Automated Wheat Disease Detection using Deep Learning: an Object Detection and Classification Approach

Sepideh Etaati, Javad Khoramdel, and Esmaeil Najafi

Faculty of Mechanical Engineering, K.N. Toosi University of Technology, Tehran, Iran
Emails: etaati@email.kntu.ac.ir, j.khorramdel96@gmail.com, najafi.e@kntu.ac.ir

***Abstract*—The wheat crop is a crucial staple in agriculture, but its yield is often compromised by various diseases. With a growing global population, adaptive agricultural practices are essential, and early detection of wheat diseases becomes vital. Utilizing deep learning techniques can offer practical solutions for disease detection and classification. This paper explores different approaches to automate wheat head disease detection and identifies the best methods. Two distinct datasets are used: the Global Wheat Head Detection (GWHD) dataset for object detection and the Large Wheat Disease Classification Dataset (LWDC) for classification. The YOLOv4 object detection network is trained on GWHD, achieving a mean Average Precision (mAP) of 91%. The trained weights are used for domain transfer for training on LWDC dataset for further exploration, where training with three classes and COCO's pre-trained weights yields superior mAP results. Additionally, five CNN models, including VGG19, ResNet50, EfficientNet-B0, NASNetMobile, and NASNetLarge, are evaluated on LWDC dataset for wheat disease classification. VGG19 emerges as the top-performing model, accurately classifying various wheat diseases, achieving an average F1 score of 95%. The combination of YOLOv4 for object detection and VGG19 for classification presents promising results, offering valuable insights for precision agriculture and early disease detection in wheat crops.**

***Index Terms*—Agricultural Artificial Intelligence, Wheat Diseases Detection, Deep Learning, Object Detection, Domain Transfer, Image Classification**

I. Introduction

With the increasing global population, the demand for food production has become more pressing than ever. As a fundamental source, agriculture plays a pivotal role in meeting this demand, with wheat being a crucial staple. Unfortunately, around 20% of the world's annual wheat production is lost to diseases and pests [1]. Consequently, implementing adaptive manufacturing techniques, including detecting and classifying various wheat diseases, becomes imperative. However, farmers face challenges in accurately recognizing all types of diseases. To address this issue, machine learning offers valuable assistance. Leveraging the progress in image processing and utilizing object detection and classification neural networks hold the promise of achieving this goal.

This paper tests different object detection and classification techniques on two distinct datasets. While previous research has primarily focused on employing classification techniques for diagnosing plant diseases, this study extends its scope by incorporating object detection for locating and classifying plant diseases. The paper employs two datasets. The first dataset is the Global Wheat Head Detection Dataset (GWHD) [2], which is utilized for training the YOLOv4 model [3] to perform object detection. The acquired weights from training on GWHD dataset will be used for domain transfer. The second dataset, the Large Wheat Disease Classification Dataset [4], initially serves as a classification dataset. Originally, five different classification models are trained using this dataset. Subsequently, the dataset is annotated for object detection, and the YOLOv4 model is trained on it. Moreover, this dataset tests a two-stage method for detecting and classifying the wheat head diseases. The YOLOv4 model is first employed to locate wheat heads within the images, subsequently cropping the images based on the detected bounding boxes. These cropped images then serve as input for the classification model, enabling the classification of wheat head diseases. To summarize, the main contributions of this article include:

- Employing domain transfer to enhance the precision of object detection in the context of wheat head disease detection.
- Attaining the state-of-the-art mean average precision (mAP) on the Global Wheat Head Detection Dataset, which served as the intermediary task for domain transfer.
- Labeling the Large Wheat Disease Classification Dataset for training object detection models.
- Implementing both single- and two-stage disease detection method on the labeled dataset.

The structure of the paper is organized as follows: Section II will review previous works for detecting and classifying diseases on plans with deep learning, Section III will illustrate the two datasets used in this research, Section IV will explain the methods used for training the datasets, Section V will present the results and discuss them, and Section VI will conclude the paper with a summary of the main findings and implications.

II. Related Work

Various approaches can be employed to detect and classify plant diseases using machine learning. One such research [5] introduces an automatic and efficient method based on K-means clustering. Initially, the color image is transformed from RGB to the Lab color space. Within this Lab color space, clustering is executed by calculating the absolute difference between each pixel and the clustering center. This method was applied to a dataset comprising three distinct diseases affecting wheat leaves. Remarkably, the accuracy rate achieved for each class was approximately 90%.

