# Uma Implementação Baseada em Mask R-CNN para Detecção de Insetos em Imagens Digitais

Telmo De Cesaro Júnior
Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSul)
Passo Fundo, RS, Brasil
Email: telmo.junior@passofundo.ifsul.edu.br

Rafael Rieder
Universidade of Passo Fundo (UPF)
Passo Fundo, RS, Brasil
Email: rieder@upf.br

***Abstract*—The manual task of counting and identifying small insects, such as aphids and parasitoids, captured in yellow traps on the field, is an exhaustive, time-consuming, and non-scalable activity in agricultural research centers. The tasks involve complexity in the screening of the elements of interest and the use of magnifying glasses and microscopes. Recent advances in artificial intelligence and high-performance computing have enabled the development of efficient computer vision solutions for monitoring pests and identifying diseases in plants. In this context, this paper presents a routine for automatic counting and identification of insects in images, scanned from samples captured by the traps in the Embrapa Wheat experimental sites. For the implementation a small data set was used, image processing techniques and the convolutional two-stage Mask R-CNN approach were applied. The preliminary results indicate a mean accuracy (mAP) of 60.4% and show some details to increase the efficiency of the proposed method.**

***Resumo*—A tarefa manual de contagem e identificação de pequenos insetos, como afídeos e parasitoides, capturados em armadilhas de campo do tipo cor, é uma atividade exaustiva, demorada e não escalável em centros de pesquisa agrícola. Essas atividades envolvem complexidade na triagem dos elementos de interesse e utilização de lupas e microscópios. Recentes avanços em inteligência artificial e computação de alto desempenho têm viabilizado o desenvolvimento de soluções de visão computacional eficientes para monitoramento de pragas e identificação de doenças em plantas. Nesse contexto, esse artigo apresenta uma rotina para contagem e identificação automática de insetos em imagens, geradas pela digitalização de amostras capturadas pelas armadilhas nas estações experimentais da Embrapa Trigo. Para a implementação foi utilizado um pequeno conjunto de imagens, técnicas de processamento de imagens e a abordagem convolucional de dois estágios Mask R-CNN. Os resultados preliminares indicam uma precisão média (mAP) de 60,4%, e mostram alguns pontos que podem incrementar a eficiência da abordagem proposta.**

## I. Introdução

O monitoramento da flutuação populacional de pragas no campo, ou em experimentos de laboratório, permite acompanhar a variação dos níveis de infestação e embasar programas de manejo integrados de pragas [1]. Nas regiões tritícolas do Brasil observa-se a ocorrência de insetos-praga. Nas culturas de trigo, cevada, triticale e aveia, o aumento da população de afídeos (pulgões), pode causar significativa redução na produtividade, em decorrência dos seus hábitos alimentares e por serem vetores de transmissão de doenças [2].

Nesse âmbito, a Embrapa Trigo criou uma rede nacional de colaboração para monitorar pragas em cereais de inverno, constituída por cooperativas, empresas, instituições de ensino e centros de pesquisa. Essas entidades utilizam armadilhas de cor para capturar afídeos e parasitoides. Com a contagem dos espécimes, é possível mensurar os níveis populacionais, correlacionar com o dano econômico causado, e gerar informações para sistemas de alerta que auxiliam a tomada de decisão do manejo da lavoura [3].

O processo de triagem dos elementos capturados nas armadilhas requer tarefas manuais, como a coleta, a eliminação da água e parte dos detritos. O restante é inserido em uma lâmina para a análise detalhada por um especialista no laboratório. Cada armadilha pode conter dezenas de afídeos de diferentes espécies, além de outros insetos como moscas e cigarras. Em seguida, para a identificação das espécies de afídeos e parasitoides, são utilizadas chaves dicotômicas que especificam a morfologia de cada espécie [2]. Em função do tamanho reduzido desses insetos, utiliza-se geralmente um microscópio estereoscópico para localizar as características morfológicas no espécime, como o número de segmentos da antena, sifúnculo, ramificações nas asas e formato do corpo.

