

## COMPARAÇÃO DO DESEMPENHO DE DIFERENTES REDES NEURAIIS NA DETECÇÃO DE DOENÇA NO ALGODOEIRO

ALEXANDRE DE FREITAS CARRASCOSA<sup>1</sup>, ARTHUR AKIRA MAMIYA<sup>2</sup>,  
RENILDO LUIZ MION<sup>3</sup>

<sup>1</sup>Graduando em Engenharia Agrícola e Ambiental, Instituto de Ciências Agrárias e Tecnológicas, UFR, Rondonópolis/MT, (66) 9 9908-7650 alexandrecarrascosa@gmail.com

<sup>2</sup>Prof. Dr., Instituto de Ciências Exatas e Naturais, UFR, Rondonópolis/MT, mamiya@ufr.edu

<sup>3</sup>Eng. Agrícola, Prof. Dr., Instituto de Ciências Agrárias e Tecnológicas, UFR, Rondonópolis/MT

Apresentado no  
XLIX Congresso Brasileiro de Engenharia Agrícola - CONBEA 2020  
23 a 25 de novembro de 2020 - Congresso On-line

**RESUMO:** Atualmente, muitas redes neurais têm sido desenvolvidas, cada uma com uma especificidade diferente, na qual técnicas são aplicadas para atender a determinada demanda. A agricultura tem acompanhado este desenvolvimento e incorporando tais tecnologias as suas atividades, contudo um desafio é a variabilidade das situações, principalmente relacionado a variação do ambiente. Neste contexto, a presente pesquisa tem por objetivo comparar o comportamento das redes YOLO V3 e Mask RCNN para a detecção da mancha de *Stemphylium* no algodoeiro, a partir de imagens obtidas em condições de campo. Foi feita investigação do efeito da taxa de aprendizagem do YoloV3 nesta aplicação específica, e da proporção de imagens para treinamento e validação, analisando os resultados através da *mean Average Precision (mAP)*, na qual os resultados foram de 0,76 para o Yolo e 0,472 para o Mask RCNN. Os resultados foram utilizados para investigar imagens em diferentes condições de desfoque, que mostraram queda no *mAP* conforme o aumento do desfoque.

**PALAVRAS-CHAVE:** algodão, inteligência artificial, segmentação de imagem.

## COMPARISON OF PERFORMANCE OF DIFFERENT NEURAL NETWORKS IN THE DETECTION OF DISEASE IN COTTON

**ABSTRACT:** Currently, many neural networks are being developed, each with a different specificity, in which techniques are applied to meet a certain demand. Agriculture has accompanied this development and incorporated such technologies into its activities, however a challenge is the variability of circumstances, mainly related to the variation of the environment. In this context, this research aims to compare the behavior of the YOLO V3 and Mask RCNN networks for the detection of *Stemphylium* spot on cotton, from images obtained under field conditions. The effect of learning rate on YoloV3 for this specific application was investigated, as well as the proportion of images used for training and validation, analyzing the results through the mean Average Precision (mAP), in which the results were 0,76 for Yolo and 0,472 for o Mask RCNN. The results were compared in the detection of images in different levels of blurring, that showed a decrease in mAP as the blur increased.

