

## IMPLEMENTAÇÃO E APLICAÇÃO DE MODELO PARA REDES NEURAIS CONVOLUCIONAIS DE MÁSCARA

ÁLVARO B. CARVALHO<sup>1</sup>, EDUARDO H. GOMES<sup>2</sup>

<sup>1</sup> Graduando em Tecnologia de Análise e Desenvolvimento de Sistemas, Voluntário PIBIFSP, IFSP, Câmpus Cubatão, alv.barros@hotmail.com.

<sup>2</sup> Mestre em Engenharia da Informação, Professor do IFSP - Câmpus Cubatão, ehgomes@ifsp.edu.br.  
Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação

10º Congresso de Inovação, Ciência e Tecnologia do IFSP ou no 4º Congresso de Pós-Graduação do IFSP

27 e 28 de novembro de 2019- Sorocaba-SP, Brasil

**RESUMO:** A área de processamento de imagens apresenta grandes avanços tecnológicos nas últimas décadas. Um exemplo disso é a crescente qualidade de produtos sendo comercializados que utilizam essa tecnologia em seus processos e a popularização dessas ferramentas na vida cotidiana, em computadores pessoais e outros aparelhos. O reconhecimento de seres humanos é um dos maiores focos de pesquisa, como por exemplo reconhecimento facial. Hoje já é comum um sistema reconhecer partes específicas do rosto de uma pessoa, definir a emoção que a face expressa, entre outros. É com o intuito de aprimorar o reconhecimento de imagens do corpo humano que aplicamos as técnicas de redes neurais convolucionais nas partes restantes do corpo. O presente trabalho tem por objetivo apresentar um estudo sobre as técnicas de reconhecimento visual, baseado nas técnicas de detecção de objetos que sejam capazes de reconhecer, delimitar, segmentar e medir o pé humano. Os experimentos foram realizados no ambiente TensorFlow em conjunto com o modelo MS-COCO. Espera-se que os resultados dessa aplicação possam ser utilizados em aplicações de comércio eletrônico como provedores virtuais para a indústria calçadista e outras aplicações que possam surgir.

**PALAVRAS-CHAVE:** reconhecimento de padrões; detecção; processamento de imagens; redes neurais convolucionais; TensorFlow.

## IMPLEMENTATION AND APPLICATION OF MODEL FOR MASKING CONVOLUTIONAL NEURAL NETWORKS

**ABSTRACT:** Image processing has shown huge technological advances in the last decades. Such example is the growing quality of products being commercialized that utilize these technologies in its processes and the popularity of these tools in everyday life, being on personal computers or other devices. Human recognition is one of the main researched topics, such as facial recognition. Today is already common for a system recognize specific face parts, define which emotion the target is feeling, and others. It is with the intente to upgrade the human body image recognition that we employ already developed convolutional neural networks in the rest of the body. The presente work studies technologies for visual recognition, taking into consideration object detection technology that is able to recognize, bound, mask and measure the human feet. Experiments were realized on TensorFlow with the MS-COCO model. The results of this application may be used in applications for eletronic commerce as virtual dressers for shoe industry and other applications that may come.

**KEYWORDS:** pattern recognition; detection; image processing; convolutional neural networks; TensorFlow

## INTRODUÇÃO

As Redes neurais surgiram a partir do estudo da maneira como seres humanos resolvem tarefas de reconhecimento de padrões e utilizam de uma estrutura composta por neurônios artificiais, pesos e bias. Hoje, as técnicas evoluíram muito de sua origem na biologia (RIPLEY, 1996). A arquitetura das

redes neurais convolucionais se diferenciam especializando-se no tratamento de imagens e matrizes, garantindo melhor performance e eficiência nas tarefas dessa natureza (KARPATY, 2016).

As redes neurais recebem um vetor como dado de entrada e o transforma passando por uma série de camadas. Cada camada é composta por neurônios, e cada neurônio é conectado com todos os neurônios da camada passada e nenhum da camada atual. A última camada leva o nome de saída e no contexto de classificação, representa os resultados do reconhecimento (BENGIO, 1997).

Segundo Karpathy (2016) e Florindo (2018), diferentemente das redes neurais, as redes convolucionais organizam seus neurônios em três camadas: largura, altura e profundidade. Além disso, os neurônios são conectados apenas à uma pequena região da camada anterior, resultando em um processamento mais eficiente e performático. São capazes de aprender utilizando matrizes complexas de muitas dimensões e não lineares partindo de grandes coleções de exemplos (BENGIO, 1997), tornando-as candidatas óbvias para o trabalho de reconhecimento de imagens (ABDULLA, 2019).

