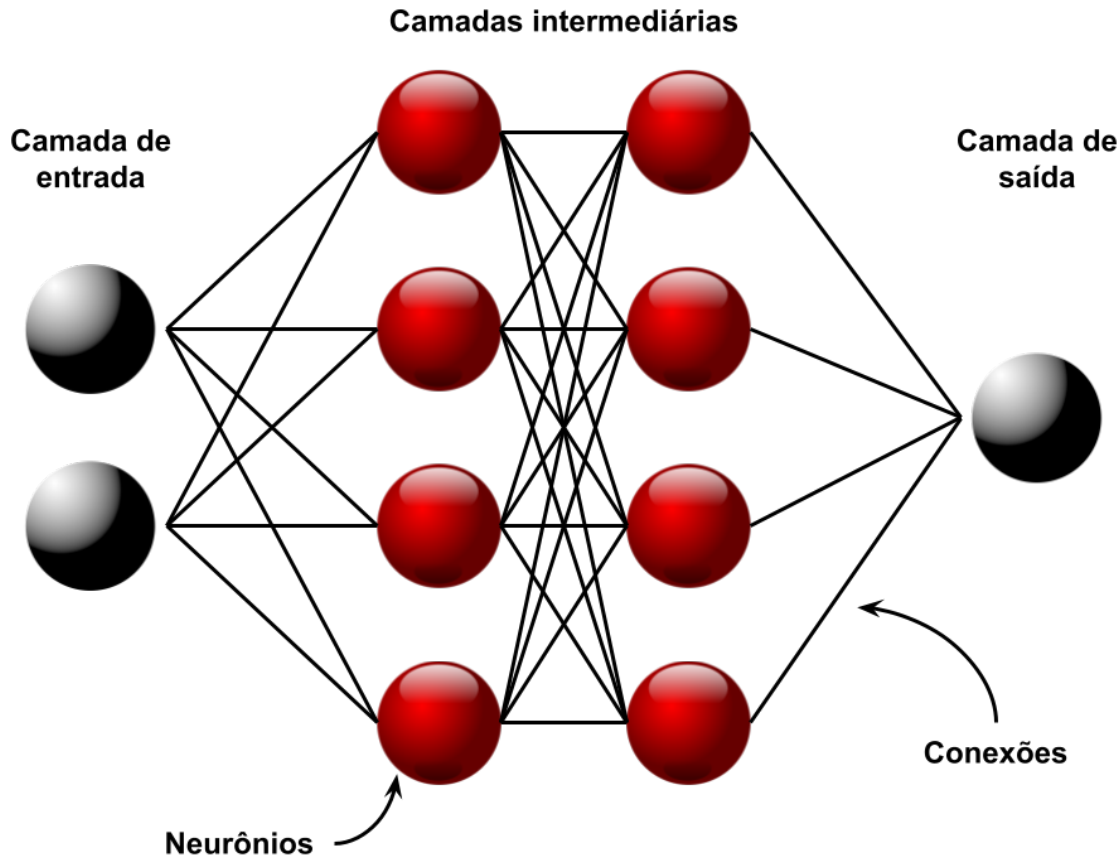

Figura 2 – Ilustração de Rede Neural Artificial com múltiplas camadas e conexões.

Matematicamente pode ser definida como uma instância de uma estrutura formal contida em outra instância. Dadas instâncias X e um espaço de saída Y, vários *embeddings* são possíveis. Seja um objeto X e um rótulo ou alvo Y, o *embedding* é mapeado de forma implícita por uma função injetora $f : X \rightarrow Y$ (LANG, 2012). Quando métodos de aprendizado profundo são utilizados, podemos escrever esse processo como um mapeamento sequencial $f : X \rightarrow Z \rightarrow Y$, em que Z é a projeção ou imersão (*embedding*) do espaço de entrada, a qual é utilizada como representação para computar o espaço de saída Y.

No contexto de reconhecimento facial, a rede neural *FaceNet* (SCHROFF; KALENICHENKO; PHILBIN, 2015) tem grande destaque ao empregar funções de custo do tipo *triplet* para aprender *embeddings* com excelente capacidade de discriminar entre indivíduos, mesmo quando considerando variações de rotação, iluminação entre outros. Podem ser utilizadas várias redes neurais profundas como base, sendo as Redes Residuais (HE *et al.*, 2016) uma opção comum na literatura.

2.4 Algoritmos de agrupamento

Os algoritmos de agrupamento são usados para identificar padrões em conjuntos de dados não rotulados. São conjuntos de técnicas computacionais de Aprendizado de Máquina e cálculos variados de distância ou semelhança que visam realizar agrupamentos automáticos. A ideia é agrupar objetos similares baseando-se em uma função de dissimilaridade, que recebe dois objetos e retorna a distância entre eles (LINDEN, 2009).

No contexto deste projeto, são usados para agrupar os *embeddings* obtidos através de redes neurais profundas, com o objetivo de avaliar a semelhança entre os exemplos de cada grupo e as diferenças entre exemplos de diferentes grupos. Exemplos de algoritmos são o *KMeans* proposto no trabalho de MacQueen (1967) e o *Affinity Propagation*, formulado no estudo de Frey e Dueck (2007).

2.5 Algoritmos de projeção

Algorítimos de projeção usam transformações lineares para realizar a projeção de um vetor sobre outro. Podem ser utilizados como técnicas de redução de dimensionalidade, que usam transformações matemáticas para encontrar as componentes principais de um vetor. São amplamente usados para permitir que grandes vetores sejam transformados em novos sistemas de coordenadas, a fim de que possam ser plotados em um gráfico com duas ou três dimensões. Exemplos são o *Principal Component Analysis (PCA)* (WOLD; ESBENSEN; GELADI, 1987) e o *t-distributed stochastic neighbor embedding (t-SNE)* (MAATEN; HINTON, 2008).

2.6 Recuperação de Imagem Baseada em Conteúdo

Recuperação de imagem baseada em conteúdo é a técnica de visão computacional utilizada no problema de busca de imagens digitais em grandes bases de dados. A busca analisa o conteúdo da imagem (ou *embedding*) e não metadados. O termo "conteúdo"nesse contexto se refere a cores, formas, texturas ou qualquer outra informação que possa ser derivada das próprias imagens (TORRES; FALCÃO, 2008).

2.7 Processamento de imagens

Conforme mencionado na Seção 2.1, uma imagem pode ser interpretada por computadores como uma matriz de *pixels* em um sistema de cores. Nesse contexto, conseguimos definir como uma função bidimensional, $f(x,y)$, em que x e y são coordenadas espaciais (plano), e a amplitude de f em qualquer par de coordenadas (x, y) é chamada de intensidade (GONZALEZ; WOODS, 2000). Esses elementos que tem coordenadas e intensidade são chamados de *pixels*. Processamento de imagens são todos processos que têm imagens como entradas e saídas, ou que

extraem outros tipos de informações através do processamento de *pixels*, como uma matriz de características ou *bounding boxes*.

Existem muitas técnicas de processamento de imagens. Aqui vamos caracterizar algumas que foram essenciais nos experimentos deste trabalho: *CLAHE*, transformação gama, equalização de histograma e mediana. Para conceituar essas técnicas vamos utilizar principalmente as contribuições de Gonzalez e Woods (2000).

A transformação gama é uma transformação de potência e apresenta a forma básica $s = cr^{\gamma}$, onde c e γ são constantes positivas, r é a imagem de entrada e s a de saída. Esse tipo de transformação pode auxiliar a representar cores com exatidão, já que altera as proporções de vermelho, verde e azul em imagens coloridas. Por isso ela é importante no contexto desta dissertação.

No contexto do processamento de imagens, mediana é um filtro de estatística de ordem não linear. A partir de um pixel central se desenha uma região, chamada de *kernel*, e é calculada a mediana das intensidades dos *pixels* dessa região, que substitui o valor da intensidade do pixel central. Esse filtro é amplamente usado na literatura para redução de ruídos e suavização. O formato do *kernel* pode ser determinada de diferentes formas. Nesse trabalho utilizamos as formas de quadrado (*square*) e disco (*disk*).

Gonzalez e Woods (2000) define o histograma de uma imagem como uma função $h(k)$, onde $k \in [0, L-1]$, e L é o número de intensidades ou cores possíveis na imagem. A equalização de histograma nesse contexto é uma técnica de mapeamento não linear que transforma intensidades de cada pixel na imagem de entrada, para novas intensidades na imagem de saída. É uma ferramenta importante para realçar faixas de intensidade.

CLAHE é uma sigla que em inglês significa "*Contrast-Limited Adaptive Histogram Equalization*". Esse tipo de equalização de histograma é chamada de adaptativa pois considera os *pixels* vizinhos para o mapeamento de novas intensidades. Além de adaptativa, essa transformação é limitada por contraste, ou seja, aplica um aprimoramento de contraste a fim de restringir a inclinação da função de mapeamento (PIZER *et al.*, 1987).

2.7.1 Sistema de cores

Nas circunstâncias desse projeto, a utilização das cores no processamento de imagens é especialmente importante, já que precisamos identificar a cor de pele nas fotos de faces das pessoas disponíveis no banco de dados. Um sistema de cores é uma maneira de especificar cores de forma padronizada através de coordenadas e níveis de intensidade (GONZALEZ; WOODS, 2000). Um dos sistemas mais utilizados na prática é o RGB, em que para cada pixel são determinados valores de intensidades de *red* (vermelho), *green* (verde) e *blue* (azul). (PREMA; MANIMEGALAI, 2012) estuda quais faixas de intensidade RGB estão relacionadas às cores de pele conhecidas. Para classificar automaticamente cor de pele, (KINYANJUI *et al.*, 2019) utiliza

outro sistema de cores, o LAB, onde L indica a luminosidade (fator importante no contexto de classificação automática de cor de pele) e A e B são coordenadas cromáticas, que indicam a presença de vermelho, verde, amarelo e azul (SZELISKI, 2010).

Nos processamentos de imagens realizados nesse projeto também utilizamos o sistema de cor HSV e imagens em escala de cinza (*grayscale*). O modelo HSV tenta se aproximar da forma como os seres humanos descrevem e interpretam cores. O canal H descreve uma cor pura (amarelo, laranja ou vermelho puros) (GONZALEZ; WOODS, 2000), o canal S indica o grau de diluição de uma cor pura na luz branca e o canal V indica a intensidade (ou nível de cinza) e é um fator importante para processamento de equalização de histogramas. O sistema HSV e os outros dois citados acima (RGB e LAB) possuem três canais, que juntos descrevem a cor de um *pixel*. Diferente deles, as imagens em escala de cinza possuem somente um canal, que como o canal V do sistema HSV, indica a intensidade de cinza em um *pixel*, onde o menor valor significa branco e o maior valor significa preto.

CAPÍTULO

3

REVISÃO DA LITERATURA

Neste Capítulo é apresentada uma síntese dos principais estudos selecionados sobre viés em grandes bases de dados e uma visão crítica sobre os possíveis problemas causados por parte desses vieses.

3.1 Sociedade, desenvolvimento tecnológico e injustiças

Existem hoje poucos artigos disponíveis que debatem o viés de seleção e ainda menos artigos que debatem esse viés no contexto da falta de representatividade étnico-racial e de gênero. Podemos enxergar esse fato como consequência do racismo e do machismo estrutural do sistema capitalista como é evidenciado por Almeida (2019) que situa o racismo como processo político de opressão e exploração e na dissertação de Oliveira *et al.* (2019) que revisita Heleieth Saffioti para mostrar que o problema do machismo que afeta as mulheres na sociedade de classes só será resolvido com a destruição do capitalismo. O desenvolvimento tecnológico não está descolado da realidade de injustiças sociais que vivemos.

O papel do pesquisador também deve ser o de repensar radicalmente como funciona a sociedade, entendendo como o desenvolvimento tecnológico está intrinsecamente conectado com a necessidade do lucro, como é exposto por Morozov e Marcondes (2018). A acumulação de capital e a busca pelo lucro determina as relações de trabalho e a reprodução da vida humana sob o sistema capitalista, inclusive integrando opressões como racismo, machismo, lgbtfobia e o capacitismo como opressões funcionais para a lucratividade (DAVIS, 2016). O modelo de sociedade e do desenvolvimento tecnológico que vivemos visa transformar cada aspecto da nossa vida, característica ou movimento que fazemos em dados. Os dados, por sua vez, são vistos como mercadoria por grandes empresas do ramo tecnológico. Isso significa que todas esferas da nossa vida estão sendo mercantilizadas. E como diz o autor da obra, "deixar o *Google*

organizar todas as informações do mundo faz tanto sentido quanto deixar a *Halliburton*[1] lidar com todo o petróleo do planeta" (MOROZOV; MARCONDES, 2018)(pag.29). O caminho do desenvolvimento tecnológico hoje aponta para o lucro e não para a resolução de desigualdades estruturais. Se levarmos essa lógica para os sistemas de reconhecimento facial, veremos que enquanto são desenvolvidas soluções que são vendidas, mesmo que perpetuem o racismo, poucos se preocuparão em corrigir a injustiça causada pelo viés presente em grandes bases de dados.

3.2 Bases de dados, viés e classificação justa

Dois exemplos de grandes bases de dados dados disponíveis publicamente na internet, que são resultados de pesquisas científicas e amplamente utilizadas para o desenvolvimento de soluções de reconhecimento facial são a *MS-Celeb-1M* (GUO *et al.*, 2016) e a *MegaFace* (NECH; KEMELMACHER-SHLIZERMAN, 2017).

Recentemente, pesquisas investigaram a presença e o impacto do “ruído” nessas grandes bases de dados revelando problemas com redundância de rótulo. Wang *et al.* (2018) expõe o impacto desse problema no reconhecimento, que gera uma classificação enviesada. A partir disso, os autores propõe um procedimento de limpeza do conjunto de dados, concluindo que 32% do *MegaFace* e 20% do *MS-Celeb-1M* são suficientes para obter resultados comparáveis em desempenho a modelos treinados em seus conjuntos de dados completos.

O trabalho de Jiang e Nachum (2019) propõe uma correção de viés de rótulo de classificadores treinados com aprendizado de máquina através de um algoritmo. Mostra, com garantias teóricas, que um classificador treinado com os dados reavaliados pelo algoritmo proposto corresponde a um treinado com um base de dados limpa, livre de viés de rótulo. Este trabalho também cita que há 3 principais momentos onde intervir para deixar a classificação mais justa: No pré-processamento, durante o processamento com a ressignificação dos pesos (proposta do artigo) e no pós-processamento.

Se preocupar com a disponibilidade de exemplos suficientes considerando diversos atributos relacionados a representatividade é uma questão importante neste contexto. O viés relacionado à representatividade da população durante a coleta de dados é raramente abordado na literatura. No caso do reconhecimento facial, os vieses de idade, gênero e étnico-raciais são particularmente relevantes. A aplicação de sistemas que usam aprendizado de máquina se tornam um grande problema quando classificadores treinados a partir dessas bases de dados são usados para criar modelos direcionados a um conjunto de indivíduos mal representados na base.

Não é novidade o fato que no desenvolvimento histórico de tecnologias de ponta, grupos e classes específicas são prejudicadas em detrimento de outros. Em Mehrabi *et al.* (2019) podemos observar um bom levantamento do impacto concreto de casos em que a Inteligência Artificial e o Aprendizado de Máquina refletem as injustiças do sistema em que vivemos. Exemplo claro disso

[1] Empresa multinacional americana de serviços à indústria de exploração e produção de petróleo

é o *Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)*, software que mede o risco de uma pessoa ser reincidente criminal, que é usado por juízes nos EUA para apoiar a decisão de julgamentos. Foi feita uma investigação nesse software que constatou que ele aponta um risco maior de reincidência para pessoas negras do que para pessoas brancas [2]. Ainda nesse artigo podemos ver exemplos trazidos pelos autores de viés de seleção em algoritmos de *word embedding* que tornam classificadores machistas e mais casos de viés relacionados a cor da pele, como na identificação de câncer de pele, já que as bases de dados usadas nesse contexto tem poucos exemplos de peles escuras.

Em Merler *et al.* (2019) os autores fazem a análise de uma série de base de dados públicas mostrando como grande parte delas não é representativa, com destaque para as distribuições de idade e de cor da pele. Ainda no artigo, é colocada a importância de considerar outros aspectos como medidas de distância, área e razão craniofaciais para serem consideradas na avaliação da representatividade de uma base de dados. A partir de análises estatísticas, os autores concluem que as medidas craniofaciais apresentam uma maior variabilidade e podem capturar informações que anotações de gênero e idade por si só não conseguem no contexto de sistemas de classificação. Em Buolamwini e Gebru (2018) a avaliação vai ainda além e por meio de uma análise interseccional de grandes bases de dados públicas/governamentais e classificadores comerciais, considerando gênero e cor da pele, as autoras deixam explícito como as mulheres negras são o grupo mais prejudicado na questão da injustiça de classificação e má representação nas bases.

O estudo Merler *et al.* (2019) estabelece as bases para a análise estatística automática de grandes bases de dados no sentido de medir atributos relacionados a representatividade. O avanço nesse tipo de análise será essencial para o desenvolvimento de aplicações com classificação justa. Além disso, o artigo aponta para a geração de imagens artificiais, que tem grande potencial para resolver o problema de falta de representatividade em grandes bases de dados e consequentemente da classificação injusta. O estudo e desenvolvimento de GANs[3] nesse sentido também é um trabalho crucial a ser feito.

Existe uma dificuldade notável para a obtenção de bases de dados para reconhecimento facial para pesquisa científica. A dificuldade é ainda maior quando buscamos bases de dados que são anotadas por cor da pele ou atributos étnico-raciais. Muitas bases não são bem estruturadas, não apresentando subdivisões que agrupam pessoas únicas. O trabalho Merler *et al.* (2019) também apresenta uma tabela onde podemos observar que são raras as bases de dados que trazem anotações por cor de pele. As bases de dados propostas nos dois últimos artigos aqui citados (MERLER *et al.*, 2019) e (BUOLAMWINI; GEBRU, 2018), que buscam ser bases representativas trazendo uma distribuição mais igualitária no que se refere a gênero e cor de pele, não estão mais disponíveis para *download*. No *Survey* recente de Mehrabi *et al.* (2021) as bases

[2] <www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
[3] Redes neurais adversárias generativas

de dados citadas acima são apontadas como importantes para o estudo de classificação justa no atributo de imagens faciais. O artigo traz também um interessante mapa de calor mostrando a distribuição de trabalhos anteriores em *fairness*, agrupados por domínio e definição de *fairness*.