As atividades manuais realizadas pelos Especialistas são consideradas um problema, pois, além de ser exaustiva, o nível de precisão está condicionada a experiência profissional e a grande quantidade de elementos por amostra requer o uso de equipamentos de suporte. Sendo assim, com base em recentes estudos [4]–[8], os quais propõem automação parcial ou total do processo de identificação de insetos ou doenças de plantas, vislumbra-se a possibilidade de auxiliar a Embrapa Trigo no controle e monitoramento dessas pragas.

Nesse contexto, este trabalho apresenta uma proposta de rotina computacional, baseada em inteligência artificial e visão computacional, para automatizar o processo de identificação de afídeos e parasitoides.

## II. Referencial teórico

Soluções baseadas em Visão Computacional (VC) e Inteligência Artificial (IA) são bem difundidas na Agricultura, contribuindo na redução de custos com a automação de tarefas repetitivas [9]. Através de IA, automatizações têm alcançado níveis satisfatórios no processo de classificação de imagens

e detecção de objetos. Identificação e contagem de insetos em armadilhas, insetos em plantas e doenças em folhas de plantas [10]–[12], por exemplo, tem apresentado uma precisão média de 90,18%, 88,50% e 83,06%, respectivamente. Estes resultados demonstram boa eficácia em comparação às técnicas manuais.

Entre as técnicas de IA aplicadas a imagens digitais, Rede Neural Convolucional Profunda (DCNN) é considerada um dos modelos mais eficazes para o aprendizado de características. Em 2018, Kamilaris e Prenafeta-Boldú [9] apresentaram uma seleção de 40 trabalhos que aplicaram diferentes técnicas de aprendizado profundo na Agricultura, destes, 42% aplicaram DCNNs. De acordo com os autores, esse recurso é superior as técnicas de processamento de imagens. No entanto, nessa pesquisa foram apontadas algumas peculiaridades, como a necessidade de um grande volume de dados (imagens) e disponibilidade de unidade(s) de processamento gráfico (GPU) para acelerar o treinamento do modelo.

Segundo Chen *et al.* [13], os níveis de precisão das DCNNs indicam que essa técnica representa atualmente o estado-da-arte para as tarefas de classificação de imagens [14], detecção de objetos [15] e segmentação [16], superando os métodos tradicionais, onde as características dos objetos de interesse são extraídas manualmente [17].

Na tarefa de classificação de imagens deve existir apenas um padrão de objeto no centro da imagem, portanto, não se aplica para casos de ocorrência de objetos conectados ou parcialmente sobrepostos. Na detecção de objetos, é possível reconhecer vários objetos em uma determinada imagem. Porém, é necessário um amplo conjunto de imagens para treinamento e teste do modelo, onde cada objeto de interesse deve ser identificado e rotulado [8]. Em termos de complexidade, essa tarefa demanda maior tempo de processamento.

Com base no atual processo de captura de imagens realizado pela Embrapa Trigo, DCNNs se apresentam como uma alternativa adequada para detecção de insetos. As amostras contém centenas de objetos de interesse, detritos e outros insetos alados, como moscas e cigarras. Não são raros os casos de sobreposição ou conexão de objetos nas imagens. O uso de DCNNs permite incorporar no modelo o conhecimento do Especialista com a marcação dos insetos de interesse nas imagens, além da extração automática das características presentes nas áreas demarcadas.

A tarefa de detecção de objetos por DCNNs teve significativo aumento de desempenho com as contribuições de LeCun, Bengio e Hinton [18]. Sun *et al.* [8] sugerem a organização de DCNNs por Classificadores de dois estágios baseados em Region Proposal Network (Faster Region-Based Convolutional Neural Network, Faster R-CNN [15]; Region-based Fully Convolutional Networks, R-FCN [19]; Mask R-CNN [20]) e Classificadores de um estágio sem Region Proposal Network (Single Shot Multibox Detector , SSD [21]; You Only Look Once, YOLO [22]; RetinaNet [23]).

Fuentes *et al.* [12] relatam uma organização semelhante, com a definição das meta-arquiteturas Faster R-CNN, R-FCN e SSD para a detecção de objetos por DCNNs. Os autores destacam a necessidade de avaliar diferentes extratores de características. O extrator é a principal parte em uma arquitetura profunda e a sua escolha depende de vários fatores, como o tipo e a quantidade de camadas e parâmetros. Quanto mais profunda a rede em camadas e parâmetros, maior é a complexidade, o que demanda mais recursos computacionais. Por outro lado, redes não profundas ( *Shallow networks*) podem ser ineficientes para extraírem características relevantes de objetos pequenos em imagens com alta resolução.