**KEYWORDS:** artificial intelligence, cotton, image segmentation.

**INTRODUÇÃO:** A escolha da cultura do algodão (*Gossypium L.*) está relacionada a sua importância para o estado do Mato Grosso, sendo este responsável por 69,54% da produção nacional (CONAB, 2020). Contudo o plantio do algodão tem altos custos, principalmente em relação aos insumos, segundo uma tabela de custos disponibilizada pela Companhia Nacional de Abastecimento (CONAB) os agrotóxicos correspondem a 40,43% dos custos de produção. Logo ao se desenvolver ferramentas que utilizam o aprendizado profundo, busca-se reduzir estes custos por meio de uma aplicação de insumos em plantas que apresentem os sintomas, principalmente em fases iniciais, desta maneira, diversas pesquisas quanto a viabilidade de uso das redes neurais na detecção de doenças vêm ocorrendo (ARSENOVIC, 2019; BARBEDO, 2018; FERENTINOS, 2018), assim como outras pesquisas que investigaram o uso de diferentes técnicas e redes neurais aplicadas à detecção de doenças na cultura do algodão (CHOWDHARY, 2020; UDAWANT, 2019; GULHANE, 2011). Esta pesquisa também abordou fatores como importância da qualidade e quantidade do conjunto de dados (BARBEDO, 2019; MOHANTY et al., 2016, GULHANE, 2011), alguns trabalhos ressaltam em diversas situações a influência das questões de qualidade do banco de imagens no resultado da rede (BARBEDO, 2019; MOHANTY et al., 2016). através dos diferentes testes com conjuntos de imagens menores e maiores. Ambas redes são de detecção, contudo a Mask RCNN, além da detecção trabalha com o processo de segmentação da imagem, criando uma máscara conforme as classes aprendidas. Esta pesquisa teve por objetivo analisar a viabilidade de uso do aprendizado profundo, através da comparação no desempenho de diferentes redes neurais, sendo elas Yolo V3 (REDMON & FARHADI, 2018) e Mask RCNN (HE et al. 2017), na detecção de doença na cultura do algodão (*Gossypium L.*)

**METODOLOGIA:** As imagens de algodão foram obtidas em uma área experimental da Universidade Federal de Rondonópolis, utilizando um *smartphone* (9.6 M *Pixels*), no meio e fim do período matutino. Estas apresentaram lesões na superfície das folhas, e adotando o manual da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), foram classificadas como mancha de *Stemphyium* (CHITARRA, 2014). As pesquisas foram realizadas utilizando a plataforma *Google Colaboratory*, para realizar os treinamentos e análises das métricas propostas. Na pesquisa realizada por Kumbhar, (2009) obtiveram-se bons resultados de acurácia, 80%, utilizando pequenos conjuntos de imagens, 141 imagens. Desta maneira esta pesquisa iniciou com um conjunto de aproximadamente 150 imagens, que foram sendo aumentadas gradativamente. Para uso com a rede Yolo V3 na implementação de Bochkovski (2016), estas imagens foram redimensionadas para 416x416 *pixels*. Para utilizar o conjunto de imagens existe a necessidade de realizar a marcação das imagens, para isso utilizou-se da ferramenta *LabelImage* (TZUTALIN, 2015). Normalmente este processo ocorre de maneira manual, o que o torna muito demorado, podendo levar dias, dependendo do tamanho do conjunto. As 150 imagens foram marcadas manualmente e com os pesos obtidos no treinamento foi realizada a marcação das imagens restantes, utilizando a biblioteca *OpenCV* (BRADSKI, 2000) para marcar parcialmente e posteriormente realizar as correções manualmente. Durante as etapas iniciais do treinamento observaram-se baixos valores para o *mAP*, como método de correção foi realizado um refinamento no conjunto de imagens, buscando seleção de imagens de melhor qualidade. Para o treinamento, iniciou-se a partir de pesos pré-treinados (REDMON & FARHADI, 2018; ABDULLA, 2017) para o conjunto de imagens MSCoco (LIN et al., 2014). Vale ressaltar que foram utilizadas duas arquiteturas para a rede Yolo, sendo elas o YoloV3-tiny e o YoloV3 *Darknet-53*, contendo 9 e 53 camadas convolucionais, respectivamente. Os testes com o YoloV3-tiny foram divididos na variação da taxa de aprendizagem entre os valores de 0,001, 0,01 e 0,1 e a variação

dos conjuntos de treinamento e validação para 90:10, 80:20, 50:50 e 20:80. As configurações utilizadas para o treinamento da rede Yolo foram *batch size* 4, taxa de aprendizagem (*learning rate*) de 0,01, *burn in* de 1000 e 90% das imagens para treinamento e 10% validação, com algumas aumentações, sendo elas transformações afins, aumento e redução do brilho e variação do matiz de cores. As configurações utilizadas para o treinamento da rede Mask, foram *batch size* 4, taxa de aprendizagem (*learning rate*) 0,001 e 0,01 e 90% das imagens para treinamento e 10% validação, sendo utilizada a implementação de Abdulla (2017).

