



## Usando a rede neural Faster-RCNN para identificar frutos verdes em pomares de laranja

Luiz Miguel **Cerqueira**<sup>1</sup>, Kleber Xavier Sampaio de **Souza**<sup>2</sup>, Sônia **Ternes**<sup>3</sup> e João **Camargo Neto**<sup>4</sup>

Nº 20605

**Resumo** - O Brasil é um país altamente dependente de sua agropecuária para a geração de Produto Interno Bruto (PIB). Entre os produtos de sua agricultura se destaca o de cultivo da laranja e seus derivados, responsável por uma grande taxa de geração de renda e empregos no País, principalmente no triângulo mineiro e São Paulo. Este trabalho descreve o processamento de imagens digitais de frutos em árvores por meio de técnicas de visão computacional e aprendizado de máquina. Essa técnica tem potencial para, no futuro, auxiliar no processo de estimativa de produção, utilizado por produtores na formulação de estratégias e planejamento de vendas.

**Palavras-chave:** Visão computacional, redes neurais, aprendizado profundo, Faster-RCNN

1 Autor, Bolsista CNPq (PIBIC): graduação em Ciência da Computação, UNICAMP, Campinas-SP.

2 Orientador, Engenheiro-eletricista, doutor em Engenharia Elétrica, pesquisador da Embrapa Informática Agropecuária, Campinas, SP

3 Matemática aplicada e computacional, doutora em Engenharia Elétrica, pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

4 Engenheiro-eletricista, doutor em Engenharia de Biossistemas, analista da Embrapa Informática Agropecuária, Campinas, SP.



**Abstract** - *Brazil is a country that is highly dependent on its agriculture for the growth of its gross domestic product. Among the products of its agriculture, the cultivation of orange and its derivatives stands out for being responsible for a great portion of the income and work generation in the country. In this work, we describe the processing of digital pictures of fruits on trees by the means of computer vision and machine learning. In the future, the techniques developed here might help yield estimation processes by producers, helping them on strategical and sales planning.*

**Keywords:** Computer vision, neural networks, deep learning, Faster-RCNN

## 1 INTRODUÇÃO

A citricultura é, dentre outros gêneros de agricultura realizados no Brasil, uma das mais importantes para o País nos aspectos econômicos e sociais, devido à sua alta capacidade de geração de empregos e Produto Interno Bruto (PIB) em toda sua cadeia produtiva, recebendo um destaque ainda maior na região de São Paulo e do Triângulo Mineiro (FUNDECITRUS, 2020). Por ser um setor de suma importância, há a necessidade de uma gestão efetiva de recursos que tenha controle sobre a produção para que haja uma melhor alocação de recursos. Normalmente, o método de derrça, que se trata do processo de retirar todos os frutos de uma árvore, é utilizado para a contagem do total de frutos de árvores selecionadas para se obter uma estimativa de produção atual e futura. A derrça não só é um processo destrutivo por necessitar remover todos os frutos da árvore a ser investigada, mas também é um processo altamente custoso em tempo e mão de obra (FUNDECITRUS, 2020). Devido a isso, há uma demanda para a pesquisa de métodos computacionais para realizar a contagem automática de forma eficiente em tempo e custo.

Dentre os diferentes métodos computacionais possíveis, há um destaque para a aplicação de redes neurais em várias áreas de pesquisa aplicada. Outros autores utilizaram redes neurais de diversas estruturas, como por exemplo a YOLO-V3 (CAMARGO NETO et al., 2019), e a Mask-RCNN (SANTOS et al., 2020). Neste trabalho, aplica-se a rede neural conhecida como Faster-RCNN (REN et al., 2017), e também por sua alta precisão na detecção e classificação de imagens sem comprometer o custo de tempo.

## 2 MATERIAIS E MÉTODOS

### 2.1 A REDE NEURAL FASTER-RCNN



RCNN, acrônimo de *Region-based Convolutional Neural Networks*, é uma arquitetura de rede neural que tem como objetivo agrupar e classificar regiões em uma imagem, por meio da obtenção de características que formam cada região de interesse, indicando se cada uma dessas regiões contém uma laranja, comparando esta resposta com a resposta da rede neural em um dado momento

