USANDO A REDE NEURAL SSD PARA IDENTIFICAR FRUTOS VERDES EM POMARES DE LARANJA

Mariana Alves de **Sousa**[1]**;** Kleber Xavier Sampaio de **Souza**[2]; João **Camargo Neto**[3]; Sônia **Ternes**[4]; Inácio Henrique **Yano**[5]

Nº 21610

RESUMO *– A agropecuária é uma das mais importantes fontes de riqueza no Brasil. Dentro desse contexto, se destaca o cultivo das laranjas, principalmente na região de São Paulo e do Triângulo Mineiro. Infelizmente, o processo de estimativa da quantidade de frutos é custoso, assim, essa pesquisa tem como objetivo analisar por meio de visão computacional e de aprendizado profundo se essas técnicas geram resultados satisfatórios para identificar os frutos por fotografias. Caso apresente um bom desempenho, esta tecnologia poderá ser utilizada para prever a quantidade de laranjas em árvores.*

Palavras-chaves: Visão computacional, redes neurais, aprendizado profundo, rede SSD.

1 Autora, Bolsista CNPq (PIBIC): Graduação em Engenharia de Computação, Unicamp, Campinas - SP; mariana81899920@gmail.com.

2 Orientador: Pesquisador da Embrapa Informática Agropecuária, Campinas - SP; kleber.sampaio@embrapa.br.

3 Analista da Embrapa Informática Agropecuária, Campinas - SP.

4 Pesquisadora da Embrapa Informática Agropecuária, Campinas - SP.

5 Analista da Embrapa Informática Agropecuária, Campinas - SP.

ABSTRACT – *Agriculture is one of the most important sources of wealth in Brazil. In this context, the cultivation of oranges stands out, mainly in the region of São Paulo and the Triângulo Mineiro. Unfortunately, the fruit quantity estimation process is costly, so this research aims to analyze through computer vision and deep learning if these techniques generate satisfactory results to identify the fruit by images. If it performs well, this technology can be used to predict the amount of oranges on trees.*

Keywords: Computer vision, neural networks, deep learning, SSD network.

1. INTRODUÇÃO

Sabe-se que a laranja é o fruto mais cultivado no Brasil, sua produção contribui para mais da metade da produção mundial de suco de laranja. A cultura da laranja também contribui fortemente com o Produto Interno Bruto, dado que o Brasil produziu em 2019 cerca de 17 milhões de toneladas em uma área de cerca de 600 mil hectares, predominantemente em São Paulo e no Triângulo Mineiro (IBGE, 2021). Nessas regiões, é importante destacar a atuação do Fundo de Defesa da Citricultura (Fundecitrus), que é uma associação privada mantida por citricultores e indústrias de suco, responsável por promover o desenvolvimento sustentável do parque citrícola. Uma de suas principais responsabilidades é a previsão da safra anual de citrus plantados nessas regiões. Esse processo de estimativa é feito com o uso da derriça, que consiste na colheita de milhares de laranjas das suas árvores. Com esses frutos, então é realizada a contagem manual, para que assim seja feita uma estimativa da quantidade total de frutos atuais e futuros da plantação. Essa prática além da necessidade de mão de obra e de tempo, também exige um custo financeiro à instituição, pois o Fundecitrus indeniza o citricultor pela derriça realizada.

Diante disso, há uma necessidade de métodos que sejam mais eficientes em tempo e em custo. Com este objetivo, a pesquisa desenvolvida no âmbito do projeto “Estimativa da quantidade de frutos em pés de laranja por meio de inteligência computacional“ (TERNES, 2019) busca avaliar

se a identificação e a contagem de laranjas por meio de visão computacional e de redes neurais em imagens consegue ter resultados bons o suficiente para auxiliar na estimativa da quantidade de frutos. A rede neural que está sendo usada no estudo desenvolvido no presente trabalho é a rede Single Shot MultiBox Detector (SSD) (LIU et al., 2016), uma rede avançada, mais rápida e de alta acurácia se comparada a outros algoritmos de detecção de objetos, a qual realiza uma única passagem para detectar objetos presentes na imagem usando múltiplas caixas (Multibox Detector).

2. MATERIAL E MÉTODOS

2.1. A rede neural SSD

A rede Single Shot Multibox Detector (SSD) é uma rede neural que gera regiões de interesse em imagens, nas quais realiza a localização de objetos e um classificador para detectar os tipos de objetos nessas regiões de interesse (LIU et al., 2016). O termo Single Shot se refere à rede dar uma única passagem pela imagem para localizar e classificar os objetos, já o termo MultiBox Detector se refere a uma técnica de regressão da caixa delimitadora para a detecção de objetos, em que as coordenadas de um objeto detectado na região de interesse são regredidas para as coordenadas reais do objeto (ground truth).