## MATERIAL E MÉTODOS

O conjunto de metodologias de investigação utilizadas neste projeto se fundamentam em duas práticas de pesquisa: a descritiva e a experimental.

A descritiva foi utilizada no levantamento do referencial teórico das técnicas de processamento digital de imagens e visão computacional de uso livre, gratuito e que atendessem às demandas da solução proposta.

A metodologia experimental foi utilizada pela própria natureza do projeto, cujo objetivo geral propõe estudar e criar uma aplicação de processamento computacional de imagens. Na sequência, iniciou-se o processo para descobrir qual tecnologia seria utilizada para que a aplicação fosse capaz de reconhecer, delimitar e modelar o pé humano.

Definiu-se que seria utilizada para este projeto, as redes neurais convolucionais que são especialistas no tratamento de imagens e matrizes, garantindo melhor performance e eficiência nas tarefas dessa natureza segundo Karpathy (2016) e Florindo (2018). Suas aplicações envolvem classificação (classification), localização (semantic segmentation), detecção (object detection) e segmentação (instance segmentation). A Classificação atribui objetos em uma imagem à uma distinta categoria, a localização moldura o objeto na imagem, a detecção localiza múltiplos objetos e a segmentação contorna e rotula o objeto de interesse (ABDULLA, 2019), como mostrado na Figura 1.

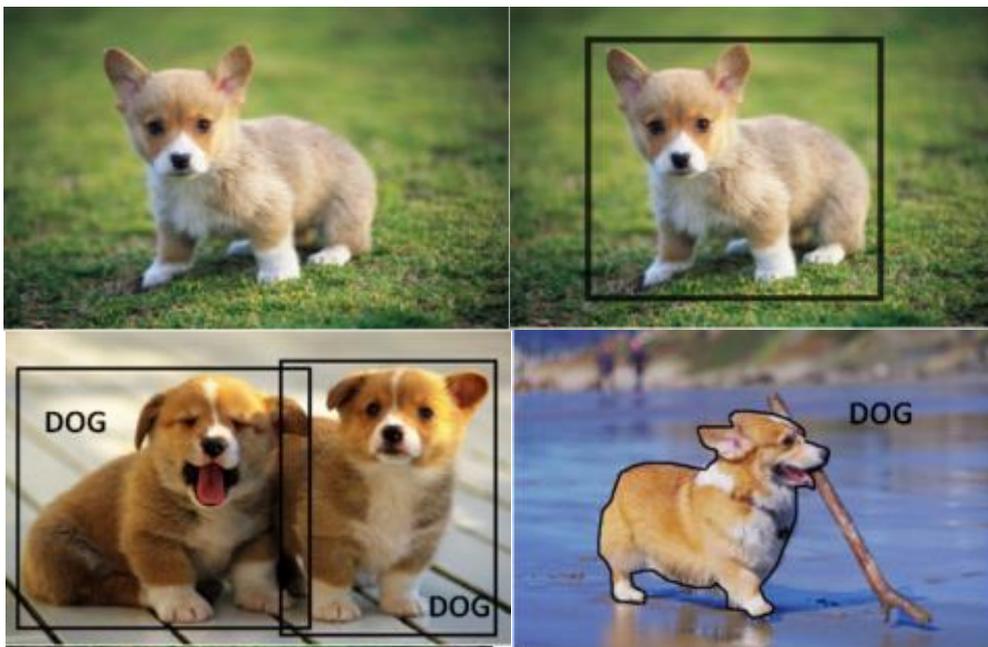


FIGURA 1. Classificação, localização, detecção e segmentação por rede convolucional (FLORINDO, 2018).

## RESULTADOS E DISCUSSÃO

Para o desenvolvimento da solução escolhemos a implementação da rede convolucional MS-COCO conforme Dollár (2014) e Ripley (1996). Utilizando a ferramenta TensorFlow (2019) e os modelos pré-treinados da MS-COCO onde os pesos (weight) são especialmente úteis pela sua capacidade de aproveitar todo treinamento já realizado, abrindo a possibilidade de interrupção da aprendizagem da rede para serem continuados em um momento posterior (TENSORFLOW, 2019), visto que esse processo é demorado.

