

Parecer InternetLab

Antonio A. Abello
Rafael Will M. de Araújo
R. Hirata Jr.

Março 2021

Sumário

I	Introdução	4
II	Reconhecimento Facial e Aprendizado de Máquina	12
II.1	Fases de reconhecimento facial	13
II.2	Da engenharia de características ao aprendizado profundo	16
II.3	DeepFace	22
II.4	Uso dos modelos e medidas de erros	24
III	Bases de dados para reconhecimento facial	32
III.1	Características físicas	33
III.2	Balanceamento entre classes	35
III.3	FERET	36
III.4	MORPH-II	38
III.5	Labelled Faces in the Wild	38
III.6	VGGFaces2	40
III.7	CelebFaces Attributes	40
III.8	Base de Dados Proprietárias	40
IV	Limitações e principais causas da perda de perfor-	

mance	42
IV.1 Resolução	43
IV.2 Envelhecimento e Idade	46
V Segurança e aspectos éticos relacionados aos sistemas de reconhecimento facial	49
V.1 Aspectos éticos	52
V.2 Efeitos Demográficos	53
VI Vieses raciais nas tecnologias de reconhecimento facial	56
VII Caso de estudo	
Metrô de São Paulo	
LPI - 10014557	60
VII.1O edital e suas justificativas	61
VII.2Infraestrutura e segurança de rede	61
VII.3Infraestrutura do sistema de vigilância	62
VII.4Software de Análise Forense	63
VII.5Outros aspectos pessoais e privacidade	67
VII.6Perguntas e Respostas	69
VII.6.1 Com os dispositivos/tecnologias que o metrô tem em mãos, seria possível, ao menos em teoria, realizar o rastreamento das pessoas que usam o metrô?	69
VII.6.2 No tocante a manifestações cívicas, um sistema como o que será implementado no metrô pode ser utilizado na classificação de pessoas por grupo político?	70

VII.6.3	Quais os principais erros que podem ocorrer no sistema apresentado pelo metrô?	71
VII.6.4	A documentação apresentada pela empresa no que concerne à capacidade técnica é suficiente para demonstrar a mi- tigaç�o dos riscos que o sistema pode apresentar para os usu�rios?	72
VII.6.5	Qual o grau de efici�ncia em rela��o � identifica��o de crian�as e adolescentes?	72
VII.6.6	No projeto, h� prote��es e indica��o de coleta e tratamento diferenciado de dados de crian�as e adolescentes?	73
VII.6.7	O projeto explica qual o banco de dados ser� utilizado?	74
VIII	Conclus�o	75
IX	Anexo I	81

Parte I

Introdução

O avanço tecnológico na área de câmeras digitais, placas especializadas para redes neurais, sistemas de armazenamento, e principalmente, a abundância de imagens e dados pessoais fez com que as tecnologias de monitoramento, vigilância e identificação começassem a ser adotadas maciçamente. Os fabricantes de Gravadores de Vídeo Digitais (DVRs) e os desenvolvedores de sistemas integrados de segurança têm conseguido penetrar suas tecnologias para o cliente mais simples como uma casa particular, passando por lojas de conveniência etc, até para os clientes mais significativos como as escolas, ou condomínios, ou os sistemas de transporte público. Essa penetração é tão intensa que estima-se que o número de câmeras ligadas em algumas cidades do mundo, como Londres, por exemplo, já ultrapasse o número de cidadãos que lá habitam [1].

Ter à disposição tantas câmeras tornou-se uma realidade orwelliana¹, a menos da questão mais importante, não há pessoas suficientes para monitorar tantas câmeras. Por isso, cada vez mais, os fabricantes de DVRs, integradores de sistemas de segurança, ou mesmo firmas de software, tentam automatizar as funções de vigilância que antes eram feitas apenas por seres humanos. Dois dos primeiros sistemas completos a serem automatizados foram o alerta de invasão de território e o controle de acesso por imagens de faces, ambos sob condições limitas de estruturação de ambientes. No caso de controle de acesso, as faces precisavam frontais e fotografadas sob condições específicas de iluminação e distância da câmera. Embora os sistemas tenham evoluído bastante ao longo dos anos, eles não são perfeitos e estão sujeitos a erros. Substituir o ser humano para certas tarefas visuais não é fácil.

O sistema visual do ser humano, assim como nossa capacidade visual para realizar as mais diferentes tarefas visuais é, sem dúvida, impressionante. Várias são as tarefas visuais que fazemos com facilidade, por exemplo, descrição de cenas, detecção e descrição de objetos (animais, aviões, bicicletas, carros, drones, faces,

¹George Orwell, no livro 1984, escreve sobre dispositivos chamados de “telescreens” que vigiam os cidadãos onde eles estiverem.

peças etc), identificação e reconhecimento de caracteres impressos e manuscritos, localização de imagens, rastreamento de objetos em movimento, vigilância, identificação e reconhecimento de faces etc.

A formação do sistema visual do ser humano começa desde quando abrimos os olhos pela primeira vez em nossas vidas. Além disso, uma das nossas primeiras experiências visuais é a visão de uma face, que logo aprendemos a localizar, reconhecer, identificar e até mesmo interpretar as expressões faciais. Por ser tão natural, achamos que essa tarefa seja muito fácil de ser executada por um computador, mas a realidade é que é justamente o contrário. Para um programa de computador, identificar uma face numa imagem que não esteja estruturada, isto é, uma imagem feita especialmente para a identificação, como a foto do passaporte, ou a foto que os sistemas de identificação de faces do telefone, é realmente muito difícil. Os passos para um sistema computacional realizar a identificação de uma face são:

1. Captura da imagem por uma câmera digital posicionada adequadamente e com resolução (tamanho da imagem) suficiente para capturar detalhes da face que tornem possível sua identificação.
2. Eventualmente, em alguns sistemas específicos de reconhecimento de face, algum sistema de detecção que o sistema está sendo atacado, por exemplo, alguém apresentando a foto do dono de um celular para destravá-lo.
3. Filtragem da imagem para avaliação da qualidade, principalmente em sistemas de liberação de catracas, celulares etc.
4. Localização da face na imagem - esta tarefa envolve percorrer a imagem toda, pixel a pixel, ou parte a parte (numa estratégia multirresolução onde cada parte é, por exemplo, um conjunto retangular de pixels) e encontrar as vizinhanças, ou partes que se parecem com uma face humana.

5. Localizadas as possíveis regiões de face, o sistema precisa extrair características da imagem para, eventualmente, orientá-la frontalmente (esse processo é feito através da modelagem 3D da imagem da face e rotação para apresentá-la como se houvesse sido capturada por uma câmera frontal à face) e escalá-la (ajustar ao tamanho requerido pelo sistema).
6. Apenas com os passos acima completados com sucesso, o sistema pode proceder para a classificação, ou ao reconhecimento da face. A classificação pode ser dar de acordo com o sexo, idade, etnia etc. Já o reconhecimento de faces pode servir para dizer a que pessoa a face pertence (identificação), ou a verificação da identidade da pessoa com a face apresentada num documento, por exemplo para controle de fronteiras. No primeiro caso, a face imageada é comparada com um conjunto de imagens faces e devolve a distância/similaridade da face capturada com as faces do conjunto.
7. Avaliação por um ser humano. Com poucas exceções, nenhuma decisão deveria ser feita baseada apenas

A figura 1 apresenta os diversos passos de um sistema de identificação facial (1 a 7 descritos acima). Alguns elementos da figura serão explorados nos próximos capítulos deste texto.

O progresso recente da área de reconhecimento de faces é animador. Por exemplo, de 2014 a 2018, há relatos de uma melhora de cerca de vinte vezes, obtendo-se às vezes, erros tão baixos quanto 0.2%. No entanto, testes reais e aplicações piloto contradizem esses números. Por exemplo, um artigo de 2018, relata que a Associação Americana de Liberdades Cívicas (American Civil Liberties Union) usou o sistema de reconhecimento facial da Amazon, o Amazon Rekognition, e identificou erroneamente 28 membros do Congresso Americano como criminosos [2]. Embora tenha sido um experimento feito de forma errada, pois o limiar de reconhecimento padrão de 80% não tenha sido mudado, ele mostra o

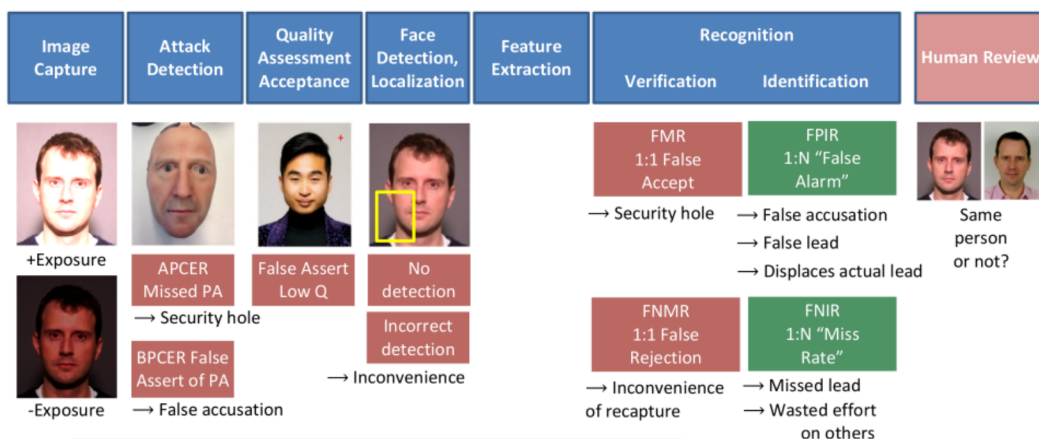


Figura 1: Esquema de um sistema de identificação de faces

Fonte: <http://www.telegraph.co.uk/technology/2016/12/07/robot-passport-checker-rejects-asian-mans-photo-having-eyes/>

quão pernicioso é o uso dessas ferramentas de reconhecimento facial disponíveis atualmente.

Talvez pior sejam as oito tentativas piloto para a implementação de reconhecimento de faces em Londres entre 2016 e 2018 que resultaram em 96% de identificações falsas como suspeitos criminosos [3]. Uma outra reportagem de 2019 [4] baseada num relatório de dois professores da Universidade de Essex [5], relata que 81% dos suspeitos reconhecidos pelo sistema de reconhecimento facial da polícia metropolitana de Londres (London Metropolitan Police) foram incorretamente identificados. Outro caso mal sucedido e que levou ao abandono do projeto piloto foi com a autoridade de transporte metropolitano da cidade de Nova Iorque (New York City's Metropolitan Transportation Authority) depois que a tecnologia teve 100% de erro [6]. Esse projeto era muito mais desafiador pois pretendia o reconhecimento facial de imagens de motoristas através do para-brisas para a identificação de criminosos nas estradas.

Há pelo menos duas razões claras para as diferenças de desempenho que são relatadas pelos artigos acadêmicos e os conseguidos na prática pela indústria na

área de reconhecimento facial, quais sejam: (1) a diferença entre a representação populacional usada para criar os modelos matemáticos e para aplicá-los; e (2) ao fato que diferentes ambientes (acadêmico e industrial) produzem incentivos distintos que levam a políticas distintas quanto ao compartilhamento de dados e de avanços tecnológicos.

Com relação à representatividade populacional, há diversas questões a considerar, mas as mais importantes são:

- gênero ou sexo - feminino, masculino, transgênero;
- cor da pele - amarela, branca, morena, negra;
- ancestralidade - africano, asiático, caucasiano etc.

Além disso, há diversas combinações possíveis dessas características, por exemplo, mulher negra de ancestralidade asiática e africana, homem moreno de ancestralidade caucasiana etc.

Como os modelos matemáticos são feitos baseados em populações majoritariamente brancas americanas, ou europeias, ou majoritariamente amarelas asiáticas, quando os modelos treinados com uma população saem-se muito mal quando aplicados em outra. Por exemplo, os modelos de reconhecimento facial desenvolvidos para asiáticos não funcionam bem, em geral, para as populações brancas europeias, ou americanas e vice-versa. Há também estudos mostrando que a maioria dos sistemas tem desempenho ruim para populações de pele negra, principalmente para o caso de mulheres [7].

Com relação às diferenças de incentivos, no ambiente acadêmico, o objetivo final é a publicação revisada pelos pares, que deve ser pelo menos a princípio reproduzível por avaliadores independentes. Esse fato faz com que os avanços acadêmicos sejam mais publicizados e até pela pressão da comunidade científica sejam mais transparentes quanto a técnicas e dados utilizados. Hoje em dia, é muito comum que os artigos disponibilizem os dados e os programas usados para que a comunidade possa reproduzir os resultados alcançados.

As aplicações industriais, tanto comerciais como estatais, tem como objetivo a elaboração de um produto final que cumpra determinadas métricas pré-estabelecidas, sem necessariamente haver um incentivo para a transparência deste produto além de seu resultado. Não só isso, há incentivos que favorecem uma maior opacidade do produto final, como a competição entre empresas ou o uso desses produtos para coação de atividades ilícitas. Por causa disso, a maior parte da literatura que iremos estudar neste parecer será proveniente do meio acadêmico, o qual cremos não estar significativamente distante do estado-da-arte. No entanto, para avaliar os resultados e a performance de aplicações industriais temos disponíveis também resultados de auditorias e avaliações externas. É o caso, por exemplo, do *Face Recognition Vendor Test* (FRVT) [8], um estudo realizado pelo *National Institute of Standards and Technology* (NIST), órgão ligado ao Departamento de Comércio dos Estados Unidos. Nesse relatório, companhias privadas e públicas poderiam submeter modelos em forma *black-box* (ou seja, num formato que produza resultados mas que não torne possível o destrinchamento de seu funcionamento) para avaliação em conjunto de dados privados. Utilizaremos principalmente o relatório de 2018, que contou com a participação de diversas empresas, como a Microsoft, a chinesa Yitu, Visionlab, entre outras. Utilizaremos principalmente a parte 2 [8], referente à identificação e a parte 3 [7], que estuda efeitos demográficos sobre os modelos.

Por causa de tudo o que foi exposto anteriormente, várias cidades e países têm discutido a validade do ponto de vista ético de usar os métodos de reconhecimento facial, alguns até tratando o assunto como se fosse uma ameaça à sociedade. Há casos de cidades que baniram completamente esse uso pelos atores governamentais pela cidade de San Francisco, CA, EUA [9], ou Oakland, CA, EUA [10] etc. Há também estados que criaram leis específicas sobre a violação de privacidade dos sistemas de reconhecimento facial, por exemplo, o estado de Washington, EUA [11], que permite o uso de reconhecimento facial apenas em casos de urgência, exige que as agências governamentais relatem o uso dessa tec-

nologia por seus agentes e cria uma comissão para o estudo do assunto. Além desse, também o estado do Texas e de Illinois, EUA, também passaram a ter leis semelhantes. Na Europa a situação é ainda mais drástica e encaminha-se para a proibição de sistemas de Inteligência Artificial para quaisquer usos "inaceitáveis". Segundo o documento, o uso de reconhecimento facial será sujeito a requisitos estritos.

No que segue, o capítulo II apresenta resumidamente a história da área de reconhecimento facial com aprendizado de máquina até o atual estado da arte com as redes neurais convolucionais, explicando como a tecnologia funciona e apresentando suas limitações. No capítulo III apresenta um apanhado das bases de dados mais famosas que foram usadas ao longo dos anos para reconhecimento facial. Apresentamos as principais características dessas bases como, por exemplo, como as imagens foram adquiridas, número de imagens, representatividade populacional etc. Apresentamos também onde e para o que as bases de faces são melhor aplicadas. No capítulo IV apresentamos quais as principais razões dos sistemas do estado da arte falharem. Tratamos as principais causas de erro e de onde vêm as limitações dos sistemas. Apresentamos as limitações dos sistemas de aquisição de imagens em ambientes não estruturados, assim como as limitações dos sistemas de imageamento presentes nos sistemas de segurança atuais. No capítulo V apresentamos alguns problemas relacionados à segurança dos dados em sistemas de reconhecimento facial. Apresentamos também os aspectos éticos relacionados aos sistemas de reconhecimento facial, principalmente as questões de vieses étnicos e efeito inibidores nos seres humanos. No capítulo VII apresentamos um estudo de caso do Metrô de São Paulo e, no capítulo VIII apresentamos algumas conclusões deste trabalho.

Parte II

Reconhecimento Facial e Aprendizado de Máquina

Seguindo a classificação de um recente estudo sobre reconhecimento facial [12], pode-se dividir o desenvolvimento de sistemas de reconhecimento facial em quatro períodos. A categorização dessas fases se dá pelo lançamento de três bases de dados de faces que ajudaram esse desenvolvimento e que serão melhor exploradas no capítulo III, que são: a “Face Recognition Technology” (FERET) em 1998 [13], a “Labeled Faces in the Wild” (LFW) [14] em 2008 e a “DeepFace” [15] em 2014.

II.1 Fases de reconhecimento facial

O **primeiro período** (1964-1995) conhecido de reconhecimento facial através de um computador foi devido a Woodrow Bledsoe em 1964 [16], quando ele trabalhava para a Panoramic Research Incorporation. O conjunto de dados usado foram fotografias de faces frontais, ou de lado (“mugshots”) e o algoritmo usava um vetor² de distâncias entre pontos fiduciais faciais. Por exemplo, distância entre os cantos da boca, distância entre o centro das narinas, distância entre o centro dos olhos etc. As ideias e os sistemas desenvolvidos por Bledsoe na Panoramic Research Inc. funcionavam bem, dada as limitações dos conjuntos de dados e das tecnologias usadas e levaram a uma segunda fase de desenvolvimento, que comparou o desempenho de seres humanos contra os métodos propostos por ele e sua equipe. Esse sucesso alavancou o desenvolvimento da área que, em conjunto com o desenvolvimento do hardware (câmeras e computadores), possibilitou a criação de novos algoritmos de processamento de imagens para encontrar as características faciais, ou para encontrar padrões em faces, ou modelar o problema usando Modelos Escondidos de Markov (Hidden Markov Models), ou usar métodos de projeção de faces em uma nova representação (auto-faces) etc. Em conjunto com esse desenvolvimento, várias técnicas de aprendizado de máquina convencional foram desenvolvidos e aplicadas ao problema de reconhecimento facial. Para o sucesso dos métodos automáticos, é fundamental ter imagens e, para

²Um conjunto ordenado de números

isso, é necessário o recrutamento de voluntários que cedessem suas faces para o estudo, a contratação de modelos humanos, ou de fotógrafos, ou pessoas para rotular os dados, inclusive corrigindo e certificando os pontos fiduciais faciais. Além disso, o acesso às câmeras digitais, ou a estúdios preparados com ambiente e iluminação controladas não era simples e tornava o trabalho ainda mais era desafiador.