Para analisar o atributo da cor da pele, grande parte dos artigos na literatura usam o sistema *Fitzpatrick Skin Type classification system* (FITZPATRICK, 1975) ou o *Individual Typology Angle (ITA)* (CHARDON; CRETOIS; HOURSEAU, 1991). Isso é particularmente importante devido à questão de que grande parte das bases de dados disponíveis não trazem anotações no que se refere a raça, etnia e cor da pele. A partir disso, estes sistemas são de grande ajuda para a análise de representatividade em grandes bases de dados.

3.3 Processamento de imagens, viés nos modelos e classificação de cor de pele

Outro assunto importante no quadro deste trabalho é a detecção automática de pele. Apesar de ter sido publicado há uma década, o artigo Prema e Manimegalai (2012) traz um levantamento dos métodos e técnicas mais utilizados para a tarefa de detecção automática de pele e seus resultados de avaliação numérica. Assim como em Harville *et al.* (2005), é apontado que a iluminação na captura da imagem interfere muito nessa tarefa e que por esse motivo, é importante o desenvolvimento de técnicas de correção através do processamento de imagens. Em ambos artigos a maioria das abordagens levantadas são baseadas em pixel: o algoritmo percorre cada pixel e faz uma classificação binária entre pele e não-pele. Existem abordagens alternativas paramétricas e não paramétricas, como o modelo de histograma e modelos estatísticos. Em Harville *et al.* (2005), é descrito um processo para detecção que leva em conta um pré processamento de correção de cor e detecção facial para que sejam considerados apenas os *pixels* da face. Esse processo é especialmente importante no contexto desta dissertação. Entre os melhores métodos citados nos dois artigos, estão os que usam o sistema de cores RGB para a tarefa de detecção automática de cor de pele. Existem artigos mais recentes que inclusive utilizam métodos complexos de aprendizado de máquina para a realização da tarefa de detecção de pele (SALAH; OTHMANI; KHERALLAH, 2021), (HE *et al.*, 2019).

Ainda no contexto de processamento de imagens, Tian e Cohen (2018) propõe um método de aprimoramento que é uma compensação entre o contraste global e o contraste local com correção de cor. Esse método não é pensado especificamente pra face, mas considera o problema de reconhecimento facial no contexto da iluminação. Consideramos uma forma de processamento eficaz, mas como o foco do trabalho não é desenvolver um método de contrate e esse método não tinha implementação optamos por usar o *CLAHE* neste projeto de mestrado.

O estudo Amini *et al.* (2019) desenvolve um algoritmo para mitigar vieses nos dados de trenamento usando um *autoencoder* variacional, abordando a questão do viés étnico-racial e de gênero em sistemas de detecção facial. Para avaliar o algoritmo, os autores usam a base

de dados *Pilot Parliaments Benchmark (PPB)*, proposta em Buolamwini e Gebru (2018). Dado um conjunto de dados de treinamento enviesado, os modelos propostos diminuem o viés de gênero e étnico-racial em comparação com classificadores padrão, além de aumentar a precisão da detecção.

Quando a comunidade científica começou a analisar o viés de seleção em situações de aprendizado de máquina, notadamente àqueles ligados à atributos étnicos-raciais, acreditava-se que as causas estavam apenas ligadas às bases de dados que alimentam os treinamentos dos algoritmos. Recentemente, muitos pesquisadores têm apontado que os modelos e algoritmos também causam esse viés e tem impacto no problema da classificação justa. Por exemplo, como é evidenciado em Hooker *et al.* (2020) e Hooker *et al.* (2019), a poda e a quantificação de redes neurais profundas podem amplificar o viés. Ainda segundo Agarwal, D'souza e Hooker (2020), o trabalho de memorização e variação de gradientes (*VoG*) mostra que exemplos difíceis são aprendidos posteriormente no treinamento, que as taxas de aprendizagem afetam o que é aprendido e que portanto, a interrupção precoce tem um impacto desproporcional em certos exemplos. Hooker (2021) aponta que uma das razões pelas quais o modelo é importante é porque as noções de justiça geralmente coincidem com a forma como os recursos sub-representados são tratados. Portanto, é incorreto afirmar que o modelo é independente de considerações de viés algorítmico. Nossas escolhas em torno da arquitetura do modelo, hiper-parâmetros e funções objetivas informam considerações de viés algorítmico.

Ruback, Avila e Cantero (2021) faz um estudo sobre vários tipos de vieses que podem se manifestar em sistemas de reconhecimento facial desde o momento do pré processamento, passando pela coleta dos dados, até o pós-processamento. No artigo é evidenciado através de exemplos como o reconhecimento facial tem reforçado problemas que já existem na sociedade como o racismo estrutural, confrontando a visão de sistemas de aprendizado de máquina como "tecnologias neutras". O trabalho dá bases para a reflexão sobre o que deve significar classificação justa na prática política da aplicação de algoritmos de reconhecimento facial.

Recentemente, pesquisadores brasileiros têm estudado a presença de viés em modelos que utilizam grandes bases de dados e aprendizado de máquina para resolver problemas que envolvem identificação de lesões de pele e que portanto levam em conta a questão da detecção de pele em imagens. Bissoto, Valle e Avila (2020) realiza uma análise de técnicas de última geração nesse contexto e propõe conjuntos de dados para testar modelos quanto à presença de viés. O estudo feito em 2020 aponta que os métodos atuais de correção de viés não estão prontos para resolver o problema em modelos de lesão de pele. Os pesquisadores dizem que trabalhos futuros devem considerar imagens mais diversas de diferentes origens e com rótulos diferentes para que soluções mais robustas e confiáveis sejam construídas.

Em Yu *et al.* (2018) é proposta uma variante da *triplet loss* (fução de custo da rede largamente usada pra reconhecimento facial) que corrige adaptativamente o viés de seleção presente na função de custo original. O *triplet loss* adaptado é testado para recuperação de

imagens baseadas em conteúdo e apresenta valores de revocação média maiores que os da função de custo original.

3.4 Considerações finais

Existem outros trabalhos cujo objetivo é mitigar viés de seleção em sistemas de aprendizado de máquina, como Heckman *et al.* (1998) que busca caracterizar o viés de seleção usando dados experimentais, Zadrozny (2004) que avalia classificadores sob viés de seleção, Huang *et al.* (2007) que visa a correção do viés de seleção de amostra por dados não rotulados, Wang *et al.* (2019) que discute os ruídos no reconhecimento facial profundo e Maughan e Near (2020) que busca uma medida de justiça individual para aprendizagem profunda.

O objetivo desta dissertação é propor e avaliar um método para detecção de tons de pele que permita auditar bases de dados de forma a minimizar problemas com viés de seleção em modelos de reconhecimento facial. Na revisão da literatura é possível observar que ainda há espaço para trabalhos nessa linha e nesse projeto escolhemos tratar esse objetivo geral e os objetivos específicos citados na Seção 1.2.

CAPÍTULO

4

PROPOSTA

É proposto um método de classificação automática de cor de pele que pretende auxiliar pesquisadores em trabalhos que tem o objetivo de auditar bases de dados de forma a minimizar problemas com viés de seleção em modelos de reconhecimento facial. Além disso, é proposto um método para analisar conjuntos de dados de imagens considerando a representatividade de suas instâncias, que podem ser usadas para avaliar o viés da seleção. A proposta também inclui um estudo de estratégias para adaptação de vieses via métodos estatísticos, de aprendizagem ou de processamento de imagens, em direção a modelos que não discriminam pessoas por traços étnico-raciais. Este trabalho também pretende analisar criticamente o desenvolvimento científico na área do reconhecimento facial e fornecer bases sólidas para o desenvolvimento de aplicações mais justas.

4.1 Etapas principais

(a) Obter modelos pré-treinados de redes neurais profundas que usam os grandes conjuntos de dados para reconhecimento disponíveis na Web para criar modelos;

(b) Realizar a análise do espaço de atributos computando as representações vetoriais de imagens de faces usando características extraídas a partir dos modelos obtidos em (a), agrupando-as e comparando com dados rotulados;

(c) Implementação de um método base de classificação automática de cor de pele e testes a partir desse método considerando as imagens originais e processadas com transformações gama, *CLAHE* e equalização de histograma;

(d) Classificação automática de cor de pele a partir de uma ponderação tripla considerando os melhores resultados obtidos em (c);

(e) Modificação do método base de classificação automática de cor de pele para um método de classificação por pixel e novos testes a partir desse método, aplicando combinações de filtros

de mediana e moda nas imagens;

(f) Classificação automática de cor de pele aplicando os melhores métodos obtidos em (c), (d) e (e) para cada faixa de cor de pele baseados em F-Score conforme será explicado a seguir na Seção 4.4;

(g) Análise da revocação média com base na metodologia relacionada à recuperação de imagens baseada em conteúdo para comparando métodos obtidos nas etapas anteriores.

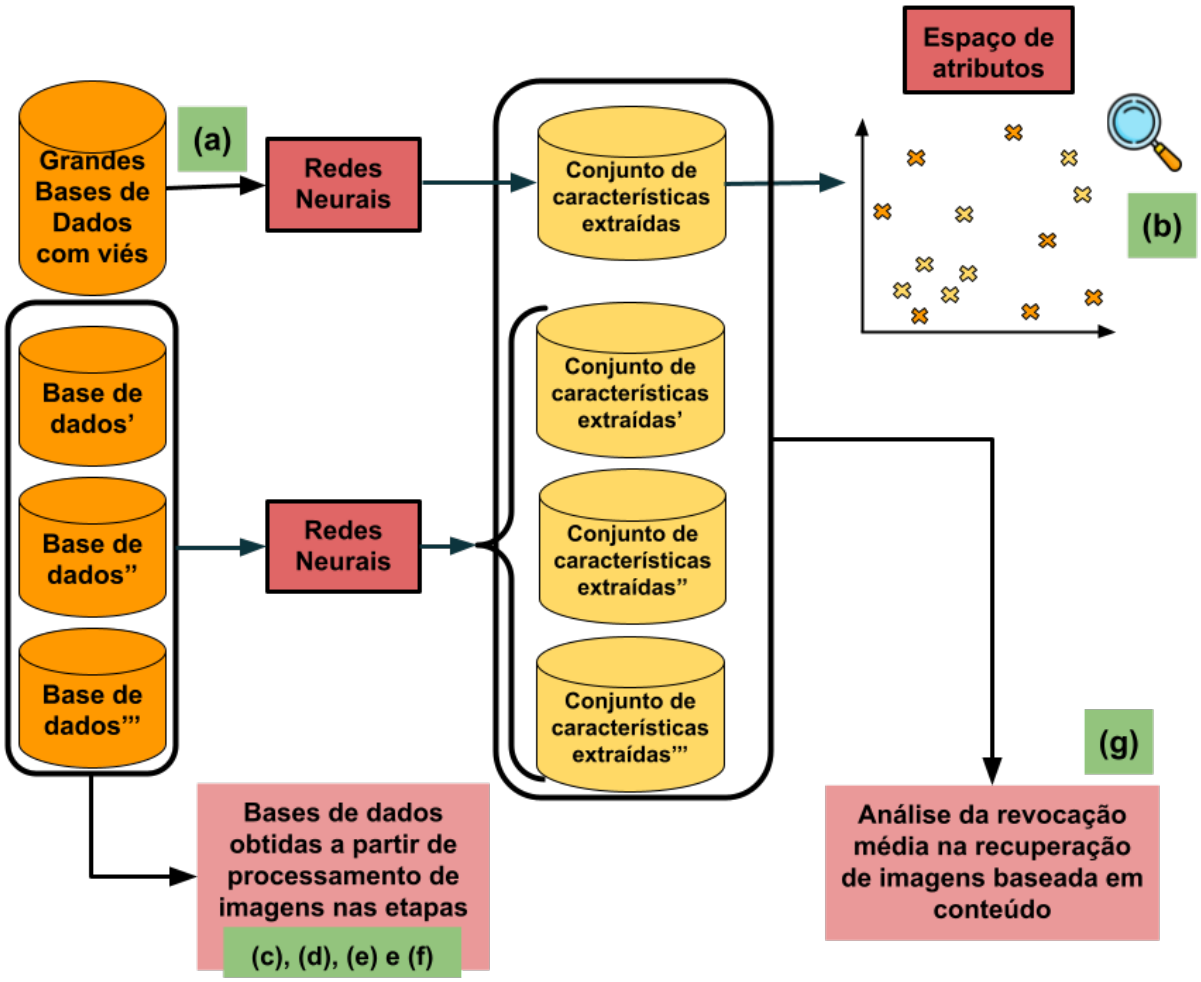

Figura 3 – Proposta geral: (a) Obter modelos pré-treinados de redes neurais profundas, (b) Realizar a análise do espaço de atributos, (c, d, e, f) Processamento de imagens e (g) Análise da revocação média

4.2 Detalhamento de cada etapa

4.2.1 Rotulação de destacamento da LFW

Como apresentado na revisão da literatura, existe uma dificuldade na obtenção de grandes bases de dados estruturadas para reconhecimento facial. Nesse contexto, a base escolhida foi a

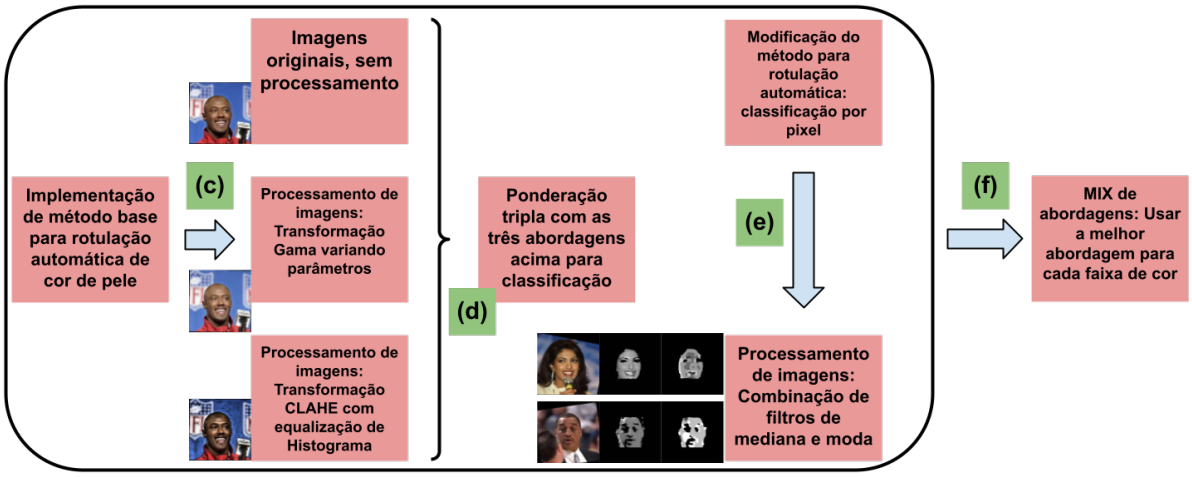

Figura 4 – Detalhamento da proposta geral: (c) Método base de classificação automática de cor de pele, (d) Ponderação tripla, (e) Método de classificação por pixel e (f) Aplicação de melhor abordagem para cada faixa de cor

Labeled Faces in the Wild (LFW). Como a base não é anotada por cor da pele e é esse atributo que queremos analisar, foi feita uma anotação manual de uma parte deste conjunto.

Para analisar o viés de seleção étnico-racial supostamente gerado através do treino de um extrator de características a partir de uma base que apresenta falta de representatividade considerando pessoas de pele mais escura, foram rotulados manualmente 150 exemplos (pessoas únicas) da base de teste (LFW).

Para rotular a cor da pele foi usada a escala de *Fitzpatrick*. Além disso, a intenção foi selecionar pessoas únicas que tenham entre 2 e 5 imagens de face na base, a fim de tornar a análise mais robusta. Porém, para garantir uma distribuição uniforme entre pessoas únicas do sexo masculino e feminino e entre grupos da escala *Fitzpatrick* não foi possível garantir que todas pessoas únicas selecionadas para rotulação manual tivessem mais que uma imagem de face disponível.

Então, para garantir uma análise justa à medida que não existiram recursos humanos disponíveis para rotular manualmente mais que 150 exemplos, estes foram divididos em 3 partições considerando a escala *Fitzpatrick*:

- **Partição A** - 50 pessoas com cor de pele níveis 1 ou 2
- **Partição B** - 50 pessoas com cor de pele níveis 3 ou 4
- **Partição C** - 50 pessoas com cor de pele níveis 5 ou 6

A escolha sobre quais elementos rotular dentre as 5749 pessoas únicas da base se deu pela quantidade de imagens por pessoa. Foram escolhidas inicialmente pessoas únicas com 5 imagens de face cada. No processo de rotulação, foi percebido que esse critério não é suficiente para garantir a distribuição das partições citadas acima. Então, foram rotuladas também pessoas

únicas com 4 ou menos imagens de face. Os exemplos também foram divididos entre 75 pessoas do sexo masculino e 75 do feminino, o que pode auxiliar em análises futuras.

4.2.2 Extração de características e analise de espaço de atributos

Para extrair as a representação vetorial de características de cada pessoa única, foram utilizados pesos pré treinados da *facenet* usando o modelo de CNN *ResNet50* (HE *et al.*, 2016) como *backbone*. Para pessoas únicas que possuem mais de uma imagem, foi extraído o *embedding* de cada imagem, o qual é a saída da penúltima camada da rede neural pré-treinada, e feita uma média aritmética. Pela arquitetura da rede, esse *embeddings* possui 2048 dimensões. Este processo foi feito duas vezes, primeiro o espaço de exemplos como todas as pessoas únicas da base e outra considerando apenas os exemplos manualmente rotulados por gênero e cor de pele. O propósito dessa etapa foi verificar como os embeddings das faces de diferentes pessoas, levando em consideração seus tons de pele conforme a rotulação descrita na Seção anterior.