**RESULTADOS E DISCUSSÕES:** As redes utilizam diferentes funções para realizar os cálculos da função de perda (*loss function*), sendo a média da entropia cruzada (*average binary cross-entropy*)(HE, 2018) para o Mask RCNN e a soma residual dos quadrados (*sum of squared error*)(REDMON, 2018) para o YoloV3, contudo neste trabalho ambas serão abordadas como função de perda. Um dos hiperparâmetros alterados foi a Taxa de Aprendizagem, estando relacionada à velocidade com que a rede irá aprender a detectar o objeto de interesse dentro de cada imagem. Os resultados da função de perda e *mean average precision (mAP)* serão utilizados para comparar as duas redes, contudo alguns outros indicadores foram obtidos para a rede. Na primeira etapa o treinamento foi realizado com um conjunto de 324 imagens, com 80% para treinamento e 20% validação, variando a taxa e aprendizagem de 0,001 até 0,1. Inicialmente os testes foram realizados com o YoloV3-*tiny*; começando as análises com a taxa de aprendizagem em 0,001, obteve-se um valor de perda de 4,88 e *mAP* de 0,32. Aumentando a taxa para 0,01, a função de perda caiu para 3,33 e o *mAP* subiu para 0,332, com pequenas oscilações no valor. Com a taxa de aprendizagem de 0,1, a perda subiu para 7,12 e o *mAP* reduziu para 0,142. O treinamento foi repetido, porém com 90% das imagens no treinamento e 10% na validação, com a taxa de aprendizagem em 0,001 a perda apresentou um valor de 1,55 e *mAP* de 0,296, em 0,01 a perda ficou em 1,0384 e *mAP* de 0,316 para a taxa de aprendizagem em 0,1 o valor da função de perda divergiu, e o *mAP* reduziu-se a 0. Os baixos valores de *mAP* poderiam ter relação com a quantidade ou qualidade das imagens, desta maneira foi realizada a expansão do conjunto de dados para 664 imagens, com isso houve uma melhora considerável no valor de *mAP*, aumentando de 0,38 para 0,76. Com a melhora obtida nesta adição de novas imagens e com base nos resultados anteriores para taxa de aprendizagem buscou-se a melhor proporção treinamento/validação. Os testes foram divididos em 4 variações a) 90% treinamento 10% validação, b) 80% treinamento 20% validação, c) 50% treinamento 50% validação e d) 20% treinamento 80% validação. Com o treinamento, observou-se que a função de perda diminuía conforme o conjunto de treinamento se tornava menor em relação a validação, desta maneira obtiveram-se os valores de a) 1,71, b) 1,41, c) 1,04 e d) 0,57. Apenas com o valor da função de perda do treinamento não é possível saber a precisão da rede, logo adotou-se o *mAP* para compreender seu comportamento. Para as mesmas condições os valores máximos de *mAP* foram: 0,76, 0,72, 0,66 e 0,58, para a), b), c) e d) respectivamente. Logo, analisando esta métrica é possível considerar que a proporção 90/10 apresentou melhor qualidade. Algumas outras medidas foram avaliadas, e servem como informações para analisar o comportamento da rede quando expostas a diferentes proporções de dados para treinamento e validação, conforme Tabela 1. É possível notar uma prevalência de valores acima de 0,70 para 90:10, em comparação a proporção 20:80 que mesmo apresentando uma perda baixa os outros valores também são menores.

TABELA 1 - Valores aproximados para diferentes proporções de treinamento e validação para rede YoloV3-tiny

Proporções	Loss	mAP	Recall	F,Score	Precisão	AP
90/10	1,71	0,70	0,74	0,75	0,75	0,70
80/20	1,41	0,56	0,65	0,65	0,65	0,55
50/50	1,04	0,54	0,64	0,65	0,65	0,54
20/80	0,57	0,47	0,54	0,60	0,65	0,46

A partir destes resultados, adotando a taxa de aprendizagem de 0,01 e a proporção 90:10, foi realizado o teste com a arquitetura de 53 camadas, para avaliar o comportamento da rede. Com o *darknet-53* ao invés do *tiny* o melhor valor de *mAP* foi de 0,78. Em relação ao *tiny* a principal diferença está relacionada a *loss function* que apresentou um valor menor na arquitetura completa. Este treinamento demandou mais tempo para ser finalizado, pois a arquitetura requer mais processamento. Além da necessidade de processamento os pesos resultantes ocupam um espaço maior; cada peso da arquitetura com 9 camadas ocupa 33 Mb enquanto a arquitetura com 53 camadas ocupa 239 Mb. Os valores para outras métricas podem ser observados na tabela 2.

