remote sensing

MDPI

Article

Evaluation of Diverse Convolutional Neural Networks and Training Strategies for Wheat Leaf Disease Identification with Field-Acquired Photographs

Jiale Jiang [1], Haiyan Liu [1], Chen Zhao [2], Can He [1], Jifeng Ma [1], Tao Cheng [1], Yan Zhu [1], Weixing Cao [1] and Xia Yao [1,*]

[1] National Engineering and Technology Center for Information Agriculture (NETCIA), MARA Key Laboratory of Crop System Analysis and Decision Making, MOE, Engineering Research Center of Smart Agriculture, Jiangsu Key Laboratory for Information Agriculture, Institute of Smart Agriculture, Nanjing Agricultural University, One Weigang, Nanjing 210095, China; jialejiang@njau.edu.cn (J.J.); 2018101173@njau.edu.cn (H.L.); 2021101184@njau.edu.cn (C.H.); 2006069@njau.edu.cn (J.M.); tcheng@njau.edu.cn (T.C.); yanzhu@njau.edu.cn (Y.Z.); caow@njau.edu.cn (W.C.)

[2] King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia; chen.zhao@kaust.edu.sa

* Correspondence: yaoxia@njau.edu.cn

Abstract: Tools for robust identification of crop diseases are crucial for timely intervention by farmers to minimize yield losses. Visual diagnosis of crop diseases is time-consuming and laborious, and has become increasingly unsuitable for the needs of modern agricultural production. Recently, deep convolutional neural networks (CNNs) have been used for crop disease diagnosis due to their rapidly improving accuracy in labeling images. However, previous CNN studies have mostly used images of single leaves photographed under controlled conditions, which limits operational field use. In addition, the wide variety of available CNNs and training options raises important questions regarding optimal methods of implementation of CNNs for disease diagnosis. Here, we present an assessment of seven typical CNNs (VGG-16, Inception-v3, ResNet-50, DenseNet-121, EfficentNet-B6, ShuffleNet-v2 and MobileNetV3) based on different training strategies for the identification of wheat main leaf diseases (powdery mildew, leaf rust and stripe rust) using field images. We developed a Field-based Wheat Diseases Images (FWDI) dataset of field-acquired images to supplement the public PlantVillage dataset of individual leaves imaged under controlled conditions. We found that a transfer-learning method employing retuning of all parameters produced the highest accuracy for all CNNs. Based on this training strategy, Inception-v3 achieved the highest identification accuracy of 92.5% on the test dataset. While lightweight CNN models (e.g., ShuffleNet-v2 and MobileNetV3) had shorter processing times (<0.007 s per image) and smaller memory requirements for the model parameters (<20 MB), their accuracy was relatively low (~87%). In addition to the role of CNN architecture in controlling overall accuracy, environmental effects (e.g., residual water stains on healthy leaves) were found to cause misclassifications in the field images. Moreover, the small size of some target symptoms and the similarity of symptoms between some different diseases further reduced the accuracy. Overall, the study provides insight into the collective effects of model architecture, training strategies and input datasets on the performance of CNNs, providing guidance for robust CNN design for timely and accurate crop disease diagnosis in a real-world environment.

Keywords: wheat leaf diseases; deep learning; scene labeling; convolutional neural networks; transfer learning

check for updates

Citation: Jiang, J.; Liu, H.; Zhao, C.; He, C.; Ma, J.; Cheng, T.; Zhu, Y.; Cao, W.; Yao, X. Evaluation of Diverse Convolutional Neural Networks and Training Strategies for Wheat Leaf Disease Identification with Field-Acquired Photographs. *Remote Sens.* **2022**, *14*, 3446. https://doi.org/10.3390/rs14143446

Academic Editors: Marian-Daniel Iordache and Stephanie Delalieux

Received: 17 June 2022
Accepted: 16 July 2022
Published: 18 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Wheat is one of the most cultivated crops and is the most important food grain for humans [1]. The 2020 World Population Data Sheet (https://interactives.prb.org/2020

-wpds/, accessed on 16 June 2022) indicates that world population will exceed nine billion people by 2050, representing an increase of more than 25% from 2020. Wheat supply must therefore increase to meet this global demand. However, wheat leaf diseases can seriously affect yield and are a major threat to food security worldwide [2]. Infected wheat leaves often show symptoms that are used by experienced agricultural experts to determine the type of disease afflicting the plant. However, the traditional visual method of diagnosis is time-consuming and laborious, requiring highly trained experts who are inherently limited in their ability to cover large regions [3].

