

Avaliando a Eficácia de Redes Neurais Artificiais para Reconhecimento de Face Utilizando Raspberry Pi

Ejziel Sampaio Santos¹, Fabrício Veloso de Jesus¹, Walber Conceição de Jesus Rocha¹
João Carlos Nunes Bittencourt¹

¹Centro de Ciências Exatas e Tecnológicas
Universidade Federal do Recôncavo da Bahia
Av. Rui Barbosa, 710 - 44380-000 - Cruz das Almas, Brasil.

{ejziel, fabricio, walber.jesus}@aluno.ufrb.edu.br, joaocarlos@ufrb.edu.br

Abstract. *The rise of mobile robotics has moved many everyday tasks to computational platforms. Computer vision applications targeting face recognition, despite being a simple task for humans, demand for a large amount of processing power in computer platforms. This paper presents a comparison of several face recognition systems using convolutional neural networks and softmax regressors, to evaluate the accuracy and performance rates of the application in a Raspberry Pi computer platform. Three different topologies of neural networks were evaluated, considering seven users. The results point to an accuracy rate of 69.23% and an average execution time of 301.45 ms in the best case.*

Resumo. *A ascensão da robótica móvel fez com que diversas tarefas do cotidiano fossem executadas com o auxílio de plataformas computacionais. Aplicações de visão computacional voltadas ao reconhecimento de faces, embora seja uma tarefa simples para os humanos, demandam um grande poder de processamento. Este trabalho apresenta um sistema de reconhecimento facial utilizando redes neurais convolucionais e regressores softmax, tendo em vista avaliar as taxas de acurácia e desempenho da aplicação na plataforma computacional Raspberry Pi. Foram avaliadas 3 diferentes topologias de redes neurais, com 7 diferentes usuários, apresentando taxa de acurácia de 69.23% e tempo de execução médio de 301,45 ms, no melhor caso.*

1. Introdução

Habitualmente, os seres humanos realizam tarefas de identificação e reconhecimento, seja de objetos, lugares ou pessoas. O reconhecimento facial expressa uma variedade de informações do indivíduo, incluindo considerações a respeito de idade, sexo, raça, identidade e estado emocional. Reconhecer faces em múltiplas situações, portanto, faz parte do cotidiano dos seres humanos. Embora o reconhecimento facial seja uma tarefa simples, não é um processo trivial de ser reproduzido em sistemas computacionais.

A maior dificuldade está na modelagem de uma face que abstraia as características que diferenciem as faces entre si, pois, embora diferentes, todos os rostos possuem características semelhantes. Além disso, imagens de diferentes dimensões dificultam o processo de reconhecimento facial, principalmente na fase de extração e classificação de traços faciais. Isso se deve ao fato de que quanto maior a dimensão da imagem, maior será o custo computacional e menor a precisão do classificador.

O presente trabalho apresenta um sistema de reconhecimento facial utilizando redes neurais convolucionais e regressores softmax, tendo em vista avaliar as taxas de acurácia e desempenho da aplicação na plataforma computacional *Raspberry Pi 4*. Foram avaliadas 3 diferentes topologias de redes neurais, com 7 diferentes usuários. O objetivo dessa pesquisa foi avaliar como uma tarefa computacionalmente intensa é capaz de performar em uma plataforma computacional de baixo custo.

Este trabalho está organizado como segue: a Seção 2 apresenta os principais aspectos teóricos relacionado ao reconhecimento facial e suas técnicas; a Seção 3 apresenta a metodologia utilizada para o desenvolvimento do modelo proposto; na Seção 4, são apresentados os resultados do estudo, como acurácia e tempo de execução; a Seção 5, por fim, apresenta as conclusões e perspectivas de trabalhos futuros.

2. Revisão da literatura

2.1. Biometria Facial

O termo “biometria” advém da junção das palavras gregas *bios* (vida) e *metron* (medir). Biometria é definida como uma ciência que compreende estudos estatísticos dos atributos físicos, fisiológicos e/ou comportamentais dos seres vivos [Lourenço 2009]. Essas características são utilizadas atualmente como um método de identificação de pessoas. Dentre os traços biométricos presentes no corpo humano, a biometria facial está presente em diversas aplicações tecnológicas. O modelo biométrico é utilizado, principalmente, devido a sua vasta gama de aplicações, como o reconhecimento facial em ambientes controlados e a identificação de pessoas em tempo real, mesmo em ambiente com uma grande presença de ruídos [Silva 2016].

