

Detecção de Desfolha de Soja Utilizando Redes Neurais Convolucionais

Patrik Olã Bressan
Federal Institute of Mato Grosso do Sul
Jardim - MS, Brazil.
patrik.bressan@ifms.edu.br

Wesley Nunes Gonçalves
Federal University of Mato Grosso do Sul
Campo Grande - MS, Brazil.
wesley.goncalves@ufms.br

Abstract— The agribusiness represents a significant portion of the global economy. In Brazil, agribusiness has a significant share of the country's economy and represented 21.6% of GDP in 2017. To increase productivity, proper management of a crop, including pest control, is of vital importance. Annually, plant pests cause losses of 20% to 40% of production. For this reason, it is important to monitor the level of defoliation to take preventive actions. Therefore, in this work an automatic methodology is proposed using Convolutional Neural Networks, to detect the level of defoliation from leaf images in the soybean crop. In addition to detecting the presence of defoliation, the proposed methodology also provides the affected regions of the leaf through the segmentation of the image. Experimental results showed 83% accuracy using the proposed methodology versus 60% of SegNet CNN. The results are promising considering that the images were captured in the field, which presents challenges such as lighting, stages of development, scale, among others.

O agronegócio representa uma parcela significativa da economia global. No Brasil, o agronegócio tem uma expressiva participação na economia do país e representou 21,6% do PIB em 2017. Para aumentar a produtividade, é de vital importância o gerenciamento adequado de uma cultura, incluindo o controle de pragas. Anualmente, as pragas de plantas causam perdas de 20% a 40% da produção. Desta forma, é importante monitorar o nível de desfolha para tomar ações preventivas. Portanto, neste trabalho é proposto uma metodologia automática utilizando Redes Neurais Convolucionais, para detectar o nível de desfolhamento a partir de imagens na cultura da soja. Além de detectar a presença da desfolha, a metodologia proposta também fornece as regiões afetadas da folha por meio da segmentação da imagem. Os resultados experimentais mostraram 83% de precisão usando a metodologia proposta versus 60% da CNN SegNet. Os resultados são promissores considerando que as imagens foram capturadas no campo, que apresenta os desafios como iluminação, estágios de desenvolvimento, escala, entre outros.

I. INTRODUÇÃO

O agronegócio é uma das atividades mais importantes para os países em desenvolvimento, contribuindo para a produção de alimentos e insumos. No Brasil, o agronegócio tem uma participação expressiva na economia do país e representou 21,6% do PIB em 2017 [1]. Apesar da importância e do crescimento da produtividade nos últimos anos, a agricultura em 2050 terá que produzir quase 50% a mais do que em 2012 para atender à demanda da população mundial [2]. No agronegócio existem produções que se destacam, como a soja, um grão empregado na alimentação humana e animal no mundo todo. A estimativa mundial da safra de soja realizada em 2017/2018 foi de aproximadamente 336,699 milhões de toneladas [1]. O cultivo dessa planta tem exigido um nível maior de conhecimento técnico e acompanhamento.

Para aumentar a produtividade, é de vital importância o gerenciamento adequado de uma cultura, incluindo o controle

de pragas. As pragas de plantas causam perdas anuais de 20% à 40% da produção [2]. As perdas causadas por insetos invasores custam à economia global cerca de US\$70 bilhões por ano [3]. A principal consequência é a herbivoria e a lesão que resulta em uma redução funcional da superfície total da planta e consequentemente impacta na produção final da cultura. Logo é importante monitorar o nível de desfolha para tomar ações preventivas. Geralmente o nível de desfolha é estimado visualmente por especialistas usando um guia [4], ou o método de contagem de grades [5]. Além disso, dispositivos de medição, como o LI-3000A e o LI-3100 [6], estão disponíveis para estimar automaticamente a área foliar. Embora a área foliar possa ser correlacionada com o nível de desfolha, ela não estima a área danificada com precisão, especialmente quando ocorre nas bordas. Portanto, a estimativa do nível de desfolha usando as técnicas mencionadas é uma tarefa demorada e subjetiva.