Customizações são geralmente aplicadas para o aprimoramento da capacidade de generalização de DCNNs, pois, possibilitam adaptar o aprendizado de acordo com as peculiaridades de um determinado contexto, como tamanho do objeto, disponibilidade de imagens para cada classe, número de camadas da imagem, etc. Como exemplo, Nazri, Mazlan e Muharam [24] aplicaram CNNs para a redução do número de classes de mil para duas, em razão da necessidade de identificar somente duas espécies de cigarras.

Com isso em mente, optou-se em aplicar o método Mask R-CNN para a proposta deste trabalho. Esse método, baseado na abordagem Faster R-CNN [20], possui uma camada adicional capaz de segmentar cada objeto identificado, por meio da técnica de segmentação de instância. Acredita-se que este é um importante recurso para resolver o problema dessa pesquisa, em função da existência de vários insetos por imagem e da ocorrência de casos de conexões ou sobreposições parciais.

## III. Materiais e Métodos

O conjunto de imagens utilizados para as etapas de treinamento de teste foi composto por 167 imagens. Cada imagem foi pré-processada com a eliminação de áreas irrelevantes, conversão para escala de cinza e aplicação do filtro *GaussianBlur* para a eliminação de ruídos. A imagem resultante possui dimensão de 6156 x 6156 pixels, resolução horizontal e vertical de 96 dpi e intensidade de 8 bits.

Partindo da necessidade de detectar parasitoides (*Hymenoptera: Aphelinidae e Braconidae, Aphidiinae*) e afídeos alados (*Hemiptera, Aphididae*), o conjunto final de imagens foi composto pelos seguintes subconjuntos: (i) afídeos alados; (ii) parasitoides; (iii) afídeos alados e parasitoides, (iv) afídeos alados, parasitoides e detritos; (v) afídeos alados e detritos. Com o auxílio da ferramenta gráfica LabelImg [1] o Especialista rotulou todos os objetos de interesse presentes nas 167 imagens, classificando-os em duas classes: afídeo ou parasitoide.

A Tabela I apresenta as quantidades de insetos rotulados e a quantidade de imagens para cada subconjunto. Apesar do pequeno conjunto de imagens para aprendizado profundo, foram rotulados 14.809 insetos, sendo 12.354 para treinamento e 2.455 para teste do modelo. Apenas os subconjuntos (iv) e (v) contém imagens geradas por digitalização das armadilhas. Os demais foram criados artificialmente com a inserção manual dos espécimes na lâmina sem a presença de detritos.

A implementação de Mask R-CNN de Abdulla [25] foi utilizada nesse trabalho, bem como, a linguagem Python e as bibliotecas OpenCV, TensorFlow e Keras. Nessa implementação

[1]Disponível em: https://github.com/tzutalin/labelImg

Tabela I
CONJUNTO DE IMAGENS PARA TREINAMENTO E TESTE

| Conj. | Afídeos | Parasit. | Ins./Img. Train. | Ins./Img. Test. | Total |
|---|---|---|---|---|---|
| 1 | 4934 | 0 | 4688/ 44 | 246/ 2 | 4934/ 46 |
| 2 | 0 | 6137 | 5942/ 43 | 195/ 2 | 6137/ 45 |
| 3 | 1745 | 1611 | 1506/ 13 | 1850/ 15 | 3356/ 28 |
| 4 | 77 | 15 | 10/ 1 | 82/ 5 | 92/ 6 |
| 5 | 290 | 0 | 208/ 33 | 82/ 9 | 290/ 42 |
| Total | 7046 | 7763 | 12354/ 134 | 2455/ 33 | 14809/ 167 |