O **segundo período** (1996-2006) é marcado pelo reconhecimento do governo americano que a imagem de uma face pode ser usada como um atributo biométrico não-invasivo e que as técnicas de reconhecimento facial tinham amadurecido o suficiente para justificar um investimento maior na criação de uma base de dados. Assim, em 1996, após um esforço de três anos e custo de US\$ 6,5 milhões de dólares, o conjunto de dados “Face Recognition Technology” (FERET) foi lançado. Por ser o maior e mais abrangente conjunto de dados, rapidamente ele tornou-se o principal referência para comparação entre algoritmos de reconhecimento facial, assim como em competições, atração de investimentos de pesquisa e interesse acadêmico.

Em termos de algoritmos, dada a quantidade de imagens disponíveis, começou-se a ser possível usar os métodos de Aprendizado de Máquina para Visão Computacional tais como: Máquinas de Vetores de Suporte (SVMs), Modelos Escondidos de Markov (HMMs), Redes Neurais Convolucionais simples (CNNs). Esses métodos obtinham bons resultados, mas eram limitados ao conjunto de dados, isto é, as aplicações em que eram usados restringiam-se a reconhecimento facial em fotografias de faces frontais, ou de lado [12, 17].

Em termos de aplicações comerciais, logo percebeu-se que as limitações eram maiores pois pequenas mudanças nas condições de iluminação, ou marcas faciais, ou adendos como perucas, ou cachecóis, ou um simples óculos, deterioravam muito os resultados [18, 19, 20]. Tecnicamente, a razão da falta de robustez não está apenas nos algoritmos de aprendizado para reconhecimento de faces, mas nos algoritmos de Visão Computacional que falham em localizar e extrair



(a) Face com os pontos fiduciais corretamente localizados.



(b) Face não localizada

Figura 2: Resultados do algoritmo de localização e extração de pontos fiduciais da face. Na figura da esquerda, o algoritmo localizou a face e os pontos fiduciais. Na figura da direita, falhou em fazer isso.

as características corretas. Um exemplo disso é ilustrado nas figuras 2a e 2b, que foram tiradas sob as mesmas condições de iluminação, distância etc. Embora os algoritmos usados sejam modernos, mesmo assim, na figura em que a face foi parcialmente coberta pela mão 2b, o algoritmo de localização (Histogram of Oriented Gradients) [21] e extração dos pontos fiduciais da face (um preditor de pontos fiduciais de faces) [22] falhou, diferentemente do que aconteceu na figura 2a, cujos pontos foram encontrados corretamente. No anexo IX, apresentamos um estudo mais variado de como o algoritmo é afetado pela iluminação, pose da face, distância etc.

O **terceiro período** (2007-2013) é marcado justamente pelos esforços para superar a falta de robustez dos algoritmos. Para isso, os criadores melhoraram um conjunto de dados já existente “Faces in the Wild” [23], que foi criado a partir de fotos do sítio de notícias “Yahoo News”, extraindo-se partes das imagens de reportagens que continham faces em diferentes combinações de pose, diferentes condições de iluminação etc. O conjunto de dados “Labeled Faces in

the Wild” (LFW) melhora esse conjunto eliminando imagens duplicadas, eliminando manualmente imagens que não eram de faces, normalizando os tamanhos e rotulando-as. O conjunto tem 13.233 imagens de 5.749 indivíduos diferentes [24] e foi disponibilizado para a comunidade publicamente.

Em termos de métodos, continuou-se aplicando os mesmos métodos que anteriormente com a esperança que os novos conjuntos de dados ajudassem a melhorar os resultados. Porém, logo viu-se que a maior diversidade não era suficiente para melhorar significativamente a robustez da identificação de faces em aplicações comerciais. Em termos de pesquisa acadêmica, os avanços também foram pequenos. No entanto, os cientistas continuaram trabalhando bastante para melhorar os resultados.

O **quarto período** (2014-até hoje) é marcado pela criação de um modelo neural profundo conhecido como DeepFace [15], desenvolvido pelos pesquisadores da companhia Facebook usando imagens dos perfis dos usuários, até onde sabemos, sem o consentimento consciente desses usuários. O conjunto de dados relatado é de 4,4 milhões de imagens de faces pertencentes a 4.030 pessoas diferentes (de 800 a 1.200 imagens de faces por pessoa). O maior impacto deste avanço é, sem dúvida, o aumento da quantidade de acertos em quase 27% do estado da arte anterior (ou, 97,35% de acurácia).

II.2 Da engenharia de características ao aprendizado profundo

Como visto na seção anterior, os métodos de Visão Computacional para reconhecimento facial³ eram, até o final do terceiro período, baseados em extração de características, escolha das características mais importantes e uso de algum método de aprendizado de máquina. Em 2013, há uma mudança importante de

³E para a maioria das outras aplicações também.

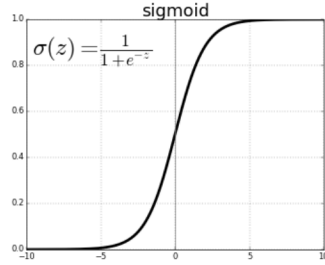


Figura 3: Gráfico de uma função sigmóide

Fonte: <https://ml4a.github.io/images/figures/sigmoid.png>

paradigma com o sucesso de uma rede neural convolucional que ficou conhecida como AlexNet, por causa do nome do seu criador, Alex Krizhevsky [25]. Para contextualizar a importância desta contribuição, vamos apresentar o que é uma rede neural, o que é uma convolução, o que é uma rede neural convolucional e, finalmente, a AlexNet.

Uma rede neural de classificação, ou um classificador, é uma função formada por uma composição de arranjos de somas ponderadas, cada uma delas seguida de uma função de ativação. Em termos matemáticos, no espaço bidimensional, isto é, com duas características para facilitar a notação, se chamarmos as características de X_1 e X_2 , uma rede de apenas um neurônio com uma função de ativação do tipo sigmóide (veja a fig. 3) pode ser expressa pela seguinte fórmula:

$$f(X_1, X_2) = \sigma(w_1 * X_1 + w_2 * X_2), w_1, w_2 \in R, \quad (1)$$

onde $\sigma(z) = \frac{1}{1+e^{-z}}$, $z \in R$.

A função sigmóide é chamada função de ativação pois o resultado dela é o valor 1, ou muito próximo de 1, se o valor da entrada z é maior do que 0, ou 0, ou próximo de 0, caso o valor da entrada z seja menor do que zero. minimizar o erro entre y e $f(X_1, X_2)$. Se o treinamento tiver exemplos suficientes, o erro da função f será pequeno.

A fig. 4 apresenta um diagrama de uma rede neural de apenas um neurônio

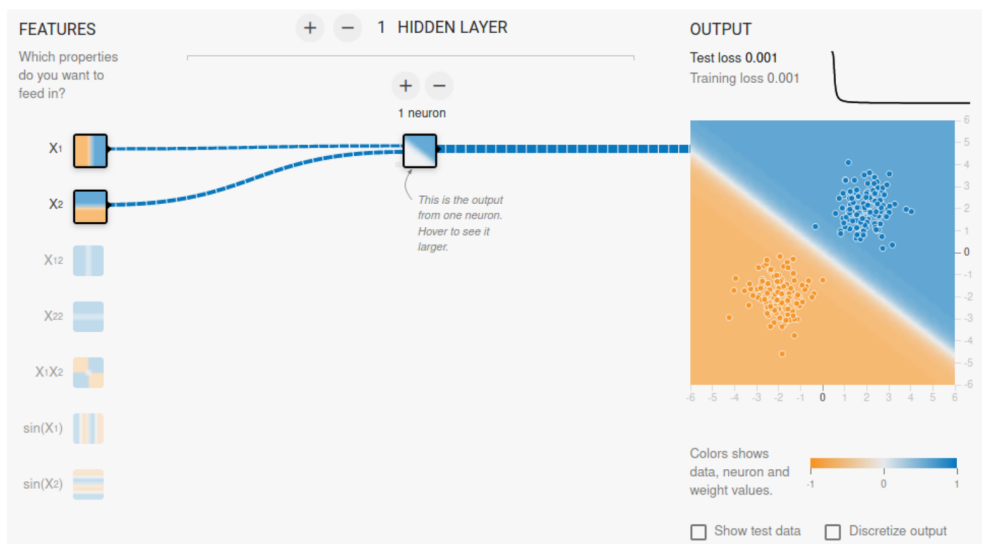


Figura 4: Exemplo de uma rede neural de apenas um neurônio - perceptron.

para classificação de duas classes (pontos azuis e laranjas). A entrada de cada membro da classe, representada por X_1 e X_2 (que poderiam representar, por exemplo, a distância entre os olhos, X_1 , e a distância entre os cantos da boca, X_2), é combinada como acima e, após um tempo de treinamento, a rede é capaz de separar perfeitamente as duas classes, que neste exemplo didático foram criadas como sendo linearmente separadas.

Uma rede de apenas um neurônio é bastante limitada e, por isso, os “neurônios” são combinados em uma rede para melhorar sua expressividade. A fig. 5 apresenta um problema de classificação mais complexo em que as classes não são linearmente separáveis, isto é, não existe uma única reta que separe as duas classes. Para resolver esse problema, precisamos mudar a *arquitetura* da rede, aumentando o número de neurônios em cada camada (colunas) e aumentando o número de camadas. No caso da fig. 5, aumentamos de um o número de neurônios, sem aumentar o número de camadas, e o resultado não é perfeito pois a rede classifica o espaço em três faixas, uma azul e duas laranjas, deixando uma parte das

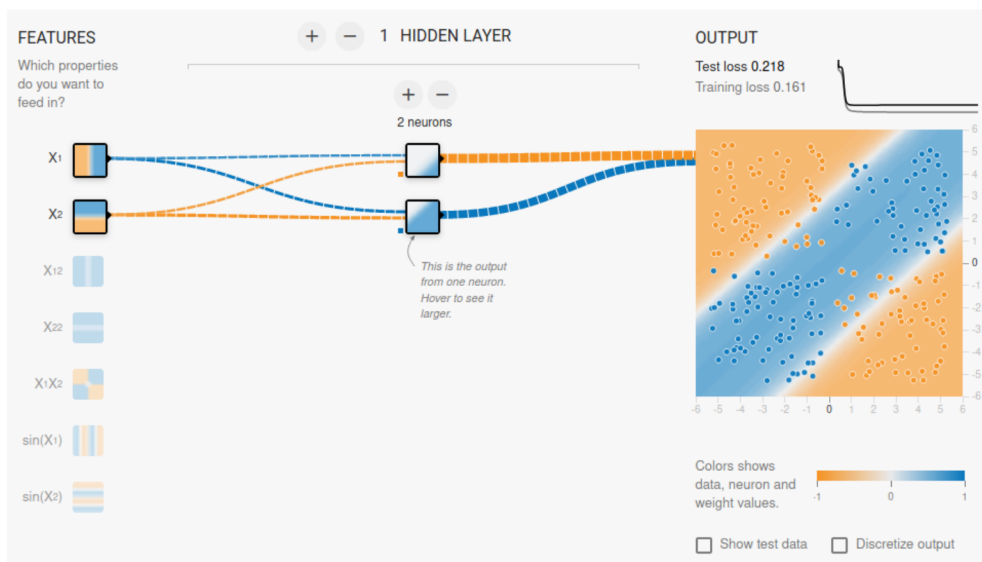


Figura 5: Exemplo de uma rede neural com uma camada de dois neurônios.

amostras laranja classificadas como azul e vice-versa.

A fig. 6 apresenta o mesmo problema da fig. 5, mas a arquitetura é mais representativa, tendo duas camadas e dois neurônios por camada. Note que, agora, o classificador separa corretamente as classes azul e laranja. Esses dois exemplos mostram como uma pequena mudança no número de neurônios afeta a representatividade da rede e pode mudar significativamente os resultados. Ao contrário do que se pode pensar, embora aumente a expressividade, acrescentar camadas e neurônios aumenta o tempo de treinamento e pode aumentar o erro, também.

As redes neurais começaram a ser exploradas computacionalmente na década de 60 e até o final da primeira década deste século, ou seja, após 50 anos de desenvolvimento, seu sucesso ainda era limitado pela quantidade de dados e de recursos computacionais necessários para elas funcionarem bem. Porém, como toda área, após um período longo de maturação, no final da primeira década deste século e início da década passada, certas contribuições começaram a mudar a área.

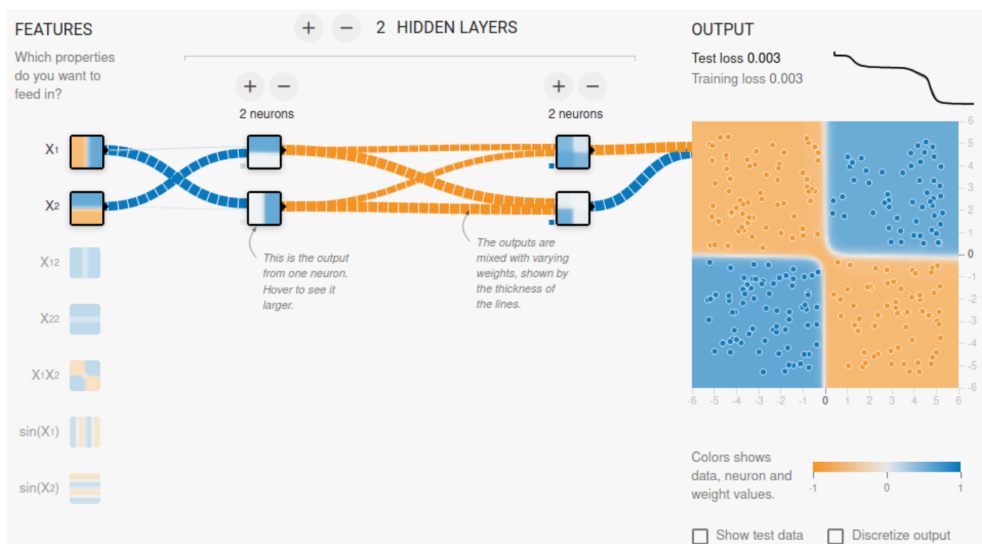


Figura 6: Exemplo de uma rede neural com duas camadas de dois neurônios cada.

A primeira delas foi por causa do trabalho do Dr. Yan LeCun [26]. O trabalho de LeCun apresenta uma rede convolucional cuja entrada é uma imagem e não mais um conjunto de características previamente extraídas da imagem. Essa ideia só foi possível por causa do aumento da capacidade de cálculo dos computadores pois a rede é composta por várias subredes. Cada rede faz o papel de um filtro de imagens que extrairá as características da imagem e, depois, terminará com uma rede neural clássica de classificação. A fig. 7 apresenta a arquitetura da rede LeNet-5

Porém, foi a contribuição do Dr. Alex Krizhevsky [25] baseada no trabalho de LeCun que mudou a área de Aprendizado de Máquina no mundo todo. O trabalho de Krizhevski apresentava uma rede neural de classificação para uma competição conhecida como ImageNet. Nessa competição, algo em torno de dez milhões de imagens de mais de dez mil classes diferentes são usadas para treinar um classificador. Os competidores são desafiados, então, a usar seus classificadores para classificar 150 mil imagens de mais de mil classes diferentes (cachorro, gato

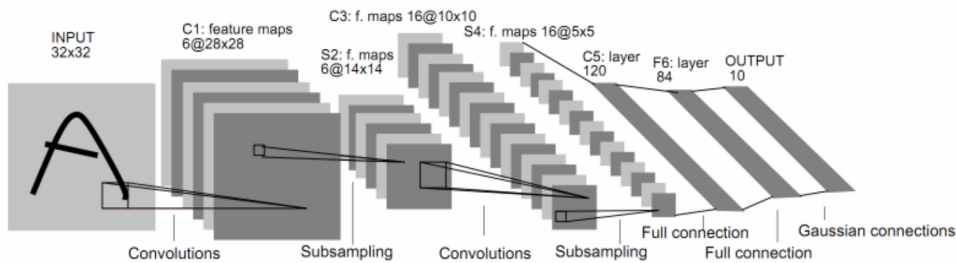


Figura 7: Arquitetura da rede neural convolucional LeNet-5.

Fonte: <https://ml4a.github.io/images/figures/sigmoid.png>

etc).

No que importa para este texto, as principais contribuições do trabalho de Krizhevski foram:

- O uso de aumentação de dados - como as entradas são imagens, elas podem ser espelhadas, rotacionadas, escaladas, transladadas etc, o que aumenta a quantidade de dados disponíveis para treinamento
- O uso de uma nova função de ativação chamada *Unidades Lineares Retificadas* (*Rectified Linear Units*) ou simplesmente, ReLu.
- O uso de placas gráficas para otimizar o tempo de treinamento e aplicação (Graphical Processing Units, ou simplesmente GPUs)

A fig. 8 apresenta a rede neural convolucional conhecida como AlexNet. A entrada da rede é uma imagem Nessa representação diagramática, os neurônios individuais deixam de ser representados individualmente pois são muitos e as camadas passam a serem representados por caixas. Assim, cada bloco de camadas de neurônios é representada por um paralelepípedo cuja largura e altura são os neurônios de cada camada e a profundidade indica o número de camadas. Por exemplo, o segundo bloco tem 48 camadas (profundidade) e a altura e a largura tem 55 neurônios, ou seja, $55 \times 55 = 3.025$ neurônios em cada uma das 48 camadas

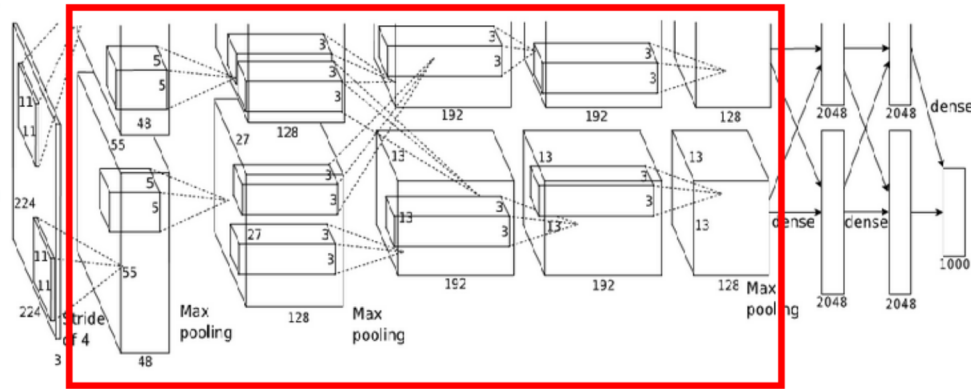


Figura 8: Arquitetura da rede neural convolucional AlexNet. A caixa em vermelho indica a parte da rede que extrai características das imagens.