O sistema de detecção das redes RCNN é dividido em dois módulos. No primeiro módulo, são encontradas regiões de interesse (regiões que apresentam a possibilidade de conter um objeto) em uma imagem, a partir de um algoritmo especializado nessa função. Originalmente, para encontrar as regiões de interesse, a rede RCNN utiliza o método de busca seletiva para agrupar pixels em tais regiões de interesse, método esse que é severamente custoso em tempo. Faster-RCNN é um aprimoramento sobre a RCNN que substitui o método de busca seletiva por uma rede para a geração de propostas de regiões, conhecida como Region Proposal Network (RPN), ganhando assim tempo de processamento. O algoritmo da RPN percorre a *feature map* (imagem processada por um filtro que tenta capturar os objetos da imagem) de cada imagem com janelas deslizantes, que por sua vez possuem ‘âncoras’ que encontram regiões de interesse em cada uma das instâncias das janelas. São aplicadas em cada janela diversas ‘âncoras’ de tamanho e proporção diferentes, para que objetos de tamanhos diferentes possam ser identificados. Assim, objetos na imagem podem ser encontrados independentemente de localização espacial e distância da câmera.

No segundo módulo de uma rede RCNN, encontra-se um detector que utiliza as regiões encontradas no primeiro módulo. Para acelerar o processo de treinamento e detecção, as redes de ambos módulos compartilham um conjunto de camadas convolucionais.

De acordo com Ren et. al. (2017), a rede Faster-RCNN pode ser treinada de dois modos diferentes. O primeiro método, que foi utilizado pelos autores em seu artigo, é chamado de método alternado, e se resume em alternar entre o treinamento da rede RPN e da rede detectora a cada quantidade de iterações previamente determinada. O segundo método, empregado na implementação do sistema descrito por este artigo, adota o nome de treinamento conjunto, por treinar ambas as redes a cada iteração do sistema, ou seja, as propostas geradas na rede RPN são imediatamente utilizadas para treinar a rede detectora na mesma iteração. Este método foi escolhido por apresentar uma velocidade de treinamento maior sem prejudicar a qualidade da detecção.

A aplicação utilizada como base para a utilização de funções de aprendizado de máquina é similar à aplicação inicialmente criada pelos desenvolvedores do artigo de Faster-RCNN, mas com preferência à utilização do TensorFlow em lugar do Caffe (CHEN; GUHPTA, 2020). Utilizamos a rede VGG16 (SIMONYAN; ZISSERMAN, 2014), que compartilha 5 camadas convolucionais entre a rede RPN e a rede RCNN.



## 2.2 FONTE E TRATAMENTO DOS DADOS

Para alimentar a rede RPN durante o treinamento, utilizamos um conjunto de 3065 imagens de árvores de laranja com marcações referentes às localizações das laranjas na imagem, cada imagem com um tamanho de 416x416 pixels. Essas imagens foram divididas em três grupos: a) 1832 imagens utilizadas para o treino; b) 203 para validar o treino; e c) 1030 que serviram como teste da rede após o treino. As imagens, em grande parte, foram fornecidas pelo Fundo de Defesa da Citricultura (Fundecitrus) - associação de produtores do estado de São Paulo, que promove o desenvolvimento da área citrícola, sendo uma de suas atribuições a execução do Programa de Estimativa de Safra (PES). Essas imagens foram tiradas de diversos dispositivos, como câmeras fotográficas e celulares, e sem nenhuma especificação de luminosidade ou resolução. Desta maneira, o banco de imagens possui uma maior diversidade nas imagens, para que o sistema esteja melhor preparado para as possíveis variações de ambiente e instrumentos que o usuário possua.