Os *embeddings* foram agrupados de 2 maneiras diferentes para analise do espaço de atributos. O *KMeans* foi aplicado para 6 e 3 grupos. Esses números de grupos foram escolhidos considerando que existem 6 níveis de cor de pele na escala *Fitzpatrick* e que dividimos nossos exemplos rotulados manualmente em 3 partições. Este processo de agrupamento também foi feito duas vezes: uma considerando a base completa e outra só os exemplos rotulados.

Por meio da aplicação do algoritmo *principal component analysis* (PCA) para redução de dimensionalidade dos *embeddings* de 2048 para 2 dimensões, foi possível plotar o espaço de características. Cada grupo derivado dos algoritmos de agrupamento foi representado por uma cor nos gráficos. Foram gerados gráficos para cada tipo de agrupamento considerando 3 contextos: a base de dados inteira, apenas os exemplos rotulados e a base inteira mais uma vez, porém só mostrando no gráfico os exemplos rotulados. Essa é a análise do espaço de atributos através dos agrupamentos.

Em seguida, foi feita a comparação dos resultados dos agrupamentos com dados rotulados. Nessa etapa analisamos em que agrupamentos foram envolvidos os exemplos rotulados a fim de entender se os *embeddings* gerados pela CNN proposta consideram atributos como a cor da pele e, portanto, se pessoas únicas de mesma cor de pele são agrupadas em um mesmo grupo específico. Em resumo, esse experimento teve o objetivo de investigar se os algoritmos de agrupamento consideraram a característica da cor de pele como central para agrupar os exemplos.

4.2.3 Implementação de método base para classificação automática de tons de pele

Considerando que a rotulação manual é mais precisa do que automática, o objetivo é encontrar o método de rotulação automática que mais se aproxima da rotulação manual. Os 150

exemplos rotulados manualmente foram usado para comparação entre métodos. Encontrado o melhor método, ele será aplicado para toda base.

Um algoritmo de detecção de face foi executado em todas as imagens da base dados completa usando a biblioteca *openCV* e o método de *Viola & Jones* (VIOLA; JONES, 2001) antes de dar início às etapas de processamento. As *bounding boxes* de cada face foram salvas e serviram como pré-processamento dos algoritmos de classificação automática de cor de pele. É um importante pré-processamento pois na base de dados existem imagens onde há mais de uma face (outra(s) pessoa(s) no fundo da imagem), mas só nos interessa a cor de pele da maior face de cada imagem.

Na etapa (c) do processamento, cada imagem única de cada pessoa única foi percorrida de três maneiras distintas:

- aplicando transformações gama (variando o parâmetro γ da fórmula explicita na Seção 2.7 entre 0.5 e 1.5), ou
- aplicando transformações *CLAHE* e de equalização de histograma no canal V da imagem convertida para *HSV* (são somadas as duas transformações e o resultado é dividido por dois), ou
- não fazendo nenhum tipo de processamento de imagem.

Após a etapa citada acima, foi percorrido cada *pixel* da região da face previamente detectada de cada imagem única verificando quais deles são pele de acordo com Prema e Manimegalai (2012) e colecionando estes em uma lista. Com essa lista completa, foi gerada uma nova imagem apenas com estes *pixels* e convertida para *LAB*. Assim são encontrados os valores necessários para extrair um valor de ITA utilizando o cálculo formulado no trabalho de Kinyanjui *et al.* (2019). O valor encontrado é normalizado a partir de valores experimentais de mínimo e máximo obtidos nos processo da qualificação deste trabalho, é discretizado em faixas de cor de pele entre 1 e 6 baseadas na escala *FitzPatrick* e salvo como rótulo para cada imagem única. A partir desse momento, são definidas as 6 faixas de tons de pele como ITA1, ITA2, ITA3, ITA4, ITA5 e ITA6.

Ainda nesta etapa foram selecionados dois valores de gama nas transformações gama (diferentes de "1" pois uma transformação com $\gamma = 1$ é a identidade) que são levados à frente nas próximas etapas: o que apresenta o menor RMSE (avaliação descrita na Seção 4.4) e o valor $\gamma = 1/1.22$ que é padrão da curva de transformações gama (PONTI; NAZARÉ; THUMÉ, 2016; KANAN; COTTRELL, 2012).

4.2.4 Ponderação tripla das abordagens obtidas em (c)

Na etapa (d) os rótulos gerados na etapa (c) são observados e é feita uma ponderação entre eles a fim de melhorar nossa abordagem de rotulação automática. Todo subconjunto rotulado da

LFW é percorrido e para cada imagem única é feita a seguinte comparação: Se o rótulo gerado pela transformação gama é igual o rótulo gerado pela transformação *CLAHE* com equalização de histograma (ou seja, a imagem única é classificada na mesma faixa de cor), a ponderação tripla copia esse rótulo. Caso contrário, o rótulo da ponderação tripla será o mesmo do classificado a partir das imagens únicas sem nenhum tipo de processamento. Esse procedimento de ponderação tripla foi realizado duas vezes, variando o valor de gama da transformação de acordo com os melhores resultados obtidos em (c).

4.2.5 Modificação do método base de classificação automática de cor de pele para um método de classificação por pixel

Na etapa (e), o método de classificação automática de cor de pele descrito em (c) foi modificado: O primeiro passo agora é criar uma cópia de cada imagem única e convertê-la para LAB a fim de encontrar os valores necessários para fazer o cálculo do valor de ITA (KINYANJUI *et al.*, 2019). Assim como no método antigo, cada *pixel* da região da face previamente detectada de cada imagem única é percorrido para verificação de quais deles são pele, mas, diferente do método (c), estes não são colecionados em uma lista; é extraído um valor de ITA para cada *pixel*. Com esses valores, é criada uma nova imagem *grayscale* onde é atribuído aos *pixels* da região da face que são pele o respectivo valor de ITA extraído a partir do cálculo com conteúdo prévio dele. Para os *pixels* que não estão na região da face e que não pele é atribuído o valor 0. O resultado desse primeiro processo gera uma máscara que pode ser vista na coluna 2 da Figura 5. Em seguida a máscara é normalizada a partir de valores experimentais de mínimo e máximo de ITA obtidos nos processo da qualificação deste trabalho (coluna 3 da Figura 5). Dando continuidade ao processo de classificação automática de cor de pele da imagem, é aplicado um filtro de mediana somente na região da face com *footprint* quadrado 3×3 (coluna 4 da Figura 5). Então, o valor de cada pixel é discretizado em faixas de cor de pele entre 1 e 6 (coluna 5 da Figura 5). Nesse momento é investigada a aplicação de uma segunda mediana (coluna 6 da Figura 5) variando o tamanho do disco usado no *footprint* entre 3 e 13 e por fim, extraímos a moda entre os valores da matriz de *pixels*, que é salva como rótulo para cada imagem única na classificação automática de cor de pele.

Ainda nessa etapa, para verificar o comportamento das métricas de avaliação escolhidas (explícitas em 4.4) em um caso extremo, foi executada uma abordagem ”absurda” de classificação onde rotulamos todas as imagens com o valor de ITA 3.

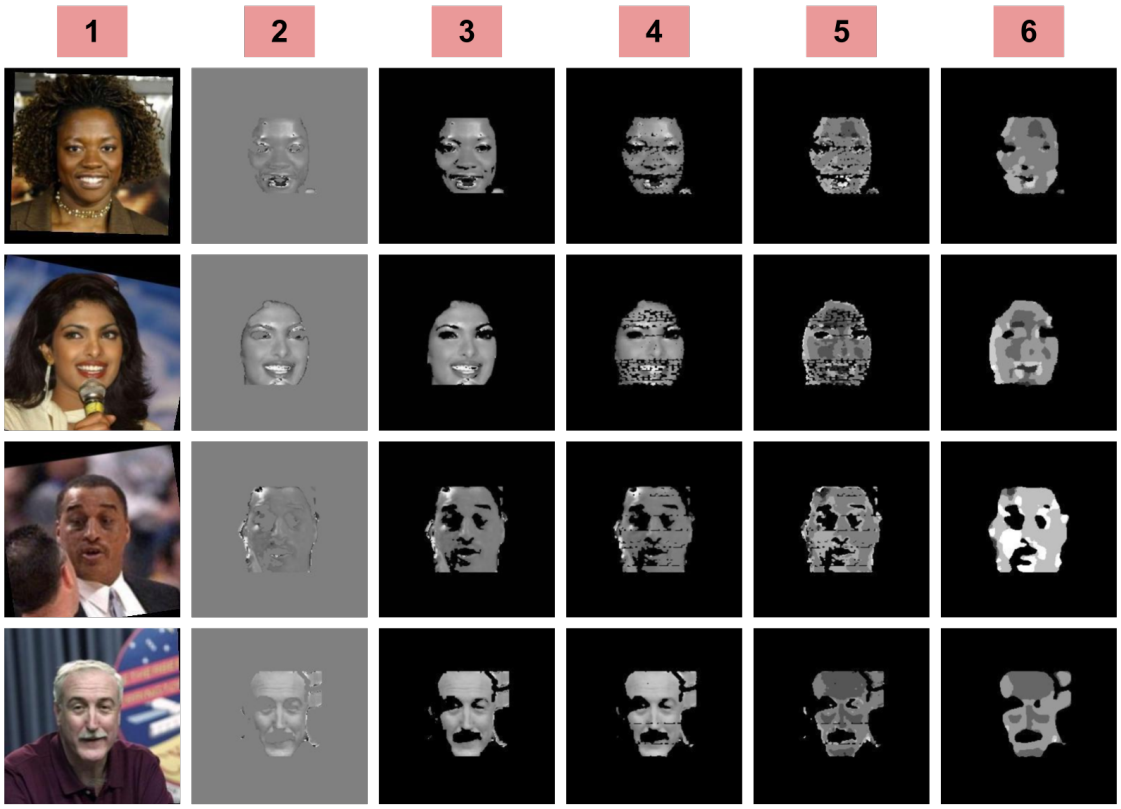

Figura 5 – Etapa (e) - método proposto de processamento de imagens: 1) Imagem original; 2) Máscara por pixel; 3) Normalização; 4) Aplicação da primeira mediana; 5) Discretização ITA e 6) Aplicação da segunda mediana.

4.2.6 Classificação automática de cor de pele aplicando os melhores métodos obtidos em (c), (d) e (e) para cada faixa de cor de pele baseados em F-Score

Durante as etapas (c), (d) e (e) foram utilizados diferentes métodos para classificação automática de cor de pele e em cada método foi possível testar mais de uma abordagem, variando parâmetros. São colecionadas 12 abordagens, que foram avaliadas segundo as métricas expostas na Seção 4.4:

- #01) Método A: Abordagem sem nenhum tipo de processamento de imagem
- #02) Método A: Abordagem aplicando transformação gama com $gama = 1.2$
- #03) Método A: Abordagem aplicando transformação gama com $gama = 1/1.22$
- #04) Método A: Abordagem aplicando transformação *CLAHE* e de equalização de histograma
- #05) Método B: Ponderação tripla entre as abordagens 1), 2) e 4)
- #06) Método B: Ponderação tripla entre as abordagens 1), 3) e 4)

- #07) Método C: Abordagem aplicando uma transformação mediana (*footprint*=quadrado 3x3)

- #08) Método C: Abordagem aplicando duas transformações mediana (na primeira com *footprint*=quadrado 3x3 e na segunda com *footprint* = disco 3x3

- #09) Método C: Abordagem aplicando duas transformações mediana (na primeira com *footprint*=quadrado 3x3 e na segunda com *footprint* = disco 5x5

- #10) Método C: Abordagem aplicando duas transformações mediana (na primeira com *footprint*=quadrado 3x3 e na segunda com *footprint* = disco 9x9

- #11) Método C: Abordagem aplicando duas transformações mediana (na primeira com *footprint*=quadrado 3x3 e na segunda com *footprint* = disco 13x13

- #12) Abordagem "absurda" onde classificamos todas as imagens mecanicamente com o valor "3" para faixa de cor de pele

Acima "Método A" se refere à aplicação do método descrito no detalhamento da etapa (c) de processamento, "Método B" à (d) e "Método C" à (e).

Na etapa (e), foi formulado um novo método que classifica automaticamente a cor de pele de uma imagem única aplicando as melhores (de acordo com o F-Score) abordagens obtidas anteriormente com os métodos descritos acima para cada faixa de cor de pele. Para aplicação desse método, é preciso definir prioridades de faixa de cor de pele para classificação. Ou seja, definida uma prioridade entre $x, y, z, w, v, u \in [1, 2, 3, 5, 5, 6]$, o método busca classificar primeiro a imagem com a abordagem (entre as 12 que já temos) que possui o melhor F-Score para a faixa de cor mais prioritária. Se a classificação resultar no valor desta faixa de cor, o novo método faz a mesma classificação. Caso contrário, utilizamos a abordagem que possui o melhor F-Score para segunda faixa de cor no nível de prioridade e comparamos o resultado da classificação com esta faixa de cor de pele. Enquanto a classificação não obtém o valor da faixa de cor de pele alvo, avançamos na lista de prioridades. Se no último nível de prioridade a classificação resultar em um valor diferente da faixa de cor que esse nível representa, usamos como última abordagem de classificação a #05, que é melhor no F-Score médio para peles mais escuras (faixas 4, 5 e 6). Assim, são colecionadas mais duas abordagens variando a prioridade de faixa de cor de pele para classificação:

- #13) Método D: Abordagem seguindo a seguinte prioridade de faixa de cor de pele para classificação: 6, 5, 4, 3, 1, 2

- #14) Método D: Abordagem seguindo a seguinte prioridade de faixa de cor de pele para classificação: 6, 1, 5, 2, 4, 3

A escolha da ordem de prioridade expressa em #13 é da cor de pele mais escura para a mais clara. Nesse caso o 1 só vem antes do 2 porque a mesma abordagem é melhor para a classificação das faixas de cor de pele 3 e 1 e com isso foi feita uma simplificação no algoritmo. A escolha da ordem de prioridade expressa em #14 é da faixa de cor de pele menos representada em número de exemplos rotulados manualmente para a mais representada.

4.2.7 Análise estatística da revocação média

Com as 14 abordagens colecionadas a partir de diferentes métodos, é escolhida a abordagem que apresenta melhores valores de F-Score para ser aplicada na base de dados completa. Outras abordagens também são escolhidas para efeito de comparação. Por fim, é analisado em quais faixas de cor de pele as *embeddings* geradas para cada imagem única de pessoas únicas mais se aproximam.

Então, é feita uma análise estatística da revocação (*recall*) considerando faces únicas: primeiro é calculado o *embedding* para cada imagem de face única (ao invés de cada pessoa única, como anteriormente). Utilizando metodologia relacionada à recuperação de imagens baseada em conteúdo, esse experimento considera o cenário de identificação de um indivíduo por meio da recuperação de imagens mais próximas em uma base de dados. Com isso, para cada imagem, é buscado quais são as 10, 5 e a imagem mais próxima desta (considerando pessoas rotuladas e com mais de uma imagem de face). Então, é calculada a revocação média para cada grupo e subgrupo levando em conta a anotação por cor de pele. Para cada imagem de face única de pessoas rotuladas, é calculada quais são as imagens mais próximas entre os novos *embeddings* gerados (considerando a base de dados completa para extração de características). A seguir, é obtido o número de faces alvo encontradas entre as recuperadas e é computada a revocação em três níveis: *Top-10*, *Top-5* e *Top-1*. Para o cálculo de distância, é usada a distância Euclidiana.

4.2.8 Aplicação do método proposto de rotulação automática de tons de pele na base Fitzpatrick17k

Como o objetivo deste projeto é propor e avaliar um método para detecção de tons de pele que permita auditar bases de dados de forma a minimizar problemas com viés de seleção em modelos de reconhecimento facial e a hipótese é de que é possível obter esse método de maneira que os resultados sejam similares para diferentes tons de pele, é importante que seja possível aplicar o método proposto em diferentes base de dados.

Então, com o intuito de testar se o método proposto é generalizável, ele foi testado também na base Fitzpatrick17k. Nesse teste foram aplicadas as etapas principais 4.1 (b), (c), (d), (e) e (f) e os resultados que estão expostos no Capítulo 5 foram analisados. Diferente da LFW, a Fitzpatrick17k tem anotações por cor de pele para grande parte das imagens. Assim, foi possível aplicar o experimento em 16525 imagens, um volume mais de 100 vezes maior do que

o aplicado na LFW.

4.2.9 Uso de aprendizado profundo na classificação de tons de pele

Em uma etapa adicional, também foi investigada a rotulação automática de cor de pele através da aplicação de técnica de aprendizado de máquina. Para isso, foi utilizada uma rede neural artificial para classificar automaticamente entre as 6 classes *Fitzpatrick* o destacamento de 150 pessoas da LFW que foram também manualmente rotuladas. A rede utilizada foi uma *ResNet50* com pesos pré treinados da *imagenet*. Para utilizá-la nesse contexto, removemos sua última camada e adicionamos nova camada de saída com 6 classes equivalentes às faixas de cor da escala *Fitzpatrick* e realizamos transferência de aprendizado por 15 épocas na base de dados *Fitzpatrick17k* a qual é descrita na Seção 4.5 e possui rótulos para diferentes tons de pele. Essa é a única base de dados rotulada com escala suficiente para permitir o treinamento de tal modelo. Com isso, é feito um classificador baseado em redes profundas, e adicionamos mais uma abordagem para a coleção de abordagens aplicadas na LFW, que foi identificada como #15.

4.3 Tecnologias utilizadas

Para a implementação dos modelos de redes neurais e computação de representações vetoriais das imagens de faces, são utilizadas as bibliotecas de aprendizado de máquina *Keras 2.1.5* [1] e *TensorFlow 1.5.0* [2], além de outras bibliotecas auxiliares como *Pandas* e *Numpy*, todas disponíveis na linguagem de programação *Python 3.4.3*.

Para realizar a análise do espaço de atributos por meio de algoritmos de agrupamento e plotagem de gráficos, são utilizadas as bibliotecas *Sklearn* e *Matplotlib*.

Para realizar a detecção facial é utilizada a biblioteca *OpenCV 4.4.0.42* [3]. Já para as tarefas de processamento de imagem são usados métodos da biblioteca *Skimage 0.15.0* [4].

4.4 Avaliação

Para avaliação do viés étnico-racial é utilizada a escala de *Fitzpatrick* (FITZPATRICK, 1975). É uma escala numérica de cor de pele humana, inicialmente pensada para medir a dose correta de radiação ultravioleta para dermatologia. A divisão da escala é definida em 6 tipos, como explicito na Tabela 1 e na Figura 6.