Visando melhorias, métodos computacionais foram propostos em [7] e [8], disponíveis para estimar a área foliar, embora não estimem o nível de desfolha. Machado et al. [9] apresentou uma aplicação móvel que estima o nível de desfolha usando curvas de Bezier para restaurar a borda original da folha, mas ainda é necessária a intervenção de um especialista, para desenhar a borda da folha usando a curva de Bezier. Silva et al. [10] adaptaram redes neurais convolucionais (CNN) para a regressão, estimando o nível de desfolha com métodos para gerar imagens com desfolhação sintética para o treinamento. Embora os bons resultados alcançados, esses trabalhos usaram imagens capturadas em laboratório com restrições e somente uma folha por imagem em um fundo branco.

Para contornar esses problemas, este artigo propõe um método baseado em CNN para detectar pixels pertencentes a desfolha em ambientes sem restrições. Para detectar a desfolha, o método proposto usa a arquitetura da SegNet [11], uma CNN proposta para segmentação de imagens. As regiões afetadas pela desfolha são menores que as regiões não afetadas (área total da imagem), e dessa forma a SegNet não é capaz de generalizar com precisão as regiões com desfolha. Assim, propomos o treinamento da SegNet com pesos diferentes para os pixels da desfolha e do fundo durante o *backpropagation*. No método proposto o valor de perda de cada pixel é ponderado de acordo com sua respectiva classe, de forma a aumentar a importância dos pixels de desfolha.

O método proposto obteve resultados de 83.18% de acurácia contra 60.51% da SegNet. Esses resultados foram obtidos em uma base de imagens capturada em lavouras com diferentes iluminações, estágios de desenvolvimento, escalas e desafios para os sistemas de visão computacional.

II. MATERIAIS E MÉTODOS

Nesta seção são apresentados os materiais utilizados no desenvolvimento do trabalho, bem como os métodos desenvolvidos.

A. Conjunto de Imagens

O conjunto de dados foi obtido por meio do PlantVillage [12], que contém várias fotografias tiradas por telefone celular. Algumas folhas possuem gotas de chuva, insetos, dedos das pessoas que fotografaram, doenças e também a desfolha. Para compor a base de imagens, foram identificadas as 325 imagens que continham desfolha. Existem imagens em que a desfolha já representava aproximadamente 22% da área total da folha e, por outro lado, existem algumas imagens em que a desfolha é mínima, como 0.22%. Dessa forma, a metodologia proposta é avaliada em vários níveis de desfolha. Cada imagem foi manualmente anotada para avaliar os resultados da detecção da desfolha, conforme mostra a Figura 1. É importante enfatizar que as imagens foram tiradas no campo e apresentam diversos desafios.

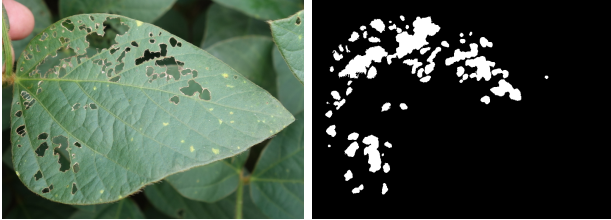


Fig. 1: Exemplo de imagens do banco de dados.

O banco de imagens foi dividido aleatoriamente em três conjuntos, do total de 325 imagens, 222 são para o treinamento, 51 para a validação e 52 para os testes. Dado que a grande maioria das imagens possuem pouca desfolha, a divisão das imagens nos conjuntos foi estratificada obedecendo a porcentagem de desfolha que varia entre 0.22% à 21.82% conforme apresentado na Tabela I. Dessa forma, os conjuntos de treinamento, validação e teste possuem exemplos que vão desde desfolha baixa até exemplos com desfolha severa.