With the development of computer vision and deep learning, image processing has achieved great success over the last decade. One of the key techniques leading to this success is convolution neural networks (CNNs), a technology that has been recently used to rapidly identify crop diseases [4]. An early exploration by [5] employed AlexNet [6] and GoogLeNet [7] to identify with an accuracy of 99.35% a total of 14 crop species (apple, bell, blueberry, cherry, corn, grape, orange, peach, potato, raspberry, soybean, squash, strawberry and tomato) and 26 diseases (17 fungal diseases, 4 bacterial diseases, 2 mold diseases, 2 viral diseases and 1 disease caused by mites) in the public dataset PlantVillage [8]. More recent studies have investigated the performance of a wide range of different CNNs in plant disease diagnosis. For example, Ferentinos et al. [9], who also studied plant disease detection based on the PlantVillage dataset, adopted five CNNs, including AlexNet, GoogLeNet, AlexNetOWTBn [10], Overfeat [11] and VGG [12]. VGG achieved the highest accuracy, 99.53%. Similarly, Too et al. [13] used the PlantVillage dataset to compare the accuracy of VGG-16, Inception-v4 [14], ResNet [15] with 50, 101 and 152 layers, and DenseNet-121 [16] for plant disease identification. In their study, the highest accuracy, 99.75%, was achieved by DenseNet-121.

However, the high accuracies of these studies were achieved using images collected under controlled conditions (i.e., a single, excised leaf photographed against a standard background with uniform lighting) without attempting to replicate the challenges of plant disease diagnosis in real-world agricultural fields. For rapid, operational use in the field, for example for a user taking photographs with a mobile phone, diagnosis should preferably use imagery of in situ plants, which implies a complex background (such as soil, weeds and shadows) as well as variable intensity and direction of illumination. Studies that have used field photographs have found lower accuracies. For example, Lu et al. [17] collected a field-based wheat disease dataset (including powdery mildew, smut, black chaff, stripe rust, leaf blotch and leaf rust) to verify the effectiveness of two VGG architectures, and achieved mean recognition accuracies over 95%. Picon et al. [18] employed ResNet-50 [15] for detection of wheat diseases (i.e., septoria, tan spot and rust) in the field, and obtained a balanced accuracy of 87%. Bao et al. [19] designed a lightweight CNN model for detecting wheat ears with scab and glume blight with an accuracy of 94.1%.

Generally, a large number of training samples is ideal to avoid overfitting and enhance generalization of CNN models [20]. Yet in practice, it is hard to collect sufficient field images of different diseases in varied field environment due to the uncertainty of wheat disease occurrence. Therefore, it is desirable to find a strategy to train CNNs on limited images collected in the field, without overfitting and while maintaining good generalization.

Transfer learning [21] is an effective strategy that addresses this issue of small training datasets [22]. In transfer learning, a network is pretrained on a large, somewhat related dataset, and then ‘transferred’ to the new application by updating the network using the small dataset of interest [20]. Ramcharan et al. [23] used transfer learning to train Inception-v3 [14] for cassava disease detection in field images (309~415 images for each disease) with an overall accuracy of 93%. In a recent study on wheat disease detection, Jiang et al. [24] collected 40 field images for each wheat leaf disease (i.e., leaf rust and powdery mildew) and adopted a modified VGG-16 model based on transfer learning to achieve 98.75% accuracy.

In summary, previous CNN studies for plant disease identification have mostly used images of single leaves, photographed under controlled conditions, which limits operational field usage. Moreover, the wide variety of available CNNs and training options raises important questions regarding optimal methods of implementation of CNNs for disease diagnosis. In this study, we assess multiple scene-labeling CNNs based on different training strategies for wheat leaf disease diagnosis using field images. We collect field images of three common wheat leaf diseases (powdery mildew, stripe rust and leaf rust) and healthy leaves to build a custom dataset. The PlanetVillage dataset with a large number of crop disease images is used to pretrain CNNs based on transfer learning. Seven typical CNNs (VGG-16, Inception-v3, ResNet-50, DenseNet-121, EfficentNet-B6, ShuffleNet-v2 and MobileNetV3) are evaluated in terms of detection accuracy, operation time and model parameters. In addition, potential factors influencing the performance of CNNs in wheat disease identification are discussed, with the aim of enhancing fast and accurate crop disease diagnosis.

2. Materials and Methods

2.1. Datasets

In this study, two datasets were used: the PlantVillage dataset and the Field-based Wheat Diseases Images (FWDI) dataset. These datasets are described below and are summarized in Table 1.

Table 1. The composition of datasets used in this study.