A arquitetura da SSD tem como base a arquitetura de VGG-16 (SIMONYAN; ZISSERMAN, 2014), em que, em vez das camadas que são totalmente conectadas, temos camadas convolucionais auxiliares, com diferentes filtros. Essas camadas convolucionais reduzem o tamanho da imagem progressivamente, gerando features maps, que são representações das características dominantes da imagem em diferentes escalas.

O funcionamento da rede se dá com a aplicação da técnica de MultiBox na imagem, em que as caixas aplicadas estão em escalas e tamanhos diferentes, o que permite que objetos de tamanhos variados sejam identificados e classificados. Vale ressaltar então que passar a técnica de MultiBox em diferentes features maps aumenta a probabilidade dos objetos serem detectados. É importante salientar dois importantes índices no estudo de redes neurais, que é o IoU, que é uma métrica que diz respeito à razão entre a interseção entre a imagem detectada e a imagem original e a união entre esses dois, e a confiança, que confirma que de fato se trata de um objeto de interesse. Essas métricas servem para estimar a probabilidade de uma previsão ser verdadeira. Como o número de caixas geradas durante a passagem é alto, é aplicado então uma técnica chamada supressão não máxima na qual caixas em que a confiança e IoU estão abaixo de limites definidos são ignoradas. Além disso, essa técnica remove previsões que se sobrepõem. Isso

garante que apenas as previsões mais prováveis sejam retidas pela rede, enquanto os objetos com índices abaixo do necessário são removidos.

As funções de aprendizado de máquina utilizadas têm como base uma aplicação similar à aplicação desenvolvida pelos criadores do artigo de Single Shot MultiBox Detector, mas com preferência à utilização do PyTorch, que é uma biblioteca bastante utilizada de aprendizado de máquina, em lugar do Caffe, um framework de aprendizado profundo (VINODABABU; SENESHEN, 2020).

2.2. Fonte e tratamento de dados

Para nossa base de dados, tivemos um conjunto de 3065 imagens de árvores com os frutos de tamanho de 416x416 pixels e também anotações com a localização de todas as laranjas presentes na respectiva imagem. O Fundecitrus forneceu esses dados, obtidos sem especificações de luminosidade nem de resolução, o que permite um treinamento com uma base de dados bem diversificada, fazendo o sistema estar preparado para diversas variações no meio ambiente ou na forma de realizar a fotografia.

Esse conjunto de imagens foi dividido em 812 imagens para testagem, 200 imagens para validação e 2053 imagens para treinamento. As anotações inicialmente continham os pontos de máximo e pontos de mínimo, em ambos os eixos. A rede no entanto precisava do ponto médio, largura e altura dos objetos em porcentagem pelo tamanho da imagem, esses novos dados foram calculados e atualizados nas anotações.

2.3. Métricas

As classificações para validar a qualidade de uma rede mais usadas são verdadeiro positivo (VP), que é caso o objeto realmente exista e o detectamos, falso positivo (FP), que é caso o objeto não exista e o detectamos ou ainda falso negativo (FN), caso o objeto exista, mas não o detectamos.

Com esses valores, calculamos as principais métricas de avaliação do modelo, que são a precisão, que avalia quão preciso é o nosso modelo em relação ao total de positivos previstos (TP/(TP+FP)), a revocação, que está relacionado a quanto realmente dos positivos reais nosso modelo detecta (TP/(TP+FN)), o F1, que avalia a relação entre a precisão e a revocação (2*precisão*revocação / (precisão+revocação)) e, por fim, a intersecção-sobre-união média (mIoU),

que é calculada através da intersecção entre a imagem original e a imagem detectada correspondente, dividida pela união da área das duas, como já discutido anteriormente.

3. RESULTADOS E DISCUSSÃO

Durante o treinamento da rede foram executadas cerca de 500.000 iterações, com uma rede gerada a cada 1.000 iterações, gerando assim 500 redes funcionais. O objetivo da criação dessas redes a cada 1.000 iterações foi avaliar e testar a qualidade da rede em diversos pontos. No início, começamos com uma taxa de aprendizado de 0,001 e configuramos para que após 150.000 iterações a taxa caia para 0,0001. Também após 300.000 iterações, ela caia novamente para 0,00001. A redução da taxa de aprendizado após um certo número de iterações é importante, porque o algoritmo de convergência é um sistema de otimização e, como tal, pode ficar preso a um mínimo local. O passo menor, determinado pela taxa de aprendizado, permite que o algoritmo procure um ponto ótimo ainda melhor do que o que ele se encontra atualmente.