Para as avaliações de representatividade, por meio da comparação de resultados de agrupamentos e análises estatísticas são usadas as métricas de precisão e revocação. Estas

[1] Documentação disponível em <https://faroit.com/keras-docs/2.1.5/>
[2] Documentação disponível em <https://github.com/tensorflow/docs/tree/r1.5/site/en/api_docs>
[3] Documentação disponível em <https://pypi.org/project/opencv-python/4.4.0.42/>
[4] Documentação disponível em <https://scikit-image.org/docs/0.15.x/>

Fitzpatrick	Cor da pele
Tipo I	Branca-pálida
Tipo II	Branca
Tipo III	Morena-clara
Tipo IV	Morena-moderada
Tipo V	Morena-escura
Tipo VI	Negra

Tabela 1 – Seis categorias da escala *Fitzpatrick*.

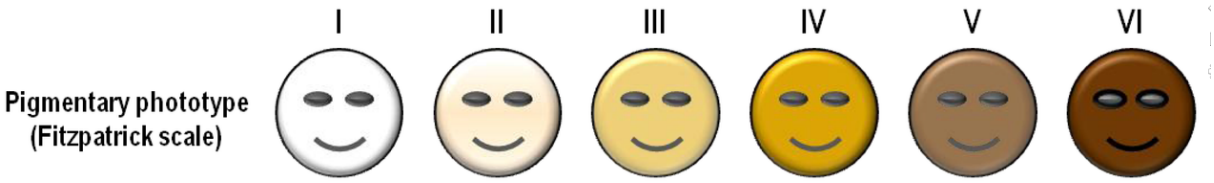

Figura 6 – Representação de fototipo pigmentar para Escala *Fitzpatrick* (D'ORAZIO *et al.*, 2013)

medidas permitem a verificação do viés de seleção na classificação gerada a partir de redes neurais treinadas em grandes bases de dados.

Para avaliar como cada abordagem de classificação automática de cor de pele mais se aproxima da classificação manual foram usadas as seguintes métricas aplicadas ao subconjunto das imagens da *LFW* rotuladas manualmente:

- Número de exemplos classificados para cada faixa de cor de pele: Quantidade de imagens únicas que cada abordagem classifica para cada uma das 6 faixas de cor de pele.

- RMSE (*root mean squared error*): é a medida que calcula a raiz quadrática média dos erros entre valores observados (rotulação manual) e predições (rotulação automática). Cada imagem única rotulada automaticamente é comparada com o rótulo manual da pessoa única.

- HIT: Quantidade de vezes que a classificação automática encontra o mesmo valor de faixa de cor de pele da classificação manual para cada imagem única.

- CLOSE: Quantidade de vezes que a classificação automática encontra o valor imediatamente abaixo ou acima da faixa de cor de pele da classificação manual para cada imagem única.

- MISS: Quantidade de vezes que a classificação automática não encontra o mesmo valor nem um valor imediatamente abaixo ou acima da faixa de cor de pele da classificação manual para cada imagem única.

- Revocação em cada faixa de cor de pele: Para cada faixa de cor de pele, é dividida a quantidade de HITS pela soma da quantidade de HITS com a quantidade de elementos falsos negativos.
- Precisão em cada faixa de cor de pele: Para cada faixa de cor de pele, é divida a quantidade de HITS pela soma da quantidade de HITS com a quantidade de elementos falso positivos.
- F-Score em cada faixa de cor de pele: É medido o F-Score (F_1) para cada faixa de cor de pele a partir dos valores de revocação e precisão em cada faixa.

Para avaliar na base de dados completa em que faixas de cor de pele as *embeddings* geradas para cada imagem única de pessoas únicas mais se aproximam é realizado o cálculo estatístico de revocação média baseado na metodologia relacionada à recuperação de imagens baseada em conteúdo. Esta medida permite a avaliação do viés de seleção na classificação gerada a partir de redes neurais treinadas em grandes bases de dados.

4.5 Base de dados de imagens

São utilizadas as seguintes bases de imagens publicamente disponíveis:

Labeled Faces in the Wild (LFW): Esta base apresenta imagens de 5749 pessoas únicas. Para cada pessoa única, traz um diretório com uma ou mais imagens da face da mesma, somando um total de 13233 imagens. Todas estão alinhadas e abrangem uma variedade de condições normalmente encontradas na vida cotidiana (HUANG *et al.*, 2008). A base não é anotada por gênero, idade ou cor da pele.

# de imagens por pessoa	# de pessoas (% de pessoas)	# de imagens (% de imagens)
1	4069 (70.8)	4096 (30.7)
2-5	1369 (23.8)	3739 (28.3)
6-10	168 (2.92)	1251 (9.45)
11-20	86 (1.50)	1251 (9.45)
21-30	25 (0.43)	613 (4.63)
31-80	27 (0.47)	1170 (8.84)
> 81	5 (0.09)	1140 (8.61)
Total	**5749**	**13233**

Tabela 2 – Distribuição de imagens e pessoas únicas na LFW.

Fitzpatrick17k: Esse banco de dados apresenta 16.577 imagens clínicas de pele colecionadas a partir de dois outros bancos de dados dermatológicos — *"DermaAmin"* e "Atlas Dermatologico" — com rótulos de faixas de cores de pele Fitzpatrick (GROH *et al.*, 2021).

CAPÍTULO

5

RESULTADOS

Este Capítulo tem o objetivo de apresentar os resultados da investigação descrita na Seção 4. Os experimentos foram realizados em um computador pessoal e também em um servidor do Instituto De Ciências Matemáticas e de Computação (ICMC) da USP de São Carlos. Os resultados serão discutidos no Capítulo 6.

5.1 Rotulação de destacamento da LFW

Na Tabela 3, se encontra a distribuição total de cada nível da escala de cor de pele entre os exemplos da LFW rotulados manualmente. Na Figura 7 (Página 52), são apresentados exemplos da rotulação manual a partir de interpretação individual da escala *Fitzpatrick*, onde cada número posicionado na diagonal superior esquerda de cada imagem representa o nível da escala rotulado.

# Escala Fitzpatrick	# de exemplos
1	19
2	31
3	30
4	20
5	28
6	22

Tabela 3 – Distribuição de exemplos manualmente rotulados representativos na escala Fitzpatrick.

5.2 Visualização de agrupamentos de *embeddings*

A seguir, são apresentadas as visualizações do espaço de atributos. Essas visualizações só foram possíveis através da aplicação do algoritmo PCA nos *embeddings*, com os exemplos classificados em grupos usando diferentes algoritmos e estratégias de agrupamento:

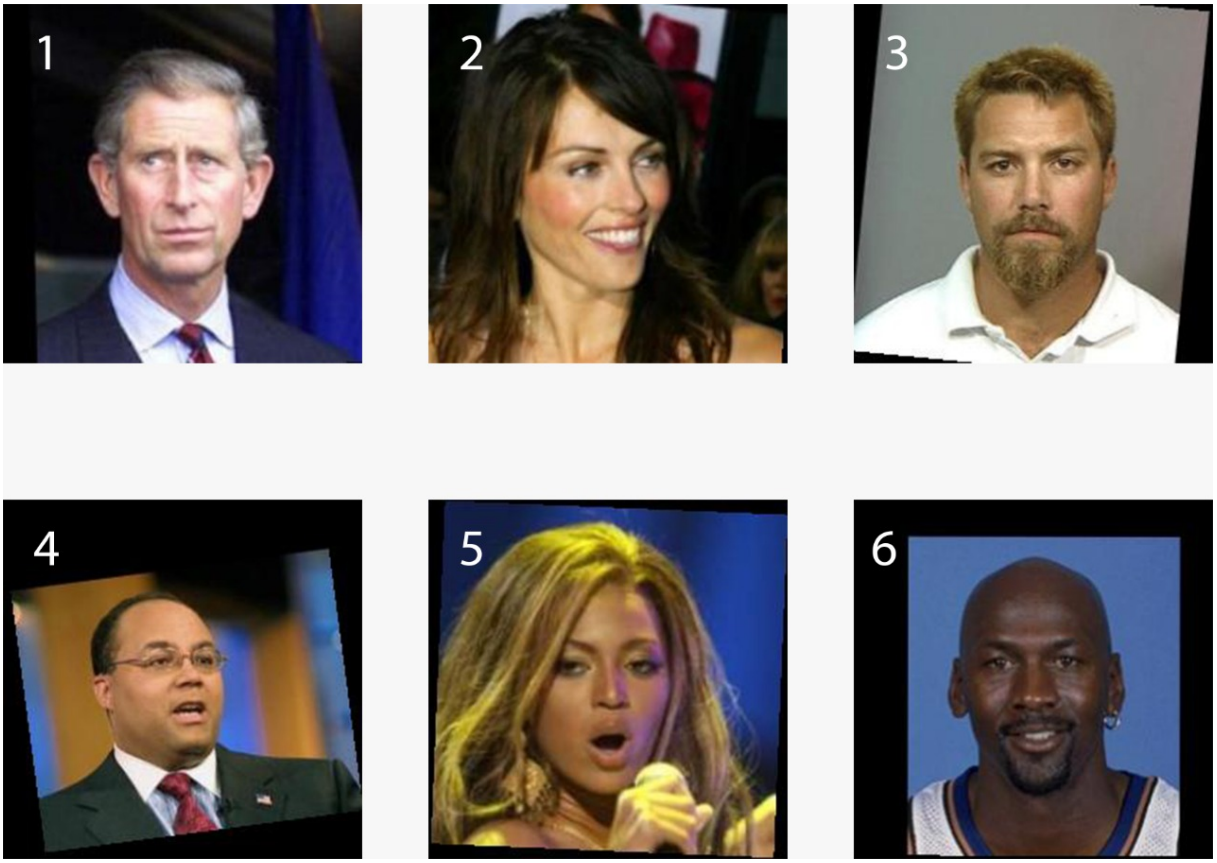

Figura 7 – Exemplos da rotulação manual a partir da interpretação da escala Fitzpatrick

- **Agrupamentos considerando apenas os exemplos manualmente rotulados para extração de características e visualização** - Figuras 8 e 9 na Página 53.
- **Agrupamentos considerando a base de dados completa para extração de características e visualização** - Figuras 10 e 11 na Página 54.
- **Agrupamentos considerando a base de dados completa para extração de características e apenas os exemplos rotulados para visualização** - Figuras 12 e 13 na Página 55.

  Nestas últimas duas visualizações citadas, os exemplos estão divididos em grupos de maneira idêntica às visualizações do item acima, porém só são "revelados"os exemplos rotulados.

5.3 Comparação dos resultados dos agrupamentos com dados rotulados

Foram criados 3 vetores de 3 posições. Cada vetor corresponde a uma partição das 3 descritas em 4.2.1 (A, B, e C). Cada posição de cada vetor corresponde aos grupos gerados a

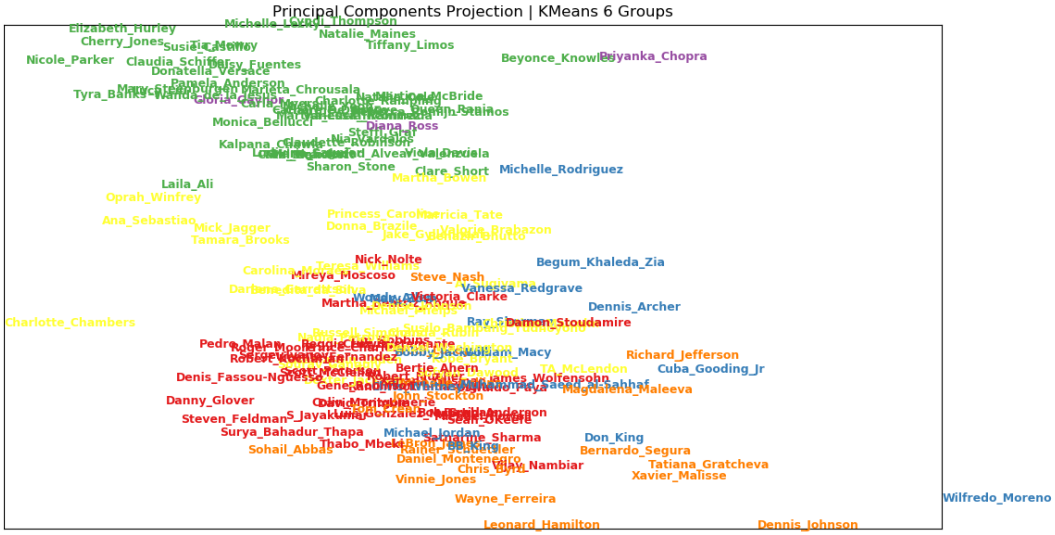

Figura 8 – KMeans com 6 grupos considerando apenas os exemplos manualmente rotulados para extração de características e visualização

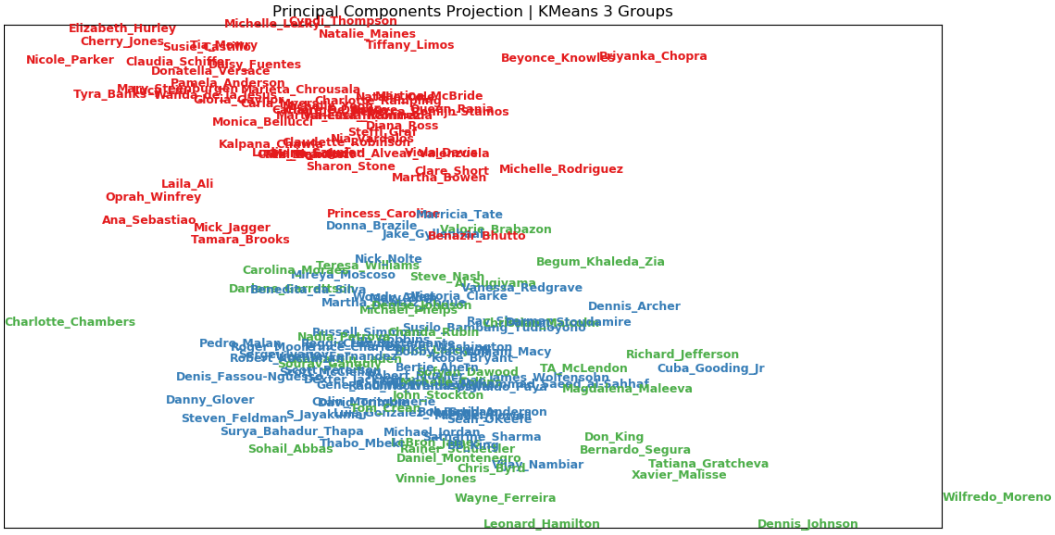

Figura 9 – KMeans com 3 grupos considerando apenas os exemplos manualmente rotulados para extração de características e visualização

partir do algoritmo *KMeans* (com K=3) considerando apenas os exemplos manualmente rotulados para extração de características.

Para cada pessoa única, é verificado através do rótulo a qual partição (A, B ou C) ela pertence:

- Se é da partição A, somamos 1 no vetor correspondente à partição A, na posição equivalente

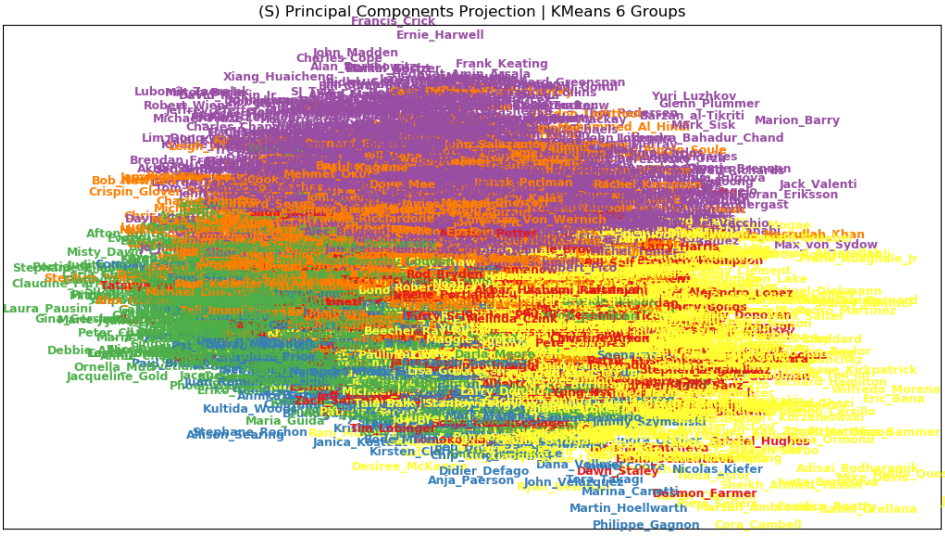

Figura 10 – KMeans com 6 grupos considerando a base de dados completa para extração de características e visualização

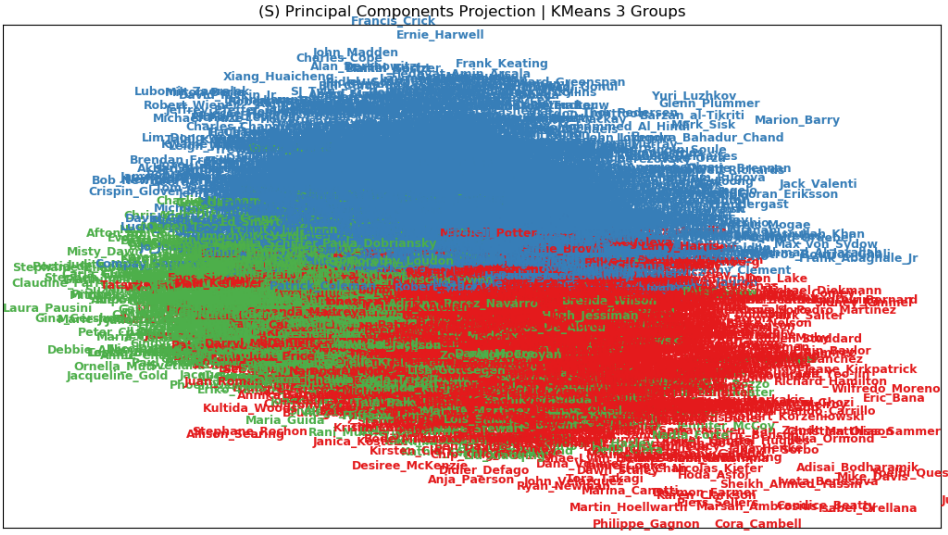

Figura 11 – KMeans com 3 grupos considerando a base de dados completa para extração de características e visualização

ao número do grupo classificado pelo *KMeans* (0, 1 ou 2)

