

# UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS NO RECONHECIMENTO FACIAL PARA AUTENTICAÇÃO BIOMÉTRICA

José Geraldo Rodrigues Fernandes<sup>1</sup>; Fernando Eduardo Resende Mattioli<sup>2,3</sup>; Rogério Bernardes Andrade<sup>4</sup>

<sup>1,2,4</sup> Faculdade de Talentos Humanos - FACTHUS, Uberaba (MG), Brasil

<sup>3</sup>Universidade Federal do Triângulo Mineiro – UFTM, Uberaba (MG), Brasil

jofer.br@gmail.com, fernando.mattioli@facthus.edu.br, rbandrade@facthus.edu.br

**RESUMO:** O reconhecimento facial é um conjunto de técnicas desenvolvidas para identificar uma pessoa por meio de imagens ou vídeos. Com o surgimento das novas tecnologias, dentre estas as redes neurais artificiais, sua aplicação se tornou uma necessidade, impulsionando a busca de melhorias nos projetos desenvolvidos por pesquisadores e profissionais do setor. Estas melhorias passaram pelo desenvolvimento de novas técnicas para minimizar as deficiências detectadas nas redes neurais artificiais tradicionais, que necessitavam muitos padrões de treinamento de um mesmo indivíduo para obter resultados muitas vezes insatisfatórios e imprecisos. Mais recentemente, com o desenvolvimento das redes neurais convolucionais e o aumento na capacidade de processamento dos novos dispositivos, foram criadas novas técnicas de processamento e algoritmos cada vez mais inteligentes que trouxeram novas perspectivas para este domínio. Dentre as inovações encontradas destacam-se as redes com arquitetura *One shot*, que necessitam de um número reduzido de padrões de treinamento obtendo resultados equivalentes ou superiores àqueles apresentados pelas arquiteturas tradicionais (estritamente sequenciais). Neste trabalho, é apresentado um estudo comparativo entre as arquiteturas sequencial e *One shot*, aplicadas ao reconhecimento facial. Os modelos de arquitetura *One shot* apresentaram resultados promissores no problema investigado, demonstrando desempenho superior ou equivalente aos modelos de arquitetura sequencial em 7 dos 10 experimentos realizados.

**PALAVRAS CHAVE:** Redes Neurais Convolucionais, *One shot*, reconhecimento facial.

## USE OF ARTIFICIAL NEURAL NETWORKS IN FACIAL RECOGNITION FOR BIOMETRIC AUTHENTICATION

**ABSTRACT:** Facial recognition is defined as a set of techniques developed to identify one person using images or videos. With the development of new technologies, such as artificial neural networks, its application became a need, boosted by the search for improvement in conducted projects by researchers and professionals. These improvements include the development of new techniques to minimize the shortcomings detected in traditional artificial neural networks, which require a large amount of training patterns (images) from the same individual to produce unsatisfactory results in many cases. More recently, with the development of convolutional neural networks and the increase in processing power of new devices, new techniques and algorithms, more robust and more intelligent were created, bringing new perspectives for this domain. Among this innovation, *One shot* networks, which require a reduced subset of training patterns, present competitive results, even outperforming traditional architectures (purely sequential) in some domains. In this work, a comparative study on sequential and *One shot* architectures on facial recognition is presented. The evaluated *One shot* models presented promising results in the investigated problem, matching or even outperforming sequential architecture models in 7 of the 10 conducted experiments.

**KEYWORDS:** Convolutional Neural Networks, *One shot*, facial recognition.

## INTRODUÇÃO

O reconhecimento facial é composto por um conjunto de técnicas para identificar uma pessoa por meio de imagens ou vídeos. Apesar de se tratar de uma tecnologia que já existe há décadas, seu uso se tornou mais perceptível e acessível nos últimos anos, uma vez que seu uso foi estendido a soluções inovadoras tais como aplicativos pessoais de foto e autenticação secundária para dispositivos de segurança (AMAZON, 2019).