Com a rede treinada e configurada, a imagem foi passada como uma matriz de pontos RGB (TENSORFLOW, 2019) e processada com as classes importadas do modelo MS-COCO treinado, presente no arquivo weight importado. Aproveitando-se das classes já presentes no dataset MS-COCO, utilizamos imagens que continham objetos reconhecidos pela rede, no caso, humanos, através da categoria person do dataset, conforme Figura 2.

A rede retornou as instâncias encontradas, os pontos pertencentes à essa mesma instância e quatro pontos para a caixa de delimitadora (TENSORFLOW, 2019). Uma exibição padrão desse resultado é a aplicação de uma "máscara" de uma cor vermelha para cada instância na imagem original, sobrepondo-a com o resultado da rede neural como mostrado na Figura 2.

A figura 2 expõe o reconhecimento do pé humano obtida pela aplicação desenvolvida pelo projeto utilizando-se do tensorflow com o modelo MS-COCO e suas técnicas de classificação, localização, detecção e segmentação.

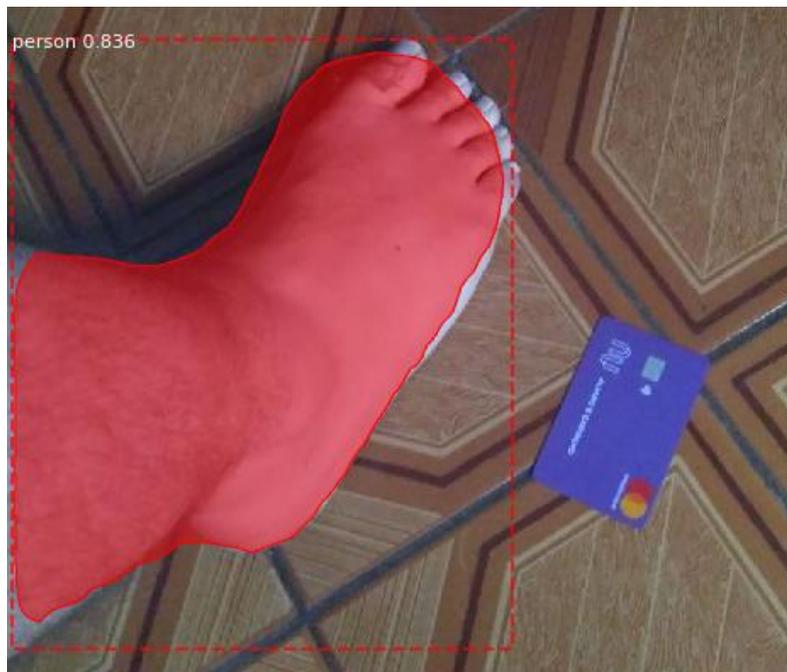


FIGURA 2. Exemplo de reconhecimento do pé humano obtida pela aplicação desenvolvida.

## CONCLUSÕES

A aplicação atendeu as expectativas iniciais de reconhecimento visual e segmentação do pé humano a partir das técnicas de redes convolucionais e se mostrou viável para futuras aplicações comerciais. Porém, para efetuar a medição do pé humano a partir de sua segmentação, seria necessário um maior aprofundamento dessa pesquisa em técnicas de fotogrametria, onde as técnicas se utilizam de objetos com medidas com padrões pré-estabelecidos como o cartão de crédito.

## AGRADECIMENTOS

Agradeço ao Professor Mestre Orientador do Trabalho por todo apoio e orientação dessa pesquisa e ao IFSP - Campus Cubatão pela oportunidade oferecida.

## REFERÊNCIAS

ABDULLA, W. Splash of Color: Instance Segmentation with Mask R-CNN and TensorFlow. Disponível em: < <https://engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238b46> > Acesso em: 30 ago. 2019.

ABDULLA, Waleed. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. Github, 2017.

BENGIO, Y. LECUN, Y. Convolutional Networks for Images, Speech and Time-Series. 1997.

DOLLÁR, Piotr. Microsoft COCO: Common Objects in Context. 2014.

FLORINDO, João B. Redes Neurais Convolucionais – Deep Learning. Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas (UNICAMP), Brasil. Disponível em: < <https://www.ime.unicamp.br/~jbflorindo/Teaching/2018/MT530/T10.pdf> >. Acesso em: 27 ago. 2019.

KARPATHY, A. CS231n Convolutional Neural Networks for Visual Recognition, 2016.

RIPLEY, B. D. Pattern recognition and Neural Networks. Cambridge University Press. ISBN 0 521 46086.7 1996.

TENSORFLOW. Disponível em: < <https://tensorflow.org> >. Acesso em: 20 ago. 2019.