(note que há dois blocos desses nessa camada de blocos). Não é difícil fazer a conta e descobrir que essa rede tem por volta de 650.000 neurônios e mais de 60 milhões de parâmetros. A caixa vermelha da fig. 8 circunda a parte dos filtros convolucionais da rede e, no final, tem-se os conjuntos de características que são passados para a parte da rede que faz a classificação, no caso, em 1.080 classes.

Como a entrada da rede convolucional é uma imagem, simplificada, a fase de treinamento da rede serve para gerar máscaras para filtros convolucionais de imagens específicos para extrair características que serão usadas nas camadas finais, onde efetivamente é feita a classificação (note as três últimas camadas “densas”) da rede da fig. 8.

II.3 DeepFace

O trabalho de Krizhevsky é seminal na medida que ele trouxe uma mudança de paradigma para a área de Visão Computacional e refocou o trabalho de toda comunidade. Na área de reconhecimento facial não foi diferente e em 2014, na conferência “Computer Vision and Pattern Recognition” (CVPR) que é conside-

rado o encontro mais importante da área de Visão Computacional. No artigo de título provocativo: “DeepFace: Closing the Gap to Human-Level Performance in Face Verification” (DeepFace: eliminando a diferença entre a habilidade humana para verificação de faces) de Yaniv Taigman e coautores [15], é apresentada a primeira rede convolucional para reconhecimento facial cujo desempenho no conjunto de dados “Labeled Faces in the Wild” (LFW) de 97,35% e reduziu o erro do classificador estado da arte em mais de 27%.

Uma das várias razões para esse sucesso, talvez a principal delas, foi o fato de parte dos autores serem funcionários da Facebook Research e, por isso, terem à sua disposição um conjunto de mais de 4 milhões imagens de faces identificadas de mais de 4.030 pessoas diferentes. Esse número representa é uma pequeníssima porcentagem dos usuários da época, que eram mais de um bilhão de usuários, mas eram usuários que tinham em torno de 800 a 1.200 imagens de si mesmos em seus perfis. Esse conjunto de dados ficou conhecido como *Social Face Classification* (SFC).

O processo tem três partes importantes antes de passar para a rede neural. Ele inicia-se pela detecção e localização da face usando seis características fiduciais da face, segue com um processo de modelagem 3D da face detectada (através de um processo de estimação e triangulação das partes da face) e um processo de rotação do modelo 3D para a face alinhar-se frontalmente (veja as duas primeiras imagens da fig. 9, da esquerda para a direita) e escalada para as dimensões 152×152 pixels. Em seguida, a imagem alinhada frontalmente passa pela rede neural que termina com uma representação de 4.096 características que, finalmente, pode ser classificada pela última camada com um dos 4.030 rótulos (identidades) (veja os dois últimos retângulos da fig. 9).

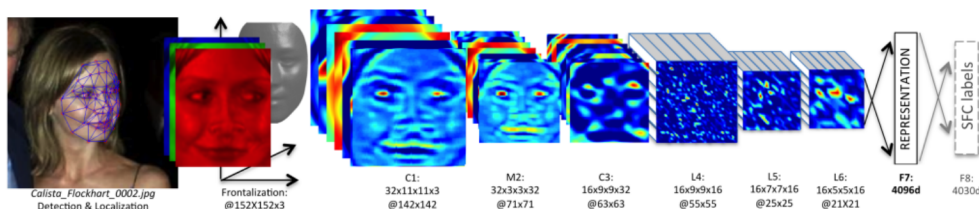


Figura 9: DeepFace: preparação da imagem e arquitetura da rede convolutional.
Fonte: Imagem do artigo[15]

II.4 Uso dos modelos e medidas de erros

Essencialmente, um modelo de reconhecimento facial diz quão próximas, ou distantes, duas imagens de faces humanas estão uma da outra. Munido dessa métrica de proximidade, pode-se realizar diferentes tarefas, as mais comuns sendo verificação e identificação.

A verificação é um problema binário, onde se compara duas imagens dadas, por exemplo, a imagem da foto do passaporte de uma pessoa e a imagem da foto de uma câmera que está apontando para ela no momento da apresentação do passaporte para um dispositivo de liberação de entrada. O resultado é a "simples" decisão de se as imagens correspondem à mesma pessoa, ou não. Se forem, o sistema libera a catraca. Se o sistema disser que não são, deve-se proceder uma verificação por um ser humano.

Em uma tarefa de identificação, o modelo possui uma lista de imagens associadas a identidades conhecidas, também chamada de galeria. A tarefa do algoritmo é, então, analisar uma nova imagem dada e compará-la, calculando-se a distância dela contra todas as faces da galeria. Essa distância é uma medida de semelhança, ou dissimilhança, entre as faces. O modelo, então, produz uma lista ordenada pela distância, em geral da menor para a maior, da face dada a cada uma das faces da galeria. Em geral, esses sistemas possuem um limiar de corte para evitar a apresentação de todas as faces da galeria. As faces que estão

abaixo (ou acima no caso de dissimilhança) do limiar são os possíveis candidatos a serem identificados com a imagem dada.

Como vimos anteriormente, em ambos os casos, verificação, ou identificação, as imagens são transformadas em representações numéricas e são essas representações que são comparadas. A figura 10 apresenta um exemplo desse processo onde o autor ajustou o limiar de forma adequada para haver poucas falhas. A figura apresenta duas imagens diferentes da artista Angelina Jolie e duas imagens diferentes da artista Scarlett Johansson. Cada duas imagens da figura apresenta a comparação da representação de duas imagens diferentes. Seguindo o sentido da leitura, os dois primeiros pares apresentam a distância entre as representações de duas faces de Jolie e duas faces de Johansson, respectivamente. A distância entre as representações foi aproximadamente 0.48 e 0.54, respectivamente. Os outros quatro pares apresentam a distância entre as imagens cruzadas, sempre comparando Jolie com Johansson. Nesse caso, a menor distância foi de aproximadamente 0.59 (segunda imagem da segunda linha) e a maior distância foi de aproximadamente 0.66 (última imagem da primeira coluna).

Há dois tipos de erros que podem ocorrer nos processos de identificação, ou verificação:

- os **falsos positivos**, ou erros de tipo I, ocorrem quando o modelo associa incorretamente dois sujeitos, isto é, atribui a identidade de um sujeito a uma imagem que não lhe corresponde.
- os **falsos negativos**, ou erros de tipo II, ocorrem quando o modelo não reconhece que uma imagem corresponde a determinada identidade.

Nos casos em que o sistema acerta, isto é, associa corretamente dois sujeitos, ou rejeita a associação entre dois sujeitos temos os **verdadeiros positivos** e os **verdadeiros negativos**, respectivamente. A figura 11 apresenta uma tabela com esses tipos de erros e outras medidas que serão explicadas posteriormente.

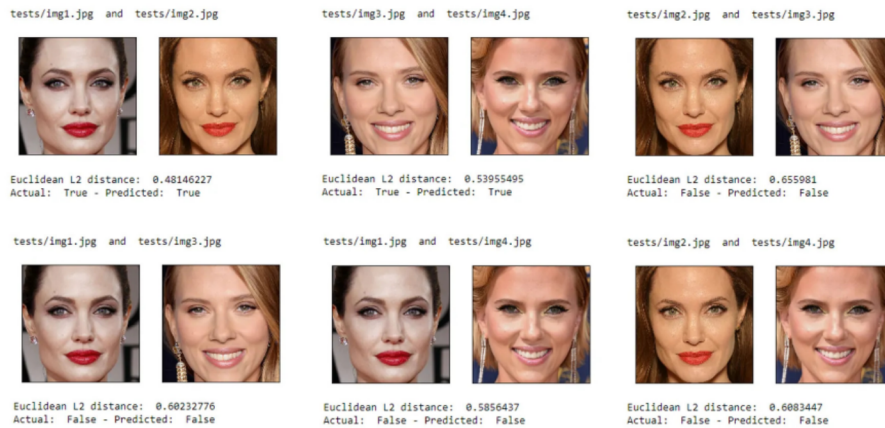


Figura 10: Exemplo do uso do programa DeepFace.

Fonte: <https://sefiks.com/2020/02/17/face-recognition-with-facebook-deepface-in-keras/>

		Verdade			
		População total	Condição positiva	Condição negativa	
Predito	Condição positiva prevista	Verdadeiro positivo	Falso positivo, Erro do tipo I	Prevalência $= \frac{\sum \text{Condição positiva}}{\sum \text{População total}}$ Valor preditivo positivo (PPV), Precisão = $= \frac{\sum \text{Verdadeiro Positivo}}{\sum \text{Condição positiva prevista}}$	Acurácia (ACC) = $= \frac{\sum \text{Verdadeiro positivo} + \sum \text{Verdadeiro negativo}}{\sum \text{População total}}$ Taxa de falsa descoberta (FDR) = $= \frac{\sum \text{Falso positivo}}{\sum \text{Condição positiva prevista}}$
	Condição negativa prevista	Falso negativo, Erro do tipo II	Verdadeiro negativo	Taxa de Falsa Omissão (FOR) = $= \frac{\sum \text{Falso negativo}}{\sum \text{Condição negativa prevista}}$	Valor preditivo negativo (NPV) = $= \frac{\sum \text{Verdadeiro negativo}}{\sum \text{Condição negativa prevista}}$
		Taxa de Verdadeiro Positivo (TPR), Revocação, Sensibilidade, probabilidade de detecção, Potência $= \frac{\sum \text{Verdadeiro positivo}}{\sum \text{Condição positiva}}$	Taxa de Falso Positivo (FPR), Fall-out, probabilidade de alarme falso $= \frac{\sum \text{Falso positivo}}{\sum \text{Condição negativa}}$	Teste da razão de verossimilhança positiva (LR+) = $= \frac{TPR}{FPR}$	Razão de possibilidades de diagnóstico (DOR) = $= \frac{LR+}{LR-}$ F1 score = $= 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}}$
		Taxa de Falso Negativo (FNR), Taxa de perda $= \frac{\sum \text{Falso negativo}}{\sum \text{Condição positiva}}$	Especificidade (SPC), Seletividade, Taxa de Verdadeiro Negativo (TNR) $= \frac{\sum \text{Verdadeiro negativo}}{\sum \text{Condição negativa}}$	Teste da razão de verossimilhança negativa (LR-) = $= \frac{FNR}{TNR}$	

Figura 11: Resultados possíveis para problemas binário e estatísticas relacionadas. Reprodução Wikipedia

Para agregar os resultados de um modelo de verificação sobre uma grande base de dados de teste algumas estatísticas são definidas. As mais comuns são a taxa de verdadeiros positivos, também conhecida como **sensibilidade** (*sensitivity*), ou **revocação** (*recall*), que é a proporção de verdadeiros positivos pela quantidade de eventos positivos na base de dados, a taxa de verdadeiros negativos, também conhecida como especificidade, que é a proporção de verdadeiros negativos sobre a quantidade de eventos negativos e a taxa de falso positivos, que é a proporção de falsos positivos sobre a quantidade de eventos negativos.

Para modelos de identificação, a definição das estatísticas é ligeiramente distinta, pois o modelo produz uma lista com número de candidatos variável. É necessário também realizar tanto buscas que contém um candidato da galeria do modelo, como que não. Para isso, fixa-se uma posição da lista de candidatos e/ou um limiar (*threshold*) de distância entre imagens. A partir disso, se define a taxa de identificações de falsos positivos como a proporção de candidatos acima do limiar em buscas que não contém uma identidade da galeria. A taxa de falsos negativos, por sua vez, é definida como a proporção de vezes que o candidato buscado aparece abaixo do limiar, ou lugar da lista, pelo número de buscas com candidatos da galeria.

Em geral, a relação entre a quantidade de falsos positivos e a quantidade de falsos negativos é inversamente proporcional e pode-se escolher um balanço adequado ajustando-se a sensibilidade do modelo. Desta forma, modelos mais "sensíveis" são mais criteriosos para inferir identidade, portanto deixando de encontrar relações possíveis e consequentemente aumentando a quantidade de falsos negativos. Esses modelos também evitam encontrar relações espúrias e consequentemente há um decréscimo de falsos positivos. Modelos menos sensíveis encontram mais relações adequadas (decrécimo de falsos negativos) ao custo de mais relações espúrias (aumento de falsos positivos).

O fato da quantidade e da natureza dos erros ser ajustável levanta a questão de como se poderia, então, avaliar e comparar diversos modelos de reconhecimento

facial. Para verificação ou identificação é comum ajustar a sensibilidade dos modelos até encontrar uma taxa de falsos positivos pequena, como 1 em 10.000, e comparar então a taxa de falsos negativos resultante. Para tarefas de identificação pode-se ainda desconsiderar o limiar e reportar quantas vezes um candidato é encontrado em determinada posição da lista. Assim, a acurácia de primeira-ordem (rank-1) representa a porcentagem de vezes que o candidato correto estava em primeiro lugar e a de quinta-ordem (rank-5) quantas vezes estava entre os 5 primeiros, assim por diante.

É também comum apresentar o balanço entre os tipos de erros graficamente, utilizando-se de curvas em gráficos. Para isso, calculam-se estatísticas relevantes para cada limiar e se desenha a relação entre essas estatísticas. Para verificação e outros problemas binários, a curva ROC (*Receiver Operating Characteristic*) mostra a taxa de verdadeiros positivos em função da taxa de falsos positivos, como se pode ver na figura 12. Assim, uma curva que compreende mais área indica um modelo melhor, pois pode atingir taxas maiores de verdadeiros positivos a um custo menor de falsos positivos. Por vezes se utiliza a estatística de área debaixo da curva (AUC). Para problemas de identificação, como no caso do NIST [8], se utiliza a curva DEC (*Detection Error Tradeoff*), que mostra a taxa de identificações falso negativas em função da taxa de identificações falso positivas. Neste caso o contrário se aplica, e curvas com menos área são mais desejáveis. Uma explicação de como interpretar uma curva DET é apresentada na figura 13

A acurácia desse tipo de sistema, portanto, é resultado da calibração das ocorrências de falsos positivos e falsos negativos. Isso tem como consequência uma característica intrínseca ao tipo de mecanismo aqui analisado: a impossibilidade de que o sistema opere sem nenhum tipo de falha, tendo em vista que a existência de erros é componente básico de seu funcionamento. Nesse contexto, é relevante ainda destacar que o aumento da acurácia não resolve, por si só, os problemas relacionados à implementação do uso do reconhecimento facial - o que será exposto

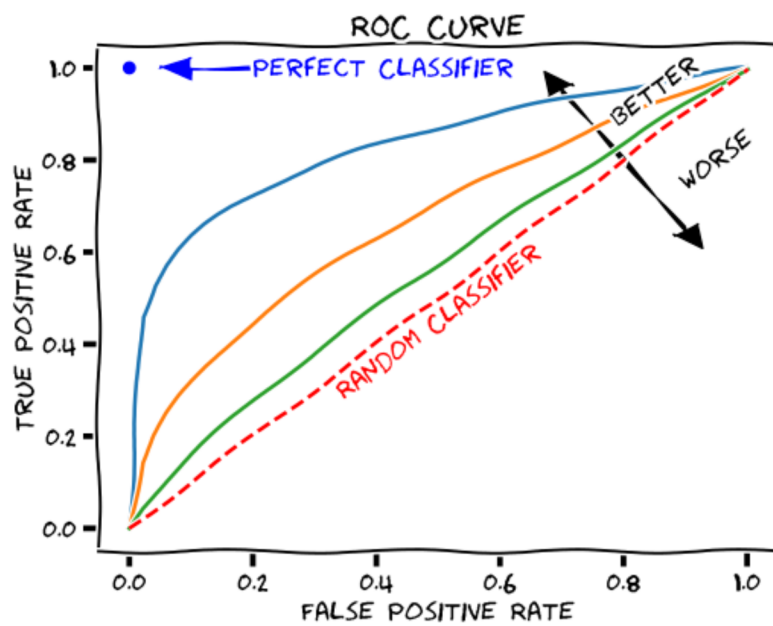


Figura 12: Exemplo de curva ROC. Reprodução Wikipedia

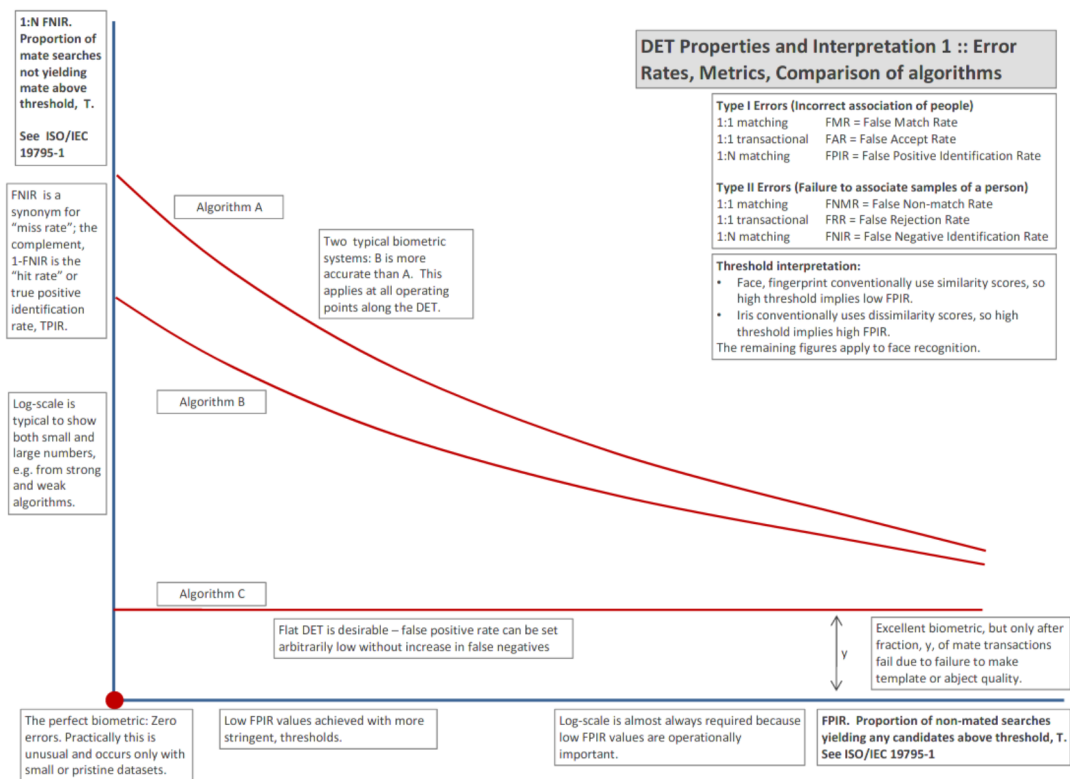


Figura 13: Exemplo de curva DET com interpretação. Reprodução [8]

em outros tópicos deste parecer, como aquele que discorre sobre vieses raciais no uso dessas tecnologias.