As métricas citadas foram calculadas para a base de dados de validação e de teste. Nas Figuras 1 e 2, temos os valores de precisão, revocação, F1 e mIoU para a base de dados de validação e de teste, respectivamente. Na Figura 3, temos os valores de VP, FP e FN em razão da quantidade de laranjas totais da base de teste. Nas figuras, os dados informados foram desenvolvidos com um IoU em 0,5 e uma confiança em 0,75. Para essas medidas, temos que a precisão é de 0,917, a revocação é de 0,867 e o F1 é de 0,891. Na Figura 4, temos uma demonstração do resultado final da rede.

Os resultados apresentaram desempenho semelhante aos obtidos em trabalhos de pesquisa anteriores, no âmbito dos quais foram testadas outras arquiteturas de redes neurais convolutivas aplicadas ao mesmo problema (CAMARGO NETO et al., 2019; CERQUEIRA et al., 2020). No entanto, é importante ressaltar que a rede SSD tem uma vantagem por ser mais rápida (LIU et al., 2016).

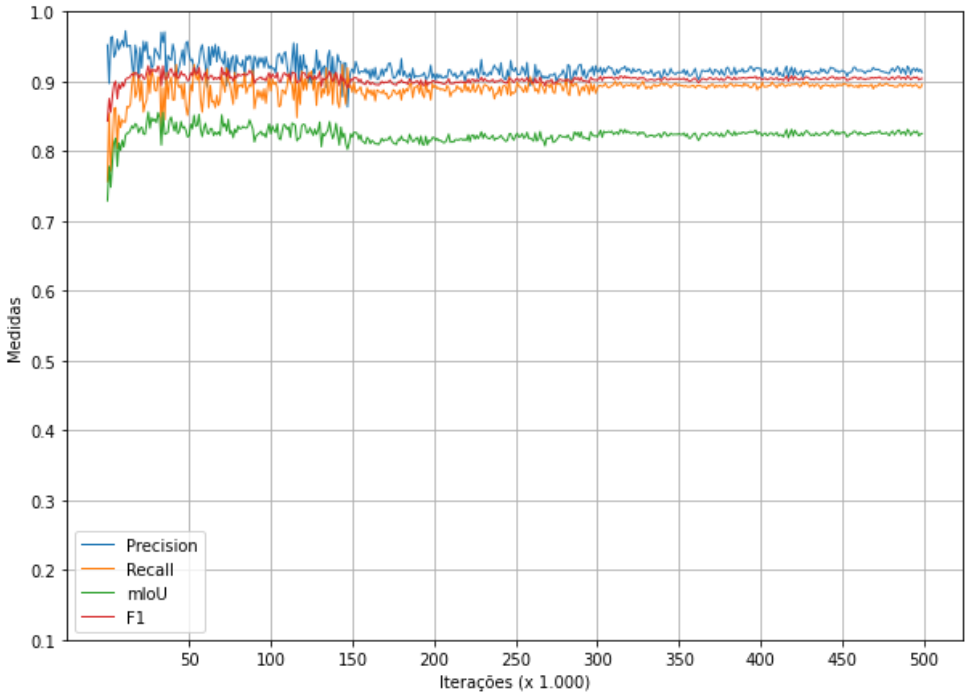

Figura 1. Análise das medidas resultantes para a base de dados de validação

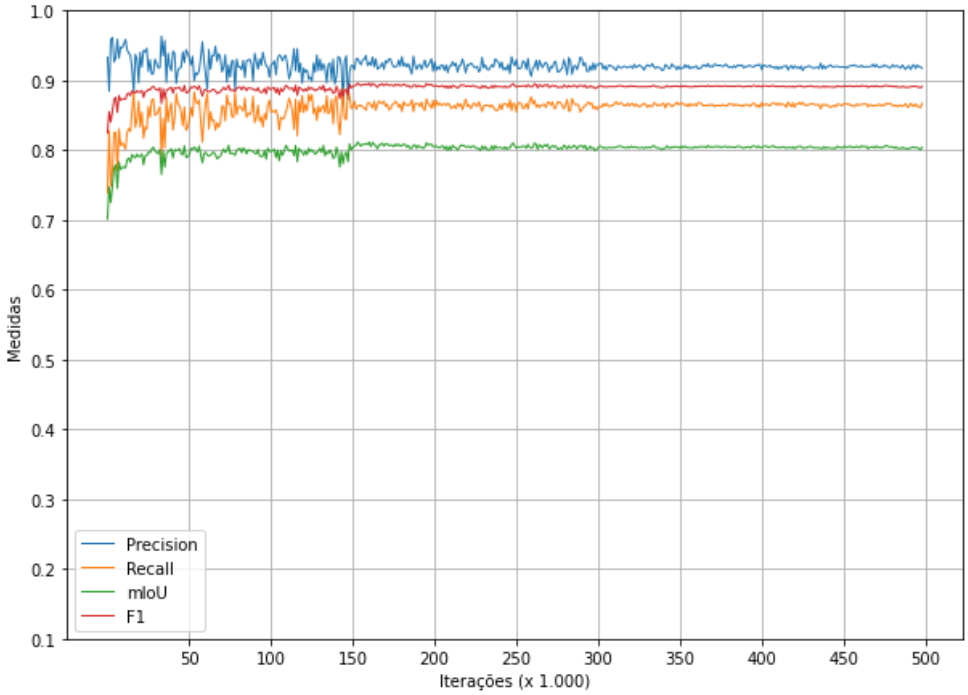

Figura 2. Análise das medidas resultantes para a base de dados de teste.

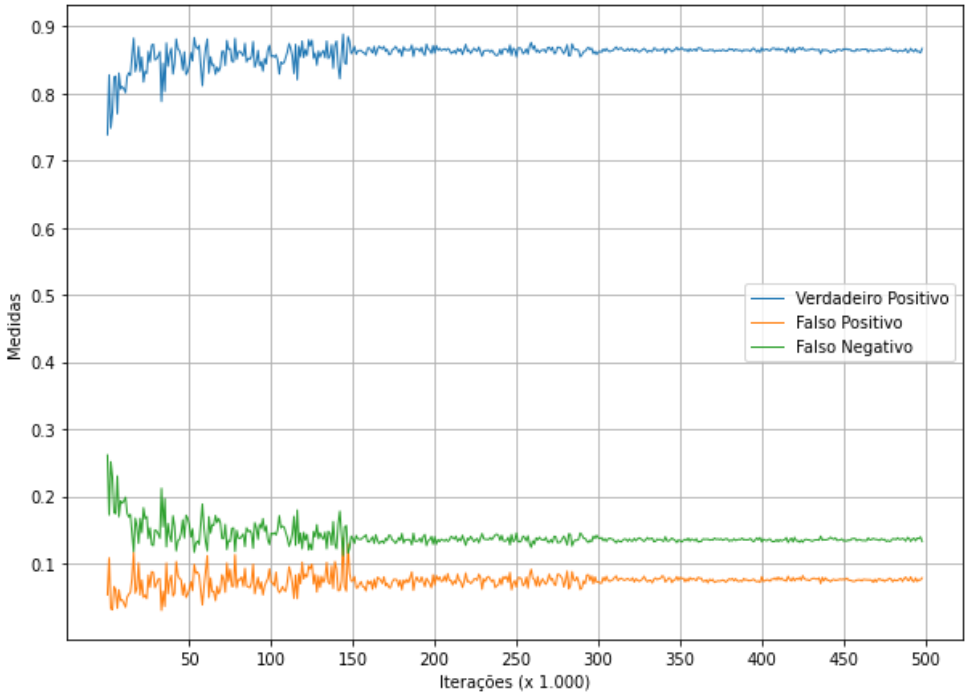

Figura 3. Análise da classificação dos resultados na base de dados de testes.

Figura 4. Demonstração do resultado da rede para imagens digitais. Fotos: Fundecitrus, processadas por Mariana Sousa.

Na Tabela 1, podemos avaliar a qualidade da rede neural comparando os valores das métricas com diferentes parâmetros de IoU e de confiança medidos para o último modelo gerado, que, por ter sido treinado durante 500 mil interações, é o modelo de melhor desempenho.

Tabela 1. Resultados encontrados após o teste da rede em seu último modelo gerado, com variação em IoU e limiar de confiança.