- Se é da partição B, somamos 1 no vetor correspondente ao partição B, na posição equivalente ao número do grupo classificado pelo *KMeans* (0, 1 ou 2)

- Se é da partição C, somamos 1 no vetor correspondente ao partição C, na posição equivalente ao número do grupo classificado pelo *KMeans* (0, 1 ou 2)

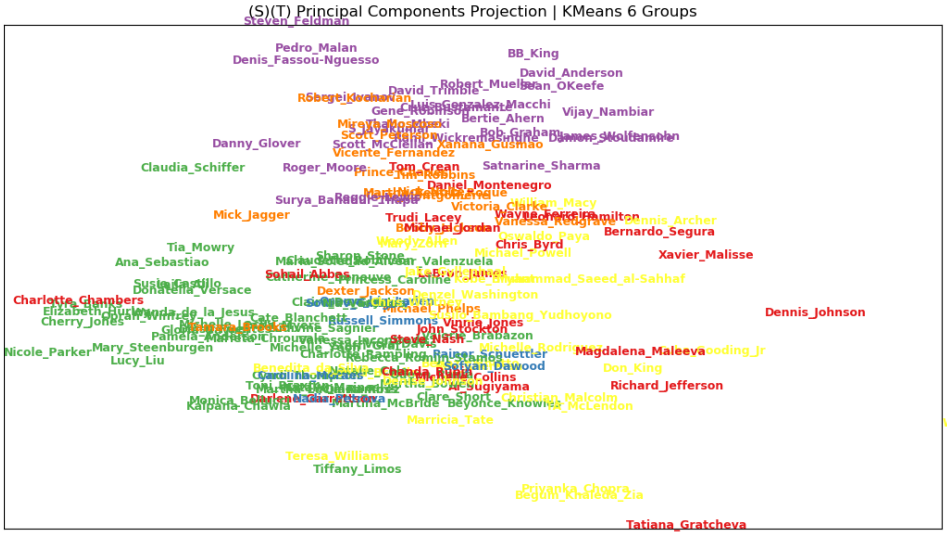

Figura 12 – KMeans com 6 grupos considerando a base de dados completa para extração de características e apenas os exemplos rotulados para visualização

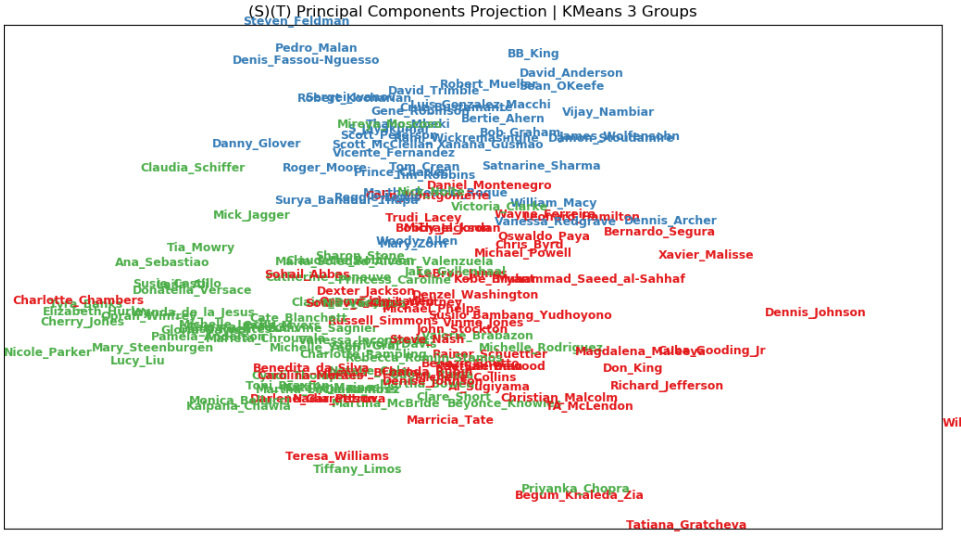

Figura 13 – KMeans com 3 grupos considerando a base de dados completa para extração de características e apenas os exemplos rotulados para visualização

Se a cor de pele fosse um fator consideravelmente determinante para o agrupamento, o resultado se aproximaria de vetores com uma das posições somando aproximadamente 50 e as outras 0.

Após sucessivas execuções do algoritmo, tirada a média e feito um arredondamento, o resultado obtido foi o seguinte:

- Vetor I: [7, 22, 21]
- Vetor II: [14, 17, 19]
- Vetor III: [15, 20, 15]

O mesmo experimento foi realizado considerando a base de dados completa para extração de características. O resultado foi:

- Vetor I: [23, 9, 18]
- Vetor II: [20, 18, 12]
- Vetor III: [15, 26, 9]

5.4 Busca por método de rotulação automática que mais se aproxima da rotulação manual

5.4.1 Experimentos na base LFW

As Tabelas 4 (Página 57), 5 (Página 58), 6 (Página 59) e 7 (Página 60) são tabelas com coleções de todas as métricas extraídas para cada uma das 14 abordagens descritas na Seção 4.4, que foram aplicadas ao subconjunto das imagens da LFW rotuladas manualmente.

Para facilitar a visualização das tabelas:

- verde: indica os melhores resultados entre as abordagens para cada métrica;
- verde mais vibrante: indica o melhor resultado para cada métrica entre as 14 abordagens que usam técnicas de processamento de imagens;
- vermelho: indica os piores resultados entre as abordagens para cada métrica;
- vermelho mais vibrante: indica o pior resultado para cada métrica entre as 14 abordagens que usam técnicas de processamento de imagens;
- preto: valores que não podem ser calculados pois são divisões por zero;
- laranja: indica valores que se destacaram como melhores resultados porém são da abordagem “absurda” (#12);
- amarelo: indica os resultados da abordagem de aprendizado profundo (#15) que são superiores aos melhores resultados obtidos com as outras 14 abordagens.

#	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6	RMSE	HIT	CLOSE	MISS
#01	42	44	182	115	62	27	1.862	103	179	192
#02	34	31	148	169	81	19	1.864	85	181	206
#03	48	74	158	100	56	36	2.018	93	163	216
#04	31	37	202	144	41	18	1.883	84	181	208
#05	37	39	177	133	61	27	1.852	105	174	195
#06	44	38	199	110	56	27	1.898	98	178	198
#07	25	54	333	28	26	6	1.684	133	179	182
#08	21	51	343	29	23	4	1.641	115	117	182
#09	22	60	343	19	24	2	1.632	119	176	179
#10	20	68	346	16	20	0	1.619	124	173	177
#11	21	77	338	19	14	0	1.598	121	176	177
#12	0	0	474	0	0	0	1.525	105	188	181
#13	15	46	193	133	58	29	1.778	112	166	196
#14	21	70	181	116	57	29	1.718	122	170	182
#15	3	206	228	37	0	0	1.440	139	212	123
GT	89	146	105	42	54	38	0	474	0	0

Tabela 4 – LFW - Métricas: Número de exemplos classificados para cada faixa de cor de pele, *RMSE*, *HIT*, *CLOSE* e *MISS*. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos; #15: abordagem com aprendizado profundo.

No gráfico da Figura 14 na Página 60 é possível visualizar com maior qualidade, e em termos proporcionais, os dados expostos nas primeiras colunas da Tabela 4, que representam a distribuição de imagens classificadas para cada abordagem em cada uma das 6 faixas de cor de pele. Nesse contexto, "GT" significa *"ground truth"*, ou seja, a informação que se sabe ser verdadeira, fornecida por observação direta da rotulação manual, em oposição às informações fornecidas por classificação automática.

No gráfico da Figura 15 na Página 61 são somados os valores de F-SCORE por faixa de cor de pele de cada abordagem.

5.4.2 Experimentos na base Fitzpatrick17k

A seguir são expostos também os resultados da execução do método proposto de classificação automática de tons de pele na base *Fitzpatrick17k*. As Tabelas 8 (Página 61), 9 (Página 62), 10 (Página 62) e 11 (Página 63) são respectivamente análogas às Tabelas 4, 5, 6 e 7 descritas na Subseção 5.4.1. O esquema de cores para facilitar a visualização das tabelas descrito em 5.4.1 também vale para as tabelas obtidas nestes experimentos. Os gráficos das Figuras 16 (Página 63) e 17 (Página 64) são respectivamente análogos aos gráficos das Figuras 14 e 15 expostos na Subseção 5.4.1.

RECALL						
#	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6
#01	0,079	0,082	0,495	0,357	0,204	0,158
#02	0,067	0,062	0,356	0,429	0,204	0,105
#03	0,101	0,200	0,308	0,310	0,056	0,184
#04	0,056	0,048	0,452	0,381	0,093	0,105
#05	0,067	0,082	0,495	0,405	0,222	0,158
#06	0,079	0,048	0,505	0,357	0,167	0,184
#07	0,135	0,151	0,724	0,048	0,019	0,000
#08	0,124	0,151	0,743	0,095	0,000	0,000
#09	0,135	0,192	0,743	0,024	0,000	0,000
#10	0,112	0,226	0,733	0,024	0,000	
#11	0,101	0,247	0,714	0,024	0,000	
#12			1,000			
#13	0,056	0,137	0,486	0,405	0,222	0,184
#14	0,090	0,212	0,457	0,381	0,222	0,184
#15	0,011	0,473	0,590	0,167		
Revocação - LFW						

Tabela 5 – LFW - Métricas: Revocação. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos; #15: abordagem com aprendizado profundo.

5.5 Análise da revocação considerando faces únicas da LFW

Revocações médias (mR@1, mR@5 e mR@10) foram calculadas para cada grupo descrito na Subseção 4.2.1 e também para cada faixa de cor de pele isolada.

As Figuras 18, 19, 20, 21 representam respectivamente os gráficos das revocações médias referentes à metodologia relacionada à recuperação de imagens baseada em conteúdo com *embeddings* extraídos das abordagens #01 (porém, sem detecção facial no pré-processamento), #01 (com detecção facial), #03 e #14. Os dados foram extraídos a partir da execução das abordagens na base de dados completa para que fosse investigado em que faixas de cor de pele as *embeddings* geradas para cada imagem única de pessoas únicas mais se aproximam, em cada abordagem. A Figura 22 mostra todos os gráficos citados acima lado a lado, onde o número 1 é o primeiro citado e 4 o último. Todas essas Figuras podem ser vistas a partir da Página 64.

As Tabelas 12 (Página 65), 13 (Página 65), 14 (Página 66) e 15 (Página 66) representam respectivamente os valores das revocações médias em cada um das três partições de faixas de cor de pele nas abordagens #01 (sem detecção facial no pré-processamento), #01 (com detecção facial), #03 e #14.

Nas Figuras 23 e 24 são apresentados exemplos reais de como a recuperação de imagens baseada em conteúdo funciona no reconhecimento facial de pessoas de pele mais clara (ITA1) e

PRECISION						
#	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6
#01	0,167	0,273	0,282	0,130	0,177	0,222
#02	0,176	0,290	0,250	0,107	0,155	0,211
#03	0,188	0,392	0,203	0,130	0,054	0,194
#04	0,161	0,189	0,233	0,111	0,122	0,222
#05	0,162	0,308	0,294	0,128	0,197	0,222
#06	0,159	0,184	0,266	0,136	0,161	0,259
#07	0,480	0,407	0,227	0,071	0,038	0,000
#08	0,524	0,431	0,226	0,133	0,000	0,000
#09	0,545	0,459	0,226	0,050	0,000	0,000
#10	0,500	0,478	0,221	0,059	0,000	
#11	0,429	0,462	0,221	0,050	0,000	
#12			0,222			
#13	0,333	0,435	0,264	0,128	0,207	0,241
#14	0,381	0,443	0,265	0,138	0,211	0,241
#15	0,333	0,335	0,272	0,189		
Precisão - LFW						

Tabela 6 – LFW - Métricas: Precisão. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos; #15: abordagem com aprendizado profundo.

pele mais escura (ITA6) usando a abordagem #01 (ou seja, sem pré-processamento de imagem além da detecção facial) para extração de *embeddings*.

F-SCORE						
#	**ITA1**	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**
#01	0,121	0,170	0,380	0,233	0,190	0,189
#02	0,119	0,165	0,301	0,248	0,179	0,156
#03	0,143	0,289	0,253	0,213	0,055	0,189
#04	0,106	0,114	0,334	0,231	0,107	0,161
#05	0,112	0,184	0,387	**0,251**	0,209	0,189
#06	0,118	0,112	0,375	0,237	0,164	**0,220**
#07	0,285	0,266	0,434	0,059	0,028	0,000
#08	0,294	0,276	**0,439**	0,114	0,000	0,000
#09	**0,309**	0,312	**0,439**	0,037	0,000	0,000
#10	0,277	0,340	0,433	0,041	0,000	0,000
#11	0,244	**0,346**	0,426	0,037	0,000	0,000
#12	0,000	0,000	0,517	0,000	0,000	0,000
#13	0,178	0,269	0,366	**0,251**	0,214	0,212
#14	0,218	0,317	0,354	0,248	**0,216**	0,212
#15	0,150	0,401	0,413	0,178	0,000	0,000
F-Score - LFW						

Tabela 7 – LFW - Métricas: F-Score. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos; #15: abordagem com aprendizado profundo.

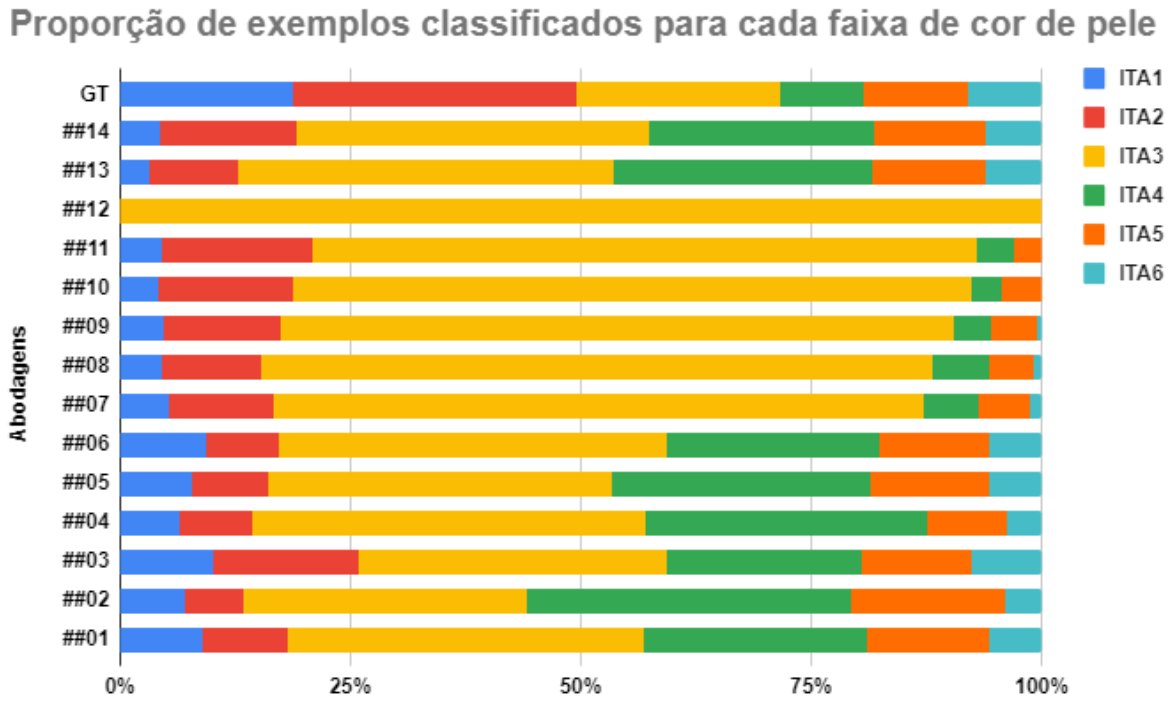

Figura 14 – LFW - Proporção de exemplos classificados para cada faixa de cor de pele em cada abordagem de classificação

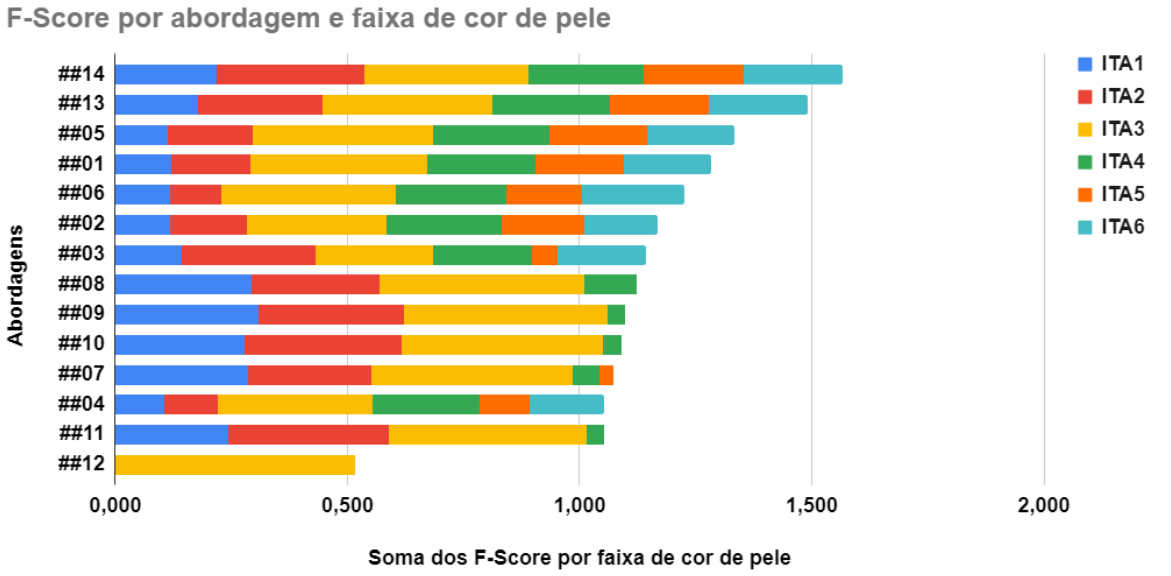

Figura 15 – LFW - Soma dos F-Score por faixa de cor de pele e por abordagem de classificação

#	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6	RMSE	HIT	CLOSE	MISS
#01	1580	2071	5449	4428	1984	988	1,903	3812	6021	6113
#02	1394	1669	4663	5385	2276	1109	1,881	3878	5925	6139
#03	1812	2602	5176	3792	2001	1120	2,037	3329	5557	7062
#04	2041	2620	4075	3751	2675	1344	1,962	3068	5489	7396
#05	1532	2085	5193	4704	2042	944	1,880	3897	6012	6037
#06	1638	2119	5446	4410	1925	962	1,918	3782	5932	6232
#07	670	2126	9890	2491	1080	239	1,669	3889	6684	5370
#08	626	2203	9897	2473	1074	203	1,657	3906	6711	5308
#09	591	2288	9856	2504	1055	164	1,645	3913	6719	5280
#10	573	2404	9824	2481	1010	127	1,624	3943	6762	5178
#11	549	2519	9787	2445	981	101	1,613	3981	6747	5127
#12	0	0	16525	0	0	0	1,395	3299	7572	5095
#13	336	1637	4679	4056	2587	3163	1,847	3709	5977	6227
#14	346	1637	4636	4056	2587	3163	1,847	3705	5972	6214
GT	2940	4796	3299	2776	1527	628	0	16525	0	0

Tabela 8 – Fitzpatrick17k - Métricas: Número de exemplos classificados para cada faixa de cor de pele, *RMSE*, *HIT*, *CLOSE* e *MISS*. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos.