Parte III

Bases de dados para reconhecimento facial

Neste capítulo apresentaremos um apanhado das bases de dados que foram usadas ao longo dos anos para reconhecimento facial. Apresentaremos as principais características dessas bases como, por exemplo, como as imagens foram adquiridas, número de imagens, representatividade populacional etc. Apresentaremos também onde e para o que as bases de faces são melhor aplicadas.

III.1 Características físicas

Ao se trabalhar com reconhecimento facial, um dos principais fatores a serem considerados é a distribuição das características dos indivíduos representados nas bases de imagens. Três características básicas que podem ser apontadas em qualquer indivíduo a partir da sua face são: idade (faixa etária), sexo e ancestralidade. A maioria das bases de faces existentes disponibilizam imagens rotuladas, geralmente em um arquivo à parte, com pelo menos uma dessas características para cada indivíduo representado.

O envelhecimento é um processo estocástico inevitável que afeta diretamente a aparência facial. Esse processo envolve não somente a troca de tecido mole, mas também acontece a nível ósseo estrutural. Ou seja, a face de um determinado indivíduo evolui com o tempo e será inevitavelmente diferente, estruturalmente, com o passar dos anos. Isso significa que algum nível de variação baseado em idade deve ser considerado ao se construir sistemas de reconhecimento facial [27]. Além disso, a nível visual, a taxa de envelhecimento de diferentes indivíduos pode variar de maneira drástica, o que mostra o quão desafiador é resolver essa questão ao se construir bases de dados de faces e os sistemas que as utilizam.

O sexo é uma das características biométricas mais comuns que nós humanos utilizamos para reconhecer uma pessoa. Também é uma característica importante que deve ser levada em consideração ao se construir uma base de faces. Uma grande concentração de indivíduos de apenas um sexo pode levar um reco-

nhecedor facial a uma condição de superajuste⁴, na qual ele reconhece com alta precisão indivíduos de uma determinada classe para a qual ele foi treinado, mas gerará muitos erros para as demais classes. Outro fator a ser considerado ao se trabalhar com faces é que vários métodos descritos na literatura empregam uma etapa de pré-processamento que inclui a detecção da face, geralmente uma região retangular [28]. Muitas vezes essa etapa “elimina” indicações importantes como comprimento do cabelo, área do pescoço ou algum tipo de acessório utilizado pelo indivíduo, o que pode afetar o desempenho do sistema, em especial quando um dos objetivos é identificar o sexo dos indivíduos.

Há também uma preocupação mais recente de alguns pesquisadores na área de reconhecimento facial em relação a indivíduos transgêneros. O trabalho de Kumar e demais [29] utilizou uma base de faces de imagens coletadas do *Youtube* de indivíduos que fazem terapia de reposição hormonal. O método proposto leva em consideração que essa terapia induz mudanças significativas nas características visuais da face ao longo do tempo. Em [30] são levantadas algumas questões sociais e recomendações sobre esse tema para desenvolvedores de sistemas de reconhecimento facial, em especial quando o sistema é capaz de fazer o reconhecimento automático do sexo do indivíduo.

A ancestralidade de um indivíduo não diz respeito apenas à cor de sua pele, mas também à forma de certas regiões da face como: olhos, nariz, boca, queixo, etc. Ou seja, a ancestralidade afeta diretamente as características de uma base de imagens de faces. O trabalho de Fu e demais [31] discute os desafios desse parâmetro em sistemas de reconhecimento facial, além de listar vários métodos automáticos para reconhecimento de ancestralidade. Um dos problemas intrínsecos ao se considerar esse tipo de informação é saber quem rotulou essa característica dos indivíduos numa base de imagens. Um problema muito conhecido e discutido nessa área é o efeito da “ancestralidade cruzada” [32], no qual um indivíduo de

⁴O superajuste acontece quando o classificador acerta muito bem nos exemplos de treinamento, mas falha quando aplicado ao mundo real.

determinada ancestralidade possui uma tendência a classificar melhor pessoas da mesma ancestralidade que ele em comparação com pessoas de outras ancestralidades. Por esse motivo, por exemplo, há uma tendência de que sistemas de reconhecimento feitos no ocidente classifiquem melhor pessoas de origem caucasiana, e sistemas de reconhecimento feitos no oriente reconheçam melhor pessoas de origem asiática [32].

III.2 Balanceamento entre classes

Outro ponto que deve ser levado em conta na construção ou escolha de um banco de dados de faces é a questão do balanceamento de classes. Na área de aprendizagem de máquina dizemos que os dados estão desbalanceados quando as classes de dados que deseja-se classificar não são representadas igualmente. Caso esse problema seja negligenciado, classes com poucas amostras contribuirão muito pouco com o modelo computacional aprendido, ao mesmo tempo em que classes com o maior número de amostras tornarão o modelo enviesado [33]. A maioria dos conjuntos de dados com um número razoavelmente grande de imagens (10.000 ou mais) não são balanceados [12]. Embora o desbalanceamento seja um problema, quando a diferença na quantidade de exemplares de uma classe para outra é pequena, seus efeitos podem ser desprezados.

Ao lidar-se com características humanas, em especial com imagens de faces, esse assunto faz-se excepcionalmente sensível, uma vez que é desejável que sistemas de reconhecimento sejam neutros em relação ao sexo, ancestralidade e idade dos indivíduos que eles reconhecem. Isso significa que a diferença na taxa média de acerto durante o reconhecimento de indivíduos de qualquer uma das classes deve ser mínima, idealmente zero.

III.3 FERET

O "Facial Recognition Technology" (FERET) [34] é um conjunto de dados criado em 1993 como parte de um programa homônimo, com colaboração de pesquisadores da Universidade George Mason e do Army Research Laboratory, o laboratório de pesquisa corporativo do Exército dos Estados Unidos. Apesar de ser um conjunto de dados antigo, foi um dos mais utilizados na literatura para comparação de métodos de reconhecimento facial e até os dias atuais serve como padrão para pesquisadores compararem seus resultados. É composto por imagens de aproximadamente 1.000 indivíduos. Essas imagens estão distribuídas em diferentes coleções, contendo imagens frontais, laterais e com diferentes poses. Também são disponibilizados rótulos contendo sexo e ancestralidade dos indivíduos. A figura 14 mostra alguns exemplos da coleção de imagens frontais do conjunto de dados FERET. As oito imagens coloridas nas duas primeiras linhas são imagens sem tratamento e as oito imagens em preto e branco são imagens tratadas de modo que possam ser usadas para classificação. O tratamento consiste em detecção da região da face, centralização, alinhamento vertical e ajuste de tamanho.

Apesar desse conjunto de dados ser usado até os dias atuais como referência para diversos trabalhos acadêmicos, a distribuição das classes não é balanceada. Aproximadamente 7,98% das imagens são de indivíduos africanos, 17,50% asiáticos, 5,42% do oriente médio, 63,25% caucasianos e 5,83% de hispânicos. Quanto ao sexo, a distribuição dentro de cada uma dessas classes é um pouco mais justa: dos indivíduos africanos, 56,41% são mulheres e 43,59% são homens; dos caucasianos 41,42% são mulheres e 58,58% são homens; dos hispânicos 52,63% são mulheres e 47,37% são homens. Indivíduos do oriente médio e asiáticos são os que possuem as piores distribuições de sexo dentro do conjunto de dados : 11,32% de mulheres e 88,68% de homens para o primeiro, enquanto que o último possui 33,33% de mulheres e 66,67% de homens.

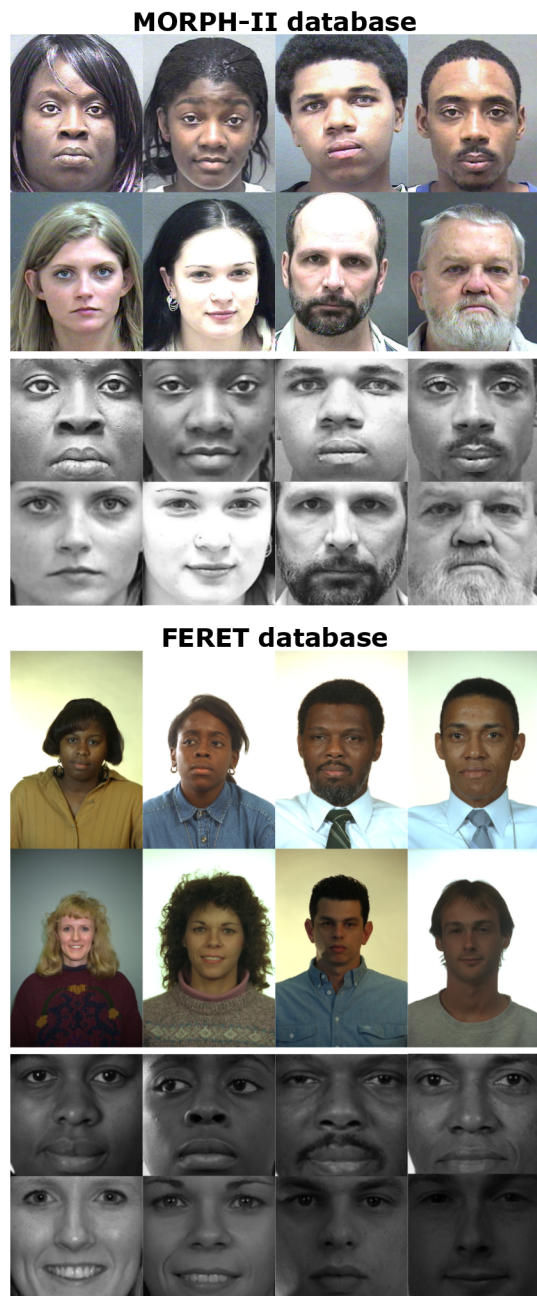


Figura 14: Uma pequena amostra das imagens de faces disponibilizadas pelos conjuntos de imagens MORPH-II e FERET [35]. As imagens coloridas são imagens sem tratamento e as imagens em preto e branco são imagens tratadas.

III.4 MORPH-II

O "MORPH Longitudinal Database" (MORPH-II) [36] é um conjunto de dados de reconhecimento facial composto por imagens obtidas de registros públicos adquiridas ao longo de cinco anos. Ele foi desenvolvido por pesquisadores do Departamento de Computação da Universidade da Carolina do Norte Wilmington (UNCW), e contém cerca de 55.000 imagens de mais de 13.000 indivíduos. A disponibilidade de imagens varia para cada indivíduo, havendo pessoas com somente uma imagem e outras com mais de cinco imagens. Todas as imagens são rotuladas por idade, sexo e ancestralidade. A Figura 14 mostra alguns exemplos de imagens do conjunto de dados MORPH-II. As oito imagens coloridas nas duas primeiras linhas são imagens sem tratamento e as oito imagens em preto e branco são imagens tratadas de modo que possam ser usadas para classificação. O tratamento consiste em detecção da região da face, centralização, alinhamento vertical e ajuste de tamanho.

Embora seja um conjunto de dados razoavelmente grande, o MORPH-II também possui desbalanceamento entre ancestralidades e sexos. Aproximadamente 77,2% dos indivíduos são de origem africana, 19,2% de origem caucasiana, 3,2% hispânicos, enquanto que asiáticos e outras ancestralidades representam menos de 0,5%. Dentro de cada uma dessas classes também há um forte desbalanceamento de sexo, com uma predominância de homens (entre 70% a 94%, a depender da ancestralidade).

III.5 Labelled Faces in the Wild

O conjunto de dados "Labelled Faces in the Wild" (LFW) é um conjunto de dados de reconhecimento facial focado para tarefas de verificação [37]. Ele foi desenvolvido pelo UMass Lab, da Universidade do Massachusetts. É composto de 13.233 imagens faciais de 5.749 identidades, além de uma lista de pares entre



Figura 15: Amostra de imagens de faces do conjunto de dados LFW.

imagens classificados como pertencentes, ou não, a mesma identidade. As imagens vêm de artigos do site Yahoo News, são extraídas a partir de algoritmos de detecção de rostos e, então, tem uma identidade associada baseada na legenda do artigo. A figura 15 mostra alguns exemplos de imagens da LFW.

Embora os autores e curadores do conjunto de dados não incluam a informação da quantidade de exemplos por ancestralidade, idade ou sexo, é reconhecido que vários grupos não estão bem representados, tais como idosos, crianças e bebês, e mulheres. Recentemente o trabalho de [38] rotulou a informação de gênero para todas as faces do conjunto de dados: 22,4% são imagens de mulheres e 77,6% imagens de homens.

III.6 VGGFaces2

VGGFaces2 [39] é um conjunto de dados de imagens faciais fornecido pelo *Visual Geometry Group* (VGG) da Universidade de Oxford. Ele é composto por imagens de celebridades e outras pessoas notórias que podem ter fotos encontradas em mecanismos de busca facilmente. Ele possui 3,31 milhões de imagens ligadas a 9.131 identidades distintas. Há uma preocupação maior sobre a distribuição equitativa de características físicas. Por exemplo, além das imagens também é disponibilizado o gênero atribuído à identidade. Há uma distribuição de 59,7% de homens contra 40,3% de mulheres no conjunto de dados. Os autores também estimaram as idades das fotos presentes, encontrando uma maior prevalência de fotos entre 25 a 34 anos (cerca de 40%). Não há uma estimativa da distribuição de ancestralidade no conjunto de dados, mas os autores recomendam cautela no uso deste, notando que ele pode não representar bem a população humana global.

III.7 CelebFaces Attributes

CelebFaces Attributes (CelebA) [40] é um conjunto de dados de imagens faciais fornecido pelo MMLab, da Universidade Chinesa de Hong Kong. Ele foi coletado por uma metodologia semelhante ao do VGGFaces2 e é composto de 202.599 imagens de rostos. Além da identidade, ele também contém atributos como cor de cabelo, gênero e idade (entendida como um atributo binário: jovem - "*young*"). A proporção de imagens de jovens é de 77,3%, e a proporção de homens é de 41,7%. Não há informação sobre ancestralidade.

III.8 Base de Dados Proprietárias

É comum que grandes companhias produzam conjunto de dados *in-house*, para uso próprio em sua pesquisa e treino de modelos. Esses conjuntos de dados ge-

almente não podem ser compartilhados por questões de segurança, ou questões legais. Por exemplo, os termos de uso de algumas redes sociais permitem o uso de imagens adicionadas à plataforma, mas não a sua distribuição. Como vimos anteriormente, esse é o caso do trabalho dos pesquisadores do Facebook [15]. Isto é uma prática relativamente comum e torna os modelos desenvolvidos dessa maneira um pouco menos transparentes, pois a reprodutibilidade dos experimentos descritos é limitada.

Parte IV

Limitações e principais causas da perda de performance

Neste capítulo apresentaremos quais as principais razões dos sistemas do estado da arte falharem. Trataremos as principais causas de erro e as razões das limitações dos sistemas. Em especial, apresentaremos as limitações dos sistemas de aquisição de imagens em ambientes não estruturados, assim como as limitações dos sistemas de imageamento presentes nos sistemas de segurança atuais.

IV.1 Resolução

Imagens ou pontos de interesse em baixa resolução são a principal causa de perda de performance para sistemas de reconhecimento facial. Para simplificar o texto, vamos estabelecer como baixa resolução não apenas a pouca quantidade de pontos na imagem em relação à região de interesse, mas também consideramos as distorções por embaçamento, ou por não estarem bem focadas. Nesse sentido, a resolução não estará necessariamente ligada às capacidades físicas de um sistema, pois mesmo em sistemas de alta-resolução, quando um ponto de interesse está localizado em uma parte pequena da imagem, ou fora de foco, a imagem resultante pode não ter resolução o suficiente para ser reconhecida acuradamente.

O problema da baixa resolução para sistemas de reconhecimento facial é conhecido pelo menos desde 2009, quando Lui e demais [41] fizeram uma meta-análise sobre as variáveis explicativas da performance desses sistemas. Como resultado, eles descobriram que uma maior resolução, medida como o tamanho da imagem, não produz um aumento de acurácia. Porém, a partir de um certo limiar inferior, uma menor resolução pode causar perdas de até 17,5% em performance, medida por diferentes métricas. Esse resultado não se alterou com o tempo. Em 2012, também focados em reconhecimento facial, Hu e demais [42] encontraram os mesmos problemas e, em 2016, Dai e demais [43], munidos de modelos mais modernos e também testando para uma gama maior de tarefas de visão computacional, registraram perdas de cerca de 3% para uma redução em três vezes e de 7% a 10% para uma redução de quatro vezes de resolução em problemas de re-

conhecimento de dígitos e classificação de cena. Em um estudo recente do nosso grupo [44], obtivemos resultados semelhantes. Para identificação, encontramos uma perda de 6% para uma perda de quatro vezes de resolução, que aumentou para 56% para uma perda de oito vezes de resolução. É notável principalmente a não-linearidade da perda em relação a resolução.

Efeitos parecidos podem ser observados em aplicações industriais. Na última versão do FVRT [8], os modelos foram testados em três conjunto de dados privados de diferentes qualidades: um de fotos de criminosos tiradas consensualmente e com cooperação dos indivíduos fotografados, em pose adequada e por um fotógrafo profissional; um outro de imagens frontais tiradas por uma câmera de baixa qualidade, como uma *Webcam*; e por fim um de fotos *in-the-wild* (não-estruturadas), tiradas por fotojornalistas e fotógrafos amadores, em condições diversas. Os testes mostraram que a performance piora significativamente nos dois últimos, com perdas de 2% a 5% em acurácia de primeiro lugar (rank-1) para identificação e um aumento de até 30% de falsos negativos para se atingir uma taxa de 1 em cada 100 falsos positivos.

Uma ideia natural seria utilizar outros sistemas de visão computacional para melhorar a imagem antes de identificá-la, por exemplo, com sistemas de super-resolução⁵. No caso de sistemas de super-resolução baseados em redes neurais profundas, eles são treinados com imagens de baixa resolução obtidas degradando-se artificialmente imagens de alta-resolução. O objetivo da rede é tentar recuperar a imagem original de alta resolução de forma que a imagem proposta pela rede (super-resolvida) seja o mais parecida possível com a original. Essa ideia foi tentada inúmeras vezes, com resultados levemente positivos. Estudos como [45], [43] ou mesmo um estudo no nosso grupo [44] conseguiram utilizar esses sistemas para atingir melhores resultados em imagens de baixa resolução, mas a melhora não foi o suficiente para sanar as perdas, e o resultado da classificação das imagens

⁵Sistemas de super-resolução aumentam a resolução de uma imagem através de operadores de imagens, ou redes neurais profundas.

super-resolvidas ficou mais próximo do das imagens de baixa resolução do que das imagens de alta resolução em todos os casos.