Os recursos de análise facial, como os disponibilizados no mercado, permitem a identificação da

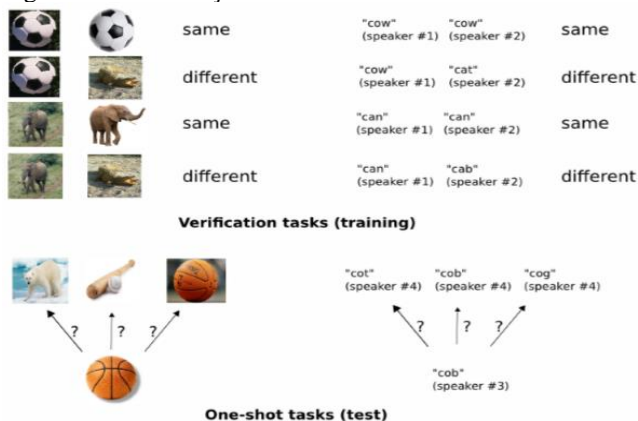
presença de rostos em imagens ou vídeos, indicando também quais atributos esses rostos têm. Como exemplos de atributos pode-se citar posição e estado dos olhos, boca, nariz, cor do cabelo e geometria visual do rosto. Esses atributos de detecção são cada vez mais úteis para aplicações onde se necessita organizar ou pesquisar milhões de imagens por segundo utilizando *tags* de metadados (estado emocional, acessórios, faixa etária etc.) ou identificar um indivíduo (AMAZON, 2019).

Silva (2017) apresenta uma técnica de reconhecimento facial utilizando pontos específicos do rosto humano. Em seus experimentos, o efeito que as

topologias das RNA possuem na eficiência de classificação de padrões foi explorado. Redes com mais camadas de neurônios tendem a apresentar tempos de processamento maiores e requerem uma grande quantidade de dados para o treinamento. Além disso, ainda que determinada rede apresente um desempenho satisfatório na classificação de padrões de treinamento, sua capacidade de generalização (processamento de dados desconhecidos) pode ser comprometida caso haja excesso (memorização dos padrões) na fase de treinamento.

No trabalho de Koch et al. (2015) é apresentado um método para treinamento de redes neurais siamesas, caracterizadas por apresentarem dois ramos de entrada, alimentados simultaneamente por 2 imagens de treinamento. Ao contrário das redes sequenciais, que visam categorizar uma entrada de acordo com as classes conhecidas em treinamento, esta arquitetura de rede (também denominada *One shot*) fornece uma medida da similaridade entre as duas entradas. Uma vez treinada, a rede *One shot* pode ser submetida à processar padrões diferentes do treinamento, de classes inteiramente novas e distribuições desconhecidas. A Fig. 1 resume o conceito de classificação *One shot*, no qual os padrões de entrada são classificados como pertencentes à mesma categoria (*same*) ou pertencentes a categorias distintas (*different*).

Figura 1: Classificação *One shot*.



Fonte: Koch et al. (2015).

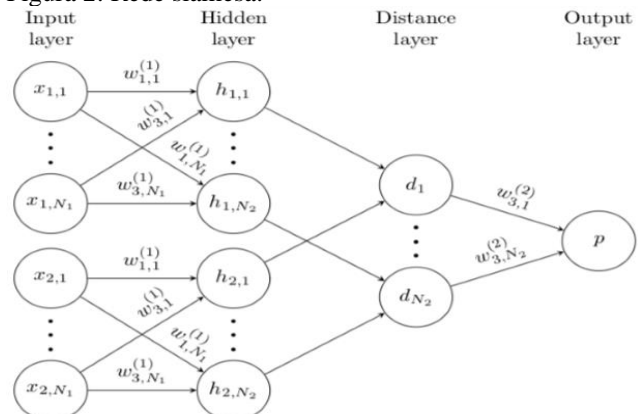
Neste contexto, o presente trabalho apresenta um estudo comparativo de arquiteturas sequenciais e *One shot* aplicadas ao reconhecimento facial. Para tal, foi utilizado um *dataset* contendo imagens de vários indivíduos, capturadas em diferentes ângulos e diferentes condições. Os dados disponíveis foram divididos em 3 conjuntos distintos, para treinamento, validação e teste dos modelos. Para cada arquitetura de rede, foram avaliados 5 modelos de topologia distintos, contendo diferentes números de camadas e de neurônios em sua formação. Para efeito de comparação, os demais parâmetros da rede (número de épocas, número de padrões de treinamento, algoritmo de otimização) foram mantidos constantes.

## REFERENCIAL TEÓRICO

Redes Neurais Convolucionais (RNC) são uma classe de redes neurais artificiais inspiradas no processo biológico da visão humana, muito utilizadas em tarefas de processamento e classificação de imagens. Por meio da operação de convolução, características são automaticamente extraídas das imagens de entrada. Estas características possuem pesos que são automaticamente ajustadas de modo a minimizar o erro entre os valores esperados (*target*) e os valores obtidos (saída da rede). As RNC têm sido utilizadas com sucesso em diferentes tarefas de processamento, classificação de imagens e visão computacional. Dada sua capacidade de extração automática de características, o pré-processamento exigido em RNC é muito menor em comparação com outros algoritmos de classificação de imagens (DATA SCIENCE ACADEMY, 2016).