B. SegNet

A SegNet [11] consiste em uma rede de codificação e outra de decodificação correspondente, seguida de uma função classificadora de pixels, conforme a Figura 2. A SegNet proposta possui apenas duas classes uma para o fundo e outra para a desfolha. As duas partes da SegNet são descritas a seguir.

- **Codificador:** tem como objetivo construir um mapa de características reduzido, por meio de camadas de convolução, ReLU e *max pooling*. A imagem de entrada é transformada em um volume composto por

TABLE I: Organização e divisão do banco de imagens.

Porcentagem de Desfolha (%)	Número de Imagens	Treinamento-Validação-Teste (70%-15%-15%)
0.2255 - 1.3057	78	54-12-12
1.3057 - 2.3860	70	48-11-11
2.3860 - 3.4662	56	40-8-8
3.4662 - 4.5464	27	19-4-4
4.5464 - 5.6266	28	20-4-4
5.6266 - 6.7068	13	9-2-2
6.7068 - 7.7870	11	7-2-2
7.7870 - 8.8672	10	6-2-2
8.8672 - 9.9474	10	6-2-2
9.9474 - 11.0276	5	3-1-1
11.0276 - 12.1078	3	1-1-1
12.1078 - 13.1880	2	1-0-1
13.1880 - 14.2682	7	5-1-1
14.2682 - 21.8297	5	3-1-1

características que descrevem o conteúdo visual da imagem. Essas características são aprendidas assim como nas redes neurais convolucionais. As camadas de convolução possuem filtros de tamanho 3×3 enquanto que as camadas de *max pooling* possuem 2×2 . Os pesos iniciais das camadas são inicializados por meio de pesos pré-treinados da VGG16 [13] no grande conjunto de dados de classificação de objetos ImageNet [14].

- **Decodificador:** reconstrução do mapa de características obtido do codificador. Composto por 13 camadas, onde é realizado *upsampling* no mapa de características. As camadas de *upsampling* tem como objetivo aumentar a resolução do mapa de características de tal forma que, ao final do decodificador, a resolução do mapa de características é a mesma da imagem de entrada. Na última camada do decodificador, utiliza-se a função Softmax que classifica cada pixel de forma independente.

Codificador e Decodificador Convencional

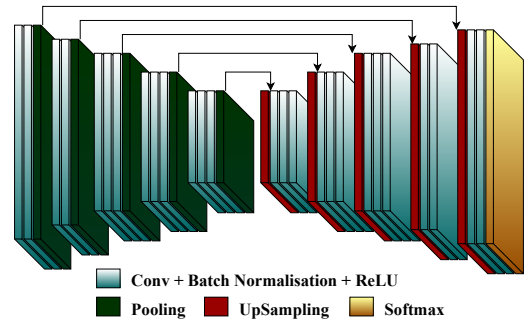


Fig. 2: Ilustração da arquitetura SegNet composta por duas partes: codificador e decodificador.

Para treinar a SegNet, é utilizado o *backpropagation* e uma função de perda chamada entropia cruzada (*cross-entropy*) [15] que mede a concordância entre as predições e as imagens anotadas. O valor de perda $L_{x,y}$ para um pixel x, y é calculado conforme a Equação 1, em que M é o número de classes, $c_{x,y}^j$ é o indicador binário (0 ou 1) se o pixel (x, y) pertence a classe j , e $p_{x,y}^j$ é a probabilidade predita pela SegNet do pixel x, y pertencer a classe j . O valor de perda dos pixels são usados para atualizar os pesos da SegNet, sendo que, quanto maior o valor de perda, maior é a mudança dos pesos durante o *backpropagation*.

$$L_{x,y} = - \sum_{j=1}^M c_{x,y}^j \log(p_{x,y}^j) \quad (1)$$

C. Método Proposto

Na detecção da desfolha, o principal problema é o desbalanceamento das classes, isto é, existem muito mais pixels de fundo (área total da imagem) do que pixels pertencentes à desfolha. Dessa forma, os valores de perda dos pixels do fundo acabam dominando o aprendizado e, portanto, o padrão da desfolha não é aprendido de forma eficaz. Dessa maneira, propomos ponderar o valor de perda dos pixels de acordo com a sua classe na imagem anotada. O valor de perda proposto nesse trabalho é calculado pela Equação 2, onde ω^j é o peso da classe j calculado previamente.