Furthermore, in the domain of plant disease detection, some researchers have employed a two-stage detection approach. Notably, [6] presented a robust method capable of real-time detection of disease classes and their corresponding locations on tomato leaves. Their work involved the utilization of three distinct detectors, namely Faster R-CNN [7], SSD [8], and R-FCN [9], which were combined with feature extractors such as VGG and ResNet. To train these models, the researchers utilized the Tomato Diseases and Pests Dataset, which comprises nine different classes of tomato diseases. Additionally, [10], [11] introduced another significant contribution to the field. They curated a dataset specifically for wheat plants, known as the Wheat Rust Classification Dataset. This dataset was designed for object detection purposes, focusing on wheat rust, and it contains two distinct classes. After locating the corresponding bounding boxes, the researchers cropped the images and employed them as input for image classification networks. Five models of Convolutional Neural Networks (CNNs) were trained on this dataset, namely, MobileNet V3 [12], Inception [13], ResNet-50 [14], EfficientNet-B0 [15], and VGG16 [16]. Remarkably, EfficientNet-B0 demonstrated the most promising results, showcasing its effectiveness in wheat rust classification.

Various classification methods have been explored for diagnosing plant diseases, including wheat stripe rust. Notably, the article [17] discusses different classification methods used for diagnosing plant diseases, focusing on wheat stripe rust. It introduces C-DenseNet, a revolutionary deep learning network incorporating the Convolutional Block Attention Module (CBAM) into DenseNet. The dataset used for training the model consists of five categories for stripe rust infection and six wheat stripe rust infection levels. Instead of down-sampling images, they applied image cropping for preprocessing to avoid object visibility issues. Data augmentation was used to prevent overfitting. The C-DenseNet architecture includes convolutional layers with dense layers in between. CBAM, comprising channel and spatial attention modules, enhances feature detection. The model achieved an impressive accuracy rate 97.99% in diagnosing wheat stripe rust. In [18], a new approach for classifying plant diseases using the Vision Transformer (ViT) was proposed. The study utilized three datasets: the Wheat Rust Classification Dataset, Rice Leaf Disease Dataset (RLDD), and Plant Village. Compared to traditional convolutional-based architectures, the ViT-based model consistently outperformed CNN or hybrid models in terms of performance while featuring significantly fewer parameters.

Another robust work for classifying plant diseases is presented in [19]. The paper introduced a novel CNN model for

(a) (b)

Fig. 1. Sample images from Global Wheat Head Detection Dataset with its corresponding bounding boxes, adopted from [2].

classifying the Cassava disease dataset, which faces the class imbalance challenge. The model's performance was enhanced using class-weight, SMOTH, focal loss techniques, and considerable input shape dimensions of images. An evaluation was conducted using k-fold cross-validation, and the model achieved the best accuracy of 93%. In [4], a novel CNN model was introduced, trained on the Large Wheat Disease Classification Dataset comprising 12,000 images in 10 different classes for various diseases in wheat. The proposed classification model achieved an impressive accuracy of 97.88%, surpassing the performance of ResNet50 and VGG16 networks. These research works demonstrate the effectiveness and potential of deep learning models, such as Vision Transformer (ViT) and CNN, in accurately classifying plant diseases and advancing the field of precision agriculture.

III. Implemented Datasets

In this paper, the evaluation of object detection and classification models is conducted using two distinct datasets. The first dataset exclusively comprises images of healthy wheat and is referred to as the Global Wheat Detection dataset [2]. The second dataset has three classes and is known as the Large Wheat Disease Classification Dataset.

A. Global Wheat Head Detection

The first dataset used in this paper is called the Global Wheat Head Detection (GWHD) dataset, as referenced in [2]. This dataset consists of wheat head images captured in farms from various countries. It is a substantial dataset, with images having a size of 1024x1024 pixels. The GWHD dataset includes over 3000 images for the training set and 1000 for the test set. A notable challenge in this dataset is the varying number of wheat heads in each image, ranging from zero to as many as 116. The primary objective of this dataset is to detect and localize all the wheat heads in each image. A sample image and its corresponding labeled image are provided in Fig. 1 for reference.