IoU	Confiança	Precisão	Recovação	F1	FP	FN
0,1	0,75	0,923	0,799	0,857	0,067	0,200
	0,50	0,906	0,813	0,857	0,084	0,187
	0,25	0,823	0,823	0,850	0,114	0,177
0,2	0,75	0,922	0,858	0,889	0,073	0,142
	0,50	0,903	0,870	0,886	0,093	0,130
	0,25	0,875	0,882	0,878	0,126	0,118
0,3	0,75	0,921	0,866	0,893	0,074	0,134
	0,50	0,902	0,878	0,890	0,096	0,122
	0,25	0,875	0,890	0,881	0,130	0,110
0,4	0,75	0,920	0,867	0,893	0,075	0,133
	0,50	0,901	0,879	0,890	0,097	0,121
	0,25	0,870	0,892	0,881	0,133	0,108
0,5	0,75	0,917	0,867	0,891	0,078	0,133
	0,50	0,896	0,880	0,888	0,102	0,120
	0,25	0,863	0,892	0,877	0,141	0,108
0,6	0,75	0,912	0,867	0,889	0,084	0,133
	0,50	0,886	0,880	0,883	0,113	0,120
	0,25	0,851	0,892	0,871	0,157	0,108
0,7	0,75	0,900	0,868	0,884	0,096	0,132
	0,50	0,871	0,880	0,875	0,131	0,120
	0,25	0,830	0,893	0,860	0,183	0,107
0,8	0,75	0,865	0,868	0,866	0,135	0,132
	0,50	0,818	0,881	0,848	0,196	0,119
	0,25	0,756	0,894	0,819	0,289	0,106
0,9	0,75	0,654	0,869	0,746	0,460	0,131
	0,50	0,568	0,881	0,691	0,670	0,119
	0,25	0,472	0,894	0,618	1,000	0,106

4. CONCLUSÃO

A pesquisa realizada desenvolveu uma rede neural para detectar laranjas em imagens capturadas sem nenhuma especificação ou requisito de alta qualidade, o que é ideal, pois assim não é necessário condições especiais ou uso de dispositivos sofisticados para a fotografia. Os resultados obtidos pela rede foram bastante promissores, o que significa um avanço na forma de prever a quantidade de frutos. Com diferentes índices de IoU e confiança conseguimos identificar que a rede funciona em uma margem de valores muito bons, tendo um F1 de 0,891. Apesar de os resultados terem sido muito próximos aos obtidos com as outras arquiteturas testadas, ainda assim a rede SSD apresenta vantagem sobre as demais por ser uma rede com execução muito mais rápida, uma vez que ela realiza uma única passagem pela imagem.

5. AGRADECIMENTOS

Devemos agradecimentos ao PIBIC/CNPq, pela concessão da bolsa de Iniciação Científica à aluna Mariana Alves de Sousa, e à equipe do PES/Fundecitrus pela disponibilização das imagens.

6. REFERÊNCIAS

CAMARGO NETO, J; TERNES, S.; SOUZA, K. X. S. de; YANO, I. H.; QUEIROS, L. R. Uso de redes neurais convolucionais para detecção de laranjas no campo. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais...** Ponta Grossa: SBIAGRO, 2019. p. 312-321. Organizadores: Maria Fernanda Moura, Jayme Garcia Arnal Barbedo, Alaine Margarete Guimarães, Valter Castelhano de Oliveira. SBIAgro 2019. . Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1125722/1/PC-Redes-neurais-SBIAGRO-2019.pdf>. Acesso em: 10 jun. 2021.

CERQUEIRA, L. M.; SOUZA, K. X. S. de; TERNES, S.; CAMARGO NETO, J. **Usando a rede neural Faster-RCNN para identificar frutos verdes em pomares de laranja.** In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 14., 2020. **Anais...** Campinas: Embrapa Informática Agropecuária, 2020. p. 1-9. Evento online. CIIC 2020. No 20605. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1127731/1/RE20605-CIIC-2020.pdf>. Acesso em: 10 jun. 2021.

IBGE. **Produção agrícola - lavoura permanente:** ano 2019. Disponível em: <https://cidades.ibge.gov.br/brasil/pesquisa/15/0>. Acesso em: 1 abr. 2021.

LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C.Y.; BERG, A.C. SSD: Single shot multibox detector. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 14., 2016, Amsterdam. **Proceedings…** Cham: Springer: 2016. p. 21–37. (Lecture notes in computer science, 9907). DOI: 10.1007/978-3-319-46448-02.

SIMONYAN, K.; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition**. 2014. Disponível em: <https://arxiv.org/abs/1409.1556>. Acesso em: 13 jun. 2021.

TERNES, S. **Estimativa da quantidade de frutos em pés de laranja por meio de inteligência computacional**. Campinas: Centro Nacional de Pesquisa Tecnológica em Informática para Agricultura, 2019. 25p. (Embrapa. Tipo I – Pesquisa e Desenvolvimento – Código SEG 10.18.03.016.00.00). Projeto em andamento: eContaFruto. Acesso em: 13 jun. 2021.

VINODABABU, S; SENESHEN, M. **Pytorch tutorial to object detection**. Disponível em: <https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Object-Detection#concepts>. Acesso em: 10 abr. 2020.