Vários fatores justificam o desempenho das RNC em problemas envolvendo processamento de imagens digitais. A conectividade local pode reduzir bastante o número de parâmetros no modelo, o que inerentemente fornece alguma forma de regularização integrada. No entanto, camadas convolucionais têm um custo computacional inerentemente mais alto quando comparadas às camadas lineares. Além disso, a operação de convolução usada nessas redes tem uma interpretação direta de filtragem, onde cada filtro busca identificar determinado padrão nas imagens processadas. Assim, as saídas de cada camada convolucional correspondem a importantes características espaciais no espaço de entrada original, oferecendo robustez a transformações simples (MNIH e HINTON, 2009; KRIZHEVSKY, SUTSKEVER e HINTON, 2012; SIMONYAN e ZISSERMAN, 2014).

Figura 2: Rede siamesa.



Fonte: Koch et al. (2015).

As redes siamesas foram introduzidas pela primeira vez no início dos anos 90 por Bromley e LeCun para resolver a verificação de assinatura como um problema de correspondência de imagem (BROMLEY et al., 1993). Uma rede neural siamesa consiste em redes gêmeas que aceitam entradas distintas, unidas em determinado ponto por

uma função de concatenação. Esta função processa a representação da camada de nível mais alto em cada um dos ramos da rede, conforme apresentado na Fig. 2. Esta arquitetura permite que imagens semelhantes produzam ativações equivalentes nas camadas mais altas da rede, enquanto imagens distintas apresentem uma diferença significativa em sua ativação. Este comportamento permite que a rede determine, em sua saída, o grau de similaridade entre as duas imagens de entrada.

O treinamento de redes siamesas é conduzido gerando-se um conjunto de padrões a partir da combinação das imagens disponíveis. Estes padrões apresentam 2 saídas possíveis, indicando se as imagens de entrada pertencem ou não à mesma categoria. Uma vez treinada, a rede processa um padrão desconhecido por meio da comparação deste com imagens de referência, pertencentes às classes de treinamento. Adicionalmente, a rede pode ser alimentada com 2 imagens completamente desconhecidas, indicando o grau de similaridade das mesmas. Este método de treinamento é denominado *One shot* (FEI-FEI, FERGUS e PERONA, 2006; LAKE et al, 2011).

## MATERIAIS E MÉTODOS

Os experimentos descritos neste trabalho foram conduzidos utilizando a linguagem de programação *Python* e as bibliotecas *Keras* e *Tensorflow*. Como ambiente de execução, foi utilizado o *Google Colab*. Para importação das imagens e processamento dos resultados foi utilizada a biblioteca *NumPy*.

*Python* é uma linguagem de programação de alto nível, *open source*, de fácil aprendizagem e multiplataforma. É uma das linguagens mais utilizadas atualmente em aplicações de inteligência artificial (PYTHON.ORG, 2019). *Keras* é um *framework* para rápida prototipação e experimentação em redes neurais artificiais. Possui uma arquitetura modular e extensível, permitindo a rápida implementação e teste de diversos modelos de arquitetura e topologias de redes neurais artificiais (DATA SCIENCE ACADEMY, 2016).

*Tensorflow* é uma biblioteca de aprendizagem de máquina, desenvolvido pela *Google* para projetar, construir e treinar de modelos de aprendizagem de máquina em redes neurais artificiais. É a tecnologia atualmente utilizada como *backend* do *framework Keras* (DATA SCIENCE ACADEMY, 2016). *NumPy* é um pacote para computação científica em *Python*, oferecendo um robusto modelo de estruturas multidimensionais além funcionalidades otimizadas de álgebra linear e de números aleatórios. (NUMPY.ORG, 2019)

O *Google Colab* (*Google Colaboratory*) é um ambiente virtualizado em nuvem que permite a execução de experimentos em aprendizagem de máquina com mínima configuração. Com suporte aos *notebooks Jupyter*, oferece ainda as funcionalidades de compartilhamento de *scripts* entre usuários, além da integração com outros serviços do *Google*, tais como autenticação e *Google Drive* (GOOGLE COLAB, 2019).

Para comparar o desempenho de arquiteturas sequenciais e *One shot* no reconhecimento facial, o *dataset* original foi dividido em conjuntos de treinamento, validação e teste. 80% das amostras disponíveis foram utilizadas para treinamento, enquanto 10% foram utilizadas em validação. Os 10% restantes foram separados no conjunto de teste, utilizado após o treinamento das redes para avaliar o desempenho de generalização dos modelos.