A aparência da face é uma característica biométrica particularmente convincente, pois é usada rotineiramente como primeiro método de reconhecimento entre pessoas. Por sua naturalidade, é a mais aceitável das biometrias. Costa *et al.* (2006) destaca as seguintes características como alguns dos pontos fortes no uso da biometria facial:

1. Fácil capacidade de coleta: a coleta dos dados pode ser realizada mesmo sem a cooperação do indivíduo;
2. Alta aceitabilidade: os indivíduos geralmente se sentem à vontade para utilizar tal método;
3. Alta universalidade: o sistema de reconhecimento facial pode ser utilizado por qualquer pessoa.

Isso implica que sistemas de reconhecimento facial precisam ser robustos o suficiente de modo que sejam capazes de apresentar um nível de eficiência adequado, mesmo em ambientes não controlados. Adicionado a isso, com o advento dos mecanismos de segurança automatizados, os sistemas de reconhecimento facial podem ser implantados em câmeras remotas, reduzindo assim a sobrecarga nas redes de comunicação.

2.2. Visão Computacional e Aprendizado de Máquina

Através da visão, os seres humanos apresentam a capacidade de extrair informações sobre os mais diversos objetos, possibilitando localizá-los e classificá-los com precisão. Ao observar, por exemplo, um cachorro numa imagem, conseguimos não apenas identificar que se trata de um cachorro, como também a posição exata a qual ele se encontra no campo de visão.

A visão computacional tem como objetivo produzir sistemas computacionais dotados de habilidades semelhante à visão dos seres humanos. O objetivo associado ao desenvolvimento deste campo do conhecimento está no desenvolvimento de técnicas que permitam a extração de informações e características relevantes para uma determinada aplicação a partir de uma imagem [Sebe et al. 2005].

Atualmente, os problemas que envolvem a visão computacional vêm se tornando mais complexos, fazendo com que os métodos clássicos não sejam suficientes para atingir resultados considerados satisfatórios. A partir da necessidade de métodos mais robustos, foram introduzidas técnicas de aprendizado de máquina com o intuito de obter resultados com maior robustez e flexibilidade [Khan et al. 2018].

O Aprendizado de Máquina é uma subárea da Inteligência Artificial (IA) que busca proporcionar aos sistemas computacionais a capacidade de realizar determinadas tarefas sem que tenham sido previamente programados. Deste modo, a partir de observações do mundo real (que são convertidas em dados digitais), as técnicas de Aprendizado de Máquina permitem que sistemas computacionais aprendam de maneira autônoma [Faceli et al. 2011].

Dentre as categorias de Aprendizado de Máquina, destaca-se o Aprendizado Supervisionado. Esta técnica baseia-se na implementação uma função que realiza um mapeamento entre os dados de entrada e os dados de saída. Deste modo, dada uma amostra de dados de entrada, é possível prever as variáveis de saída. Este aprendizado pode ser realizado de diversas formas, a saber: regressão logística; árvores de decisão; máquina de vetores de suporte; redes neurais; entre outras. Nas técnicas baseadas em redes neurais, os estudos se concentram no aprendizado profundo, que envolve estruturas computacionais complexas. As redes neurais convolucionais são amplamente empregadas em técnicas de visão computacional, sendo uma das técnicas mais difundidas em meio a literatura [Khan et al. 2018].

2.3. Redes Neurais Artificiais e Reconhecimento Facial

Os avanços da neuro-computação, permitem que problemas complexos, como a detecção e classificação em tempo real de objetos, sejam resolvidos utilizando a aprendizagem profunda. Uma rede neural pode ser definida como um processador que é maciçamente e paralelamente distribuído, formado por outras unidades de processamento mais simples, que armazenam conhecimento experimental e o tornam disponível para uso [Haykin 1999].

Redes Neurais Artificiais (RNAs) são sistemas paralelos distribuídos compostos por unidades de processamento simples (nós) que calculam determinadas funções matemáticas (normalmente não-lineares). Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais [Braga et al. 2000].