RECALL						
#	**ITA1**	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**
#01	0,118	0,152	0,428	0,352	0,184	0,109
#02	0,113	0,134	0,350	0,477	0,221	0,147
#03	0,125	0,180	0,379	0,217	0,121	0,096
#04	0,084	0,187	0,168	0,325	0,261	0,108
#05	0,116	0,159	0,393	0,411	0,194	0,107
#06	0,118	0,155	0,422	0,349	0,180	0,096
#07	0,092	0,185	0,714	0,127	0,016	0,003
#08	0,088	0,191	0,714	0,127	0,015	0,003
#09	0,086	0,197	0,711	0,128	0,013	0,002
#10	0,085	0,207	0,709	0,128	0,013	0,002
#11	0,085	0,218	0,707	0,127	0,011	0,002
#12			1,000			
#13	0,080	0,170	0,366	0,334	0,288	0,149
#14	0,081	0,171	0,364	0,334	0,289	0,149
Revocação - FITZ17K						

Tabela 9 – Fitzpatrick17k - Métricas: Revocação. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos.

PRECISION						
#	**ITA1**	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**
#01	0,230	0,370	0,267	0,226	0,146	0,073
#02	0,251	0,405	0,254	0,251	0,154	0,089
#03	0,209	0,344	0,251	0,163	0,096	0,056
#04	0,127	0,359	0,142	0,244	0,152	0,052
#05	0,235	0,383	0,257	0,248	0,150	0,075
#06	0,221	0,367	0,264	0,225	0,148	0,066
#07	0,411	0,427	0,246	0,152	0,023	0,009
#08	0,420	0,427	0,246	0,152	0,022	0,010
#09	0,434	0,424	0,245	0,150	0,019	0,006
#10	0,448	0,421	0,245	0,150	0,019	0,008
#11	0,463	0,424	0,245	0,152	0,018	0,010
#12			0,207			
#13	0,336	0,417	0,230	0,233	0,146	0,089
#14	0,335	0,417	0,230	0,233	0,146	0,089
Precisão - FITZ17K						

Tabela 10 – Fitzpatrick17k - Métrica: Precisão. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos.

F-SCORE						
#	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6
#01	0,171	0,252	0,343	0,286	0,165	0,091
#02	0,178	0,255	0,300	**0,355**	0,187	0,117
#03	0,165	0,257	0,312	0,189	0,108	0,076
#04	0,105	0,267	0,155	0,283	0,204	0,079
#05	0,172	0,261	0,322	0,325	0,172	0,091
#06	0,167	0,252	0,338	0,284	0,164	0,081
#07	0,231	0,295	**0,443**	0,139	0,019	0,006
#08	0,232	0,298	**0,443**	0,139	0,018	0,006
#09	0,236	0,301	0,441	0,139	0,016	0,004
#10	0,240	0,305	0,441	0,139	0,016	0,005
#11	**0,246**	**0,313**	0,440	0,139	0,014	0,006
#12	0,000	0,000	0,505	0,000	0,000	0,000
#13	0,194	0,282	0,294	0,282	**0,213**	**0,118**
#14	0,195	0,282	0,294	0,282	**0,213**	**0,118**
F-Score - FITZ17K						

Tabela 11 – Fitzpatrick17k - Métrica: F-Score. De #01 a #12: métodos independentes; #13 e #14: combinação de métodos.

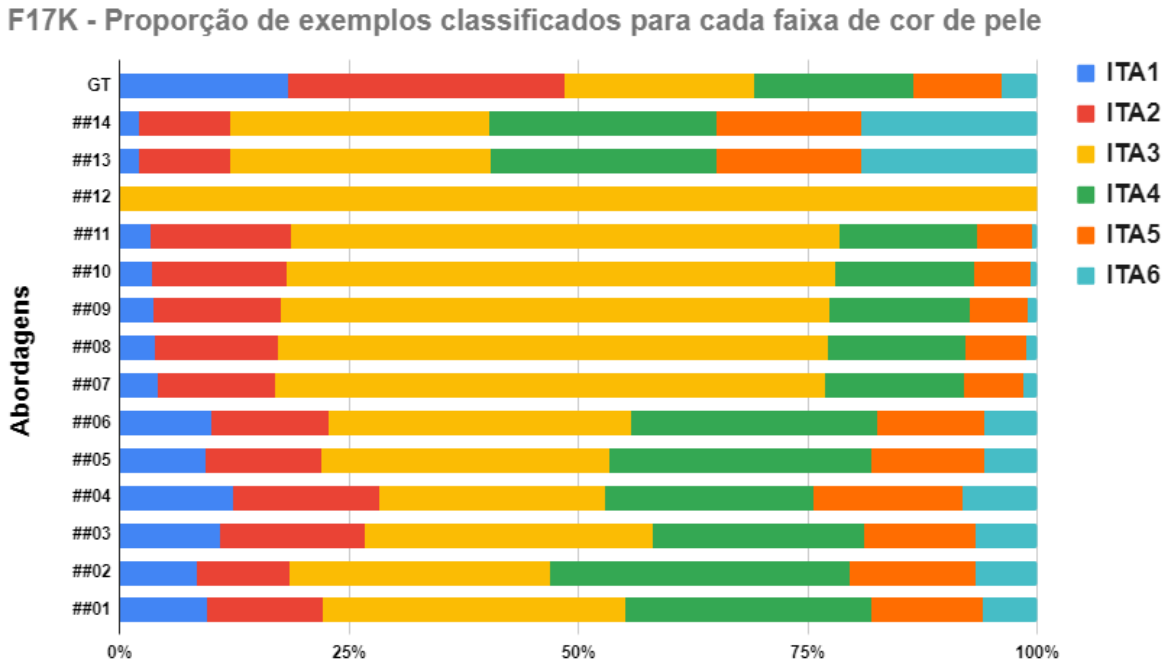

Figura 16 – Fitzpatrick17k - *Ground truth* (GT) e proporção de exemplos classificados para cada faixa de cor de pele em cada abordagem de classificação

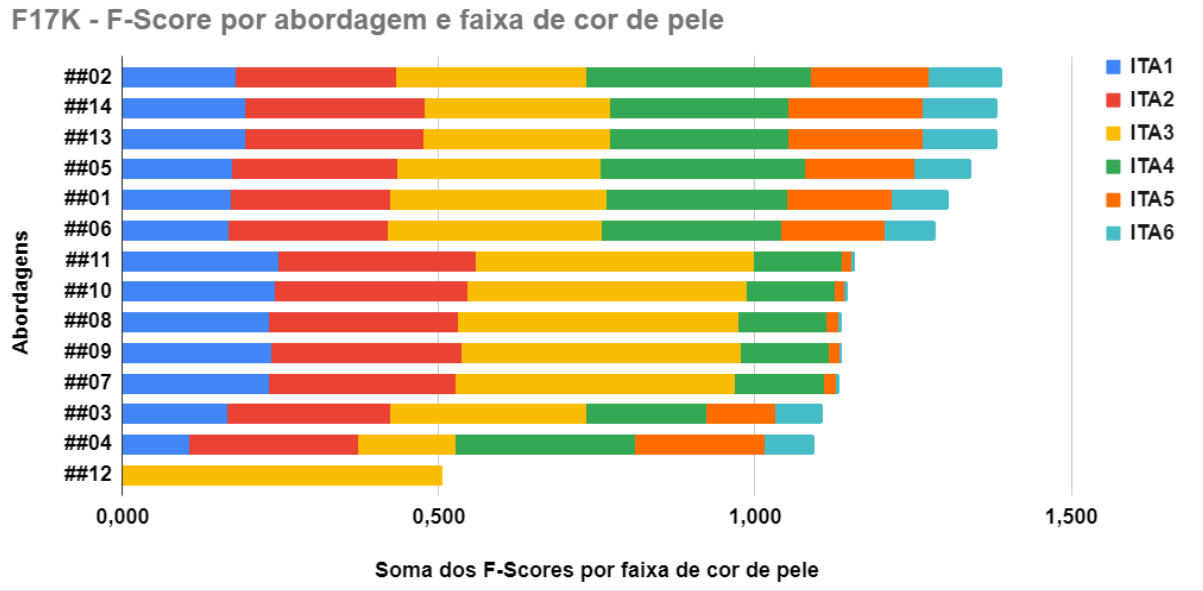

Figura 17 – Fitzpatrick17k - Soma dos F-Score por faixa de cor de pele e por abordagem de classificação

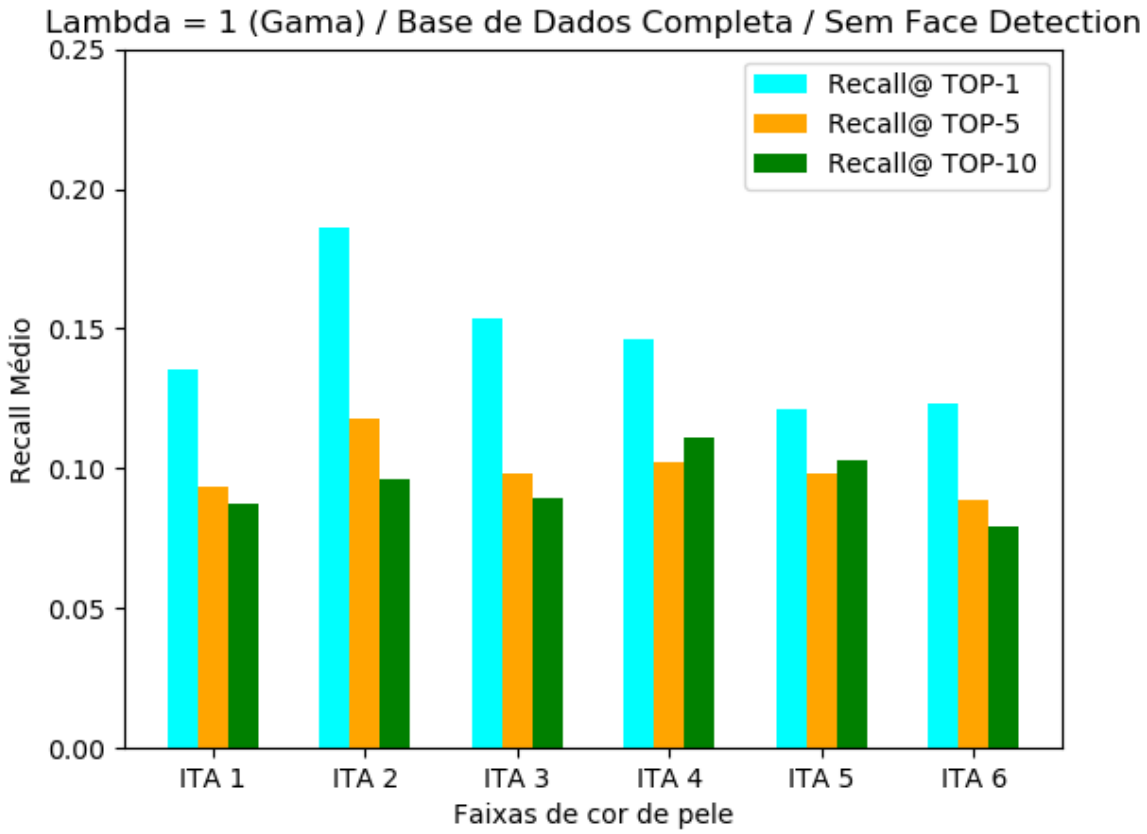

Figura 18 – Gráfico das revocações médias referentes à metodologia relacionada à recuperação de imagens baseada em conteúdo com *embeddings* extraídos da abordagem #01 (porém, sem detecção facial no pré-processamento)

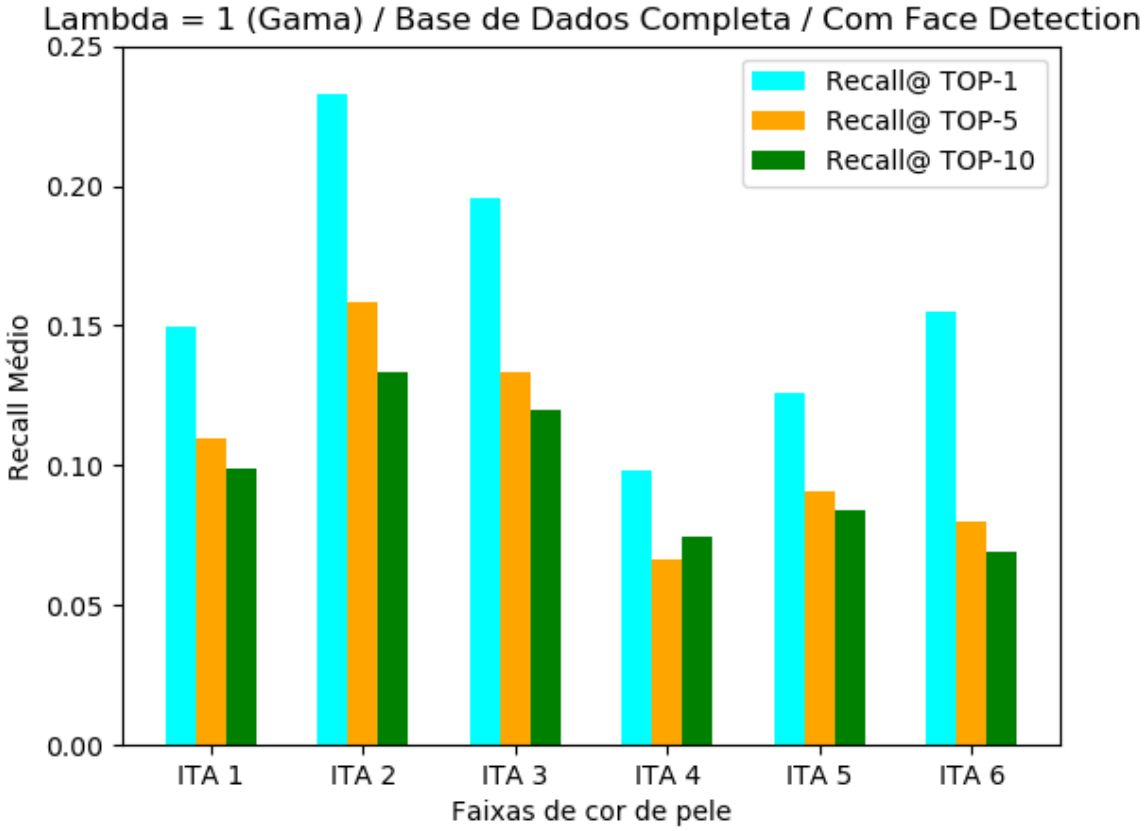

Figura 19 – Gráfico das revocações médias referentes à metodologia relacionada à recuperação de imagens baseada em conteúdo com *embeddings* extraídos da abordagem #01

Revocação média	Partição A	Partição B	Partição C
mR@1	0.161	0.150	0.122
mR@5	0.106	0.100	0.094
mR@10	0.092	0.100	0.091

Tabela 12 – Revocação média das partições na abordagem #01 (sem detecção facial no pré-processamento).

Revocação média	Partição A	Partição B	Partição C
mR@1	0.191	0.147	0.140
mR@5	0.134	0.100	0.086
mR@10	0.116	0.097	0.077

Tabela 13 – Revocação média das partições na abordagem #01 (com detecção facial no pré-processamento).