Tabela II
RESULTADOS GERADOS PELOS EXPERIMENTOS

| Exp. | Tamanho | Aumento | Tempo | Prec. train. | Prec. teste |
|---|---|---|---|---|---|
| 1 | 1024x1024 | Não | 1d 4h 47m | 87.0% | 49.8% |
| 2 | 1024x1024 | Sim | 1d 15h 39m | 71.1% | 53.6% |
| 3 | 2048x2048 | Não | 3d 19h 16m | 84.5% | 59.4% |
| **4** | 2048x2048 | Sim | 16d 1h 15m | 85.2% | **60.4%** |

é possível aplicar os extratores de recursos ResNet101 e RestNet50, a utilização do modelo pré-treinado MS COCO ou ImagetNet e recursos para a geração de novas imagens em tempo de execução, através de transformações geométricas. O TensorFlow possibilita a utilização de GPU. Basicamente, a finalidade do Mask R-CNN é gerar caixas delimitadoras, classificar e segmentar cada objeto identificado na imagem [25].

Em razão do número restrito de imagens em nosso conjunto de treinamento e testes, aplicou-se duas técnicas compensatórias para evitar o problema de *over-fitting*. A primeira foi inicializar todas as camadas da rede com os pesos do modelo MS COCO, exceto as camadas: *mrcnn_class_logits*, *mrcnn_bbox_fc*, *mrcnn_bbox*, *mrcnn_mask* e *conv1*, que foram inicializadas com pesos aleatórios. O subconjunto de camadas retreinadas foi o seguinte: *conv1.**, *res4.**, *bn4.**, *res5.**, *bn5.**, *mrcnn_.**, *rpn_.** e *fpn_.**. A camada de entrada *conv1* foi retreinada pelo fato de imagens em tons de cinza (GrayScale) serem utilizadas. A segunda técnica foi a aplicação das transformações geométricas randômicas de rotação (-90, 90), inversão horizontal (0.5) e vertical (0.5), para triplicar a quantidade de imagens ( *data augmentation* ).

Para o treinamento da rede, utilizou-se um computador com o processador Intel Core I7-6950X, 32 GB de RAM e uma GPU GeForce GTX Titan X com 12 GB de memória. A dimensão original das imagens de entrada (6156x6156) inviabilizou o seu processamento por inteiro, em virtude dos recursos computacionais disponíveis. Sendo assim, optou-se pelo redimensionamento das imagens para 1024x1024 e 2048x2048. O extrator de características utilizado foi o ResNet50. O Batch Size foi fixado em 1. Em função do tamanho reduzido dos insetos em relação ao tamanho da imagem, definiu-se o menor valor possível para o parâmetro *RPN_ANCHOR_SCALES* (valores utilizados nos quatro experimentos: 2, 4, 8, 16, 32). Os demais parâmetros não foram alterados.

Foram realizados quatro experimentos com o conjunto de imagens, utilizando transferência de aprendizado com o modelo MS COCO. Cada experimento executou de 40 a 140 épocas (treinamento). Cada época tem 623 passos. Após, a época que alcançar a maior precisão (imagens de teste) é selecionada. O objetivo desses experimentos foi de avaliar duas dimensões de imagens, a técnica de aumento de dados, o tempo de processamento, a capacidade de resolver casos de conexão e a precisão do modelo. Os experimentos estão detalhados na Tabela II.

## IV. RESULTADOS E DISCUSSÃO

Conforme a tabela II o quarto experimento obteve a maior precisão média (60.4% mAP) para as imagens de teste. Em comparação com o primeiro, houve um incremento de 10% com o uso de aumento de dados e a dimensão de 2048x2048. No entanto, o tempo de processamento para uma época com imagens de 2048x2048 foi de 4h, enquanto que com imagens de 1024x1024 foi de 1h. Ao considerar somente a técnica de aumento de dados a melhora da precisão não foi significativa.

Considerando o modelo em questão, selecionou-se duas novas imagens para avaliação. Verificou-se que a presença de detritos e o posicionamento do espécime interfere significativamente na capacidade de reconhecimento. No teste ilustrado pela Figura 1 foram identificados quatro objetos em um cenário composto por onze elementos. Os objetos 1 e 2 foram identificados corretamente como afídeos, sendo que o 1 está de lado. Os objetos 3 e 4 são falsos positivos, possivelmente em razão da semelhança com a classe afídeo. Os demais objetos foram descartados adequadamente.