As Redes Neurais Convolucionais (RNCs), representa uma das categorias mais utilizadas dentre as técnicas de redes neurais, principalmente em tarefas que tratam dados multidimensionais (imagens e vídeos, por exemplo). Uma RNC funciona de maneira similar às outras redes neurais. A principal diferença é que cada unidade numa camada da RNC corresponde a um filtro multidimensional no qual é efetuada a operação de convolução com a entrada da camada, conforme mostra a Figura 1. Ao percorrer

as camadas, a rede aprende extraindo padrões mais complexos acerca dos dados de entrada [Khan et al. 2018].

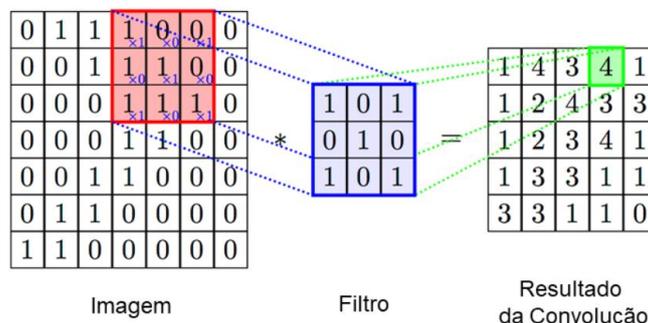


Figura 1. Operação de convolução aplicado em uma imagem.
Fonte: [Silva and Zampirolli 2020]

No diagrama apresentado na Figura 1, a matriz do lado esquerdo é uma imagem em preto e branco, que passa pela aplicação de um filtro de convolução de tamanho 3×3 . À direita está representada na matriz resultante após a aplicação da convolução.

Um algoritmo de RNC voltado para o reconhecimento facial deve ser treinado a partir de um conjunto de imagens rotuladas (faces e identidades) de indivíduos diferentes. Para cada imagem deve haver um sub-conjunto de imagens com diferentes padrões de rotação, pose e características de iluminação. Ao longo da última década foram desenvolvidos diversos sistemas de reconhecimento facial com base em RNC, a saber: FaceNet [Schroff et al. 2015]; VGG face [Parkhi et al. 2015]; Deep Face [Taigman et al. 2014]; OpenFace [Amos et al. 2016]; e L2 softmax [Ranjan et al. 2017, Silva and Zampirolli 2020].

Os sistemas supracitados são utilizados para identificar as faces presentes em imagens, extrair características da imagem, e determinar um vetor de características da face (*face embedding*). A partir do vetor de características os sistemas realizam a comparação com outros vetores, utilizando, por exemplo, a distância euclidiana entre os vetores para definir a classe com maior probabilidade da face pertencer [Serengil and Ozpinar 2020]. Essa análise também pode ser realizada utilizando outras redes neurais mais simples, como por exemplo, um regressor *softmax*. A regressão *softmax* é uma generalização da regressão logística, a qual pode ser utilizada em processos de classificação multi classe [Silva and Zampirolli 2020].

3. Metodologia

3.1. Construção dos Modelos

Sistemas modernos de reconhecimento facial costumam ser divididos em 4 etapas de processamento: (i) detecção; (ii) alinhamento; (iii) representação; e (iv) classificação [Serengil and Ozpinar 2020]. A Figura 2, apresenta os dois principais processos implementados. A Figura 2(a) descreve o processo de treinamento da rede neural, enquanto a Figura 2(b) apresenta um fluxograma do processo de reconhecimento facial a partir de uma imagem de entrada.