TABELA 2 - Resultados de diversas métricas para o YoloV3 com 53 camadas convolucionais

	Recall	Precision	Ap	F <sub>1</sub> - Score
Máximos	0,81	0,79	0,78	0,80
Médio	0,79	0,761	0,70	0,77

Após a avaliação dos resultados apresentados para a rede YoloV3, foram adaptados os mesmos parâmetros para a rede Mask RCNN. Logo para a rede Mask foi utilizado um *batch* de 4, taxa de aprendizagem (*learning rate*) de 0,01 (com *burn in*). Pode-se assim identificar o comportamento do *mAP* para diferentes padrões de imagem variando o tamanho das imagens conforme apresentado na Figura 1 e realizando a comparação das inferências.

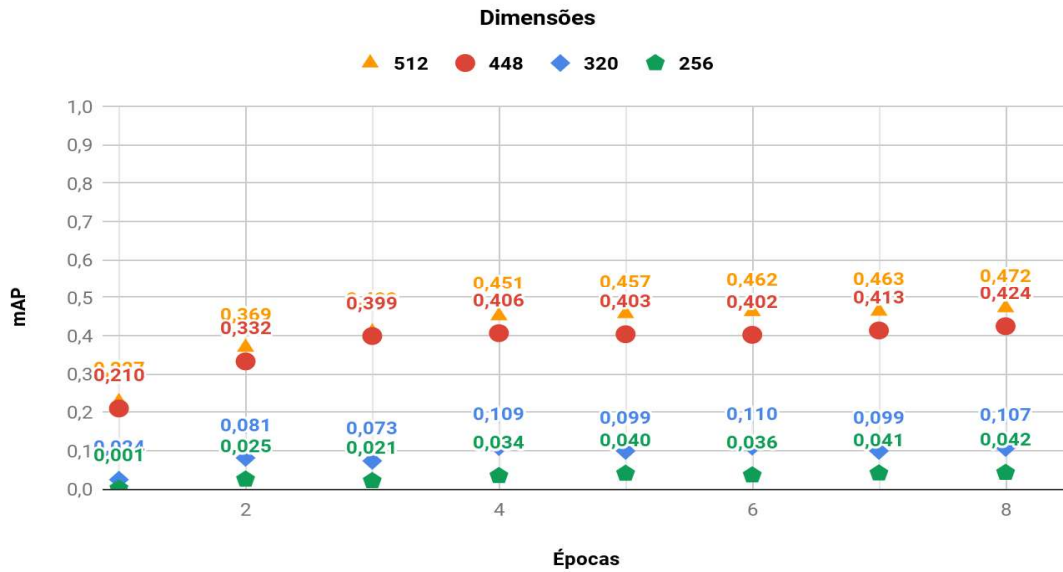


FIGURA 1 - Resultados de  $mAP$  para a rede Mask RCNN para diferentes dimensões de imagem. (512, 448, 320 e 256 pixels)

A Figura 1 mostra o comportamento do  $mAP$ , conforme a variação das dimensões das imagens, observou-se que ocorre uma pequena variação entre as dimensões de 512x512 e 448x448, enquanto nas dimensões de 320x320 e 256x256 ocorre uma redução para 0,107 e 0,042. A diminuição nos percentuais pode estar relacionada à dificuldade de identificar as manchas em resoluções muito pequenas. Como etapa final, para comparar os resultados foram realizadas inferências com um conjunto de imagens, conforme Figura 2, a detecção das redes (b e c) em comparação as marcações manuais (a).