No segundo teste apresentado pela Figura 2 existem dezesseis insetos e nenhum detrito. Nesse caso, todos parasitoides foram corretamente identificados (rótulos em verde: 1,2,5,6,9 e 11), sendo que os objetos 1 e 2 estão conectados. Entre os dez afídeos na imagem, apenas cinco foram corretamente classificados (rótulos em vermelho: 3,4,8,10 e 7). Os objetos 15,14,13,12 são falsos negativos, possivelmente em razão de estarem conectados. O 16 também é um falso negativo.

Ao considerar os subconjuntos de imagens (iv) e (v) utilizados para o treinamento e teste do modelo, que corresponde ao cenário do primeiro teste (parasitoides, afídeos alados e detritos), verifica-se que a quantidade de imagens foi menor em relação aos demais subconjuntos. Esse fato pode ter relação com a baixa precisão constada no primeiro teste. No segundo teste, o modelo conseguiu identificar corretamente a maioria dos insetos. Portanto, fica evidente a necessidade de incremento do número de imagens para esses subconjuntos.

## V. CONCLUSÃO E TRABALHOS FUTUROS

A utilização de Mask R-CNN para a detecção de afídeos alados e parasitoides em imagens digitais gerou resultados promissores, levando em consideração o número restrito de imagens utilizadas para o treinamento e teste. O incremento da resolução e a utilização de aumento de dados possibilitou a elevação da precisão média (mAP) em 10%. Nesse sentido, projeta-se a inclusão de novas imagens nos subconjuntos (iv) e (v), a utilização da técnica de recorte aleatório em substituição ao redimensionamento para evitar a redução do tamanho dos objetos e a avaliação de outros extratores de características e

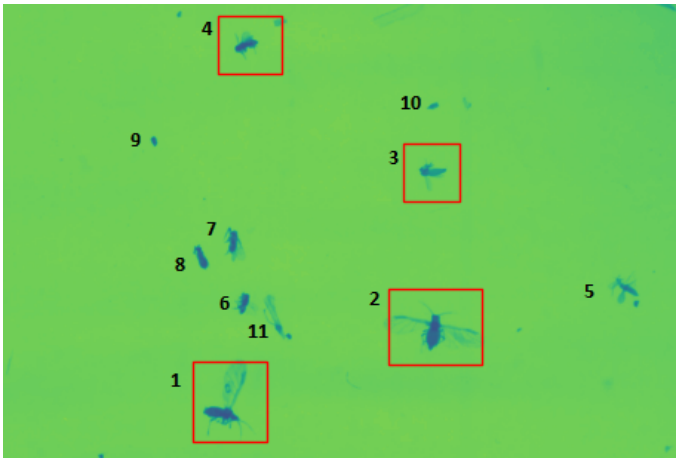


Figura 1. Exemplo de Reconhecimento com detritos

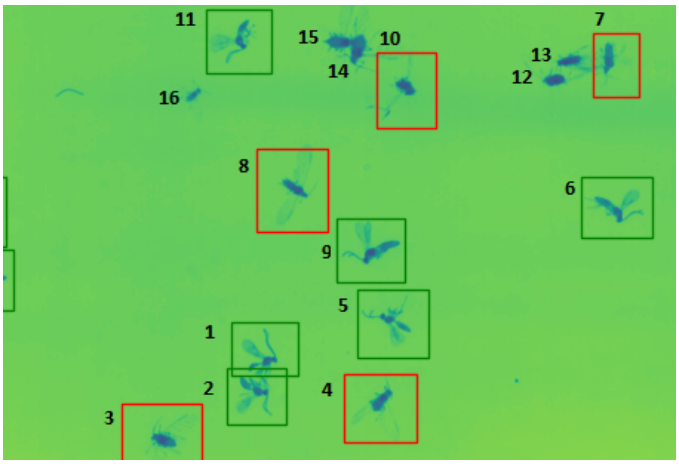


Figura 2. Exemplo de Reconhecimento sem detritos

parâmetros. Com essas alterações pretende-se elevar a precisão do modelo e a sua capacidade de detectar insetos conectados.

## Referências


[1] L. C. Wright and W. W. Cone, "Population Dynamics of Brachycorynella asparagi (Homoptera: Aphididae) on Undisturbed Asparagus in Washington State," *Environmental Entomology*, vol. 17, no. 5, pp. 878–886, 10 1988.