Foram criados 5 modelos diferentes de topologias de redes, sendo que cada modelo foi treinado por 5 vezes coletando-se a acurácia de cada modelo nos conjuntos de treinamento, validação e teste. Estes modelos são apresentados em detalhe na Tab. 1.

Tabela 1: Modelos utilizados nos experimentos.

Modelo	Características
1	1 camada convolucional com 16 filtros 5x5 ativada por ReLU.
2	1 camada convolucional com 16 filtros 5x5 ativada por ReLU + 1 camada de pooling 2x2.
3	1 camada convolucional com 16 filtros 5x5 ativada por ReLU + 1 camada convolucional com 16 filtros 3x3 ativada por ReLU + 1 camada de pooling 2x2.
4	1 camada convolucional com 16 filtros 5x5 ativada por ReLU + 1 camada convolucional com 16 filtros 3x3 ativada por ReLU + 1 camada de pooling 2x2 + 1 camada densa de 16 neurônios ativada por ReLU.
5	1 camada convolucional com 16 filtros 5x5 ativada por ReLU + 1 camada convolucional com 16 filtros 3x3 ativada por ReLU + 1 camada de pooling 2x2 + 1 camada densa de 16 neurônios ativada por ReLU + 1 camada densa de 16 neurônios ativada por ReLU.

Fonte: Os autores, 2019.

As imagens foram redimensionadas para o tamanho 50x50, preservando-se os 3 canais RGB. Cada modelo foi treinado por até 50 épocas, interrompendo-se o treinamento ao não se observar melhoria na acurácia de validação por 10 épocas consecutivas. Foram utilizados *mini batches* de 20 amostras, com algoritmo de otimização *Adam*. A camada de saída foi configurada com a função de ativação *softmax*.

Para a arquitetura *One shot*, os padrões de treinamento, validação e teste foram gerados selecionando-se aleatoriamente pares de imagens a partir dos conjuntos originais. Os dois ramos da rede siamesa foram configurados simetricamente, utilizando o mesmo modelo (Tab. 1).

Tabela 2: Resultados para 2 indivíduos.

Arquitetura	Modelo	Treinamento			Validação			Teste		
		MIN	AVG	MAX	MIN	AVG	MAX	MIN	AVG	MAX
Sequencial	1	90	90,8	91	90	90	90	80	84	85
	2	91	91,8	93	90	90	90	75	79	85
	3	48	85,6	98	50	82	90	50	82	95
	4	48	58,6	74	50	6	90	50	60	75
	5	49	59,4	100	50	58	90	50	59	95
One shot	1	63	79	88	55	65	70	45	73	80
	2	58	62,6	73	55	59	65	40	53	70
	3	59	69,8	81	55	69	75	50	68	80
	4	87	95,6	100	75	86	95	85	90	95
	5	94	97,6	100	90	92	95	80	91	95

Fonte: Os autores, 2019.

Tabela 3: Resultados para 10 indivíduos.

Arquitetura	Modelo	Treinamento			Validação			Teste		
		MIN	AVG	MAX	MIN	AVG	MAX	MIN	AVG	MAX
Sequencial	1	78	92,4	96	67	67	67	56	57,2	62
	2	87	88,2	93	70	71,6	72	52	56,4	58
	3	79	93	100	66	69,2	72	56	59	60
	4	10	10	10	10	10	10	10	10	10
	5	10	10	10	10	10	10	10	10	10
One shot	1	62	63,6	64	59	61,4	62	56	62,6	68
	2	67	68,4	70	63	64	66	58	58,8	60
	3	56	64,6	72	59	63,6	68	53	57,4	63
	4	50	66	90	50	59,2	63	49	59,2	69
	5	55	64	76	53	59,4	65	61	62,4	65

Fonte: Os autores, 2019.

## RESULTADOS E DISCUSSÃO

Para cada arquitetura e modelo foram realizadas 5 repetições do treinamento da rede neural artificial. Ao final de cada treinamento, foram calculadas as acurácias nos conjuntos de treinamento, validação e teste, utilizando a Equação 1:

$$Acc (\%) = 100 * \frac{n_c}{n_c + n_e} \quad (1)$$

A Tab. 2 apresenta os valores mínimo, médio e máximo de acurácia considerando as 5 repetições de cada experimento, no processamento de imagens de 2 indivíduos.