B. Large Wheat Disease Classification Dataset

The Large Wheat Disease Classification Dataset (LWDCD2020) is a comprehensive dataset encompassing ten different classes of wheat diseases. These classes include

Fig. 2. Sample images from LWDC Dataset: (a) Healthy Wheat, (b) Fusarium Head Blight, and (c) Loose Smut, adopted from [4].

nine distinct types of diseases affecting wheat, namely Black Chaff, Karnal bunt, Crown and Root Rot, Loose Smut, Fusarium Head Blight, Powdery Mildew, Leaf Rust, Wheat Streak Mosaic, and Tan Spot. The last class in the dataset represents images of healthy wheat. This study selects three classes from the LWDC dataset, explicitly focusing on wheat head-related diseases. These three chosen classes are Fusarium Head Blight, Loose Smut, and Healthy Wheat. Each class contains a representative sample of images, which is illustrated in Fig. 2. Note that the images for Fusarium Head Blight class have watermarks, and all the training is done with this images.

For this study, the LWDC dataset is utilized in two different ways. Firstly, it is employed for training classification models to categorize images into their respective disease classes accurately. Secondly, the images in the dataset are labeled for training object detection networks, specifically targeting wheat head-related diseases. A sample of the labeled dataset is shown in Fig. 4. Utilizing the LWDC dataset for classification and object detection tasks contributes to a comprehensive analysis of wheat diseases, providing valuable insights and potential solutions for improving disease diagnosis and management in wheat crops.

IV. Experimental Setups

This section an explanation of the various models and techniques utilized in the training of the two datasets is provided. Initially, the Global Wheat Head Detection (GWHD) dataset undergoes training using YOLOv4 with COCO [20] pre-trained weights, employing the K-fold cross-validation method to ensure robust results. Subsequently, we evaluate the performance of five distinct classification models on the Large Wheat Disease Classification (LWDC) dataset. Once this evaluation is complete, the dataset is further processed to facilitate training with an object detection model. For the object detection phase, we employ YOLOv4 with two different pre-trained weights: COCO's weights and GWHD weights. This allows us to detect and classify wheat head diseases within the dataset. In the final step, we adopt a two-stage approach. Initially, YOLOv4 is utilized to locate wheat heads within the images. Subsequently, the detected bounding boxes are used to crop the images, creating inputs for the VGG19 classification model. This model enables the precise classification of various wheat head diseases.

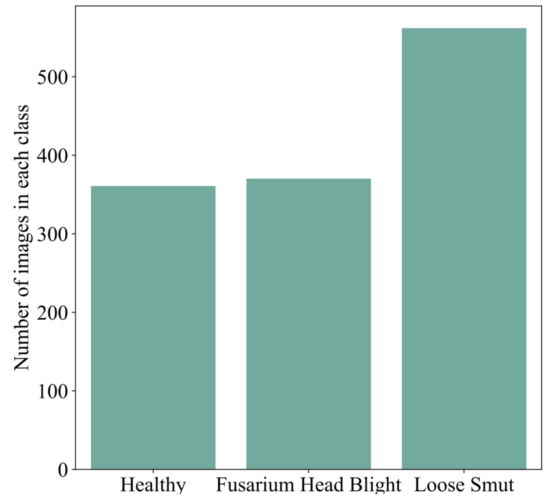


Fig. 3. Distribution of samples within the Large Wheat Disease Classification Dataset.

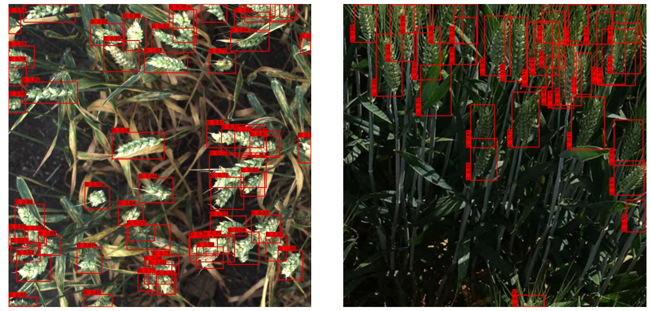
A. Object Detection

In this work, the computer vision technique called object detection is employed to locate and identify wheat heads in images. YOLOv4, the object detection model, is used for this purpose, as introduced in [3]. YOLOv4 comprises a backbone, neck, and head in its structure. The authors chose CSPDarknet53 [21] as the backbone network for the YOLOv4 object detector. The features generated in the backbone are then combined using PANet [22] and SPP [23] in the neck to prepare them for the object detection phase. YOLOv4 adopts the same YOLO head as YOLOv3 [24], employing anchor-based detection steps and providing three degrees of detection granularity. Various techniques are utilized to enhance the network's performance, including Mosaic data augmentation, CutMix [25], and Mish [26] as an activation function. The first training phase involves YOLOv4 being trained with the GWHD dataset to detect wheat heads. A k-fold cross-validation method with k=5 is utilized to validate the results. Transfer learning with COCO's pre-trained weights is implemented, and the network trains for 20 epochs, six of which are warmup epochs, to prevent instability in the loss function.