[2] P. R. V. d. S. Pereira, J. R. Salvadori, and D. Lau, "Identificação de adultos ápteros e alados das principais espécies de afídeos (hemiptera: Aphididae) associadas acereais de inverno no brasil," Embrapa Trigo, Comunicado Técnico online, 258, Tech. Rep., 2009. [Online]. Available: http://www.cnpt.embrapa.br/biblio/co/p_co258.htm

[3] E. A. Lins, J. P. M. Rodriguez, S. I. Scoloski, J. Pivato, M. B. Lima, J. M. C. Fernandes, P. R. V. da Silva Pereira, D. Lau, and R. Rieder, "A method for counting and classifying aphids using computer vision," *Computers and Electronics in Agriculture*, vol. 169, p. 105200, 2020.

[4] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Computers and Electronics in Agriculture*, 2018. [Online]. Available: https://doi.org/10.1016/j.compag.2018.04.002

[5] L. Liu, R. Wang, C. Xie, P. Yang, F. Wang, S. Sudirman, and W. Liu, "PestNet: An End-to-End Deep Learning Approach for Large-Scale Multi-Class Pest Detection and Classification," *IEEE Access*, vol. 7, pp. 45 301–45 312, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2909522

[6] M. C. Bakkay, S. Chambon, H. A. Rashwan, C. Lubat, and S. Barsotti, "Automatic detection of individual and touching moths from trap images by combining contour-based and region-based segmentation," *IET Computer Vision*, vol. 12, no. 2, pp. 138–145, 2018. [Online]. Available: https://doi.org/10.1049/iet-cvi.2017.0086

[7] V. Partel, L. Nunes, P. Stansly, and Y. Ampatzidis, "Automated vision-based system for monitoring Asian citrus psyllid in orchards utilizing artificial intelligence," *Computers and Electronics in Agriculture*, vol. 162, pp. 328–336, jul 2019.

[8] Y. Sun, X. Liu, M. Yuan, L. Ren, J. Wang, and Z. Chen, "Automatic in-trap pest detection using learning for pheromone-based Dendroctonus valens monitoring," *Biosystems Engineering*, vol. 176, pp. 140–150, dec 2018.

[9] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018. [Online]. Available: https://doi.org/10.1016/j.compag.2018.02.016

[10] Y. Zhong, J. Gao, Q. Lei, and Y. Zhou, "A vision-based counting and recognition system for flying insects in intelligent agriculture," *Sensors*, vol. 18, no. 5, p. 1489, 2018.

[11] W. Li, P. Chen, B. Wang, and C. Xie, "Automatic Localization and Count of Agricultural Crop Pests Based on an Improved Deep Learning Pipeline," *Scientific Reports*, vol. 9, no. 1, 2019.

[12] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors (Switzerland)*, vol. 17, no. 9, 2017.

[13] J. Chen, Y. Fan, T. Wang, C. Zhang, Z. Qiu, and Y. He, "Automatic Segmentation and Counting of Aphid Nymphs on Leaves Using Convolutional Neural Networks," *Agronomy*, vol. 8, no. 8, 2018.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, 01 2012.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[17] P. Fischer, A. Dosovitskiy, and T. Brox, "Descriptor matching with convolutional neural networks: a comparison to SIFT," *CoRR*, vol. abs/1405.5769, 2014, withdrawn. [Online]. Available: http://arxiv.org/abs/1405.5769

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[19] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 379–387. [Online]. Available: https://dl.acm.org/doi/10.5555/3157096.3157139

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988. [Online]. Available: https://doi.org/10.1109/ICCV.2017.322

[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. [Online]. Available: https://doi.org/10.1109/CVPR.2016.91

[23] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2018. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2858826

[24] A. Nazri, N. Mazlan, and F. Muharam, "PENYEK: Automated brown planthopper detection from imperfect sticky pad images using deep convolutional neural network," *PLOS ONE*, vol. 13, no. 12, p. e0208501, dec 2018. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0208501

[25] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow," 2017. [Online]. Available: https://github.com/matterport/Mask_RCNN