$$L_{x,y} = - \sum_{j=1}^M \omega^j c_{x,y}^j \log(p_{x,y}^j) \quad (2)$$

Podemos interpretar essa ponderação da seguinte forma: quanto maior o peso da classe ω^j , maior será o valor de perda independentemente da predição da SegNet. Assim, com o valor de perda maior, os pesos das camadas tendem a se ajustar mais para a classe com maior peso, diminuindo o desbalanceamento. Se usarmos ω^j igual a 1 para todas as classes, o treinamento ocorre da forma tradicional. Para determinar o peso da classe fundo e da desfolha, utilizamos o conjunto de treinamento conforme a Equação 3: quanto menor o número de pixels de uma determinada classe, maior será o resultado da equação e consequentemente o seu peso.

$$\omega^j = \frac{m}{n_c * n^j} \quad (3)$$

onde m é o número de pixels de todas as imagens de treinamento, n_c é o número de classes do problema (nesse trabalho são duas classes, fundo e desfolha) e n^j é o número de pixels nas imagens de treinamento que pertencem à classe j .

D. Experimentos

Os experimentos foram realizados tanto para o método proposto quanto para a SegNet, utilizando as imagens com o tamanho de 1024×1024 pixels. Para o método proposto, a Tabela II apresenta o peso das classes calculados com o conjunto de treinamento. Esses valores são usados na ponderação da Equação 2. Observa-se que o valor do peso (ω) da classe *Fundo* é inferior ao da classe *Desfolha*, pois a classe *Fundo* possui um número muito maior de pixels. Para o treinamento, foi utilizado o gradiente descendente estocástico (*Stochastic Gradient Descent* - SGD) com a taxa de aprendizado de 10^{-3} e *momentum* igual a 0.9 por 150 épocas.

TABLE II: Pesos das classes calculados a partir do conjunto de treinamento.

Tamanho da imagem	ω	
	<i>Fundo</i>	<i>Desfolha</i>
1024×1024	0.508	29.525

Durante os experimentos, usamos o aumento de dados. Essa técnica de aumento de dados amplia o conjunto de treinamento por meio de transformações específicas do domínio. Para dados de imagem, transformações comumente usadas incluem corte aleatório, perturbação aleatória de brilho, saturação, matiz e contraste.

Para quantificar os resultados do método proposto e da SegNet, utilizou-se a acurácia pixel-a-pixel (APP) que calcula a média de pixels corretamente classificados e a intersecção sobre a união (*Intersection over Union* - IoU). Essas duas métricas são amplamente utilizadas na avaliação de algoritmos de segmentação [15].

III. RESULTADOS E DISCUSSÕES

Inicialmente, o método proposto e a SegNet foram treinados usando o conjunto de treinamento. Para determinar os hiper parâmetros (e.g., taxa de aprendizado e número de épocas) e evitar o *overfitting*, avaliamos a generalização dos métodos treinados no conjunto de validação. Os resultados de ambos métodos sugeriram pouco *overfitting*, pois a diferença da acurácia no conjunto de treinamento e validação foi menor que 2%.

Após o treinamento, avaliamos os métodos no conjunto de teste usando as métricas APP e IoU conforme apresentado na Tabela III. As métricas foram calculadas considerando somente os pixels pertencentes à desfolha. Com relação aos pixels do fundo, ambos os métodos apresentaram APP superior a 0,99, mostrando que os pixels do fundo são detectados com precisão. Com o método proposto, os pixels da desfolha são detectados com 83,1% de acurácia, enquanto que somente 60,5% dos pixels da desfolha são detectados com a SegNet. Com relação a IoU, podemos observar que o método proposto também superou a SegNet, com valores de 0,549 contra 0,528.