In the second phase, YOLOv4 is trained on the LWDC dataset, where its distribution has been visualized in Fig. 3. Before training, the dataset is prepared, and all the images are labeled for object detection, as shown in Fig. 4. Approximately 200 images from the GWHD dataset are added to the Healthy Wheat class to balance the healthy wheat head class. The training of YOLOv4 encompasses two approaches: one that focuses solely on the detection of wheat heads, and another that involves the identification of three classes – Fusarium Head Blight, Loose Smut, and Healthy Wheat. The parameters include an image size of 704 pixels, a batch size of four, and

(a) (b) (c)

Fig. 4. Sample of labeled images from LWDC Dataset: (a) Healthy Wheat, (b) Fusarium Head Blight, and (c) Loose Smut



(a) (b)

Fig. 5. Prediction results achieved upon training the GWHD Dataset using YOLOv4.

a training duration spanning 35 epochs. To enhance model accuracy, transfer learning is being employed by leveraging pre-trained weights from both COCO and GWHD datasets.

Considering the dissimilarity between COCO images and those relevant to wheat, a deliberate decision is made to effectuate domain transfer. This is achieved by training YOLOv4 on the GWHD dataset, which aligns more closely with the characteristics of wheat-related imagery. The knowledge gained from this adaptation is subsequently being employed to train the model on the target dataset. Notably, the computational infrastructure for these training endeavors relies upon NVIDIA Tesla T4 GPUs.

B. Classification Models

The experiment evaluates five different Convolutional Neural Network (CNN) architecture-based models for classification using the LWDC dataset. The models under consideration are VGG19, ResNet50, EfficientNet-B0, NASNetMobile, and NASNetLarge [27]. Transfer learning with ImageNet weights is applied to these networks to enhance the training process. The input size of the images is set to 244x224 pixels. A dropout layer is incorporated into the networks to prevent overfitting, and data augmentation is employed during training. The Adam optimizer is used for optimization; all networks have a batch size of 64. Training is performed for 50 epochs using NVIDIA Tesla T4 GPU.

C. Two-Stage Detection

In addition to utilizing object detection and classification models, we are also exploring a fusion of object detection and classification tasks. This approach involves the development of a two-stage defect detection model, incorporating both YOLOv4 and VGG19. In the initial stage of this proposed method, the newly curated dataset is subjected to training with two distinct pre-trained weights: once with COCO's pre-trained weights and subsequently with GWHD's pre-trained weights. Furthermore, the dataset is trained separately, considering both detecting wheat head and detecting wheat head diseases scenarios.

In this phase, the emphasis is placed on training the dataset with a single class, specifically "wheat head". Upon completing YOLOv4 training, the resulting output images are meticulously cropped based on the corresponding bounding boxes. These cropped images are then employed as inputs for the subsequent VGG19 classification model. This integrated methodology showcases a comprehensive approach to harnessing the strengths of both object detection and classification techniques, enabling accurate defect detection and classification within the task's context.

V. Results and Discussion

A comprehensive overview of the results obtained on the desired dataset has been presented. The results focus on the performance evaluation of object detection, classification, and two-stage detection models.

A. Object Detection Results

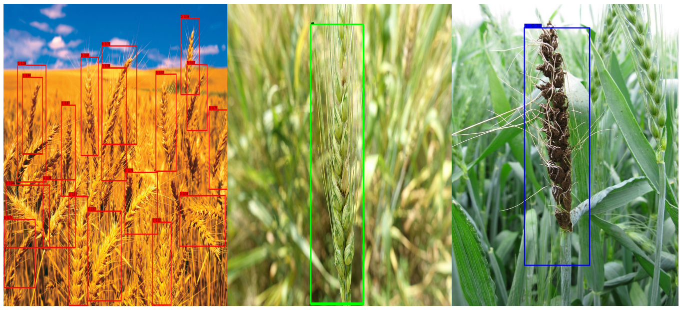
We begin training our object detection model by training the GWHD dataset using the YOLOv4 framework, incorporating k-fold cross-validation for added accuracy. This thorough approach results in an impressive average mean Average Precision (mAP) of 91% for each fold, showcasing the effectiveness of our method. The results of this intensive training process are visually showcased through images displaying the detected wheat heads. A clear illustration of these outcomes is presented in Fig. 5, offering a concise demonstration of our model's skill in accurately detecting wheat heads. For the second dataset, LWDC, we train the YOLOv4 model using two different pre-trained weights from COCO and GWHD datasets. This shows how we are using transfer learning. We then analyze and put together the results in Table I. We have pictures of the output images in Fig. 6 to understand better how well the model works. These pictures help show how effective the model is and why transfer learning is essential in our study.