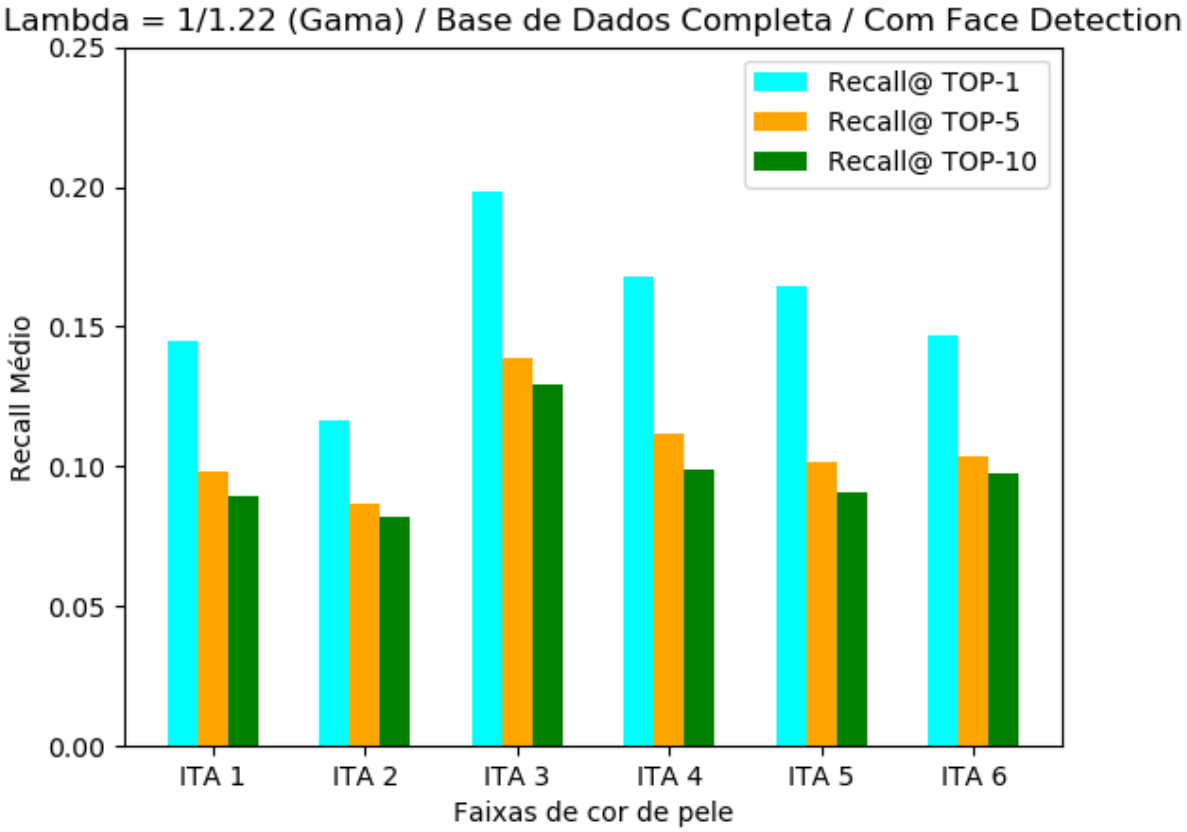

Figura 20 – Gráfico das revocações médias referentes à metodologia relacionada à recuperação de imagens baseada em conteúdo com *embeddings* extraídos da abordagem #03

Revocação média	Partição A	Partição B	Partição C
mR@1	0.131	0.183	0.156
mR@5	0.093	0.125	0.103
mR@10	0.086	0.114	0.094

Tabela 14 – Revocação média das partições na abordagem #03.

Revocação média	Partição A	Partição B	Partição C
mR@1	0.118	0.160	0.176
mR@5	0.080	0.106	0.119
mR@10	0.077	0.099	0.111

Tabela 15 – Revocação média das partições na abordagem #14.

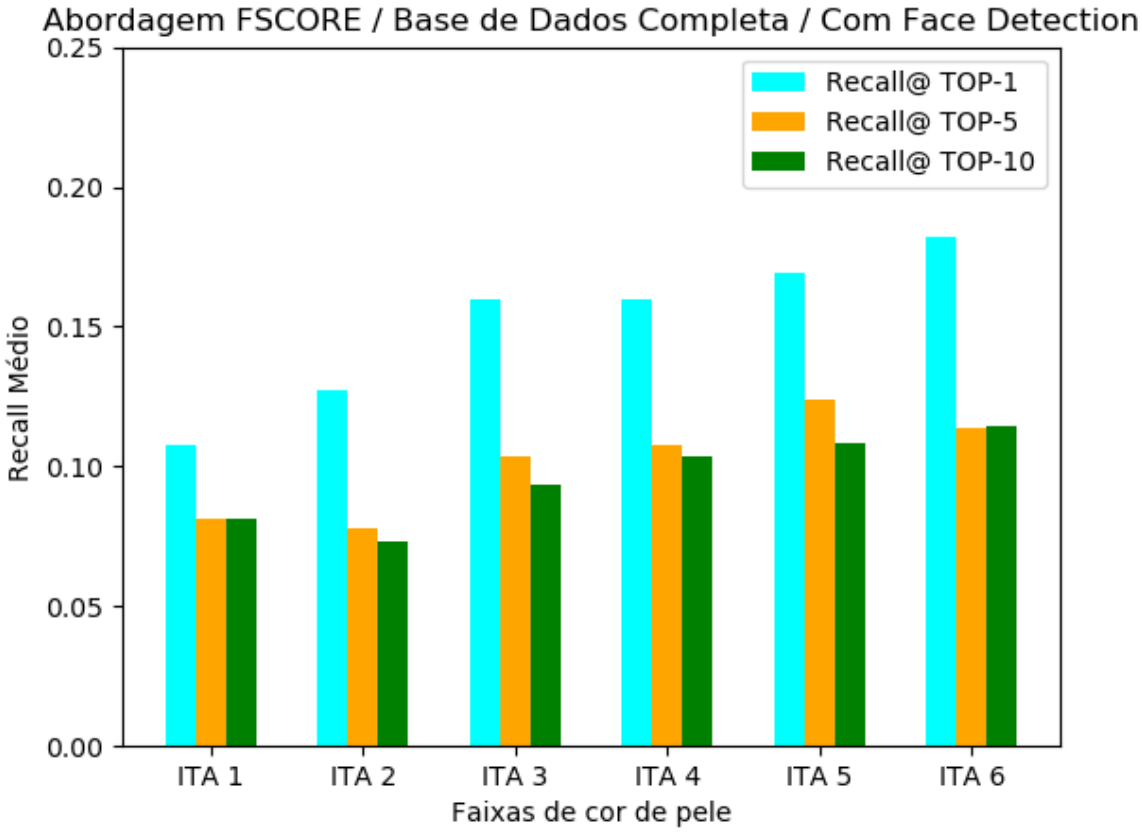

Figura 21 – Gráfico das revocações médias referentes à metodologia relacionada à recuperação de imagens baseada em conteúdo com *embeddings* extraídos da abordagem #14

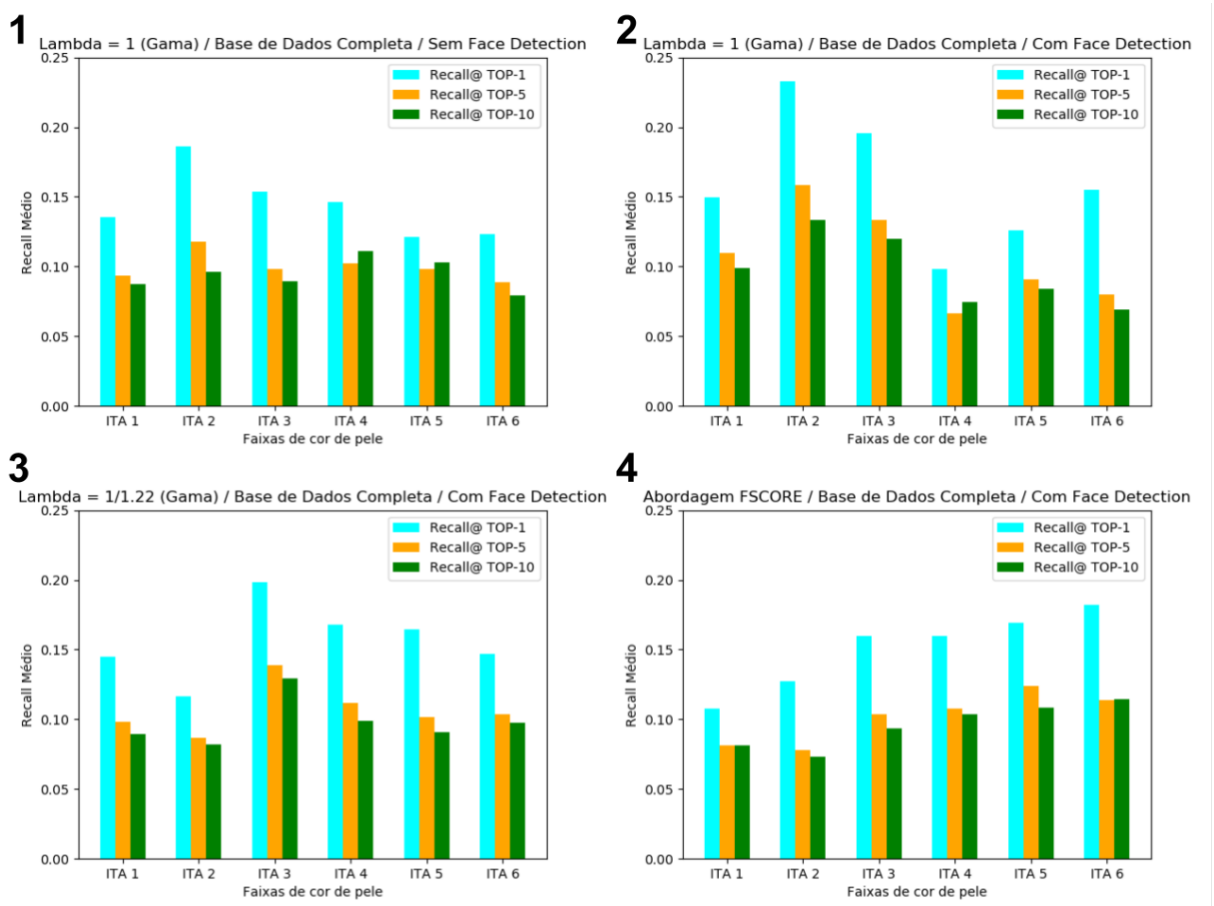

Figura 22 – Gráficos das revocações médias referentes à metodologia relacionada à recuperação de imagens baseada em conteúdo com *embeddings* extraídos das abordagens #01 (sem e com detecção facial), #03 e #14 lado a lado

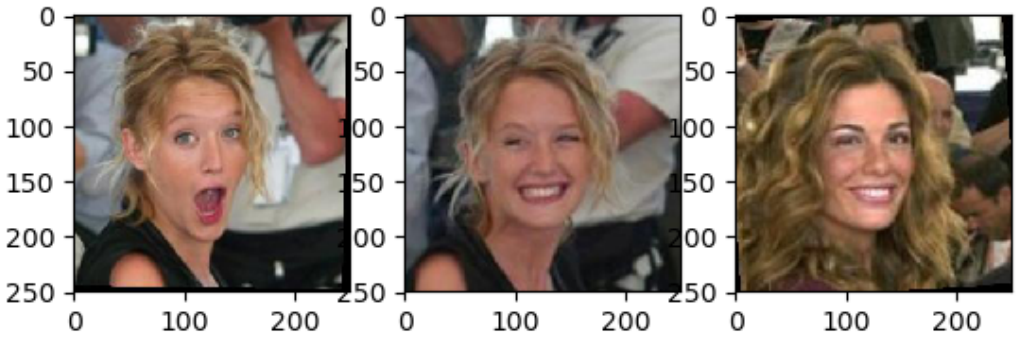

Figura 23 – Exemplo real de como a recuperação de imagens baseada em conteúdo funciona no reconhecimento facial de pessoas de pele mais clara (ITA1)

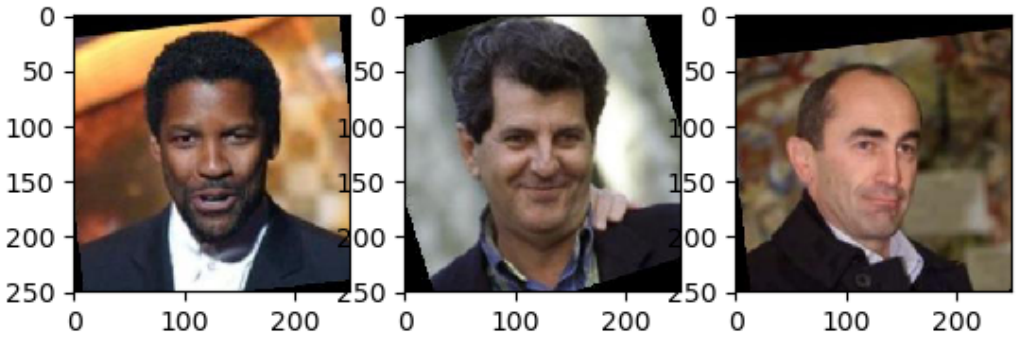

Figura 24 – Exemplo real de como a recuperação de imagens baseada em conteúdo funciona no reconhecimento facial de pessoas de pele mais escura (ITA6)

CAPÍTULO

6

DISCUSSÃO

Nessa Seção vamos discutir os resultados apresentados no Capítulo 5 que foram obtidos a partir do processo descrito no Capítulo 4.

De início, é importante observar que, durante o processo de rotulação manual de exemplos da LFW entre faixas de cor de pele de 1 a 6, foi difícil encontrar manualmente na base pessoas negras (notadamente mulheres negras) para rotulação. Mesmo quando encontrados exemplos de pessoas de pele mais escura para rotulação manual, esses exemplos continham número reduzido de imagens. Essa dificuldade trouxe um reflexo prático: entre os exemplos anotados, pessoas únicas de pele mais escura tem menos imagens únicas do que pessoas de pele mais clara no espaço de exemplos rotulados.

Extraídas as características das pessoas por meio de redes neurais e analisadas as visualizações dos agrupamentos de *embeddings* expostas na Seção 5.2 foi possível observar que em todos estes experimentos o algoritmo *KMeans* aplicado para 3 grupos separou visualmente de maneira satisfatória os exemplos no espaço. Com essas visualizações, porém, não foi possível identificar quais características centrais os algoritmos estão usando para agrupar os exemplos.

Já no experimento seguinte, que comparou os resultados dos agrupamentos citados acima com dados rotulados, foi observado um indício que a característica da cor da pele não foi levada em conta para o agrupamento. Isso também trouxe indícios que essa característica pode não ter sido relevante na própria extração de características aplicada pela rede neural escolhida no contexto da base de dados LFW sem pré-processamento de imagens.

Resultados de classificação de tons de pele: pela análise das Tabelas 4, 5, 6 e 7 (tons de pele detectados e métricas para LFW) e Tabelas 8, 9, 10 e 11 (tons de pele detectados e métricas para Fitzpatrick17k), foi perceptível que não existe uma abordagem que seja melhor em todas as métricas e tons de pele.

Pode ser observado nas Figuras 14 e 16 que todas abordagens aplicadas em ambas bases

de imagens subestimaram a quantidade de exemplos das faixas de cor de pele mais clara (ITA1 e ITA2). No caso das abordagens aplicadas na LFW, também subestimaram a quantidade de exemplos da faixa de cor de pele mais escura (ITA6). Além disso, nas duas bases de dados, todas abordagens executadas superestimaram a quantidade de exemplos da faixa de cor média ITA3.

No processo de extração do RMSE, considerando as abordagens que usam métodos de processamento de imagem aplicadas para ambas bases de dados, foi observado que o menor (e portanto melhor) resultado foi obtido na abordagem #12, que é a "absurda" onde todas as imagens foram mecanicamente classificadas com o valor "3" para faixa de cor de pele. Isso se deve pelo fato que essa métrica favorece a média (WANG; BOVIK, 2009). A abordagem #03, que usa o "Método A"aplicando transformação gama (com $gama = 1/1.22$) teve o maior (e portanto pior) RMSE também para a aplicação na LFW e na Fitzpatrick17k.

As abordagens de processamento de imagens #7, #8, #9, #10 e #11 que são baseadas no "Método C", que aplicam transformações de mediana com os parâmetros descritos na Seção 4.2, foram as que apresentam em ambas as bases os melhores resultados nas métricas de HIT e MISS, ou seja, acertaram mais a faixa de cor de pele que corresponde a cada exemplo e erraram menos.

Em relação às métricas de revocação e precisão aplicadas no contexto da classificação automática de cor de pele, também não houve abordagem que apresentou melhores resultados considerando todas as faixas de cor da pele. A partir da observação do F-SCORE das diferentes abordagens nas diferentes faixas de cor de pele, o "Método D" (#13 e #14) apresentou resultados que, no contexto de ambas as bases, estão entre os melhores. É importante observar que a abordagem de aprendizado profundo #15 aplicada na LFW apresenta o melhor F-SCORE na faixa de ITA2, que é a faixa de cor com mais exemplos na base onde foi treinada (Fitzpatrick17k).

Em particular, a Figura 15 mostra que o Método D alcançou os melhores resultados em relação ao F-SCORE na base LFW, e na análise da Figura 17 também foi notável que alcançou bons resultados no conjunto de dados FITZ17K, colocando-se no *top* 3. É importante notar que, independente da abordagem, os melhores F-SCORE obtidos para cores de pele mais clara foram maiores do que os obtidos para peles mais escuras.

Uso de aprendizado profundo: houve indícios que a abordagem #15 que usa aprendizado profundo e foi aplicada na LFW não aprendeu a identificar pessoas de pele mais escura (ITA5 e ITA6). Mesmo assim, a execução desta atingiu os melhores resultados nas métricas RMSE, HIT e MISS. Isso pode acontecer devido à baixa quantidade de exemplos desses tons de peles na LFW. Assim, o que esse tipo de método pode ter aprendido é a proporção de exemplos disponíveis de cada tom de pele, ao invés dos padrões necessários para detectá-los.

Com relação ao método proposto, é possível generaliza-lo da seguinte forma: escolha N abordagens diferentes de processamento de imagens e M diferentes abordagens de algoritmo para rotulação automática de cor de pele; Em seguida, selecione os melhores métodos para cada faixa de cor de pele de acordo com o F-SCORE obtido e aplique-as em uma ordem de prioridade.

Análise estatística da revocação considerando faces únicas: Pode ser observado nas Tabelas 12 e 13 que nas faixas de cor de pele mais escuras (grupo III e escalas *Fitzpatrick* 5 e 6) as medidas de revocação média aplicadas à base LFW original foram consideravelmente menores. Isso mostrou que as *embeddings* geradas para pessoas de cor pele escura foram mais similares entre si do que *embeddings* geradas para pessoas de cor de pele clara entre elas. Ou seja, o reconhecimento facial se "confundiu"mais na classificação de pessoas de pele escura neste caso. Aplicando técnicas de processamento de imagens, esse cenário pode mudar, como foi evidenciado nas Tabelas 14 e 15.

Quando colocadas as Figuras 18, 19, 20 e 21 lado a lado (Figura 22) foi possível observar que a abordagem #14 (indicada com o número 4 na Figura 22) apresentou maiores valores de revocação média para faixas de cor de pele mais escuras (notadamente ITA5 e ITA6).