Essa limitação levou ao desenvolvimento de técnicas de reconhecimento que funcionassem mesmo em baixa resolução, ou que funcionem independente de resolução. No mesmo estudo, Wang e demais [45] modificam iterativamente um modelo para que ele funcione em alta e baixa resolução. Para um problema de classificação de faces, por exemplo, esse modelo consegue uma melhora de 10% de acurácia frente a uma simples aplicação de métodos de super-resolução. Isso leva a uma performance de cerca de 60% de acurácia, ainda distante da acurácia de 78% em imagens de resolução alta. Essas técnicas foram sucessivamente desenvolvidas em trabalhos como os de [46], [47] e [48], cuja minúcia técnica está além do escopo do presente trabalho.

No entanto, em [49], notou-se que a maioria desses estudos sobre reconhecimento em baixa resolução usa um conjunto de dados de alta qualidade artificialmente degradados em vez de dados de baixa qualidade real, o que pode tornar os resultados mais otimistas do que seriam na realidade. Considerando isso, Li e demais [49] agruparam várias técnicas e propuseram algumas novas para serem testadas em conjunto de dados de baixa qualidade "in-the-wild". Em um primeiro experimento, eles mostram que redes treinadas para fazer super-resolução em imagens artificialmente degradadas não superaram modelos mais simples quando utilizados para melhorar a performance de classificadores em situações de baixa resolução real. Em um segundo experimento de particular interesse para nosso caso, eles testaram vários algoritmos do estado-da-arte no conjunto de dados SCface, um conjunto de dados real de vigilância consistindo em retratos de criminosos ("mugshots") em alta qualidade e imagens tiradas por câmeras de vigilância a diversas distâncias (de 1 a 4,7 metros). Um dos modelos mais avançados testado [50], por exemplo, reportava em seu artigo original uma acurácia de até 100% em conjunto de dados de baixa resolução sintética, mas conseguiu somente uma acurácia de primeiro lugar (rank-1) de 12,2% a 25,5% (a

depende da distância) em um conjunto de dados real. O melhor dentre os modelos propostos, dos autores do artigo, conseguiu entre 20,4% a 31,7% de acurácia de primeiro lugar (rank-1).

Além disso, uma crítica recente que se estende também aos trabalhos de super-resolução é que o conceito de baixa qualidade utilizado é simples demais e que para alcançar uma utilidade em casos reais é necessário estudar degradações que vão além da simples projeção de uma imagem com mais pixels para uma imagem com menos pixels [51]. De fato, para estudar melhor este assunto, foi criado o campeonato *UG²* [52] em 2019. O desafio é melhorar a classificação, tanto humana como automatizada, de imagens provenientes de diferentes fontes (satélite, drones e câmeras comuns no chão), possuindo diferentes distorções não sabidas de antemão, tais como baixa resolução, artefatos de compressão, erros de câmera, entre outros. Apesar de várias equipes de laboratórios e universidades de alto-nível participarem da competição, nenhuma conseguiu um sistema que melhorasse a performance dos classificadores uniformemente em todas as condições, muito embora se tenha conseguido melhoras específicas em alguns casos particulares.

IV.2 Envelhecimento e Idade

Como já dito anteriormente, o envelhecimento é um processo natural que produz mudanças significativas nas características faciais. Não é surpreendente então que, além da resolução e qualidade da imagem, a diferença de idade entre as imagens de busca e as imagens de galeria seja mais um outro fator significativo para explicar erros apontado pela última aplicação do FRVT [8]. Os autores utilizam uma base de dados proprietária de fotos frontais de pessoas encarceradas tiradas em intervalos regulares. Agregando os dados em pares pela diferença de tempo entre a primeira e a segunda imagem, reporta-se que há uma tendência constante de piora dos algoritmos de reconhecimento facial conforme essa diferença aumenta, levando a perdas de 3% a 5% de acurácia no pior dos casos, com uma diferença

de 14 a 18 anos.

Best-Rowden e Jain [53] realizaram um estudo parecido, mas com melhoras na metodologia. Eles utilizam um sistema comercial de reconhecimento facial e duas bases de dados privadas provenientes de fontes policiais, uma semelhante com a do FRVT e outra de criminosos reincidentes de um distrito da Flórida. Os autores melhoram a metodologia ao estimar a perda de performance por indivíduo antes de agregar os dados, o que torna possível estimar com mais precisão o efeito do envelhecimento. Com isso, estima-se que a performance do modelo comercial se mantém relativamente estável com uma diferença de cerca de cinco anos e meio a oito anos e meio, com a acurácia caindo até 5% para sete anos e meio a dez anos e meio. Uma outra conclusão a partir dos estudos é que o efeito do envelhecimento não é uniforme por indivíduo, mas pode afetar diferentes pessoas em magnitudes diferentes.

Como esperado, o efeito do envelhecimento é acentuado para crianças e adolescentes pois as mudanças de fisionomia são mais acentuadas. Deb e demais [54] construíram uma base de dados própria de 919 indivíduos dessas idades com fotos frontais tiradas voluntariamente e de maneira recorrente por um período de alguns anos. Eles avaliam o mesmo tipo de sistema comercial de [53] e incluem também um sistema de código aberto (*open-source*) baseado em uma publicação acadêmica, a FaceNet [55], treinada em uma base de dados aberta. Eles testam não apenas a performance de ambos modelos separadamente, como também em conjunto, o que melhora a performance final. Usando análise estatística semelhante ao estudo anterior, eles chegam a resultados mais restritos: a performance só é estável por até dois anos e meio, decrescendo em 17% a partir de três anos. Uma outra descoberta interessante é que o sistema treinado em na base de dados aberta teve um desempenho inicialmente muito pior, sendo necessário adaptá-lo previamente a base de dados dos autores para que pudesse ser útil ao estudo.

Essa dificuldade da maior parte dos sistemas de reconhecimento facial em reconhecer crianças e adolescentes é um fato já conhecido, e se estende também

para adultos idosos. No estudo sobre efeitos demográficos do FRVT [7], pessoas com idades menores (12 a 20) e maiores (mais de 65) foram mais dificilmente distinguidas, consistentemente e independente de gênero ou ancestralidade. Isso pode se dever ao fato já previamente mencionado de que bases de dados de reconhecimento facial são tipicamente compostas por fotos de adultos jovens, devido principalmente a complicações éticas e legais de se obter e disponibilizar fotos de menores. Sawant e Bhurchandi [56] compilam resultados da literatura para concluir também que faces jovens são de mais difícil reconhecimento que faces de adultos.

Esse mesmo artigo compila o histórico de técnicas utilizadas para lidar com envelhecimento em aplicações de reconhecimento facial. As mais recentes e efetivas envolvem o uso de Deep Learning, ou misturando redes profundas com características já previamente conhecidas da literatura como sendo informativas e invariantes à idade [57], ou usando técnicas estatísticas para separar a informação relacionada à idade da informação relacionada à identidade [58]. Ambos os casos conseguem uma modesta melhora em relação ao estado-da-arte em uma base de dados própria e no MORPH-II, respectivamente.

Parte V

Segurança e aspectos éticos relacionados aos sistemas de reconhecimento facial

O projeto de um sistema de reconhecimento facial de grande escala deve incorporar elementos de segurança compatíveis com a sensibilidade das informações coletadas, processadas e armazenadas. Dentre os itens que devem ser levados em consideração, listamos:

- Acesso ao sistema:
 - Credenciais para acessar a sala de controle/monitoramento;
 - Hierarquia e papéis dos usuários;
 - Política de criação e armazenamento de senhas;
 - Possibilidade do sistema permitir acesso remoto.
- Armazenamento dos dados (imagens) coletados:
 - Forma de armazenamento em banco de dados de imagens ou armazenamento direto em arquivos;
 - Como se dará o acesso aos dados;
 - Possibilidade de uso de criptografia para o banco de dados usado ou para o sistema de arquivos.

Qualquer sistema computacional que armazena dados sensíveis deve prover uma camada mínima de autenticação. Em geral, o processo de autenticação é executado através de um *login*, o meio pelo qual um usuário diz quem ele é, e uma *senha*, a prova de que o usuário é quem diz ser. Esse procedimento é amplamente conhecido pelos desenvolvedores de sistemas, mas nem sempre é bem implementado pelo desconhecimento de boas práticas e entendimento sobre o funcionamento de algoritmos criptográficos [59]. Há uma série de fatores a serem considerados durante a construção de um sistema, desde a sua arquite-

tura, padrões de projetos, APIs⁶, linguagens de programação, e também rotinas criptográficas. Vários estudos apontam que boa parte dos desenvolvedores não possuem conhecimento básico sobre a utilização correta e segura de algoritmos de criptografia [59, 60, 61]. Dentre os problemas mais comuns, estão:

- Uso de algoritmos e funções *hash* inseguros/obsoletos (DES, RC4, MD4, MD5, SHA1, por exemplo):
 - Esses algoritmos continuam disponíveis na maioria das linguagens de programação para permitir o funcionamento dos sistemas legados. Há também vários tutoriais e exemplos disponíveis na Internet que mostram o funcionamento desses algoritmos sem deixar explícito que são obsoletos.
- Uso de algoritmos no modo de operação ECB (*Electronic Codebook*) no lugar de um modo mais seguro como o CBC (*Cipher-Block Chaining*):
 - No modo ECB dois blocos de dados idênticos serão criptografados em blocos de texto cifrado idênticos. Já no modo CBC, mesmo blocos idênticos serão criptografados de maneiras distintas.
- Uso de geradores de números pseudo-aleatórios criptograficamente não-seguros:
 - Essa foi uma das falhas apontadas no trabalho de [62] sobre a urna eletrônica brasileira, revelando que até mesmo técnicos do Tribunal Superior Eleitoral desconheciam algumas dessas boas práticas. A falha permitia revelar o voto de uma pessoa em uma dada seção eleitoral, caso o atacante soubesse qual a ordem daquela pessoa na fila de votação.

⁶Do inglês *Application Programming Interface*: Interface de Programação de Aplicações, um conjunto de rotinas e padrões estabelecidos por um sistema para a utilização das suas funcionalidades por outros aplicativos e sistemas.

- Uso de senhas "codificadas" (armazenadas) no código fonte:
 - Uma falha semelhante causou o vazamento de dados de 243 milhões de brasileiros em 2021, em um sistema de notificações de COVID-19 do Ministério da Saúde [63].
- Uso de chaves criptográficas estáticas/constantes;
- Uso de chaves criptográficas curtas;

Além do mecanismo de autenticação, espera-se que sistemas de reconhecimento facial com foco em monitoramento possuam papéis bem definidos de usuários, cada um com níveis diferentes de acesso. Um determinado usuário deve conseguir acessar somente as funcionalidades do sistema que possuam relação com o seu trabalho. Um gerente, por exemplo, deve ser capaz de criar contas para novos usuários. Já um agente de segurança deve conseguir acessar as imagens das câmeras ao vivo, e a depender do seu nível de acesso, acessar imagens de alta resolução e mesmo imagens armazenadas que foram capturadas no passado. O acesso físico ao local do monitoramento e sala de servidores também deve seguir políticas bem definidas e normas rígidas. Dada a sensibilidade dos dados coletados, apenas pessoas credenciadas devem ter acesso a cada um desses locais.

V.1 Aspectos éticos

Alguns anos atrás tornou-se público que o trabalho de Mahalingam e demais [64] sobre reconhecimento facial em imagens de transgêneros utilizou um conjunto de dados com imagens obtidas sem o consentimento dos indivíduos [65]. A ideia principal dos autores era analisar e entender como a terapia de reposição hormonal (TRH) afeta as características faciais e quais áreas da face são mais resilientes a essas mudanças, com o objetivo de desenvolver algoritmos de reconhecimento mais robustos. Para isso, eles criaram um conjunto de dados próprio contendo mais de

1,2 milhão de imagens obtidas de vídeos públicos na plataforma Youtube, de 38 indivíduos diferentes que faziam TRH. As imagens cobrem períodos de sete meses a três anos, dependendo do indivíduo, e o conjunto de dados foi disponibilizado para fins acadêmicos, tendo sido utilizado em outros trabalhos [29].

O trabalho de Raji e Fried [12] aponta outros casos mais recentes de violação de privacidade na construção de conjunto de dados de imagens faciais. Um dos mais notórios foi a base de faces LFW+ [66], na qual seus criadores pretendiam melhorar a distribuição etária do conjunto de dados LFW [37], uma vez que o número de pessoas jovens (entre 0 e 20 anos de idade) disponíveis era de 209, o que representa somente 3,6% dos indivíduos representados. Para isso, eles desenvolveram um algoritmo para coletar em torno de 5.000 imagens a partir do Google Images utilizando termos como: "*baby*", "*kid*" e "*teenager*" (bebê, criança e adolescente, respectivamente). A criação do conjunto de dados envolveu também a contratação de alguns colaboradores através da ferramenta MTurk⁷ para filtragem e rotulação das imagens obtidas, os quais deveriam averiguar a qualidade e determinar a ancestralidade e sexo de cada indivíduo representado através de um esquema de votação majoritária. Ao todo, 2.466 imagens foram aprovadas pelos colaboradores, ou seja, são imagens de indivíduos entre 0 e 20 anos de fato, com boa resolução. Em nenhum momento os autores da LFW+ falam se obtiveram consentimento dos pais para participação dos indivíduos menores de idade no conjunto de dados, mas pela natureza na qual essas imagens foram adquiridas muito provavelmente essa preocupação foi ignorada [12].

V.2 Efeitos Demográficos

Entendemos aqui por efeitos demográficos qualquer tipo de variação de performance que possa ser atribuída a fatores demográficos como gênero, ancestralidade,

⁷Amazon Mechanical Turk (MTurk): uma ferramenta de colaboração coletiva onde os participantes são pagos para resolver uma tarefa específica.

nacionalidade, idade entre outros. Para sistemas de reconhecimento facial isso tipicamente significa que o sistema teria menos capacidade de reconhecer, ou maior probabilidade de reconhecer erroneamente indivíduos de uma certa demografia. Isto gera problemas éticos para o funcionamento do sistema de reconhecimento facial ao tornar a distribuição dos riscos e danos do sistema desigual, enquanto suas vantagens são distribuídas uniformemente. Dado que, como veremos, esses tipos de vieses provém majoritariamente da falta de representação de grupos minoritários nas bases de dados utilizadas no treino de modelos, esses efeitos podem fazer com que o sistema perpetue preconceitos e opressões estruturais prévias a ele.

É um fato amplamente documentado que modelos de visão computacional podem não só aprender como amplificar vieses contidos em suas bases de dados de origem. Wang e demais [67] encontram uma distribuição de gênero assimétrica em bases de dados de reconhecimento de uma ação, por exemplo, praticando esportes. Eles mostram que há desproporcionalmente mais imagens de mulheres cozinhando do que homens, por exemplo. Em bases de dados de classificação de múltiplos objetos, isso também acontece. Por exemplo, homens são mais mostrados juntos de raquetes de tênis, e mulheres são mais mostradas juntas de utensílios culinários. Além disso, os autores mostram como os modelos treinados nestas bases de dados amplificam esses vieses ao serem usados quando utilizados em novas imagens, prevendo erroneamente o gênero do agente dependendo da atividade realizada ou do ambiente em que se encontra.

Por vezes os autores conseguem desenvolver técnicas que impeçam que modelos não aprendam os vieses contidos nas suas bases de dados em tarefas específicas. Zhao e demais [67] desenvolvem uma penalização⁸ para garantir que o modelo não associe a previsão de gênero com a de atividade sendo realizada. Já [68] consegue treinar classificadores de gênero sem viés de idade utilizando uma base de

⁸Chama-se penalização ao método de dar um peso, ou um valor negativo (ou positivo) quando se está maximizando-se (ou minimizando-se) uma medida de erro.

dados de celebridades, onde as mulheres tendem a ser muito mais jovens do que os homens. [69].