B. Classification Models Results

The experimental efforts highlight the significance of the VGG19 network concerning the LWDC dataset. To provide a comprehensive view of our findings, we draw attention to Table II, which outlines the outcomes of our classification model comparisons. Furthermore, the learning process of the VGG19 model for training and validation are visually explained through depictions of loss and accuracy trajectories.

TABLE I
OBJECT DETECTION RESULTS OBTAINED BY TRAINING YOLOV4 ON THE LWDC DATASET

model	Healthy AP	Loose Smut AP	Fusarium Head Blight	mAP
three classes with COCO's weights	0.81	0.94	0.98	0.91
three classes with GWHD's weights	0.63	0.68	0.93	0.75



(a) (b) (c)

Fig. 6. Prediction results achieved upon training the LWDC Dataset using YOLOv4: (a) Healthy Wheat, (b) Fusarium Head Blight, and (c) Loose Smut

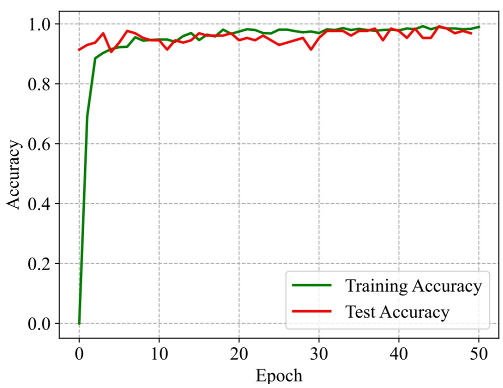


Fig. 7. Accuracy plot of the VGG19 network for both validation and training data on the LWDC Dataset.

The visual representations have been given in Fig. 7 and Fig. 8, respectively.

C. Two-Stage Detection Results

The outcomes reveal that when the one class "wheat head" dataset is trained with GWHD's pre-trained weights, the achieved mAP is 87.16%. However, when trained with COCO's pre-trained weights, the mAP slightly drops to 83.59%. These results indicate that utilizing weights pre-trained on a similar dataset can improve performance. Conversely, when the three classes dataset "wheat disease" is trained with COCO's pre-trained weights, the results outperform GWHD's pre-trained weights. The use of COCO's pre-trained weights yields superior performance in this scenario.

A potential reason for the lower accuracy observed during domain transfer could be the insufficient number of samples in the target dataset. Instead of fully retraining the entire network, a more effective approach might involve freezing more layers and selectively fine-tuning only a few layers when training on both the GWHD and target datasets. Another explanation for the disparity in results between one-class object detection and wheat disease detection could be the adequacy of samples. While there might be sufficient samples to achieve acceptable accuracy in detecting wheat heads, the same might not be true for distinguishing between different classes in the case of wheat disease.

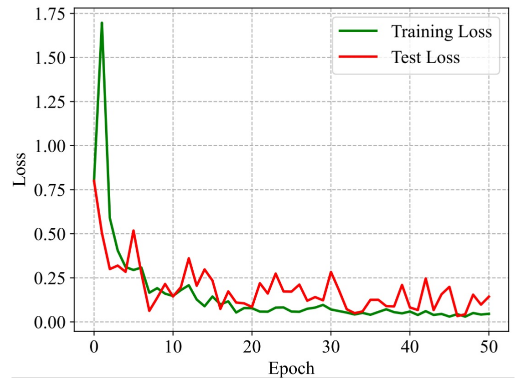


Fig. 8. Loss plot of the VGG19 network for both validation and training data on the LWDC Dataset.

In practice, this involves employing YOLOv4 for the initial detection and localization of head wheat instances. Once the head wheat instances are identified, the corresponding images are cropped according to the bounding boxes. These cropped images are then utilized as input for the VGG19 classification model. The outcomes of this approach are documented comprehensively in Table III.