Também foi possível observar (Figura 23) que para pessoas de pele mais clara as imagens recuperadas tenderam mais a ser outras imagens da mesma pessoa ou de pessoas visivelmente parecidas, enquanto para pessoas de pele mais escura as imagens recuperadas tenderam mais a ser pessoas diferentes da pessoa a quem desejamos recuperar (Figura 24).

CAPÍTULO

7

CONCLUSÃO

Nesse Capítulo são apresentadas algumas conclusões e limitações com o objetivo de apontar os principais aprendizados, e novos caminhos para futuras pesquisas na área de classificação justa no contexto do viés de seleção étnico-racial em sistemas de reconhecimento facial.

Os resultados mostraram vantagens em combinar abordagens de processamento de imagem que são melhores em cada faixa de cor de pele para obter uma abordagem de rotulação automática de grandes bases de dados (no contexto do reconhecimento facial) que se melhor se aproximam da rotulação manual.

Durante o estudo, também foi observado que rotulação automática de cor de pele baseada em regras ou aprendizado de máquina tende a obter resultados melhores para imagens de pessoas de pele mais clara ou média em termos de taxa de detecção e F-score. Os melhores resultados nos tons de pele mais escuros foram piores que os relativos aos tons de pele mais claros. Isso pode ter conexão com o método escolhido para identificar o que é cor de pele em cada imagem única. Adicionalmente, algoritmos de reconhecimento facial treinados na LFW com pesos pré treinados da *facenet* usando o modelo de CNN *ResNet50* tenderam a confundir mais pessoas negras no processo de classificação.

Nesse trabalho, foi visto que é possível mitigar esse efeito priorizando abordagens que funcionam melhor em diferentes tipos de pele, atenua esse problema. Este procedimento pode ser facilmente realizado usando um conjunto de validação em cenários do mundo real. O fato do método proposto ter atingindo melhores resultados na LFW do que na Fitzpatrick17k pode estar relacionado ao pré-processamento de detecção facial feito nas imagens da primeira, antes da identificação dos *pixels* que são da cor da pele.

Por privilegiar valores médios no calculo do erro, no processo de desenvolvimento do trabalho foi entendido que outras métricas podem ser mais adequadas que o RMSE. Métricas de avaliação como a precisão, revocação e f-score se apresentaram mais adequadas para avaliar as

informações no contexto da representatividade étnico-racial nas bases de dados.

A falta de bases de dados públicas de imagens de faces para reconhecimento facial rotuladas com autodeclaração étnico-racial das pessoas dificulta muito o trabalho de avaliação de impacto do viés de seleção racista no reconhecimento facial. Para lidar com essa limitação foi formulado nesse trabalho táticas para rotulação automática de cor de pele. Porém, mesmo essa rotulação esbarra em outra limitação: para a sociologia, os conceitos de raça e etnia não levam em conta apenas a cor de pele das pessoas, mas também aspectos histórico-sociais. Por isso, para uma avaliação com maior rigor científico, é importante que essa seja feita a partir de uma base de dados com rótulos de auto-declaração.

Outra importante limitação que deve ser citada é que o processo de rotulação manual da cor de pele foi feito a partir do julgamento de apenas um pesquisador e foram rotuladas 150 pessoas para o dataset de faces. Maiores bases de dados podem beneficiar estudos futuros.

O foco desse trabalho de mestrado foi obter um método com eficácia similar em diferentes tons de pele de forma a permitir auditar se bases de dados apresentam ou não viés de seleção considerando as característica de representatividade étnico-racial, o que ataca uma parte do problema pra que as pessoas e organizações possam utilizar modelos justos de reconhecimento facial. Porém, esse esforço não buscou resolver nem os problemas dos modelos de reconhecimento facial e seria incapaz de resolver o problema maior, que é do racismo estrutural no contexto da sociedade capitalista.

7.1 Trabalhos futuros

Três desdobramentos práticos desse trabalho que podem ser usados por próximos cientistas pesquisadores e pesquisadoras são: um método robusto de classificação automática de cor de pele com pré-processamento de imagens; um destacamento da LFW com anotações manuais sobre cor de pele e anotações de cor de pele da base LFW completa feitas a partir do método de classificação automática proposto. Os dois últimos podem ser acessados através de repositório no *github*[1].

É importante que outros métodos que identificam quais *pixels* são cor de pele em cada imagem única sejam investigados em trabalhos futuros. Além disso, é ideal que em próximas pesquisas na área onde haja a necessidade de rotulação manual de cor de pele, ela seja feita de maneira colaborativa por mais pesquisadores e pesquisadoras e que sejam rotulados uma maior quantidade de exemplos.

Podemos concluir também que é importante a investigação mais aprofundada de métodos de aprendizado profundo como redes neurais tanto para a rotulação automática de cor de pele quanto para a geração de matrizes de características para pessoas únicas para reconhecimento

[1] <https://github.com/LuizAVManoel/LFW-annotated-by-skin-tones>

facial. No âmbito da detecção de cor de pele o uso de redes neurais pode ser promissor se forem treinadas em bases de dados com classes (tons de pele) mais equilibradas. Para a geração de *embeddings*, é importante a investigação de outros tipos de redes neurais com variação de parâmetros, como as *Vision Transfomers* (DOSOVITSKIY *et al.*, 2020).

Dentro desse mesmo assunto, para verificar se os *embeddings* gerados levam em consideração os tons de pele, próximas pesquisas podem usar outros algoritmos de agrupamento, como os de densidade, e de redução de dimensionalidade diferente do que foram usados neste trabalho. Além disso, futuros estudos podem realizar a análise do espaço de atributos para conjuntos de características geradas a partir do pré-processamento de imagens nas bases com o método proposto nesta dissertação para verificar se a cor de pele se torna mais determinante para os algoritmos de agrupamento.

REFERÊNCIAS

AGARWAL, C.; D'SOUZA, D.; HOOKER, S. Estimating example difficulty using variance of gradients. **arXiv preprint arXiv:2008.11600**, 2020. Citado na página 37.

ALMEIDA, S. **Racismo estrutural**. [S.l.]: Pólen Produção Editorial LTDA, 2019. Citado na página 33.

ALTHUSSER, L. Aparelhos ideológicos de estado. **Rio de janeiro: Graal**, v. 2, 1985. Citado na página 22.

AMINI, A.; SOLEIMANY, A. P.; SCHWARTING, W.; BHATIA, S. N.; RUS, D. Uncovering and mitigating algorithmic bias through learned latent structure. In: **Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society**. [S.l.: s.n.], 2019. p. 289–295. Citado na página 36.

BISSOTO, A.; VALLE, E.; AVILA, S. Debiasing skin lesion datasets and models? not so fast. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**. [S.l.: s.n.], 2020. p. 740–741. Citado na página 37.

BRANDÃO, R.; OLIVEIRA, J. L. Reconhecimento facial e viés algorítmico em grandes municípios brasileiros. In: SBC. **Anais do II Workshop sobre as Implicações da Computação na Sociedade**. [S.l.], 2021. p. 122–127. Citado na página 23.

BUOLAMWINI, J.; GEBRU, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: **Conference on fairness, accountability and transparency**. [S.l.: s.n.], 2018. p. 77–91. Citado nas páginas 35 e 37.

CHARDON, A.; CRETOIS, I.; HOURSEAU, C. Skin colour typology and suntanning pathways. **International journal of cosmetic science**, Wiley Online Library, v. 13, n. 4, p. 191–208, 1991. Citado na página 36.

CRENSHAW, K. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. **The University of Chicago Legal Forum**, v. 140, p. 139–167, 1989. Citado na página 27.

DAVIS, A. **Mulheres, raça e classe**. [S.l.]: Boitempo Editorial, 2016. Citado na página 33.

DAWSON, C. W.; WILBY, R. An artificial neural network approach to rainfall-runoff modelling. **Hydrological Sciences Journal**, Taylor & Francis, v. 43, n. 1, p. 47–66, 1998. Citado na página 27.

D'ORAZIO, J.; JARRETT, S.; AMARO-ORTIZ, A.; SCOTT, T. Uv radiation and the skin. **International journal of molecular sciences**, Multidisciplinary Digital Publishing Institute, v. 14, n. 6, p. 12222–12248, 2013. Citado nas páginas 15 e 49.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. **arXiv preprint arXiv:2010.11929**, 2020. Citado na página 77.

FITZPATRICK, T. B. Soleil et peau. **J Med Esthet**, v. 2, p. 33–34, 1975. Citado nas páginas 36 e 48.

FREY, B. J.; DUECK, D. Clustering by passing messages between data points. **Science**, American Association for the Advancement of Science, v. 315, n. 5814, p. 972–976, 2007. ISSN 0036-8075. Disponível em: <https://science.sciencemag.org/content/315/5814/972>. Citado na página 29.

GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais**. [S.l.]: Editora Blucher, 2000. Citado nas páginas 29, 30 e 31.

GORDON, D. F.; DESJARDINS, M. Evaluation and selection of biases in machine learning. **Machine learning**, Springer, v. 20, n. 1-2, p. 5–22, 1995. Citado nas páginas 22 e 26.

GROH, M.; HARRIS, C.; SOENKSEN, L.; LAU, F.; HAN, R.; KIM, A.; KOOCHEK, A.; BADRI, O. Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. **arXiv preprint arXiv:2104.09957**, 2021. Citado na página 50.

GUO, Y.; ZHANG, L.; HU, Y.; HE, X.; GAO, J. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In: SPRINGER. **European Conference on Computer Vision**. [S.l.], 2016. p. 87–102. Citado na página 34.

HARVILLE, M.; BAKER, H.; BHATTI, N.; SUSSTRUNK, S. Consistent image-based measurement and classification of skin color. In: IEEE. **IEEE International Conference on Image Processing 2005**. [S.l.], 2005. v. 2, p. II–374. Citado na página 36.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778. Citado nas páginas 28 e 42.

HE, Y.; SHI, J.; WANG, C.; HUANG, H.; LIU, J.; LI, G.; LIU, R.; WANG, J. Semi-supervised skin detection by network with mutual guidance. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. [S.l.: s.n.], 2019. p. 2111–2120. Citado na página 36.

HECKMAN, J.; ICHIMURA, H.; SMITH, J.; TODD, P. **Characterizing selection bias using experimental data**. [S.l.], 1998. Citado na página 38.

HOOKER, S. Moving beyond "algorithmic bias is a data problem". **Patterns**, Elsevier, v. 2, n. 4, p. 100241, 2021. Citado na página 37.

HOOKER, S.; COURVILLE, A.; CLARK, G.; DAUPHIN, Y.; FROME, A. What do compressed deep neural networks forget? **arXiv preprint arXiv:1911.05248**, 2019. Citado na página 37.

HOOKER, S.; MOOROSI, N.; CLARK, G.; BENGIO, S.; DENTON, E. Characterising bias in compressed models. **arXiv preprint arXiv:2010.03058**, 2020. Citado na página 37.

HUANG, G. B.; MATTAR, M.; BERG, T.; LEARNED-MILLER, E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In: . [S.l.: s.n.], 2008. Citado na página 50.

HUANG, J.; GRETTON, A.; BORGWARDT, K.; SCHÖLKOPF, B.; SMOLA, A. J. Correcting sample selection bias by unlabeled data. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2007. p. 601–608. Citado na página 38.

JIANG, H.; NACHUM, O. Identifying and correcting label bias in machine learning. **arXiv preprint arXiv:1901.04966**, 2019. Citado na página 34.

KANAN, C.; COTTRELL, G. W. Color-to-grayscale: does the method matter in image recognition? **PloS one**, Public Library of Science San Francisco, USA, v. 7, n. 1, p. e29740, 2012. Citado na página 43.

KINYANJUI, N. M.; ODONGA, T.; CINTAS, C.; CODELLA, N. C.; PANDA, R.; SATTIGERI, P.; VARSHNEY, K. R. Estimating skin tone and effects on classification performance in dermatology datasets. **arXiv preprint arXiv:1910.13268**, 2019. Citado nas páginas 30, 43 e 44.

LANG, S. **Fundamentals of differential geometry**. [S.l.]: Springer Science & Business Media, 2012. v. 191. Citado na página 28.

LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**, n, v. 4, n. 4, p. 18–36, 2009. Citado na página 29.

MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. **Journal of machine learning research**, v. 9, n. Nov, p. 2579–2605, 2008. Citado na página 29.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics**. Berkeley, Calif.: University of California Press, 1967. p. 281–297. Disponível em: <https://projecteuclid.org/euclid.bsmsp/1200512992>. Citado na página 29.

MAUGHAN, K.; NEAR, J. P. Towards a measure of individual fairness for deep learning. **arXiv preprint arXiv:2009.13650**, 2020. Citado na página 38.

MEHRABI, N.; MORSTATTER, F.; SAXENA, N.; LERMAN, K.; GALSTYAN, A. A survey on bias and fairness in machine learning. **arXiv preprint arXiv:1908.09635**, 2019. Citado na página 34.

____. A survey on bias and fairness in machine learning. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 6, p. 1–35, 2021. Citado na página 35.

MELLO, R. F.; PONTI, M. A. **Machine Learning: A Practical Approach on the Statistical Learning Theory**. [S.l.]: Springer, 2018. Citado nas páginas 21, 25 e 26.

MERLER, M.; RATHA, N.; FERIS, R. S.; SMITH, J. R. Diversity in faces. **arXiv preprint arXiv:1901.10436**, 2019. Citado nas páginas 22, 27 e 35.

MOROZOV, E.; MARCONDES, C. **Big Tech: a ascensão dos dados e a morte da política**. UBU EDITORA, 2018. ISBN 9788571260122. Disponível em: <https://books.google.com.br/books?id=dHePvwEACAAJ>. Citado nas páginas 22, 33 e 34.

NECH, A.; KEMELMACHER-SHLIZERMAN, I. Level playing field for million scale face recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2017. p. 7044–7053. Citado na página 34.

OLIVEIRA, V. M. *et al.* Revisitando heleieth saffioti: a construção de um conceito de patriarcado. Universidade Federal de São Carlos, 2019. Citado na página 33.

O'NEIL, C. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. New York, NY, USA: Crown Publishing Group, 2016. ISBN 0553418815, 9780553418811. Citado na página 21.

PIZER, S. M.; AMBURN, E. P.; AUSTIN, J. D.; CROMARTIE, R.; GESELOWITZ, A.; GREER, T.; ROMENY, B. ter H.; ZIMMERMAN, J. B.; ZUIDERVELD, K. Adaptive histogram equalization and its variations. **Computer vision, graphics, and image processing**, Elsevier, v. 39, n. 3, p. 355–368, 1987. Citado na página 30.

PONTI, M.; NAZARÉ, T. S.; THUMÉ, G. S. Image quantization as a dimensionality reduction procedure in color and texture feature extraction. **Neurocomputing**, Elsevier, v. 173, p. 385–396, 2016. Citado na página 43.

PONTI, M. A.; RIBEIRO, L. S. F.; NAZARE, T. S.; BUI, T.; COLLOMOSSE, J. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: IEEE. **2017 30th SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)**. [S.l.], 2017. p. 17–41. Citado na página 27.

PONTI, M. A.; SANTOS, F. P. dos; RIBEIRO, L. S.; CAVALLARI, G. B. Training deep networks from zero to hero: avoiding pitfalls and going beyond. In: IEEE. **2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.], 2021. p. 9–16. Citado na página 21.

PREMA, C.; MANIMEGALAI, D. Survey on skin tone detection using color spaces. **International Journal of Applied Information Systems**, Citeseer, v. 2, n. 2, p. 18–26, 2012. Citado nas páginas 30, 36 e 43.

RUBACK, L.; AVILA, S.; CANTERO, L. Vieses no aprendizado de máquina e suas implicações sociais: Um estudo de caso no reconhecimento facial. In: SBC. **Anais do II Workshop sobre as Implicações da Computação na Sociedade**. [S.l.], 2021. p. 90–101. Citado na página 37.

RUSSELL, S. J. **Artificial intelligence a modern approach**. [S.l.]: Pearson Education, Inc., 2010. Citado na página 21.

SALAH, K. B.; OTHMANI, M.; KHERALLAH, M. A novel approach for human skin detection using convolutional neural network. **The Visual Computer**, Springer, p. 1–11, 2021. Citado na página 36.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. Citado nas páginas 22 e 28.

SZELISKI, R. **Computer vision: algorithms and applications**. [S.l.]: Springer Science & Business Media, 2010. Citado na página 31.

TIAN, Q.-C.; COHEN, L. D. A variational-based fusion model for non-uniform illumination image enhancement via contrast optimization and color correction. **Signal Processing**, Elsevier, v. 153, p. 210–220, 2018. Citado na página 36.

TORRES, R.; FALCÃO, A. X. Recuperação de imagens baseada em conteúdo. In: **Workshop de Visão Computacional**. [S.l.: s.n.], 2008. v. 4. Citado na página 29.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: IEEE. **Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001**. [S.l.], 2001. v. 1, p. I–I. Citado na página 43.

WANG, F.; CHEN, L.; LI, C.; HUANG, S.; CHEN, Y.; QIAN, C.; LOY, C. C. The devil of face recognition is in the noise. In: **Proceedings of the European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2018. p. 765–780. Citado na página 34.

WANG, X.; WANG, S.; WANG, J.; SHI, H.; MEI, T. Co-mining: Deep face recognition with noisy labels. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2019. p. 9358–9367. Citado na página 38.

WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. **IEEE signal processing magazine**, IEEE, v. 26, n. 1, p. 98–117, 2009. Citado na página 72.

WECHSLER, H. **Reliable face recognition methods: system design, implementation and evaluation**. [S.l.]: Springer Science & Business Media, 2009. v. 7. Citado na página 21.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 2, n. 1-3, p. 37–52, 1987. Citado na página 29.

WOLF, S. **Sexuality and socialism: History, politics, and theory of LGBT liberation**. [S.l.]: Haymarket Books, 2009. Citado na página 23.

YU, B.; LIU, T.; GONG, M.; DING, C.; TAO, D. Correcting the triplet selection bias for triplet loss. In: **Proceedings of the European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2018. p. 71–87. Citado nas páginas 22 e 37.

ZADROZNY, B. Learning and evaluating classifiers under sample selection bias. In: **Proceedings of the twenty-first international conference on Machine learning**. [S.l.: s.n.], 2004. p. 114. Citado na página 38.


ICMC
USP
SÃO CARLOS