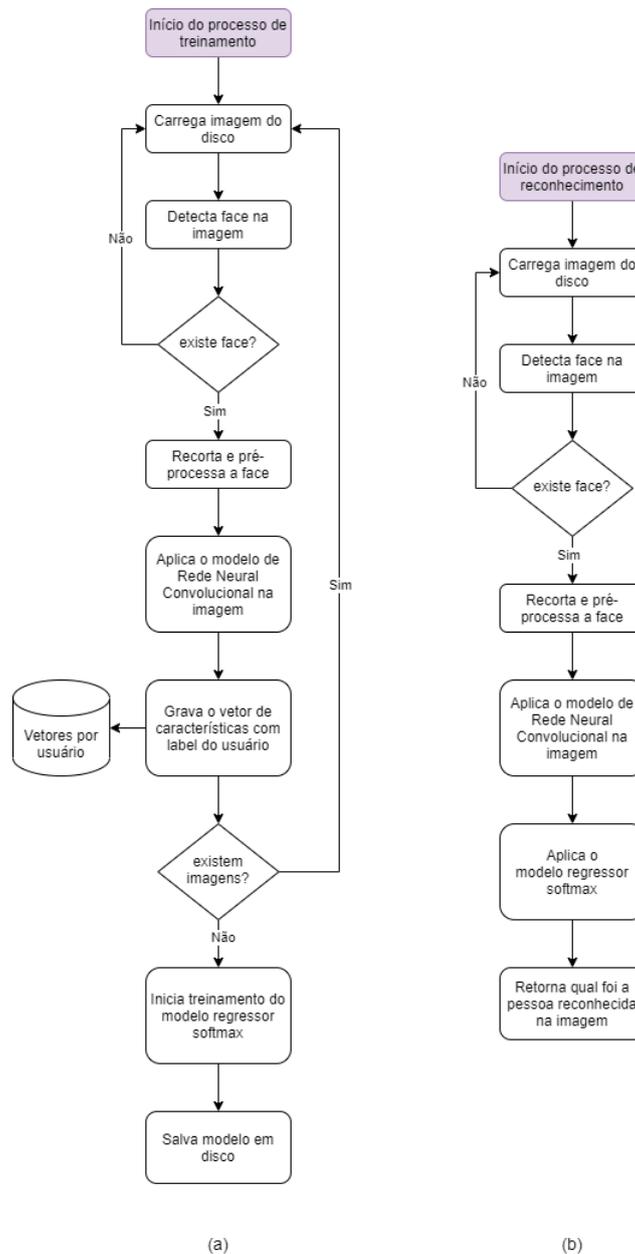


Figura 2. Descrição dos processos de treinamento e reconhecimento.

Os modelos a seguir apresentados foram implementados utilizando a linguagem Python na versão 3.8.5 usando a biblioteca de *machine learning* TensorFlow na sua versão 2.3.1 para construção e treinamento dos modelos. Para realização dos experimentos, foi utilizado um conjunto de bibliotecas de suporte, dentre as quais destacam-se a OpenCV e SkLearn, usadas para aquisição e pré-processamento das imagens. A etapa de treinamento foi realizada em um computador com um processador Core i7-9750H, dotado de uma GPU NVIDIA GTX 1660 Ti e 16GB de memória RAM.

Para o treinamento de cada modelo foi considerado o conjunto de etapas apresentadas a seguir. A primeira etapa consiste na implementação de uma RNC responsável por determinar um vetor de características a partir da imagem original. Neste experimento,

foram utilizados três modelos de RNC: OpenFace, FaceNet e VGG face.

OpenFace é um modelo leve e minimalista para reconhecimento facial desenvolvido pela Carnegie Mellon University no Reino Unido. Sua estrutura possui 3,7 milhões de parâmetros treináveis [Amos et al. 2016]. O FaceNet é um modelo de reconhecimento facial baseado em aprendizado profundo desenvolvido pela Google. Esse modelo foi concebido com base no modelo Inception com cerca de 22,8 milhões de parâmetros em sua estrutura [Schroff et al. 2015]. Construído pelo grupo de geometria visual de Oxford, o VGG face possui uma arquitetura de reconhecimento facial profunda com mais de 145 milhões de parâmetros treináveis [Parkhi et al. 2015].

A utilização de um sistema de detecção de face é importante para que um sistema de reconhecimento facial utilizando redes neurais obtenha maiores taxas de acurácia quanto seja possível. Para desenvolver o sistema de detecção de face, foi utilizado o módulo *Deep Neural Networks* (DNN Module) da biblioteca OpenCV.

Uma vez detectada e selecionada a região do quadro que contém a face, foram necessárias algumas etapas de pré-processamento (i) transformar a imagem em um vetor; (ii) expandir a dimensão do vetor; e (iii) normalizar os valores dos pixels entre -1 e 1. A imagem pré-processada é utilizada, então, como entrada para o modelo de RNC, criando assim um vetor de características. Este conjunto de etapas é repetido para cada imagem do *dataset*, fornecendo os dados necessários para o treinamento da rede de classificação.

O alinhamento facial é um estágio inicial da linha de reconhecimento facial moderna. O Google declarou que o alinhamento facial aumenta a precisão de seu modelo de reconhecimento facial FaceNet em quase 1%, o que significa muito para estudos de engenharia [Schroff et al. 2015]. Para todos os modelos utilizados no presente estudo, foram desenvolvidos dois experimentos: um deles com alinhamento de faces e o outro sem o alinhamento. Para tanto, foi utilizado um algoritmo de alinhamento de face 2D utilizando classificadores em cascata, disponível na biblioteca OpenCV.