Parte VI

Vieses raciais nas tecnologias de reconhecimento facial

Vieses também podem afetar a performance de sistemas de reconhecimento de maneira desigual. E, quando se trata de sistemas que serão aplicados em contextos marcados por desigualdades étnico-raciais estruturais, como é o caso do Brasil, as potencialidades de discriminação racial são fatores importantes a serem levados em consideração. Nesse contexto, este tópico aponta a literatura acadêmica que se debruça sobre o tema e que, por vezes, relata também a identificação de vieses de gênero. Buolamwini e Gebru [68] utilizam uma escala fenotípica de tons de pele para avaliar a distribuição de diversas bases de dados de reconhecimento de gênero e a performance de classificadores de gênero comerciais. Os autores reportam que as bases de dados tendem a ser mais populadas por homens, e por indivíduos de tons mais claros de pele. Isso se reflete diretamente na performance dos classificadores: homens tendem a ser mais corretamente classificados que mulheres, pessoas de tom de pele claro tendem a ser mais corretamente classificadas que pessoas de tom de pele escuro e mulheres de tons de pele escuro são de forma unânime o grupo mais incorretamente classificado. Vangara e demsis [69] testam modelos abertos e comerciais de reconhecimento facial na base de dados do MORPH-II, estudando a diferença de reconhecimento entre ancestralidade. Os autores descobrem que afro-americanos têm maior probabilidade de serem confundidos entre si do que outros grupos. Eles também concluem que grande parte desse efeito é reduzido quando controlado pela qualidade da imagem, uma vez que as imagens de afro-americanos são mais propensas a terem defeitos de iluminação. Um experimento parecido foi feito na parte 3 do FRVT, dedicado a estudar efeitos demográficos dos sistemas comerciais de reconhecimento facial [7]. Como parte do governo estadunidense, os pesquisadores tiveram acesso a bases de dados de aplicações de imigração e entrada no país, contendo então fotos de diversos indivíduos de várias partes do mundo. Com isso, eles conseguiram criar uma base de dados de teste agrupada por grupos geográficos (tal como leste europeu, África ocidental, entre outros) contendo um número razoável de exemplos para cada grupo. Para estudar os efeitos demográficos contidos nos sistemas inscritos

eles então compararam a probabilidade de um indivíduo ser confundido com outro indivíduo do mesmo grupo (ou, em termos técnicos, a quantidade de falsos positivos) para cada uma das regiões geográficas, bem como compararam com a probabilidade de um indivíduo qualquer ser confundido com outro qualquer, a taxa nominal de erro. Essa probabilidade é estimada simplesmente produzindo pares aleatórios de fotos de indivíduos do mesmo grupo e verificando quantas vezes o sistema erroneamente diz que se tratam da mesma pessoa. Na maior parte das vezes, os modelos de reconhecimento facial mantiveram o erro nominal para os grupos europeus e tiveram um erro maior para os grupos do leste e oeste da África e para a Ásia Oriental. Esse aumento por vezes é bem agudo, chegando entre uma e duas ordens de grandeza. Também se observa um efeito semelhante ao descrito por Phillips e demais [32] como “efeito de ancestralidade cruzada”, em que os sistemas de reconhecimento provenientes da China conseguem manter um erro nominal para diferenciar indivíduos chineses de outros indivíduos da Ásia oriental. Esse efeito não se mantém para os sistemas provenientes da Índia, no entanto. Isso significa que, nesses sistemas, a probabilidade de não-europeus serem confundidos com outras pessoas é bem maior do que para europeus. Os pesquisadores do FRVT também repetem esse experimento para uma base de dados de fotos de indivíduos encarcerados, desta vez todos de nacionalidade americanas e divididos por ancestralidade e gênero. Usando a mesma metodologia anterior eles chegam a resultados semelhantes de [68]: o erro é maior entre mulheres e especialmente maior entre mulheres pretas e asiáticas. Diferentemente do artigo previamente citado, no entanto, o erro agregado por ancestralidade é só ligeiramente maior para pessoas pretas ou asiáticas, mas é muito maior para os de origem indígena. A maioria dos artigos e estudos que estudam o tema de efeitos demográficos propõe soluções para o problema, mas elas tem limitações. Para poder avaliar melhor o diferencial de erro por tom de pele, Buolamwini e Gebru [68] constroem uma base de dados com distribuição equitativa entre fenótipos de tom de pele. Alvi e demais [70] notam que a maior parte das base de dados

disponíveis provêm de celebridades e pessoas públicas notórias, portanto sendo viesado para este universo. Como alternativa, eles produziram a "Labeled Ancestral Origins Faces in the Wild" (LAOFIW), que contém imagens de indivíduos de maneira balanceada e anotada segundo uma divisão de origem ancestral. No entanto, essas bases de dados precisam de curadoria muito mais intensa e focada, ao contrário das bases maiores que podem ser feitas automaticamente por robôs buscadores (crawlers) na Internet. Isso as torna úteis para testes, mas não o suficiente para treinar modelos inteiramente. Ao se ajustar a sensibilidade dos algoritmos de reconhecimento facial para cada demografia consegue-se diminuir consideravelmente a diferença de performance entre as diferentes ancestralidades. Essas técnicas têm em comum o fato de serem específicas para cada problema, serem novas na literatura e, por vezes, acarretar em uma perda de performance quando esta é medida da forma tradicional. Portanto, é legítimo crer que, a não ser que especificado o contrário, sistemas comerciais atuais de reconhecimento facial não as implementam por padrão. Assim, estudos sobre o tema divulgados até agora apontam as tendências existentes de que a acurácia dos sistemas possui falhas nas quais vieses raciais são notáveis, fato que se torna mais preocupante quando se leva em consideração fatores que ainda precisam ser superados na sociedade brasileira, como a discriminação racial. Nesse contexto, é relevante destacar que a popularização dos usos de reconhecimento facial na segurança pública traz especial preocupação em relação a questões raciais, diante da reprodução de racismo estrutural identificada por especialistas nas políticas públicas de segurança no Brasil. Isso porque há o risco de que o uso dessas tecnologias reproduza discriminações de natureza racial já existentes nos âmbitos nos quais elas são implementadas e, nesse contexto, a melhora na acurácia do sistema não resolve os problemas identificados, mas sim pode facilitar a perseguição de determinados grupos, uma vez que a implementação da tecnologia no mundo concreto depende também de escolhas de quem é responsável por gerir as políticas públicas.

Parte VII

Caso de estudo Metrô de São Paulo LPI - 10014557

Neste capítulo examinamos os documentos do edital de Licitação Pública Internacional (LPI) de número 10014557 - "Concepção do Sistema de Monitoração Eletrônica (SME) - Etapa 3 das linhas 1-Azul, 2-Verde e 3-Vermelha da Companhia do Metropolitano de São Paulo" à luz do que foi exposto anteriormente. A estrutura deste capítulo começará com uma apresentação dos pontos importantes do edital de licitação e terminará com um formato de perguntas e respostas, fundamentadas pelas partes anteriores deste parecer.

VII.1 O edital e suas justificativas

Segundo o edital LPI de número 10014557 e segundo Petição Metrô [70], o Metrô já possui um sistema de monitoramento eletrônico em diversas estações e pátios de estacionamento que precisa ser modernizado, pois tem muitas câmeras analógicas, e integrado, pois é composto de várias plataformas de software diferentes.

O edital é bastante detalhado, principalmente em algumas partes mais ligadas a infra-estrutura física de rede para suporte para as câmeras. Também é bastante detalhado no que tange às exigências da quantidade e tipos de câmeras e apresenta até algum detalhe sobre o posicionamento de algumas câmeras.

No entanto, há pouco detalhe com relação ao software de gerenciamento SecurOS [71], especificando apenas que esse é o software que deverá ser utilizado (atualizado) e para o qual novas licenças deverão ser compradas.

VII.2 Infraestrutura e segurança de rede

O edital detalha com bastante cuidado como deve ser a infraestrutura física e lógica da rede de comunicação, e quais os protocolos que o sistema devem satisfazer. Por exemplo, o documento IC-9.83.ME.XX/7XX-001 apresenta diversos protocolos que devem ser obedecidos.

Por exemplo, o protocolo 802.1x é um protocolo de camada dois da rede

TCP/IP que pede autenticação de qualquer novo dispositivo inserido na rede e, dependendo da resposta, pode impedir o tráfego de todos os pacotes para aquela porta. Isso é refletido na pág. 80 do documento, itens 6.7.12.1 e 6.7.12.2, que estabelecem que qualquer dispositivo, ou usuário precisam ser autenticados antes de terem acesso aos sistemas. No item 6.7.12.3, complementa-se a informação de que protocolo pode ser usado e até sugerem um protocolo de camada de aplicação para isso.

Todo esse cuidado não elimina a possibilidade de invasão, ou roubo de dados, mas dificulta bastante. Uma coisa que o edital não menciona é como será a conexão do sistema SecurOS com a Internet. Segundo documento [], o sistema do Metrô não estará ligado à Internet, o que é estranho pois sistemas como o SecurOS precisam ser atualizados etc.

VII.3 Infraestrutura do sistema de vigilância

O edital também detalha os servidores de armazenamento onde serão armazenadas as imagens. Esse detalhamento refere-se à quantidade de espaço necessária para armazenar até 30 dias de imagens. No que interessa para este parecer, do referido edital e das respostas do Metrô a questionamentos externos se depreende que o sistema de vigilância armazenará somente as imagens sem tratamento, da forma como foram adquiridas, e sem fazer a priori reconhecimento ou atribuição de identidade. Após cerca de 30 dias, as imagens serão apagadas por mera sobreposição de imagens novas, isto é, como o sistema tem capacidade de armazenamento de aproximadamente 30 dias (esse número depende de vários fatores), após esse tempo, as imagens mais novas irão substituir as mais antigas. No entanto, isso não quer dizer que algum supervisor de segurança não possa armazenar sequências de determinadas câmeras arbitrariamente, muito pelo contrário, o sistema tem capacidade de separar qualquer sequência que esteja sendo capturada e armazenar em outro sistema que esteja interligado ao SecurOS, que não

os servidores usuais.

De princípio, a identificação, reconhecimento ou outras análises podem ser feitas ao mesmo tempo que são capturadas, ou a posteriori, pelos módulos (programas agregados com funcionalidades extras) do SecurOS, ou outro software de análise forense. Em qualquer um dos casos, outros bancos de dados de informações podem ser usados para auxiliar na tarefa desejada. Por exemplo, o software pode consultar uma galeria de identidades para fazer o reconhecimento de faces.

VII.4 Software de Análise Forense

O documento CS-9.83.ME.XX/7XX-001 detalha como deve ser a implantação de um "software de análise forense"⁹ a ser utilizado pelo Metrô, que permita o reconhecimento e busca de objetos e pessoas por diversos critérios, tais como cor, vestimenta e reconhecimento facial. Em resposta a questionamentos do IDEC (Instituto de Defesa do Consumidor) e ao processo 1006616-14.2020.8.26.0053, se esclarece que o objetivo primário é a detecção de objetos abandonados e/ou suspeitos, violação de áreas de risco/proibidas, identificação de casos de violência física e vandalismo, entre outros, sendo o reconhecimento facial tal como descrito neste parecer uma atividade secundária e esporádica. A priori, se limita o reconhecimento a um suposto "banco de dados local do Metrô", de composição não especificada. No entanto, se aventa a possibilidade de colaboração com sistemas externos de órgãos de segurança pública utilizando de suas bases de dados de "contraventores, criminosos, procurados, desaparecidos, entre outros". Essa possibilidade é um dos requisitos documentados no documento de "Concepção de Sistema" referido. Assim, apesar do reconhecimento facial não ser o principal objetivo atual do Metrô, não pode-se descartar que há uma expectativa de um

⁹Usamos a notação do edital, que deve ter sido herdada da documentação do SecurOS, isto é, *forensic = retrospective*

uso futuro.

No edital não são especificadas características da operação do software de análise forense além do fato de que ele deve ser compatível com o software de vigilância SecurOS, que é desenvolvido pela empresa russa ISS (Intelligent Security Systems) ¹⁰. A empresa comercializa atualmente módulos de reconhecimento de faces ¹¹ e o software já provê um subsistema de visão computacional ¹². Ambos sistemas têm capacidade para detecção de diversos tipos de eventos. Por exemplo, o subsistema de visão computacional provê [72]:

- Detecção de uma pessoa correndo;
- Detecção de objetos largados, ou removidos;
- Detecção de aglomerações;
- Detecção de intrusão;
- Contagem de objetos/pessoas;
- Detecção de cruzamento de limites;
- Detecção de pessoas paradas a mais de um certo tempo em algum local;
- Detecção de movimento em uma direção errada/proibida.

No edital também não são especificados se os módulos mencionados acima são o software a serem utilizados, ou se outros serão escolhidos, e essa informação não consta em nenhum documento a que tivemos acesso. Embora não tenhamos acesso a um sistema SecurOS implantado, não é comum que sistemas de software sejam abertos a módulos externos, principalmente quando o fabricante comercializa seus próprios módulos. Assim, assumamos por hipótese que serão usados

¹⁰<https://en.iss.ru/company/about>

¹¹<https://en.iss.ru/products/securos-facex>

¹²<https://en.iss.ru/products/securos-computer-vision>

os módulos do mesmo fabricante na análise que segue. Entretanto, como não temos como obter informações de sensibilidade (a taxa atuante de falsos-positivos) ou performance (a taxa de falsos negativos que o modelo incorre nessa taxa de falsos-positivos) desses programas, a comparação com outros modelos industriais e acadêmicos, tais como os testados pelo NIST [8] não será possível.

O módulo de reconhecimento de faces é baseado em modelos treinados com redes neurais convolucionais, mesma tecnologia descrita no presente parecer. Como pode ser visto nestes dois vídeos publicados pela empresa no Youtube [73, 74], o programa não é perfeito, mas consegue detectar faces com razoável sucesso, desde que as faces tenham um tamanho mínimo (40 a 60 pixels de distância entre os olhos, 160 pixels de largura da face) de acordo com documento do fabricante no Anexo II) e que o número de faces não ultrapasse o número de faces cada núcleo do processador consegue processar (20 faces segundo o documento do anexo II). A figura 16 apresenta um quadro com uma representação de uma imagem de tamanho equivalente a 1024×768 pixels. Na imagem, apresentamos, da esquerda para a direita, fotos de faces de, aproximadamente, 140×140 , 100×100 , 160×160 e 80×80 pixels. Assim, de acordo com o manual do módulo de faces, o sucesso no reconhecimento de uma face aconteceria apenas em imagens do tamanho da terceira foto da esquerda para a direita. Como pode-se imaginar, é grande a chance do sistema detectar faces desde que as câmeras tenham lentes adequadas (que estão especificadas no edital) e estejam posicionadas em lugares como (especificado na pág. do documento XXXX), por exemplo:

- saídas de vagões (duas câmeras posicionadas opostamente uma em relação à outra para cada vagão);
- guichês de venda bilhetes;
- entrada e saída de escadas, ou escadas rolantes.

Uma vez que a imagem de uma face é detectada, o módulo permite que um banco

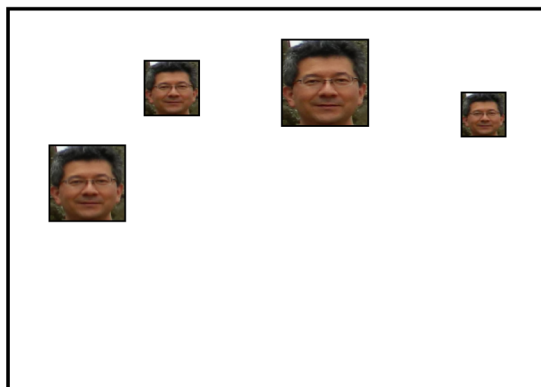


Figura 16: Diversos tamanhos de uma face em uma representação de uma imagem de 1024×768 pontos. Da esquerda para a direita, aproximadamente, 140×140 , 100×100 , 160×160 e 80×80 pixels.

de imagens e dados externo seja usado para a busca o pareamento e reconhecimento de uma face, como explicado anteriormente.

Finalmente, em resposta ao processo 1006616-14.2020.8.26.0053, o Metrô declara não possuir base de dados de larga escala que armazenem dados pessoais. Conclui-se disso que, caso a empresa que ganhou o edital resolva usar um outro sistema que não o módulo da SecurOS, o treino dos modelos de reconhecimento facial deverão ser realizados pelo provedor do software de análise forense, em bases de dados escolhidas a sua discrição. Em qualquer caso, não há como avaliar se a base de dados utilizada para treino é adequada para o contexto utilizado. Lembremos que, como descrito na parte de aspectos éticos deste parecer, a composição étnica ou de gênero da base de treino pode ter consequências significativas para a performance do modelo de reconhecimento facial em aplicações reais. Dada a conhecida diversidade étnica brasileira e a dificuldade de se montar bases de dados com distribuições equitativas de etnias e gêneros que descrevemos, é razoável assumir que podem haver problemas no uso prático de reconhecimento facial, incluindo, mas não restrito a, maior quantidade de falsos positivos para minorias

étnicas e de gênero.

VII.5 Outros aspectos pessoais e privacidade

A privacidade dos usuários de uma área monitorada por câmeras de segurança não está restrita apenas às imagens capturadas. Outros aspectos tecnológicos devem ser levados em consideração, e alguns deles podem passar despercebidos a uma análise superficial das consequências da implantação de um sistema de reconhecimento facial.

O primeiro aspecto diz respeito ao uso de algum identificador único para acessar a área monitorada pelas câmeras. No caso do transporte público, em especial na cidade de São Paulo, o uso do cartão do Bilhete Único permite a identificação dos usuários, uma vez que a confecção de cartões anônimos está sendo descontinuada pelo órgão responsável (SPTrans) [75]. Essa característica facilitaria muito uma possível rotulação das imagens¹³ das faces no futuro, sendo necessário apenas posicionar as câmeras corretamente nas catracas de acesso. Além disso, a captura de imagens nesses locais favorece o reconhecimento facial, visto que apenas uma pessoa pode passar por vez em uma catraca, e portanto as imagens obtidas teriam uma única face, evitando outros problemas conhecidos como oclusões, por exemplo. A vinculação do rótulo com uma imagem poderá ser realizada em tempo real, com alguma integração (através de APIs) entre o sistema de identificação e autorização das catracas e o sistema de reconhecimento facial, como também pode ser feita manualmente através de *timestamps*, sendo necessário apenas as informações da pessoa que passou em determinada hora, minuto e segundo de um dia qualquer, e vinculá-las às imagens obtidas naquele mesmo momento. É importante esclarecer que o SecurOS também tem um módulo que pode ser adquirido separadamente que pode integrar as informações dos cartões do Bilhete

¹³Processo em que um rótulo contendo certas características (sexo, idade, raça, nome, etc) é atribuído a uma imagem, facilitando buscas posteriores.

Único, ou dos cartões funcionais dos empregados, com o sistema de reconhecimento facial [76].

Outro aspecto normalmente negligenciado é a presença de redes Wi-Fi (sem fio). Essas redes são onipresentes hoje em dia pelo seu baixo custo de implantação, mas em um ambiente monitorado é possível obter informações importantes sobre os usuários e sua movimentação. O trabalho de [77] mostra que é possível estimar a posição de um dispositivo em um ambiente fechado através da força do sinal recebido (RSS, do inglês *Received Signal Strength*). É importante ressaltar que a tecnologia Wi-Fi não foi desenvolvida com esse objetivo, mas o seu uso em ambientes de trânsito de pessoas permitiria o rastreamento de dispositivos com grande precisão e baixo custo de implantação ao mantenedor da infraestrutura de redes. Em [78] é proposto um método que utiliza imagens de câmeras de segurança em conjunto com técnicas de rastreamento por RSS de dispositivos Wi-Fi para estimar a localização de indivíduos em ambientes fechados.

Além disso, todo dispositivo Wi-Fi possui um valor único que o identifica, conhecido como endereço de controle de acesso de mídia (MAC, do inglês *Media Access Control*). Esse valor fica associado aos pontos de acesso (*access points*) que o usuário se conecta. Ou seja, com base nas datas e horários nos quais um dispositivo se conectou à rede sem fio, é possível inferir qual o endereço MAC do dispositivo móvel (telefone) de uma determinada pessoa. Essa associação também pode ser explícita, pois em redes como a do Metrô de São Paulo os usuários precisam cadastrar seus dispositivos móveis informando nome, e-mail, número de telefone [79]. Em resumo, é possível que os administradores de uma rede sem fio saibam da presença de um determinado dispositivo no perímetro coberto pelo sinal da rede e associem o dispositivo a um indivíduo específico, além de, com um esforço mínimo, ser possível calcular a localização precisa do indivíduo na edificação.

VII.6 Perguntas e Respostas

Nesta seção tratamos de algumas questões que foram a nós colocadas durante a escrita do parecer.

VII.6.1 Com os dispositivos/tecnologias que o metrô tem em mãos, seria possível, ao menos em teoria, realizar o rastreamento das pessoas que usam o metrô?

Por exemplo: “seguir” qualquer pessoa em seu trajeto na cidade, mesmo em que diferentes dias e locais?