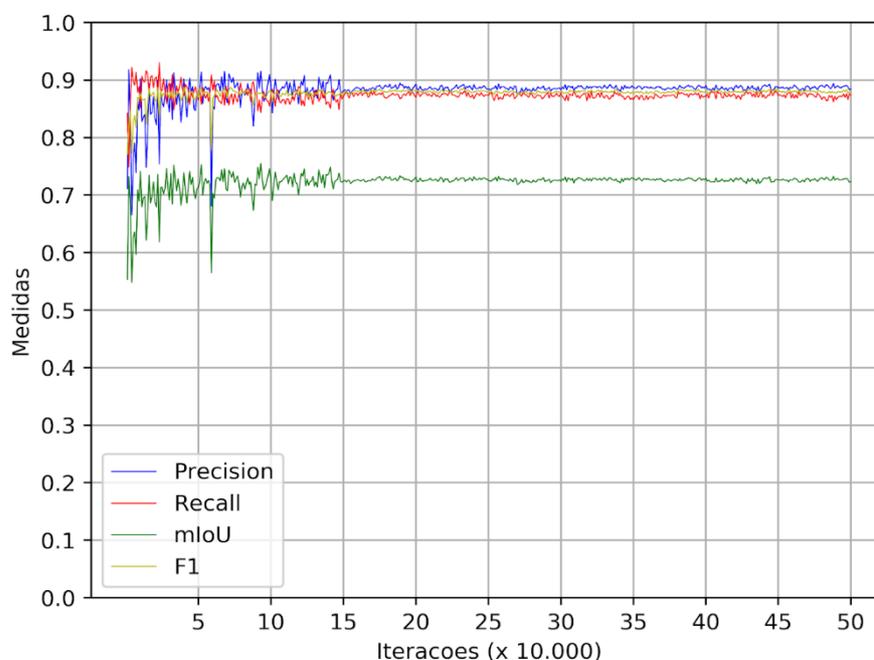
Por padrão da implementação do Faster-RCNN, cada imagem utilizada no treino é redimensionada para o tamanho de 600x600 pixels e duplicada por meio de um espelhamento horizontal. Foram utilizadas 'âncoras' de tamanhos 128, 256 e 512, nas proporções vertical-horizontal 1:1, 1:2 e 2:1. O tamanho do passo da janela deslizante é 16 pixels, o que, se compararmos com o tamanho das imagens antes do redimensionamento, equivale a aproximadamente 11 pixels da imagem original.

## 3 RESULTADOS E DISCUSSÃO

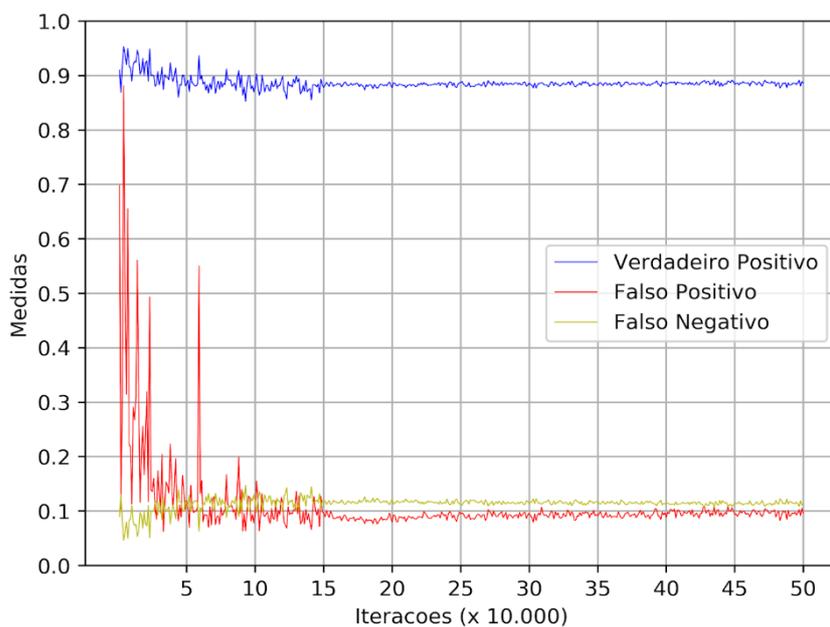
No treinamento da rede neural foram executadas 500.000 iterações, com uma rede completamente funcional sendo gerada a cada 1.000 iterações, o que significa que 500 redes foram geradas. Isso foi feito para possibilitar o teste da rede em diversos pontos, e verificar como ela está progredindo com a passagem de tempo. Originalmente iniciamos com uma taxa de aprendizado de 0,001, que após 150.000 iterações caiu para 0,0001.

Ao detectar objetos em imagens podemos classificá-los como verdadeiro positivo (VP), caso o objeto tenha sido acuradamente detectado; falso positivo (FP), caso o sistema encontre um objeto que não existe realmente; ou falso negativo (FN), caso o sistema não consiga encontrar um objeto na imagem. Calculamos esses valores para a detecção na base de validação e na base de testes, e os utilizamos para encontrar os índices de precisão ( $VP/(VP+FP)$ ), que indica quantas detecções corretas o modelo obteve, revocação ( $VP/(VP+FN)$ ), que indica a frequência com que o modelo identifica corretamente o objeto, e F1 ( $(2 * \text{Precisão} * \text{Revocação})/(\text{Precisão} + \text{Revocação})$ ), que