A rede de classificação Softmax foi treinada por 40 épocas para cada um dos modelos de RNC. As Figuras 3, 4 e 5 descrevem como cada uma das redes se comportam com o passar das épocas de treinamento realizadas durante os experimentos. Em cada imagem, (a) corresponde a rede sem alinhamento de faces e (b) representa o experimento realizado com o alinhamento de faces.

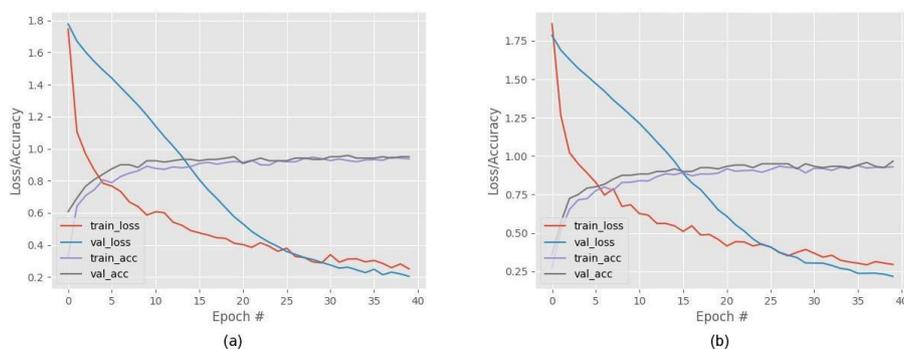


Figura 3. Treinamento do modelo de classificação utilizando OpenFace.

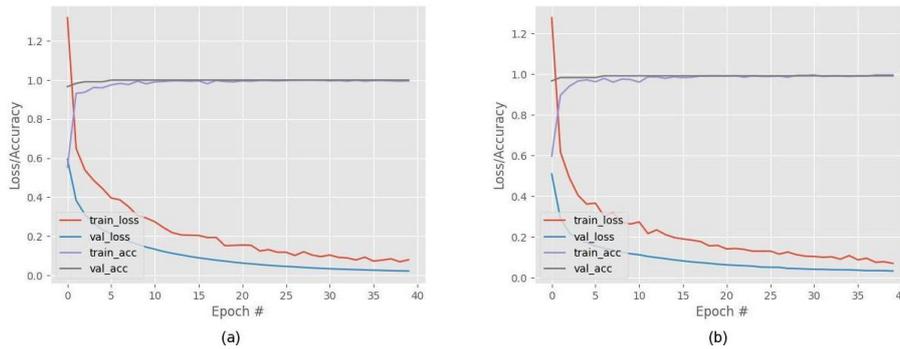


Figura 4. Treinamento do modelo de classificação utilizando FaceNet.

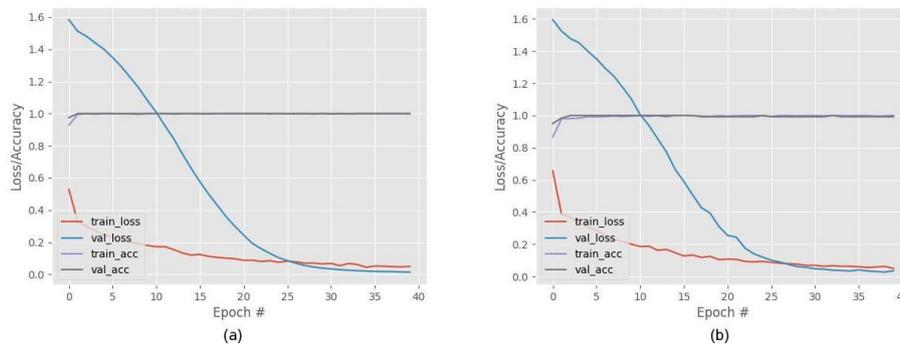


Figura 5. Treinamento do modelo de classificação utilizando VGG Face.