VI. CONCLUSION

In conclusion, this work focuses on training new models for both object detection and classification tasks using two different datasets. The first dataset, GWHD, aims to detect wheat heads in images. The YOLOv4 object detection network achieved a remarkable mean Average Precision (mAP) of 91% on this dataset. The trained weights from this detection task are then utilized for domain transfer to train the second dataset. Transferring the acquired knowledge, the attention turns to the LWDC dataset, tailored for classifying diverse wheat diseases. An observation emerges when experimenting with transfer learning scenarios: employing COCO's pre-trained weights enhances mAP performance for multi-class classification. In contrast, GWHD's pre-trained weights show better results in single-class detection. Five CNN models were tested on the LWDC dataset for the classification task. Among these models, VGG19 demonstrated the best performance, showcasing its effectiveness in accurately classifying different diseases of wheat. Finally, a two-stage detection is applied to the dataset

TABLE II
CLASSIFICATION RESULTS OBTAINED ON THE LWDC DATASET

models	Avarage F1 score	Avarage Recall	Avarage Precision
VGG19	0.95	0.92	0.98
ResNet50	0.94	0.92	0.97
EfficientNet-B0	0.93	0.92	0.95
NASNetLarge	0.88	0.87	0.91
NASNetMobile	0.77	0.76	0.82

TABLE III
OBJECT DETECTION RESULTS USING A TWO-STAGE APPROACH WITH YOLOV4 AND VGG19 MODELS TRAINED ON THE LWDC DATASET

model	Healthy AP	Loose Smut AP	Fusarium Head Blight	mAP
one class with COCO's weights	0.72	0.81	0.83	0.83
one class with GWHD's weights	0.76	0.84	0.87	0.87

to enhance the accuracy, this is a combination of the YOLOv4 and VGG19 models. This amalgamation of techniques holds the promise of refined disease identification in agricultural settings, signifying the potential for substantial advancements in this domain.

REFERENCES

[1] P. Langridge, "Wheat diseases: an overview albrecht serfling, doris kopahnke, antje habekuss, fluturë novakazi and frank ordon, julius kühn-institute (jki), institute for resistance research and stress tolerance, germany," *Achieving sustainable cultivation of wheat Volume 1*, pp. 319–326, 2017.

[2] E. David, S. Madec, P. Sadeghi-Tehran, H. Aasen, B. Zheng, S. Liu, N. Kirchgessner, G. Ishikawa, K. Nagasawa, M. A. Badhon *et al.*, "Global wheat head detection GWHD dataset: a large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods," *Plant Phenomics*, vol. 2020, 2020.

[3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[4] L. Goyal, C. M. Sharma, A. Singh, and P. K. Singh, "Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture," *Informatics in Medicine Unlocked*, vol. 25, p. 100642, 2021.

[5] X. Niu, M. Wang, X. Chen, S. Guo, H. Zhang, and D. He, "Image segmentation algorithm for disease detection of wheat leaves," in *Proceedings of the 2014 International Conference on Advanced Mechatronic Systems*, 2014, pp. 270–273.

[6] 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*, vol. 17, no. 9, p. 2022, 2017.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[9] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, vol. 29, 2016.

[10] B. Safarijalal, Y. Alborzi, and E. Najafi, "Automated wheat disease detection using a ROS-based autonomous guided UAV," *arXiv preprint arXiv:2206.15042*, 2022.

[11] Y. Alborzi, B. S. Jalal, and E. Najafi, "ROS-based SLAM and navigation for a Gazebo-simulated autonomous quadrotor," in *Proceedings of the 21st International Conference on Research and Education in Mechatronics*, 2020, pp. 1–5.

[12] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[17] Z. Mi, X. Zhang, J. Su, D. Han, and B. Su, "Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices," *Frontiers in Plant Science*, vol. 11, p. 558126, 2020.

[18] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Scientific Reports*, vol. 12, no. 1, pp. 1–10, 2022.

[19] G. Sambasivam and G. D. Opiyo, "A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks," *Egyptian Informatics Journal*, vol. 22, no. 1, pp. 27–34, 2021.

[20] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[21] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.

[22] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[24] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[25] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[26] D. Misra, "Mish: A self regularized non-monotonic activation function," *arXiv preprint arXiv:1908.08681*, 2019.

[27] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.