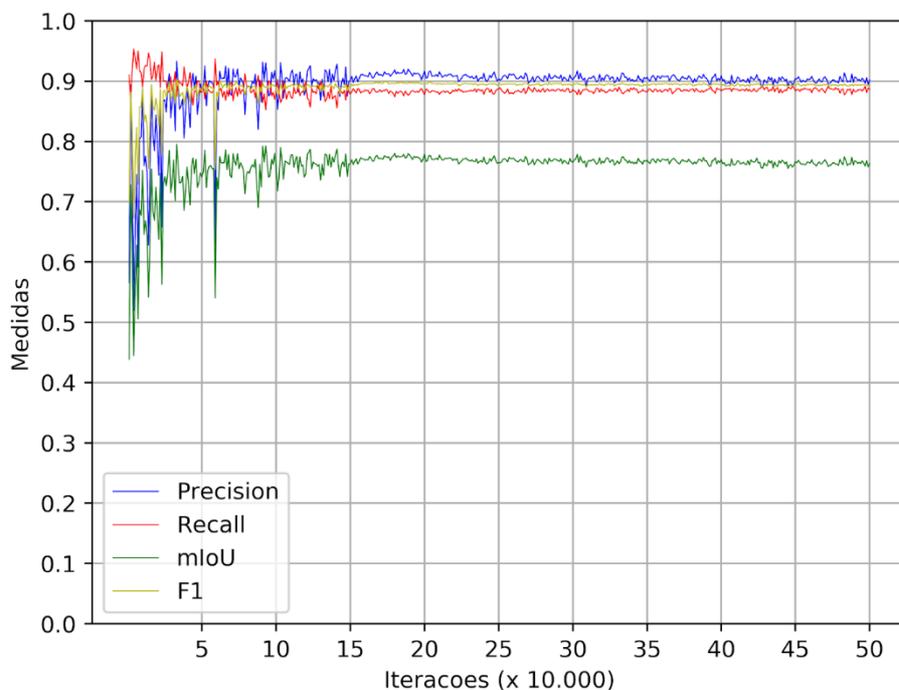
mede o compromisso entre a precisão e a revocação, além da intersecção-sobre-união média (Mean Intersection over Union, ou mIoU, em inglês), que é calculada pela intersecção entre a imagem, que é calculada pela intersecção entre a imagem original e a suposta imagem detectada correspondente, dividida pela união da área das duas. Estes indicadores são amplamente utilizados ao avaliar a qualidade de uma rede neural, conforme ilustrado nas Figuras 1, 2 e 3 encontradas abaixo. As Figuras 1 e 3 apresentam os valores de precisão, revocação, mIoU e F1 para as bases de validação e teste respectivamente, e a Figura 2 ilustra os valores de VP, FP e FN em relação ao total de laranjas existentes nas imagens originais (VP + FN). Estas medições foram feitas para IoU em 0.5 e Confiança em 0.75, ou seja, a área de sobreposição da imagem de um fruto tem que ser maior que 50% e a confiança que a rede tem nesta predição seja maior que 75%.



**Figura 1.** Análise da classificação dos resultados dos testes sobre a base de validação, para IoU em 0.5 e Confiança 0.75.



**Figura 2.** Análise da classificação dos resultados dos testes, para IoU em 0.5 e Confiança em 0.75. Eixo Y indica a porcentagem do total de frutos em cada categoria em relação ao total de frutos reais na base de teste.



**Figura 3.** Análise da performance dos resultados dos testes para IoU em 0.5 e Confiança em 0.75.



**Figuras 4 e 5.** Demonstração do resultado da rede para imagens digitais.

Um método para determinar a qualidade de uma rede neural é comparar os seus índices através de vários parâmetros de teste. Na Tabela 1, utilizamos o índice de confiança que determina a certeza com que a rede precisa acreditar que uma determinada detecção representa de fato uma laranja, e o limiar de IoU, que determina o valor mínimo do índice de IoU previamente descrito para saber se uma detecção foi correta ou errada. Com um limiar de confiança em 0.75 e um limiar de IoU em 0.5, temos que, na média, Precisão = 0,911; Revocação = 0,880 e F1 = 0,895.

## 4 CONCLUSÃO

Neste trabalho foi desenvolvida uma rede neural para a detecção de frutos em condições de imagem comuns, ou seja, imagens que são capturadas sem a necessidade de um dispositivo especial, tornando-o uma alternativa eficiente para as condições comuns do local de cultivo, em que a pessoa responsável por obter as imagens pode fazê-lo a partir de qualquer dispositivo, como uma câmera de celular. Os resultados apresentados foram muito promissores considerando o estado-da-arte da detecção de imagens, porém não apresentou um resultado tão diferente do que apresentado previamente por outras redes neurais para o mesmo trabalho, como a YOLO-v3. Várias variáveis do sistema Faster-RCNN foram alteradas para a implementação deste trabalho, mas é possível que ainda hajam variáveis a ser alteradas que poderiam melhorar o rendimento da rede neural.