3.2. Inferência e Testes

A metodologia de testes do sistema foi aplicada num computador de placa modelo Raspberry Pi 4, com 4GB de RAM, aliado ao uso do TensorFlow Lite. O TensorFlow Lite é um *framework* de aprendizado profundo com código aberto capaz de realizar inferências especialmente utilizado para aplicações em dispositivos móveis ou aqueles com menor poder de processamento. Ao todo foram utilizadas quatro métricas de avaliação de acurácia do sistema, a saber:

1. **Taxa de verdadeiros positivos (VP):** demonstra a porcentagem de pessoas cadastradas que foram reconhecidas corretamente;
2. **Taxa de verdadeiros negativos (VN):** indica o percentual de pessoas não cadastradas que foram reconhecidas corretamente;
3. **Taxa de falsos positivos (FP):** aponta para a taxa de pessoas não cadastradas que foram reconhecidas incorretamente;
4. **Taxa de falsos negativos (FN):** indica a porcentagem de pessoas cadastradas que foram reconhecidas incorretamente.

De maneira geral, um sistema de reconhecimento facial eficiente busca maximizar as taxas 1 e 2, e diminuir as taxas 3 e 4. A fim de estabelecer uma análise quanto ao rendimento do sistema, foram utilizados dois conceitos principais: (i) a especificidade; e (ii) a sensibilidade. Tais métricas são obtidas a partir das Equações 1 e 2.

$$Especificidade = \frac{VN}{(VN + FP)} \quad (1)$$

A especificidade é aqui definida como a capacidade do sistema em identificar que um indivíduo não está cadastrado na base de dados.

$$Sensibilidade = \frac{VP}{(VP + FN)} \quad (2)$$

A sensibilidade, por sua vez, corresponde à taxa com que indivíduos cadastrados são identificados corretamente.

4. Resultados

Nos testes preliminares, foram utilizadas 26 imagens, dentre as quais 13 apresentavam faces de indivíduos desconhecidos, e 13 de indivíduos cadastrados no sistema. A Tabela 1 mostra as taxas de acurácia alcançadas pelo sistema.

Tabela 1. Resultados de acurácia sem alinhamento de face.

	OpenFace	FaceNet	VGG Face
VP	46.15%	50.00%	50.00%
VN	15.38%	11.54%	19.23%
FP	34.62%	38.46%	30.77%
FN	3.85%	0.00%	0.00%

A Tabela 2 descreve a taxa geral de acurácia, bem como a especificidade e sensibilidade alcançada durante os experimentos realizados.

Tabela 2. Métricas sem alinhamento de face.

	OpenFace	FaceNet	VGG Face
Taxa de Acurácia	61.54%	61.54%	69.23%
Especificidade	0,308	0,231	0,385
Sensibilidade	0,923	1	1

As tabelas acima mostram que a porcentagem de usuários com identificação verdadeira positiva é em média 48.72% entre os modelos, e 15.38% de verdadeiro negativo. Estes resultados mostram que o sistema se comporta bem quando o indivíduo está cadastrado na base de dados, porém quando o indivíduo é um desconhecido os resultados são menos satisfatórios. A partir disso justifica-se a alta sensibilidade do sistema, em detrimento de uma baixa especificidade. Tais métricas podem ser alteradas de acordo com a aplicação do sistema, a partir da alteração do limiar para reconhecimento, que neste caso foi de 50%, ou seja, o usuário só é reconhecido caso a confiabilidade do reconhecimento seja superior a 50%.

As Tabelas 3 e 4 demonstram as taxas de acurácia individuais e a taxa geral de acurácia, bem como a especificidade e sensibilidade com e sem o uso do algoritmo para alinhamento de faces, respectivamente.

Além das análises de acurácia do sistema, é importante observar o uso de recursos de hardware para processamento, bem como a latência de execução das principais etapas do reconhecimento. A plataforma Raspberry Pi 4 possui um conjunto de componentes de hardware com poder computacional limitado, sobretudo quando comparado à um computador de mesa. Deste modo, embora o algoritmo chegue a um nível de acurácia similar

Tabela 3. Resultados de acurácia com alinhamento de face.

	OpenFace	FaceNet	VGG Face
VP	50.00%	50.00%	50.00%
VN	19.23%	15.38%	11.54%
FP	30.77%	34.62%	38.46%
FN	0.00%	0.00%	0.00%

Tabela 4. Métricas com alinhamento de face.