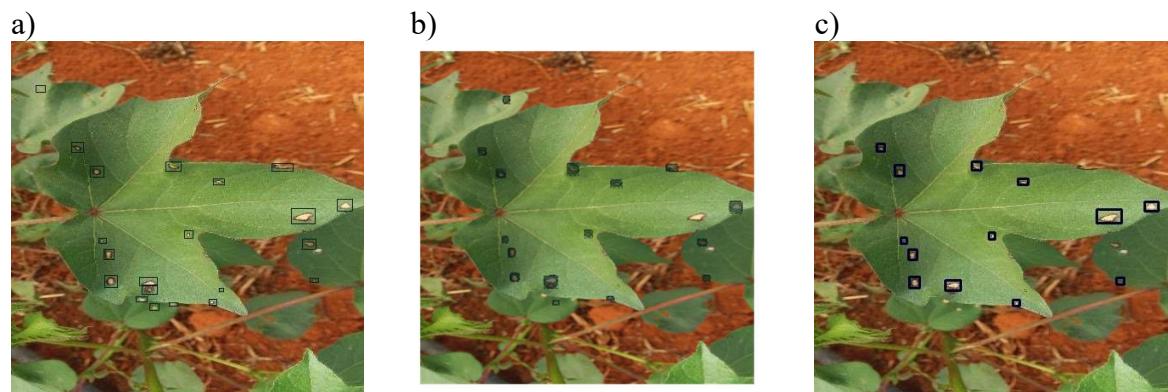


FIGURA 2 - Resultado de inferência, a) Marcações manuais, b) Mask RCNN e c) YoloV3-tiny

Como abordado neste artigo e em alguns trabalhos (BARBEDO, 2019; MOHANTY, 2016) a qualidade das imagens pode afetar a qualidade da rede, com base nisto, aplicou-se uma escala de suavização gaussiana em um novo conjunto de 54 imagens, que não foram utilizadas em nenhuma etapa anterior, e com base nesta variação observou-se a alteração no valor do  $mAP$ , conforme observado na Figura 4. Devido os resultados de  $mAP$  do Yolo terem sido maiores utilizou-se esta rede para realizar a inferência e verificar o comportamento das detecções conforme ocorre o aumento no valor da suavização

gaussiana, Figura 3. Verifica-se que para cada variação no desfoque a capacidade de detecção da rede diminui, Figura 3, sendo possível avaliar com base no gráfico da figura 4 uma drástica redução nos valores de  $mAP$ .