**Tabela 1:** Índices encontrados após o teste da rede na sua última iteração, com variação em IoU (*Intersection over Union*) e limiar de confiança.

IoU	Confiança	Precisão	Revocação	F1	FP	FN
0,1	0,75	0,925	0,885	0,904	0,072	0,115
	0,5	0,906	0,891	0,899	0,092	0,109
	0,25	0,905	0,898	0,902	0,094	0,102
0,2	0,75	0,917	0,884	0,900	0,080	0,116
	0,5	0,905	0,890	0,898	0,093	0,110
	0,25	0,896	0,897	0,897	0,104	0,103
0,3	0,75	0,914	0,883	0,898	0,083	0,116
	0,5	0,905	0,890	0,897	0,093	0,110
	0,25	0,893	0,897	0,895	0,107	0,103
0,4	0,75	0,912	0,882	0,897	0,085	0,118
	0,5	0,903	0,888	0,896	0,095	0,112
	0,25	0,891	0,895	0,893	0,109	0,105
0,5	0,75	0,911	0,880	0,895	0,086	0,120
	0,5	0,902	0,887	0,894	0,096	0,113
	0,25	0,877	0,894	0,892	0,110	0,106
0,6	0,75	0,902	0,872	0,887	0,095	0,128
	0,5	0,893	0,878	0,885	0,105	0,122
	0,25	0,880	0,884	0,882	0,121	0,116
0,7	0,75	0,869	0,840	0,854	0,127	0,160
	0,5	0,859	0,844	0,851	0,139	0,156
	0,25	0,845	0,849	0,847	0,156	0,151
0,8	0,75	0,711	0,687	0,699	0,280	0,313
	0,5	0,702	0,689	0,695	0,293	0,311
	0,25	0,688	0,691	0,690	0,313	0,309
0,9	0,75	0,186	0,180	0,183	0,787	0,820
	0,5	0,182	0,179	0,181	0,804	0,821
	0,25	0,179	0,180	0,180	0,824	0,820

## 5 AGRADECIMENTOS

Os autores agradecem ao programa CNPq/PIBIC pela concessão da bolsa de Iniciação Científica para o aluno Luiz Miguel Cerqueira e à equipe do PES/Fundecitrus pela disponibilização das fotos de pés de laranja.

## 6 REFERÊNCIAS



CAMARGO NETO, J. et al. Uso de redes neurais convolucionais para detecção de laranjas verdes. CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais...** Disponível em: <<https://drive.google.com/file/d/1p7AjyrrNhfOQLTkyh05MGpU2KgvTKFpw/view>>. Acesso em: 30 set. 2020.

CHEN, X.; GUHPTA, A. **An implementation of Faster RCNN with study for region sampling.** Disponível em: <<https://arxiv.org/pdf/1702.02138.pdf>>. Acesso em: 30 set. de 2020.

FUNDECITRUS. **Estimativa da safra de laranja 2020/21 do cinturão citrícola de São Paulo e Triângulo/Sudoeste Mineiro:** sumário executivo. Araraquara, [2020]. 14 p. Disponível em: <[https://www.fundecitrus.com.br/pdf/pes\\_relatorios/2020\\_05\\_11\\_Sumario-Executivo-da-Estimativa-da-Safra-de-Laranja-2020-2021.pdf](https://www.fundecitrus.com.br/pdf/pes_relatorios/2020_05_11_Sumario-Executivo-da-Estimativa-da-Safra-de-Laranja-2020-2021.pdf)>. Acesso em: 30 set..2020.

REN, S. et al. Faster R-CNN: towards real-time object detection with region proposal networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence, New York**, v. 39, n. 6, p. 11307-1149, June 2017.

SANTOS, T. T. et al. Grape detection, segmentation and tracking using deep neural networks and three-dimensional association. **Computers and Electronics in Agriculture, Amsterdam**, v. 170, p. 1-17, Mar. 2020.

SIMONYAN, K.; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition.** 2015. Disponível em: <<https://arxiv.org/abs/1409.1556>>. Acesso em: 25 set. 2020.