TABLE III: APP e IoU do método proposto e da SegNet.

Método	APP	IoU
SegNet	0,605($\pm 0,23$)	0,528($\pm 0,22$)
Método Proposto	0,831($\pm 0,23$)	0,549($\pm 0,19$)

Buscando verificar estatisticamente se o método proposto é superior, foi realizado o teste de Tukey sobre APP e IoU, onde os valores p foram, respectivamente, 0,0001 e 0,96. Assim, podemos inferir que o método proposto é estatisticamente superior a SegNet em relação a acurácia pixel-a-pixel, e que não há diferenças estatísticas em relação a IoU. A Figura 3 apresenta os diagramas de caixa e bigode (*boxplot*) da acurácia, onde observa-se que o “fio do bigode” da SegNet abrange praticamente todo o conjunto de resultados, indicando que existe uma grande variação nesses valores e, a mediana do método proposto (0,98) é superior a mediana da SegNet.

A Figura 4 apresenta exemplos do conjunto de teste após a detecção da desfolha usando o método proposto e a SegNet. Os pixels de desfolha são pintados de vermelho no *Ground Truth*, e os pixels classificados como desfolha são

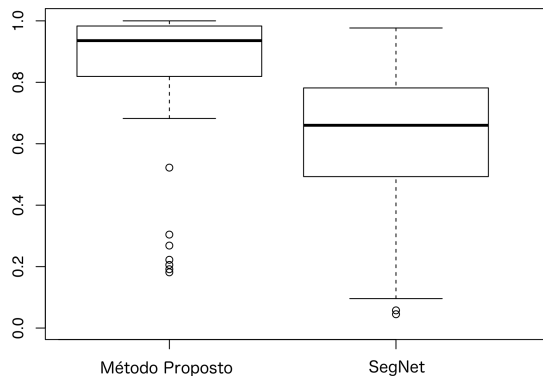


Fig. 3: *Boxplot* de acurácia.

pintados de verde. Esses resultados qualitativos mostram a robustez e acurácia do método proposto em comparação com a metodologia tradicional. É importante enfatizar que esses resultados foram obtidos em uma base de imagens capturada em lavouras (ambiente externo) com diferentes iluminações, escalas e desafios para os sistemas de visão computacional.

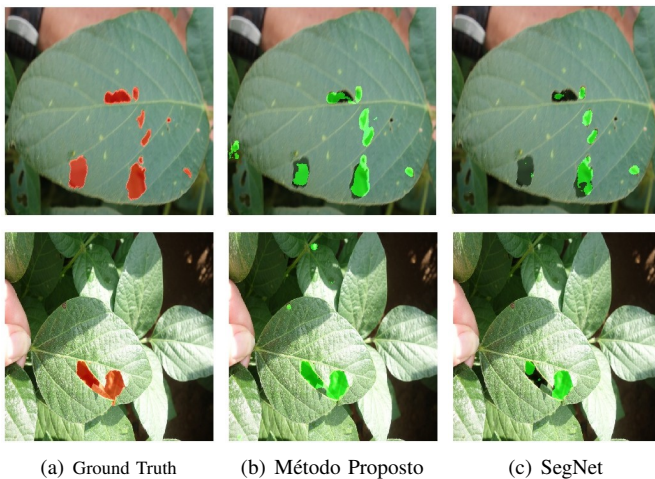


Fig. 4: Exemplos de detecções de desfolha usando o método proposto e a SegNet.