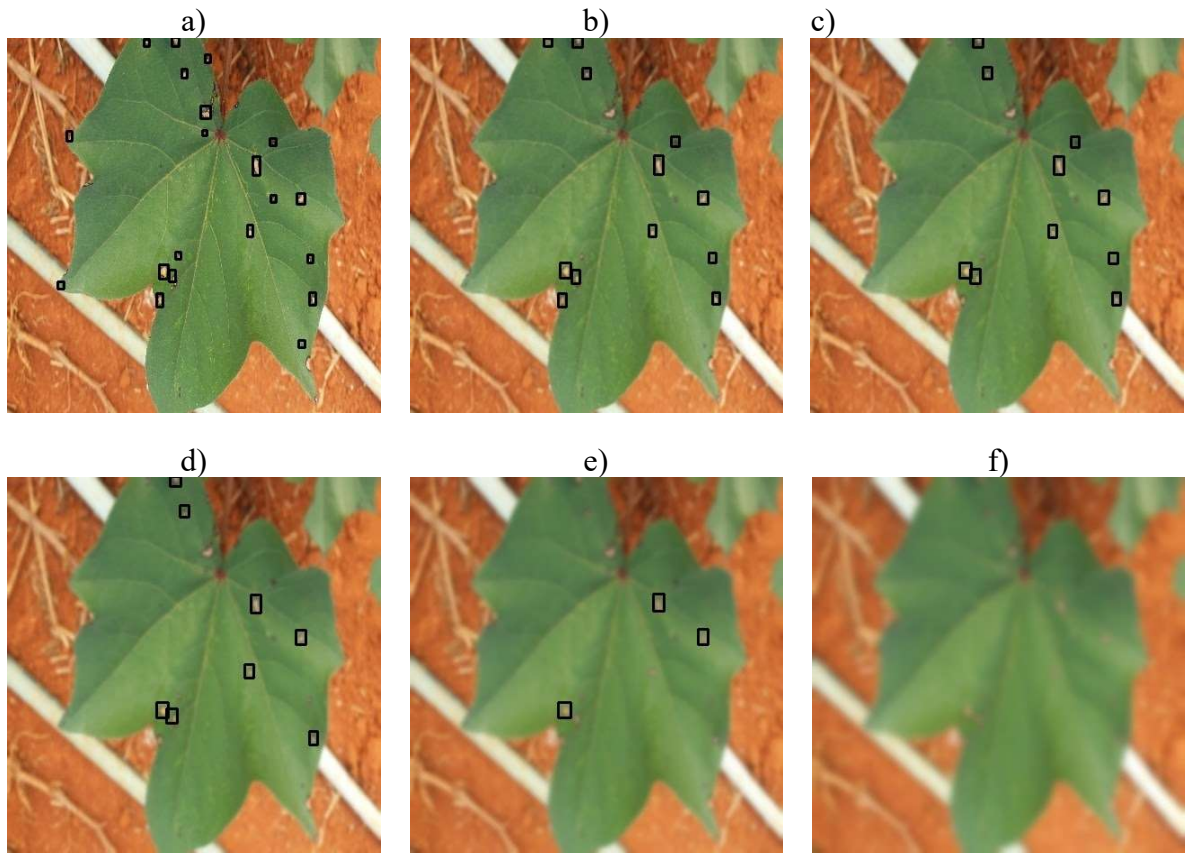


FIGURA 3 - Inferência com YoloV3-*tiny* para variações da suavização gaussiana (*blur*) para os valores de *kernel* a) Original, b) 5, c) 7, d) 9, e) 15, f) 25

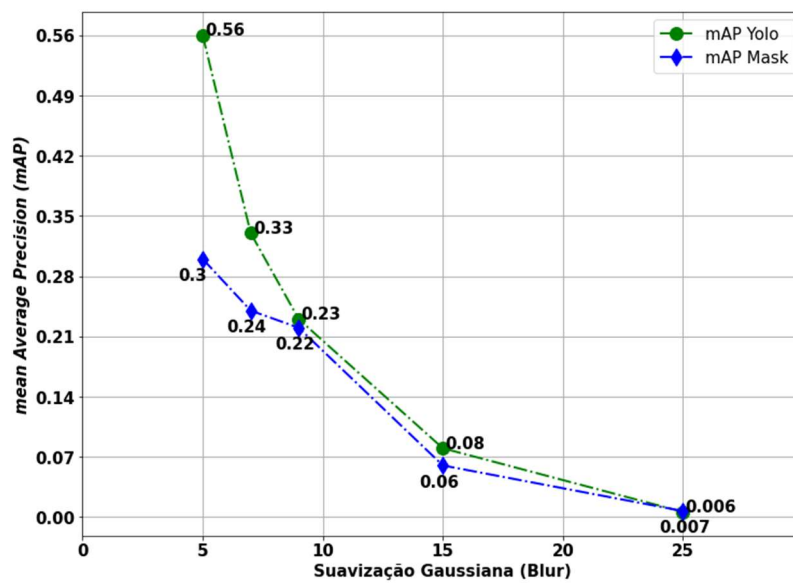


FIGURA 4 - Valores de  $mAP$  para variações da Suavização Gaussiana (*Blur*)

Um dos aspectos mais importantes abordados na pesquisa que trouxeram os melhores resultados foi o aumento e refinamento do conjunto de dados, demonstrando que a qualidade e quantidade de imagens são fatores determinantes para um bom resultado. Com as devidas mudanças o valor do *mAP* aumentou consideravelmente, mostrando a importância deste fator, principalmente devido à variabilidade de características apresentadas nas imagens. Uma das características mais importantes do YoloV3-*tiny*, em relação ao Mask RCNN, é uma menor necessidade de processamento, sendo este um aspecto de grande impacto nas implementações do Aprendizado Profundo no campo, viabilizando a utilização em campo com menor custo computacional. Um dos pontos que diferencia as duas redes é o processo de segmentação e criação de máscara do Mask RCNN que para este caso não apresentou uma grande importância, contudo em pesquisas futuras é possível analisar a eficiência da máscara em demarcar doenças de diferentes formatos. Logo tomando como base a métrica *mAP* pode-se concluir que para detecção da mancha de *Stemphylium* no algodoeiro com um conjunto de dados reduzidos, a rede YoloV3-*tiny* é eficiente, com bons resultados e menor requerimento de recursos computacionais. Com base nos testes realizados com a suavização gaussiana (*blur*) é possível notar que com os pequenos aumentos no desfoque a qualidade da rede é comprometida, e pode-se observar na Figura 3 que conforme os valores do *kernel* para suavização aumentam a capacidade de detecção diminui.