Analisando a acurácia dos modelos no processamento dos dados de teste (desconhecidos pelas redes durante o treinamento), os resultados da Tab. 2 indicam a superioridade da arquitetura sequencial nos 3 modelos mais simples (1, 2 e 3), compostos apenas por camadas convolucionais. Nos 2 modelos mais complexos (4 e 5), formados por camadas convolucionais e densas, a arquitetura *One shot* apresentou um resultado equivalente ou superior, atingindo a melhor acurácia da arquitetura sequencial (95%).

Visando avaliar a escalabilidade dos modelos propostos foi realizada uma nova rodada de experimentos,

agora utilizando imagens de 10 indivíduos distintos. Os resultados destes experimentos podem ser visualizados na Tab. 3.

Observa-se, pelos resultados obtidos neste experimento, o melhor desempenho das redes *One shot*, quando comparadas aos modelos sequenciais, na classificação dos 10 indivíduos avaliados. Os modelos sequenciais mais complexos (4 e 5) apresentaram dificuldade em classificar os mesmos, o que pode ser justificado pela pequena quantidade de amostras de cada indivíduo. Neste caso, seria necessário proceder a coleta de novas amostras das categorias avaliadas, a fim de se obter um melhor desempenho do classificador. Em situações onde não seja factível, as redes *One shot* se apresentam como uma importante alternativa, tendo obtido um desempenho próximo de 70%, mesmo com o conjunto reduzido de amostras disponíveis.

## CONCLUSÃO

Os experimentos conduzidos e descritos neste trabalho permitiram a verificação da potencial aplicação das redes neurais artificiais com arquitetura *One shot* no reconhecimento facial. Os modelos experimentados apresentaram desempenho satisfatório, em alguns casos superando as redes sequenciais na classificação de padrões diversos dos padrões de treinamento. Os resultados obtidos

indicam que o aumento no número de categorias avaliadas (indivíduos) necessita ser compensado por um aumento no número de padrões de treinamento, sob pena de queda no desempenho dos classificadores. Em situações onde a coleta de novos padrões seja complexa ou proibitiva, as redes *One shot* se apresentam como uma promissora alternativa, uma vez que novos padrões podem ser gerados a partir de combinações dos padrões existentes, sem a necessidade de novas coletas.

Como sugestões para trabalhos futuros destacam-se a investigação mais aprofundada do *dataset* estudado (com a utilização de outros indivíduos) e de novas topologias de rede. Finalmente, a avaliação do desempenho do classificador ao processar imagens de indivíduos que não estejam no conjunto inicial de treinamento também constitui uma importante linha de investigação para trabalhos futuros.

## REFERÊNCIAS

AMAZON. Os fatos sobre a tecnologia de reconhecimento facial com inteligência artificial.

Disponível em:

<<https://aws.amazon.com/pt/rekognition/the-facts-on-facial-recognition-with-artificial-intelligence/>>. Acesso em 20/03/2019.

BROMLEY, Jane et al. Signature Verification using a “Siamese” Time Delay Neural Network. Int.]. **Pattern Recognit.** Artzf Intell, v. 7, 1993.

DATA SCIENCE ACADEMY. Deep Learning Book, 2019: **Capítulo 40 - Introdução as Redes Neurais Convolucionais**. 2019. Disponível em:

<<http://deeplearningbook.com.br/introducao-as-redes-neurais-convolucionais/>>. Acesso em: 02/09/2019.

FEI-FEI, Li; FERGUS, Rob; PERONA, Pietro. One-shot learning of object categories. **IEEE transactions on pattern analysis and machine intelligence**, v. 28, n. 4, p. 594-611, 2006.

Google Colab. Welcome To Colaboratory, 2019.

**Introducing Colaboratory**, 2019. Disponível em:

<<https://colab.research.google.com/notebooks/welcome.ipynb>>. Acesso em: 09/09/2019.

KOCH, Gregory; ZEMEL, Richard; SALAKHUTDINOV, Ruslan. Siamese neural networks for one-shot image recognition. In: **ICML deep learning workshop**. 2015.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. 2012. p. 1097-1105.

MNIH, Andriy; HINTON, Geoffrey E. A scalable hierarchical distributed language model. In: **Advances in neural information processing systems**. 2009. p. 1081-1088.

NUMPY.ORG, D&I Grant from CZI. **NumPy**, 2019.

SIMONYAN, Karen; ZISSERMAN, Andrew. **Very deep convolutional networks for large-scale image recognition**. arXiv preprint arXiv:1409.1556, 2014