	OpenFace	FaceNet	VGG Face
Taxa de Acurácia	69.23%	65.38%	61.54%
Especificidade	0,385	0,308	0,231
Sensibilidade	1	1	1

dentre os modelos utilizados, o tempo de execução difere. As Tabelas 5 e 6 apontam o tempo médio de execução das principais etapas do algoritmo em cada um dos modelos sem e com a presença do alinhamento de faces, respectivamente.

Tabela 5. Resultados de tempo (ms) sem alinhamento de face.

	OpenFace	FaceNet	VGG Face
Detecção de Face	218.52	221.22	218.03
RNC para Vetor de Características	37.83	235.72	2169.29
Rede Neural de Classificação Softmax	0.07	0.07	0.22

Tabela 6. Resultados de tempo (ms) com alinhamento de face.

	OpenFace	FaceNet	VGG Face
Detecção e Alinhamento de Face	263.77	279.38	309.94
RNC para Vetor de Características	37.61	231.15	2153.05
Rede Neural de Classificação Softmax	0.07	0.07	0.22

5. Conclusão

Os resultados alcançados mostram que é viável a utilização de um hardware como o que está presente nas placas Raspberry Pi modernas para aplicações robustas, como a de reconhecimento facial. Apresentando tempos de resposta aceitáveis para aplicações cotidianas. Todos os algoritmos demonstraram boa sensibilidade. Dessa forma, o sistema é capaz de reconhecer com eficiência indivíduos cadastrados na base de dados. No entanto, a capacidade de reconhecer e evitar fraudes é baixa por conta da baixa taxa de especificidade.

A utilização do alinhamento de face promoveu melhorias nas métricas para a OpenFace e FaceNet, onde a taxa de falsos positivos sofreram uma redução. Entretanto, para a VGG Face não houve indícios de melhoria. Pelo contrário, na VGG Face, as métricas sofreram um decréscimo pois a taxa de falsos positivos aumentou com o uso do alinhamento. A limitação de hardware existente fez com que o uso do alinhamento de faces aumentasse o tempo de execução. Neste sentido trabalhos futuros podem utilizar outras bases de dados e topologias de RNC com o objetivo de minimizar os impactos das limitações de hardware nas métricas finais.

Referências

- Amos, B., Ludwiczuk, B., Satyanarayanan, M., et al. (2016). Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2).
- Braga, A. d. P., Ludermir, T. B., and Carvalho, A. C. P. d. L. F. (2000). *Redes neurais artificiais: teoria e aplicações*. LTC.
- Costa, L. R., Obelheiro, R. R., and Fraga, J. S. (2006). Introdução á biometria. *Livro-texto de Minicursos-VI SBSeg*.
- Faceli, K., Lorena, A. C., Gama, J., and Carvalho, A. (2011). Inteligência artificial. *Uma Abordagem de Aprendizado de Máquina*. Editora LTC.
- Haykin, S. (1999). *Neural networks a comprehensive foundation*. ed. sl: Pearson prentice hall.
- Khan, S., Rahmani, H., Shah, S. A. A., and Bennamoun, M. (2018). A guide to convolutional neural networks for computer vision. *Synthesis Lectures on Computer Vision*, 8(1):1–207.
- Lourenço, G. F. d. F. (2009). Reforço da segurança das biométricas utilizando codificação de fonte distribuída. Master's thesis, Dissertação de Mestrado, Instituto Superior Técnico, Universidade Técnica de
- Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep Face Recognition. In *Proceedings of the British Machine Vision Conference 2015*, pages 41.1–41.12, Swansea. British Machine Vision Association.
- Ranjan, R., Castillo, C. D., and Chellappa, R. (2017). L2-constrained Softmax Loss for Discriminative Face Verification. *arXiv:1703.09507 [cs]*. arXiv: 1703.09507.
- Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823. arXiv: 1503.03832.
- Sebe, N., Cohen, I., Garg, A., and Huang, T. S. (2005). *Machine Learning in Computer Vision*, volume 29. Springer Science Business Media.
- Serengil, S. I. and Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–5. IEEE.
- Silva, A. L. (2016). Redução de características para classificação de imagens de faces. Master's thesis, Dissertação de Mestrado, Universidade Federal Rural do Semi-Árido.
- Silva, D. G. d. and Zampiroli, F. d. A. (2020). Reconhecimento facial para validação de usuário durante um questionário no Moodle. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação (WCBIE)*, pages 124–131. SBC. ISSN: 0000-0000.
- Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, Columbus, OH, USA. IEEE.