**CONCLUSÃO:** A melhor taxa de aprendizagem foi de 0,01 e com 90% das imagens para treinamento e 10% para validação, para o Yolo-*tiny*. Com base nos resultados obtidos pode-se concluir que a rede Yolo apresentou melhor resultado, considerando o *mAP*, na qual a arquitetura *Darknet-53* foi um pouco superior, contudo para as condições desta pesquisa, quanto a recursos computacionais, a arquitetura *tiny* se mostrou suficientemente adequada. A adição de novas imagens se mostrou um diferencial, recomendando-se então para *mAP* superior a 50% pelo menos um conjunto de 554 imagens, 499 para treinamento e 55 para validação. As redes YoloV3 e Mask RCNN perderam a capacidade de detecção conforme ocorreu um aumento no desfoque.

**AGRADECIMENTOS:** Ao programa de Pós-Graduação em Engenharia Agrícola (PGEAgri), da Universidade Federal de Rondonópolis (UFR) por permitir o uso da área cultivada para realizar a captura de imagens.

#### **REFERÊNCIAS:**

ABDULLA, W.. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. [https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN). 2017

ARSENOVIC, Marko et al. Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry*, v. 11, n. 7, p. 939, 2019.

BARBEDO, J. G. A. Factors influencing the use of deep learning for plant disease recognition. *Biosystems engineering*, v. 172, p. 84-91, 2018.

BARBEDO, J. G. A.. Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*, 180, 96-107, 2019.

BOCHKOVSKIY, A. <https://github.com/AlexeyAB/darknet>, 2016.

BRADSKI.G. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.Xperience AI, OpenCV. Copyright (C) 2019-2020, all rights reserved.

CHITARRA, L. G. Identificação e controle das principais doenças do algodoeiro. **Embrapa Algodão-Fôlder/Folheto/Cartilha** (INFOTECA-E), 2014.

CHOWDHARY, K. Praveen et al. Detection of Cotton Leaf Diseases Using Image Processing and Machine Learning Approach. **International Journal of Research in Engineering, Science and Management**, v. 3, p. 127-129, 2020.

CONAB. Algodão no Brasil. **Associação Brasileira dos Produtores de Algodão**, Agosto de 2020. Disponível em: <https://www.abrapa.com.br/Paginas/Dados/Algod%C3%A3o%20no%20Brasil.aspx>. acesso em ago. 2020.

FERENTINOS, K. P. Deep learning models for plant disease detection and diagnosis. **Computers and Electronics in Agriculture**, v. 145, p. 311-318, 2018.

GULHANE, Viraj A.; GURJAR, Ajay A. Detection of diseases on cotton leaves and its possible diagnosis. **International Journal of Image Processing (IJIP)**, v. 5, n. 5, p. 590-598, 2011.

HE, Kaiming et al. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2961-2969.

KUMBHAR, Shantanu et al. Farmer Buddy-Web Based Cotton Leaf Disease Detection Using CNN. **International Journal of Applied Engineering Research**, v. 14, n. 11, p. 2662-2666, 2019.

MOHANTY, S. P., HUGHES, D. P., e SALATHÉ, M. Using deep learning for image-based plant disease detection. *Frontiers in plant science*,7:1419, 2016

REDMON, J.; FARHADI, A.. Yolov3: An incremental improvement. **arXiv preprint arXiv:1804.02767**, 2018.

TZUTALIN. LabelImage. **MIT, Computer Science and Artificial Intelligence Laboratory**. Bryan Russell, Antonio Torralba, William T. Freeman. 2015.LIN, T. et al. Microsoft coco: Common objects in context. In: **European conference on computer vision**. Springer, Cham., p. 740-755, 2014