IV. CONCLUSÃO

A detecção da desfolha em folhas de soja é uma importante etapa para aumentar a produtividade das lavouras. Esse trabalho apresentou um método que usa a ponderação das classes durante o treinamento para detecção da desfolha. Com a metodologia proposta, acurácia de 83% foi obtida em um conjunto de imagens capturadas em uma lavoura que apresenta os desafios como iluminação, estágios de desenvolvimento, escala, etc. O teste estatístico corroborou a superioridade em reconhecer a desfolha da metodologia proposta. Contudo, a técnica proposta não consegue reconhecer a desfolha quando esta representa uma grande porcentagem da área total da folha, concentrando-se em uma única área. Outras formas de ponderação de classes estão sendo desenvolvidas e testadas,

bem como o aumento no conjunto de treinamento. Como trabalhos futuros pretendemos utilizar as técnicas de multi-escala, pois a combinação de recursos em várias escalas pode melhorar o desempenho e ajudar a reconhecer a desfolha severa.

AGRADECIMENTOS

Agredecemos a FUNDECT - Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul; ao CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico; a UFMS - Fundação Universidade Federal de Mato Grosso do Sul e; a NVIDIA pela placa de vídeo GeForce GTX TITAN X doada e utilizada nesse trabalho.

REFERENCES

- [1] "Soja em números (safra 2017/2018)," <https://www.embrapa.br/en/soja/cultivos/soja/dados-economicos>, accessed: 2018-11-28.
- [2] FAO, *The future of food and agriculture - Trends and challenges*. Rome: Food and Agriculture Organization of the United Nations, 2017.
- [3] C. J. Bradshaw, B. A. Leroy, C. A. Bellard, D. A. Roiz, C. A. Albert, A. A. Fournier, M. A. Barbet-Massin, J.-M. A. Salles, F. A. Simard, and F. A. Courchamp, "Massive yet grossly underestimated global costs of invasive insects," *Nature Communications*, vol. 7, p. 12986, 2016.
- [4] M. Kogan, S. Turnipseed, B. Shepard, E. B. De Oliveira, and A. Borgo, "Pilot insect pest management program for soybean in southern Brazil," *Journal of Economic Entomology*, vol. 70, no. 5, pp. 659-663, 1977.
- [5] J. Kvet and J. Marshall, "Assessment of leaf area and other assimilating plant surfaces," *Plant Photosynthetic Production. Manual of Methods*, pp. 517-555, 1971.
- [6] H. Barclay, J. Trofymow, and R. Leach, "Assessing bias from boles in calculating leaf area index in immature Douglas-fir with the li-cor canopy analyzer," *Agricultural and Forest Meteorology*, vol. 100, no. 2, pp. 255 - 260, 2000.
- [7] A. Gong, X. Wu, Z. Qiu, and Y. He, "A handheld device for leaf area measurement," *Computers and Electronics in Agriculture*, vol. 98, pp. 74-80, 2013.
- [8] X. Fan, K. Kawamura, W. Guo, T. Xuan, J. Lim, N. Yuba, Y. Kurokawa, T. Obitsu, R. Lv, Y. Tsumiyama, T. Yasuda, and Z. Wang, "A simple visible and near-infrared (v-nir) camera system for monitoring the leaf area index and growth stage of Italian ryegrass," *Computers and Electronics in Agriculture*, vol. 144, pp. 314-323, 2018.
- [9] B. B. Machado, J. P. Orue, M. S. Arruda, C. V. Santos, D. S. Sarath, W. N. Gonçalves, G. G. Silva, H. Pistori, A. R. Roel, and J. F. Rodrigues-Jr, "Bioleaf: a professional mobile application to measure foliar damage caused by insect herbivory," *Computers and Electronics in Agriculture*, vol. 129, pp. 44-55, 2016.
- [10] L. A. d. Silva, P. O. Bressan, D. N. Gonçalves, D. M. Freitas, B. B. Machado, and W. N. Gonçalves, "Soybean defoliation using convolutional neural networks and synthetic images," *WVC - Workshop de Visão Computacional*, 2017.
- [11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.
- [12] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing," *CoRR*, vol. abs/1511.08060, 2015.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
- [14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," *CoRR*, vol. abs/1409.0575, 2014. [Online]. Available: <http://arxiv.org/abs/1409.0575>
- [15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038, 2014.