Teoricamente, os métodos de Visão Computacional atual são suficientes para que se faça um rastreamento limitado de pessoas, mesmo que não façam o reconhecimento facial do usuário. Se eles integrarem as informações do bilhete único (mesmo que não esteja sendo usado pelo seu proprietário), com as informações das câmeras, eles vão conseguir um rastreamento ainda melhor.

O novo sistema pode facilitar sim, esse rastreamento. Uma das especificações da Concepção de Sistema é a fácil integrabilidade do software de análise forense com imagens ou bases de dados externas. Por meio de uso indevido ou não-autorizado, uma pessoa pode rastrear alguma outra nas dependências do metrô, desde que em mãos de uma galeria de imagens pertencendo a ela. O ator mal-intencionado conseguiria identificar todas as aparições da pessoa registradas pelo sistema de vigilância nos últimos 30 dias, podendo depreender seu trajeto, horário, rotina entre outros. A Lei n. 14.123/2021, que trata do crime de perseguição, descreve que esta se daria mediante a seguinte ação: “Perseguir alguém, reiteradamente e por qualquer meio, ameaçando-lhe a integridade física ou psicológica, restringindo-lhe a capacidade de locomoção ou, de qualquer forma, invadindo ou perturbando sua esfera de liberdade ou privacidade”.

VII.6.2 No tocante a manifestações cívicas, um sistema como o que será implementado no metrô pode ser utilizado na classificação de pessoas por grupo político?

Sendo assim, o sistema pode ser utilizado para acompanhar ou localizar grupos específicos pela cidade? Como o Metrô pretende lidar com dados considerados ‘sensíveis’ (como opinião política) pela LGPD, sendo este exemplo um possível caso?

De acordo com a especificação do sistema de análise forense (documento CS-9.83.ME.XX/7XX-001, parágrafo 6.16) o sistema deverá poder refinar buscas de pessoas e objetos por atributos como ”tipo de roupa” ou ”cor” (supondo-se aqui que cor se refere a cor da vestimenta). Sendo assim, é possível utilizar o sistema para acompanhar grupos que são identificados por essas atribuições, tais como torcidas organizadas ou organizações políticas. Não há nos documentos providos pelo Metrô declaração de estratégias para lidar com esses dados considerados sensíveis.

A viabilidade técnica da possibilidade de rastreamento ao qual se refere a resposta VI.6.1 também pode ser utilizada na análise sobre a possibilidade de classificação de pessoas por grupo político. Isso imaginando que o sistema poderia realizar o perfilamento de pessoas pertencentes a grupos que estariam se deslocando para reunião em um local específico a fim de realizar uma manifestação cívica, por exemplo, o que poderia ser repetido em diferentes ocasiões em que ocorram aglomerações com a mesma pauta política ou social. Assim, há viabilidade técnica de que determinadas pessoas sejam associadas a determinado comportamento, o que inclui a possibilidade de perfilamento e identificação de grupos de indivíduos segundo deslocamentos específicos.

VII.6.3 Quais os principais erros que podem ocorrer no sistema apresentado pelo metrô?

Qual impacto esses erros teriam para os usuários? Quais as consequências que um falso positivo ou falso negativo pode ter em relação a um usuário?

O sistema de reconhecimento facial do metrô incorre no risco de apresentar falhas na execução: o sistema executa normalmente mas, no entanto, retorna uma resposta incorreta. Vale lembrar que esse é um risco inerente a todo sistema de aprendizado computacional e nunca pode ser completamente mitigado.

O sistema de reconhecimento contratado pelo metrô não possui objetivos de verificação, mas somente identificação. Nesse caso, como descrevemos na parte introdutória, ele pode incorrer em dois tipos de erros: falsos negativos, onde o sistema falha em identificar um indivíduo de interesse, e falsos positivos, onde o sistema confunde alguém com uma pessoa de interesse. Erros de falsos negativos tendem a impactar mais o sistema e o metrô, pois representa momentos em que o sistema falha em conseguir entregar seus benefícios (deixa de identificar um fugitivo ou uma pessoa desaparecido). Já falsos positivos tendem a impactar mais os usuários, pois ao ser identificado com alguma pessoa suspeita ou foragida, o usuário pode sofrer de um leve incômodo a uma prisão infundada. Como explicamos, as taxas de falsos negativos e falsos positivos tendem a ser inversamente correlacionadas, podendo-se ajustar a sensibilidade dos modelos para alcançar um equilíbrio mais adequado. Não há entre a documentação do metrô uma análise sobre esse equilíbrio ou referência a taxa de sensibilidade utilizada.

É importante notar também, como descrito no capítulo V, que modelos de reconhecimento facial podem incorrem em efeitos demográficos, onde a performance do modelo é pior para certas demografias. Entre efeitos tipicamente observados é a maior taxa de falsos positivos entre indivíduos de pele escura e indivíduos de etnias que não são apropriadamente representadas na base de dados. Não havendo referência a tratamento do sistema para mitigar esses vieses, é razoável supor que

eles ocorram no sistema contratado pelo metrô. Isso faz com que o sistema potencialmente seja um perpetuador de desigualdades presentes na sociedade brasileira, selecionando erroneamente mais vezes indivíduos negros e imigrantes.

VII.6.4 A documentação apresentada pela empresa no que concerne à capacidade técnica é suficiente para demonstrar a mitigação dos riscos que o sistema pode apresentar para os usuários?

Consideramos que, para o caso de análise forense, não. A documentação apresentada não é clara se o software de reconhecimento de faces será o módulo da ITS para o SecurOS, ou se deve ser contratado, ou se é um software pré-existente, ou ainda se foi desenvolvido especificamente para este projeto. Assim sendo, não temos informações acerca da base de dados usada para treino dos modelos, taxa de falsos positivos operante e taxa de falsos negativos correspondente à essa taxa de falsos positivos, entre outras informações e características importantes para avaliar a performance do sistema tanto em si como relativa a outros sistemas comerciais disponíveis.

VII.6.5 Qual o grau de eficiência em relação à identificação de crianças e adolescentes?

Ademais, como a tecnologia se comporta com o envelhecimento das pessoas ou eventuais mudanças de aparência física?

Como descrito no capítulo IV, uma das principais causas de erro nos sistemas comerciais avaliados pelo NIST é a idade. Crianças e adolescentes (de 12 a 20 anos) e pessoas idosas (mais de 65) foram os grupos de idade com pior performance no geral, conclusão compartilhada por diversos estudos na literatura, como citado. Isso se deve muito provavelmente a dificuldade de se obter grandes bases de dados

com fotos de crianças ou idosos por questões legais ou éticas.

O envelhecimento é também uma grande limitação para sistemas de reconhecimento facial. Mesmo em faces de adultos, uma diferença de mais de 5 anos entre a imagem da galeria e a imagem atual já é o suficiente para degradar a performance em cerca de 5%. Para crianças e adolescentes, o tempo necessário para degradação é ainda menor, 2 anos e meio a três anos, e a perda de performance é maior, chegando até a 17%.

Como dito anteriormente, existem desenvolvimentos na literatura que permitem melhorar o reconhecimento sobre efeito de envelhecimento ou o de crianças e adolescentes. No entanto, é necessário frisar que são alguns estudos dispersos, que não atingiram consenso. Desse modo, a não ser que alguma medida para mitigar esses erros relacionados a idade ou envelhecimento seja explicitada na documentação do sistema de reconhecimento facial, é válido assumir que ele será inadequado para a identificação de crianças e adolescentes, e sujeito a falhas por envelhecimento de sua base de dados. Em especial, deve-se considerar inadequado para realizar identificação de crianças e adolescentes desaparecidas, onde as imagens de busca são tipicamente desatualizadas.

VII.6.6 No projeto, há proteções e indicação de coleta e tratamento diferenciado de dados de crianças e adolescentes?

Existem riscos específicos relacionados a incidentes relacionados ao vazamento de dados de crianças e adolescentes?

Não há tal indicação. O sistema do Metrô armazenará a priori somente imagens puras capturadas dos últimos 30 dias. Sem discriminação, essas imagens podem conter imagens de crianças e adolescentes também. O sistema de análise forense pode então ser utilizado para reconhecer crianças ou adolescentes baseando-se em alguma outra galeria de imagens. Os riscos de uso indevido ou não autorizado

do software que falamos anteriormente se aplicam aqui, com o agravante de se tratar de menores de idade.

VII.6.7 O projeto explica qual o banco de dados será utilizado?

Em caso negativo, essa seria uma informação importante para analisar a eficiência do projeto? Caso haja essa indicação, esse banco de dados é o mais adequado para os fins almejados pelo metrô?

O projeto não descreve qual banco de dados será utilizado para treinar os modelos de reconhecimento facial. Como mostramos, essa informação é de grande importância para avaliar a eficiência do projeto, uma vez que um modelo treinado em uma base de dados de rostos pouco diversa, ou com composição étnica muito distinta da brasileira pode acarretar em perdas significativas de performance ou efeitos demográficos que prejudiquem desproporcionalmente minorias.

Ademais, para galeria de identidades a serem buscadas, se alude na resposta a questionamentos do IDEC, a um "banco de dados local do Metrô". Não se especifica sua composição, se é, por exemplo, de funcionários ou de cidadãos no geral, nem o propósito dessa busca. Há também o prospecto de utilização de bases de dados de entes de segurança pública, com a possibilidade de integração sendo um requisito descrito na Concepção de Sistema (CS).

Parte VIII

Conclusão

A tecnologia avançou muito nos últimos anos, barateou-se e está mais acessível tanto para o cidadão comum, como para as empresas privadas. Os sistemas de vigilância valeram-se muito desse avanço e cada vez menos estranha-se que existam câmeras nos filmando a todo momento, principalmente em ambientes públicos. Como todo sistema social, assim que começam a surgir os problemas, as pessoas começam a se questionar das armadilhas e perigos das novidades. O presente parecer procurou então destrinchar os diferentes aspectos referentes a esse tipo de projeto, tanto de maneira geral, quanto especificamente em relação ao caso do metrô de São Paulo. Na parte I, foi realizado um panorama sobre os atuais usos de sistemas de reconhecimento facial pelo mundo e resumido o mecanismo de funcionamento de um sistema de reconhecimento facial. Além disso, o tópico aponta marcos alcançados no desenvolvimento desse tipo de tecnologia, ao passo que cita pesquisas que mostram as incorreções e erros decorrentes da aplicação desse tipo de mecanismos na prática, no âmbito das políticas públicas. São enumeradas ainda as principais razões identificadas para falhas no desempenho desses sistemas, quais sejam: a falta de representatividade populacional no desenvolvimento desses mecanismos e a diferença de transparência, discussão e escrutínio externos entre sistemas desenvolvidos industrialmente - sejam eles comerciais ou estatais - e sistemas desenvolvidos no meio acadêmico. Adicionalmente, são citadas iniciativas nos Estados Unidos e na Europa que, diante dos erros e riscos identificados nesse tipo de sistema, visam restringir e até mesmo proibir determinados usos dessas tecnologias. Subsequentemente, a parte II do parecer se debruça sobre quatro principais fases do desenvolvimento de sistemas de reconhecimento facial, mencionando, nesse contexto, a mudança de paradigma que inaugurou os atuais modelos desse tipo de sistema. Em resumo, tal tópico resume o atual funcionamento desse tipo de tecnologia, que transforma imagens de faces em representações numéricas e compara umas às outras, buscando uma compatibilidade. Nesse mesmo contexto, o parecer se dedica a explicar a existência de erros intrínsecos a esse tipo de sistema, os falsos positivos e falsos negati-

vos. São expostos ainda os mecanismos estatísticos utilizados para mensurar a eficácia desse tipo de projeto - o qual, por princípio, apresenta uma taxa de identificações equivocadas no seu funcionamento. A parte III, por sua vez, se dedica às bases de dados que têm sido utilizadas para o reconhecimento facial ao longo do tempo. Nesse sentido, o tópico aponta primeiramente que as principais características físicas (classes) extraídas de um indivíduo a partir da face: a idade (faixa etária), o sexo e a ancestralidade. Ao mesmo tempo, o capítulo ainda aponta as situações fáticas que levam a erros na identificação dessas características, como por exemplo o envelhecimento das pessoas, as limitações dos sistemas em identificar características relacionadas à expressão da identidade de gênero e falta de consideração de diversidade étnico-racial no desenvolvimento dos sistemas. Além disso, são listados e analisados bancos de dados específicos utilizados para reconhecimento facial, apontando a ocorrência de falhas no sistema que impactam negativamente sua eficácia. Finalmente, é destacado ainda que existem bases de dados proprietárias que têm sua transparência - e, como consequência, seu escrutínio por terceiros - comprometida por questões legais ou de segurança. A parte IV se dedica a enumerar e analisar as razões que levam às falhas identificadas atualmente na execução desse tipo de sistema. São destacadas e aprofundadas questões referentes à resolução de imagens e ao envelhecimento/idade da população como características que se tornam dificuldades na execução dos sistemas de reconhecimento facial. A parte V do parecer ocupa-se em explicar fatores de segurança pertinentes a a sistemas de reconhecimento facial e a aspectos éticos a eles referentes. A respeito da segurança desses mecanismos, o trecho aponta itens que devem ser considerados em relação à proteção contra acesso indevido ao sistema e também de armazenamento das imagens coletadas. Adicionalmente, são citados estudos que apontam uma lacuna de conhecimento de criptografia identificada nos desenvolvedores desse tipo de sistema e os problemas decorrentes desse quadro. Além de mencionar falhas de natureza ética identificadas no desenvolvimento de sistemas de reconhecimento facial, o capítulo

se dedica também a apontar como questões demográficas - como gênero, ancestralidade, nacionalidade e idade, por exemplo - se relacionam com o desempenho dos sistemas e, de acordo com pesquisas já publicadas, podem impactar negativamente e/ou discriminar minorias de gênero ou étnico-raciais. Nesse sentido, tal parte ainda destaca a questão de vieses raciais, tendo em vista que trata-se um fator relevante diante do contexto social brasileiro, apontando como o uso do reconhecimento facial pode representar discriminações contra pessoas negras, por exemplo. Finalmente, a parte VI do parecer se dedica especificamente a um caso de estudo, o projeto referente ao metrô de São Paulo. O parecer neste momento então passa a analisar aspectos técnicos do edital referente ao projeto, apontando por exemplo: (i) que há poucos detalhes nos documentos oficiais em relação ao software a ser utilizado, o SecurOS; (ii) que há informações sobre as medidas tomadas no tocante à segurança das informações do sistema, o que não elimina a possibilidade de respectiva invasão ou roubo de dados; (iii) ainda em relação a esse tópico, é pontuado que o edital não menciona como se dará a conexão do sistema SecurOS com a Internet; (iv) que as possibilidades de armazenamento das imagens capturadas pelas câmeras e a possibilidade utilização destas para o reconhecimento facial, além da viabilidade técnica para que o sistema identifique outros eventos, como diferentes tipos de movimentos realizados por um indivíduo ou por um grupo de pessoas; (v) que o edital não apresenta informações suficientes sobre o banco de dados para que se possa analisar se este é adequado para o contexto utilizado, sua respectiva performance e os riscos apresentado pelo sistema frente a minorias étnicas e de gênero; (vi) que questões de privacidade trazidas pelo uso do sistema, para além das imagens capturadas, como a potencialização da identificação dos usuários por meio de reconhecimento facial associado ao uso do Bilhete Único e de captura de informações das pessoas por meio de redes de Wi-Fi. A parte VI traz também respostas a perguntas formuladas, informando que o sistema objeto de análise apresenta algumas características e possibilidades referentes: (i) à viabilidade técnica, por exemplo, de rastrear pessoas ou grupos

de pessoas pelas dependências do metrô; (ii) a falhas na execução serem inerentes a esse tipo de sistema, não sendo possível mitigar completamente esse tipo de erro (falsos negativos e falsos positivos); (iii) à não existência, na documentação do metrô, de uma análise sobre a taxa de sensibilidade utilizada ou o equilíbrio entre as taxas de falsos positivos e falsos negativos; (iv) à potencialidade de o sistema perpetuar desigualdades presentes no Brasil, selecionando de maneira errada com maior frequência indivíduos negros e imigrantes; (v) à ausência de informações referentes à análise forense que sejam capazes de demonstrar a mitigação dos riscos apresentados pelo sistema frente aos usuários do metrô; (vi) à inadequação de uso do sistema para identificação de crianças e adolescentes, incluindo pessoas desaparecidas pertencentes a esses grupos, levando em consideração que medidas para mitigar erros relativos à idade ou ao envelhecimento não estão explicitadas na documentação; (vii) à ausência de especificações relativas à coleta e ao tratamento diferenciado de dados de crianças e adolescentes; (viii) à ausência de indicação do banco de dados a ser utilizado pelo metrô para treinar os modelos de reconhecimento facial, ainda que tal informação seja de grande importância para avaliar a eficiência do projeto. Assim, diante da complexidade de tantos fatores e tantas variáveis, uma das principais conclusões do texto é que essa tarefa de reconhecimento facial é muito intrusiva e sensível, além de poder causar danos pessoais para as pessoas individualmente, principalmente no caso de reconhecimento facial para encontrar pessoas procuradas pela justiça. Nesse sentido, talvez devamos tirar uma lição muito importante do experimento realizado pela Associação Americana de Liberdades Cívicas [2]: antes de colocarmos um sistema de reconhecimento facial para encontrar criminosos, deveríamos testar se o sistema diz que encontrou criminosos apresentando as faces dos nossos representantes eleitos, desde vereadores, até o presidente; ou se o sistema diz que encontrou criminosos se apresentamos as faces de todas as advogadas, advogados, juízas e juizes; ou se o sistema diz que encontrou criminosos se apresentamos as faces de artistas etc. Assim, se casos de falsos positivos acontecerem nesses experimentos,

e com cidadãos comuns, principalmente de grupos menos privilegiados acontecer, a sociedade entenderá melhor quão sensível e complexa é essa tarefa.

- Antonio Augusto Abello - Bacharel e mestrando em Ciência da Computação no IME-USP. O Antonio está se especializando em Aprendizado de Máquina e Visão Computacional. O trabalho dele envolve o estudo do impacto da resolução das imagens no erro dos classificadores e do impacto da “perda” de frequências que compõe uma imagem no erro dos classificadores.
- Rafael Will Macedo de Araújo - Bacharel em Ciência da Computação, mestre e doutor em Ciência da Computação pelo IME-USP. O trabalho de mestrado do Rafael foi sobre segurança de informação em autenticação de dispositivos móveis. Já o trabalho de doutorado foi sobre reconhecimento de gênero e ancestralidade usando técnicas de aprendizado de máquina baseadas em dicionários.
- Roberto Hirata Jr. - Bacharel em Física e Matemática, mestre e doutor em Ciência da Computação pelo IME-USP, professor livre-docente do departamento de Ciência da Computação do IME-USP. Pesquisa e orienta na área de Aprendizado de Máquina e Visão Computacional desde 1993. Será o coordenador do parecer.

Parte IX

Anexo I



(a)



(b)



(c)

Figura 17: Hirata olhando para a câmera (17a) e olhando ligeiramente de lado (17b) e (17c). O algoritmo foi bem sucedido em todos os casos.



(a)



(b)



(c)

Figura 18: Hirata olhando para a câmera (18a) e olhando ligeiramente de lado (18b) e (18c), mas com a boca e parte do nariz escondidos. O algoritmo não encontrou a face em todos os casos.

Neste anexo, apresentamos algumas situações em que os algoritmos de localização de faces funcionam e não funcionam.

Este anexo está incompleto



Figura 19: Hirata vestindo óculos, nas mesmas condições que anteriormente. Apenas em (19a) o algoritmo encontrou a face.

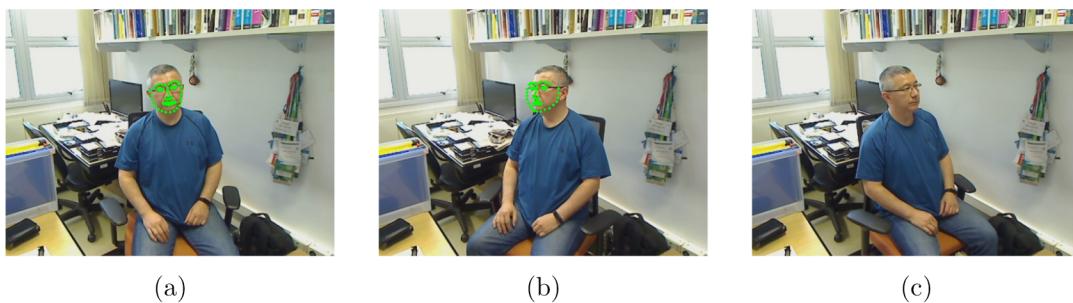


Figura 20: Hirata vestindo óculos, mas afastado do computador. O algoritmo encontrou a face nas imagens 20a e 20b, mas não em 20c. No caso da imagem 20b, os pontos fiduciais estão mal-posicionados, o que vai acarretar em uma deformação maior na rotação da imagem.

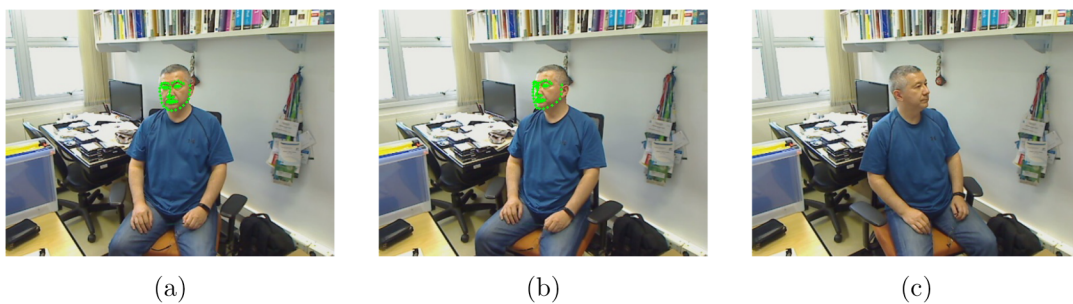
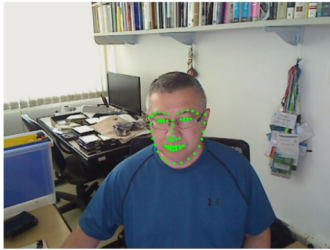
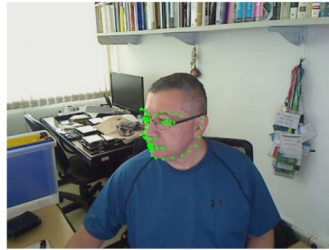


Figura 21: Hirata afastado do computador, sem óculos. O algoritmo encontrou a face nas imagens 21a e 21b, como anteriormente, mas não em 21b. No caso da imagem 21b, os pontos fiduciais também estão mal-posicionados.



(a)



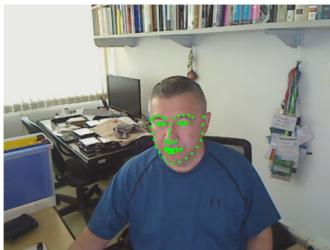
(b)



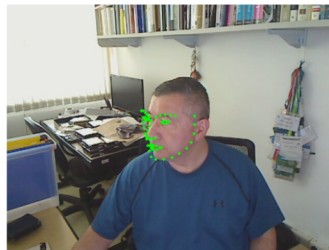
(c)

Figura 22: Hirata perto do computador, vestindo óculos, mas mudando-se ligeiramente as condições de iluminação. Desta vez o algoritmo encontrou a face nas imagens 22a e 22b.

Figura 23: Hirata (22c) e Hirata (24a)



(a)



(b)



(c)

Figura 24: Hirata perto do computador, sem óculos, mas mudando-se ligeiramente as condições de iluminação. O algoritmo encontrou a face nas imagens 24a e 24b, mas nesta última a posição dos marcadores fiduciais não está boa.



(a)



(b)

Figura 25: Mesmas condições que anteriormente, mas sem óculos e escondendo o nariz e a boca. O algoritmo falhou em ambos os casos.

Referências Bibliográficas

- [1] Silkie Carlo. Britain has more surveillance cameras per person than any country except china. that's a massive risk to our free society. <https://news.yahoo.com/britain-more-surveillance-cameras-per-151641361.html>, 2019 (Acesso em 25/04/2019).
- [2] S. Ingber. Facial recognition software wrongly identifies 28 lawmakers as crime suspects. <https://www.npr.org/2018/07/26/632724239/facial-recognition-software-wrongly-identifies-28-lawmakers-as-crime-suspects>, 2018 (Acesso em 3/05/2019).
- [3] L. Dearden. Facial recognition wrongly identifies public as potential criminals 96% of time, figures reveal. <https://www.independent.co.uk/news/uk/home-news/facial-recognition-london-inaccurate-met-police-trials-a8898946.html>, 2019 (Acesso em 3/05/2019).
- [4] R. Manthorpe and A. J. Martin. 81% of 'suspects' flagged by met's police facial recognition technology innocent, independent report says. <https://news.sky.com/story/met-polices-facial-recognition-tech-has-81-error-rate-independent-report-says-1175> 2019 (Acesso em 3/05/2019).

- [5] P. Fussey and D. Murray. Independent report on the london metropolitan police service’s trial of live facial recognition technology. Technical report, University of Essex, 2019.
- [6] P. Berger. Mta’s initial foray into facial recognition at high speed is a bust. <https://www.wsj.com/articles/mtas-initial-foray-into-facial-recognition-at-high-speed-is-a-bust-11554642000>, 2019 (Acesso em 3/05/2021).
- [7] Patrick J Grother, Mei L Ngan, and Kayee K Hanaoka. Face recognition vendor test part 3: demographic effects. 2019.
- [8] Patrick J Grother, Mei L Ngan, and Kayee K Hanaoka. Ongoing face recognition vendor test (frvt) part 2: Identification. 2018.
- [9] Shamann Walton Norman Yee, Hillary Ronen and Matt Haney. Acquisition of surveillance technology. https://www.eff.org/files/2019/05/07/leg_ver3.pdf, 2019 (Acesso em 3/05/2021).
- [10] Caroline Haskins. Oakland becomes third u.s. city to ban facial recognition. oakland just joined san francisco, ca and somerville, ma in banning the use of facial recognition. <https://www.vice.com/en/article/zmpaex/oakland-becomes-third-us-city-to-ban-facial-recognition-xz>, 2019 (Acesso em 3/05/2021).
- [11] Joe Nguyen et. al. Bob Hasegawa. Washington state legislature. <https://app.leg.wa.gov/billsummary?billnumber=6280&year=2019&initiative=False#documentSection>, 2019 (Acesso em 3/05/2021).
- [12] Inioluwa Deborah Raji and Genevieve Fried. About face: A survey of facial recognition evaluation, 2021.

- [13] P.Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.
- [14] Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, October 2008. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie.
- [15] Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [16] Woodrow W. Bledsoe. The model method in facial recognition. Technical Report PRI 15(47):2, Panoramic Research Inc., Paulo Alto, CA, 1964.
- [17] Rabia Jafri and Hamid Arabnia. A survey of face recognition techniques. *JIPS*, 5:41–68, 06 2009.
- [18] M. Sharif, F. Naz, Mussarat Yasmin, M. Shahid, and A. Rehman. Face recognition: A survey. *Journal of Engineering Science and Technology Review*, 10:166–177, 2017.
- [19] Paweł Forczmański and Magdalena Furman. Comparative analysis of benchmark datasets for face recognition algorithms verification. In Leonard Bolc, Ryszard Tadeusiewicz, Leszek J. Chmielewski, and Konrad Wojciechowski, editors, *Computer Vision and Graphics*, pages 354–362, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

- [20] Ming-Hsuan Yang, D.J. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
- [21] G. Bradski. The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*, 2000.
- [22] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.
- [23] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Yee-Whye Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–II, 2004.
- [24] Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In *Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition*, Marseille, France, October 2008. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie.
- [25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS’12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.
- [26] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In David Touretzky, editor, *Advances in Neural Information Pro-*

- cessing Systems (NIPS 1989)*, volume 2, Denver, CO, 1990. Morgan Kaufman.
- [27] Raphael Angulu, Jules R Tapamo, and Aderemi O Adewumi. Age estimation via face images: a survey. *EURASIP Journal on Image and Video Processing*, 2018(1):1–35, 2018.
 - [28] Choon-Boon Ng, Yong-Haur Tay, and Bok-Min Goi. A review of facial gender recognition. *Pattern Analysis and Applications*, 18(4):739–755, 2015.
 - [29] Vijay Kumar, Ramachandra Raghavendra, Anoop Namboodiri, and Christoph Busch. Robust transgender face recognition: Approach based on appearance and therapy factors. In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–7. IEEE, 2016.
 - [30] Foad Hamidi, Morgan Klaus Scheuerman, and Stacy M Branham. Gender recognition or gender reductionism? the social implications of embedded gender recognition systems. In *Proceedings of the 2018 chi conference on human factors in computing systems*, pages 1–13, 2018.
 - [31] Siyao Fu, Haibo He, and Zeng-Guang Hou. Learning race from face: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 36(12):2483–2509, 2014.
 - [32] P Jonathon Phillips, Fang Jiang, Abhijit Narvekar, Julianne Ayyad, and Alice J O’Toole. An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception (TAP)*, 8(2):1–11, 2011.
 - [33] Biao Leng, Kai Yu, and QIN Jingyan. Data augmentation for unbalanced face recognition training sets. *Neurocomputing*, 235:10–14, 2017.
 - [34] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE*

- Transactions on pattern analysis and machine intelligence*, 22(10):1090–1104, 2000.
- [35] Rafael Will Macêdo de Araujo. *A bag of features approach for human attribute analysis on face images*. PhD thesis, Universidade de São Paulo, 2019.
 - [36] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345. IEEE, 2006.
 - [37] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
 - [38] Mahmoud Afifi and Abdelrahman Abdelhamed. Afif4: deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. *Journal of Visual Communication and Image Representation*, 62:77–86, 2019.
 - [39] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74. IEEE, 2018.
 - [40] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.
 - [41] Yui Man Lui, David Bolme, Bruce A Draper, J Ross Beveridge, Geoff Givens, and P Jonathon Phillips. A meta-analysis of face recognition covariates. In

- 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–8. IEEE, 2009.
- [42] Shuowen Hu, Robert Maschal, S Susan Young, Tsai Hong Hong, and P Jonathon Phillips. Face recognition performance with superresolution. *Applied optics*, 51(18):4250–4259, 2012.
 - [43] Dengxin Dai, Yujian Wang, Yuhua Chen, and Luc Van Gool. Is image super-resolution helpful for other vision tasks? In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.
 - [44] Antonio Augusto Abello and Roberto Hirata. Optimizing super resolution for face recognition. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 194–201. IEEE, 2019.
 - [45] Zhangyang Wang, Shiyu Chang, Yingzhen Yang, Ding Liu, and Thomas S Huang. Studying very low resolution recognition using deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4792–4800, 2016.
 - [46] Ze Lu, Xudong Jiang, and Alex Kot. Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters*, 25(4):526–530, 2018.
 - [47] Ding Liu, Bowen Cheng, Zhangyang Wang, Haichao Zhang, and Thomas S Huang. Enhance visual recognition under adverse conditions via deep networks. *IEEE Transactions on Image Processing*, 28(9):4401–4412, 2019.
 - [48] Erfan Zangeneh, Mohammad Rahmati, and Yalda Mohsenzadeh. Low resolution face recognition using a two-branch deep convolutional neural network architecture. *Expert Systems with Applications*, 139:112854, 2020.
 - [49] Pei Li, Loreto Prieto, Domingo Mery, and Patrick J Flynn. On low-resolution face recognition in the wild: Comparisons and new techniques. *IEEE Transactions on Information Forensics and Security*, 2019.

- [50] Mohammad Haghighat and Mohamed Abdel-Mottaleb. Low resolution face recognition in surveillance systems using discriminant correlation analysis. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 912–917. IEEE, 2017.
- [51] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. To learn image super-resolution, use a gan to learn how to do image degradation first. In *Proceedings of the European conference on computer vision (ECCV)*, pages 185–200, 2018.
- [52] Rosaura G VidalMata, Sreya Banerjee, Brandon RichardWebster, Michael Albright, Pedro Davalos, Scott McCloskey, Ben Miller, Asong Tambo, Sushobhan Ghosh, Sudarshan Nagesh, et al. Bridging the gap between computational photography and visual recognition. *arXiv preprint arXiv:1901.09482*, 2019.
- [53] Lacey Best-Rowden and Anil K Jain. Longitudinal study of automatic face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(1):148–162, 2017.
- [54] Debayan Deb, Neeta Nain, and Anil K Jain. Longitudinal study of child face recognition. In *2018 International Conference on Biometrics (ICB)*, pages 225–232. IEEE, 2018.
- [55] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
- [56] Manisha M Sawant and Kishor M Bhurchandi. Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging. *Artificial Intelligence Review*, 52(2):981–1008, 2019.

- [57] Simone Bianco. Large age-gap face verification by feature injection in deep networks. *Pattern Recognition Letters*, 90:36–42, 2017.
- [58] Yandong Wen, Zhifeng Li, and Yu Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4893–4901, 2016.
- [59] Alexandre Braga and Ricardo Dahab. A longitudinal and retrospective study on how developers misuse cryptography in online communities. In *XVII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais*, pages 28–41. 2017.
- [60] Eman Salem Alashwali. Cryptographic vulnerabilities in real-life web servers. In *2013 Third International Conference on Communications and Information Technology (ICCIT)*, pages 6–11. IEEE, 2013.
- [61] Alexia Chatzikonstantinou, Christoforos Ntantogian, Georgios Karopoulos, and Christos Xenakis. Evaluation of cryptography usage in android applications. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pages 83–90, 2016.
- [62] Diego F. Aranha, Marcelo M. Karam, André Miranda, and Felipe Scarel. *Software vulnerabilities in the Brazilian voting machine*, pages 149–175. IGI Global, 2014.
- [63] Fabiana Cambricoli. Nova falha do ministério da saúde expõe dados pessoais de mais de 200 milhões de brasileiros. <https://saude.estadao.com.br/noticias/geral,nova-falha-do-ministerio-da-saude-expoe-dados-pessoais-de-mais-de-200-milhoes,70003536340>, 2021 (Acesso em 30/04/2021).

- [64] Gayathri Mahalingam, Karl Ricanek, and A Midori Albert. Investigating the periocular-based face recognition across gender transformation. *IEEE Transactions on Information Forensics and Security*, 9(12):2180–2192, 2014.
- [65] James Vincent. Transgender youtubers had their videos grabbed to train facial recognition software. <https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset>, 2017 (Acesso em 25/04/2021).
- [66] Hu Han, Anil K Jain, Fang Wang, Shiguang Shan, and Xilin Chen. Heterogeneous face attribute estimation: A deep multi-task learning approach. *IEEE transactions on pattern analysis and machine intelligence*, 40(11):2597–2609, 2017.
- [67] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017.
- [68] Mohsan Alvi, Andrew Zisserman, and Christoffer Nellåker. Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.
- [69] Kushal Vangara, Michael C King, Vitor Albiero, Kevin Bowyer, et al. Characterizing the variability in face recognition accuracy relative to race. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.
- [70] Documento 9 - petição metrô.
- [71] Intelligent Security Systems. Sítio de internet da empresa que produz e comercializa o sistema seguros. <https://issivs.com/>, Acesso em 3/06/2021.

- [72] Intelligent Security Systems. Securos administration guide. <https://en.iss.ru/support/documentation>, (Acesso em 7/07/2021).
- [73] Intelligent Security Systems. Vídeo de divulgação do sistema securos para reconhecimento de faces. <https://www.youtube.com/watch?v=A1xTtZrQGrc>, 2019 (Acesso em 3/06/2021).
- [74] Intelligent Security Systems. Vídeo de divulgação do sistema securos para detecção de máscaras. <https://www.youtube.com/watch?v=lFex5Bxf7sI>, 2020 (Acesso em 3/06/2021).
- [75] Bom dia SP. Transporte público de sp deixará de aceitar 2,6 milhões de bilhetes Únicos anônimos a partir de setembro. <https://g1.globo.com/sp/sao-paulo/noticia/2021/06/17/transporte-publico-de-sp-deixara-de-aceitar-26-milhoes-de-bilhetes-unicos-anonimos-ghhtml>, 2021 (Acesso em 20/06/2021).
- [76] Intelligent Security Systems. Vídeo de divulgação do sistema securos para integração com sistema de cartões de identificação. https://www.youtube.com/watch?v=qFU0653_rfw, 2019 (Acesso em 3/06/2021).
- [77] Gints Jekabsons and Vadim Zuravlyov. Refining wi-fi based indoor positioning. In *Proceedings of 4th International Scientific Conference Applied Information and Communication Technologies (AICT)*, Jelgava, Latvia, pages 87–95, 2010.
- [78] Christopher Nielsen, John Nielsen, and Vahid Dehghanian. Fusion of security camera and rss fingerprinting for indoor multi-person tracking. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–7. IEEE, 2016.
- [79] Camila Boehm. Metrô de são paulo terá wifi grátis em 40 estações. <https://agenciabrasil.ebc.com.br/geral/noticia/2017-02/>

metro-de-sao-paulo-tera-wifi-gratis-em-40-estacoes, 2017 (Acesso em 29/06/2021).