Dataset	Disease Type	Original Images	Original Training Set	Augmented Images	Augmented Training Set	Test Set
PlantVillage	26 types	37,721	32,739	-	-	4982
FWDI	Powdery mildew	561	449	2806	2694	112
	Leaf rust	808	647	4043	3882	161
	Stripe rust	1015	812	5075	4872	203
	Healthy wheat	259	208	1299	1248	51
	Total	2643	2116	13,223	12,696	527

2.1.1. PlantVillage Dataset

Due to its large number of images and free availability, the PlantVillage dataset has been widely used for plant disease classification based on deep learning [5,25,26]. The dataset consists of color images of 14 crop species (apple, bell, blueberry, cherry, corn, grape, orange, peach, potato, raspberry, soybean, squash, strawberry and tomato), with examples of both healthy and unhealthy leaves, representing 26 types of diseases (17 fungal diseases, 4 bacterial diseases, 2 mold diseases, 2 viral diseases and 1 disease caused by mites) [8]. Each image is a single leaf against a standard background. The labeled PlantVillage dataset used in this study is offered by AI Challenger 2018, a global AI contest (https://ai.chuangxin.com/ai_challenger?lang=en-US, accessed on 18 June 2021) that includes 32,739 images for training and 4982 images for testing (Table 1). Given the varying resolutions of images in the PlantVillage dataset, we resized the selected images to 224 × 224 pixels while preserving the details of the lesions to adapt the input size of CNN models [13,17–19]. Finally, the resized images were normalized by scaling all pixel values from [0, 255] to [0, 1].

2.1.2. Field-Based Wheat Diseases Images (FWDI) Dataset

We collected the images during the wheat-growing season of 2019–2020 at three sites (Pukou: 118°63′E, 32°06′N; Yixing: 119°82′E, 31°36′N; Rugao: 120°56′E, 32°3′N) located in the Yangtze-Huai Plain, one of the major agricultural regions in China. We collected 2643 wheat disease images under field conditions to create the FWDI dataset. The FWDI dataset contains wheat images in complex environmental backgrounds, different capture conditions, different stages of diseases development and similar appearance between different wheat diseases (e.g., stripe and leaf rust). The dataset is comprised of 259 images of healthy leaves and 2384 images of leaves exhibiting evidence of three common fungal diseases of wheat: powdery mildew (561 images), leaf rust (808 images) and stripe rust (1015 images) (Figure 1).

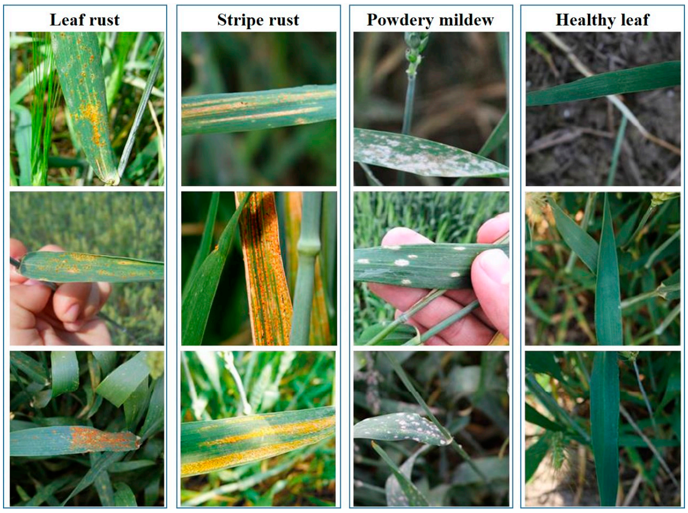


Figure 1. Sample images from our Field-based Wheat Diseases Images (FWDI) dataset.

Powdery mildew, caused by the obligate biotrophic ascomycetous fungus (*Blumeria graminis* f. sp. *Tritici*), is characterized by spots or patches of white-to-gray talcum-powder-like growth [27]. Stripe rust, commonly known as yellow rust, is caused by Puccinia *striiformis* [28]. Leaf rust, also known as brown rust or orange rust, is caused by the fungus Puccinia *triticina* [29]. Stripe and leaf rust have similar characteristic colors but different shapes. Stripe rust has yellow pustules arranged in a linear, stripe-like pattern along the leaf, whereas leaf rust has orange-brown pustules in circular to oval patterns randomly distributed on the leaf.

The images were captured between 8 h and 17 h local time using a Canon EOS 6D camera (Canon Inc., Tokyo Japan) and saved in JPEG format. The images have the following characteristics: (1) every image contains only one class of wheat disease, (2) the diseased wheat leaf in every image is unfolded, (3) the images capture a range of disease development, and (4) every image maintains all the complexity and clutter of the field environment. The images were annotated as healthy or with disease type labels, and were double-checked by experts.

We randomly selected 80% of images as the training set and 20% of images as the test set (Table 1). The resolution of the raw images in the FWDI dataset is 4000 × 5328 pixels. To preserve the major characteristics of the lesions, all images were center-cropped to 1024 × 1024 pixels and were then resized to 224 × 224 pixels using a nearest-neighbor algorithm (Figure 1). The images were then normalized to [0, 1].

To increase the size and variability of the training set, all resized images underwent geometric transformation, including horizontal flip, vertical flip and clockwise rotation of 90°, 180° and 270° (Figure 2). Geometric transformation does not change the labeled disease in the image and has been widely used for augmented training sets of CNNs [6,12]. The augmented training set is comprised of 12,696 images in total.

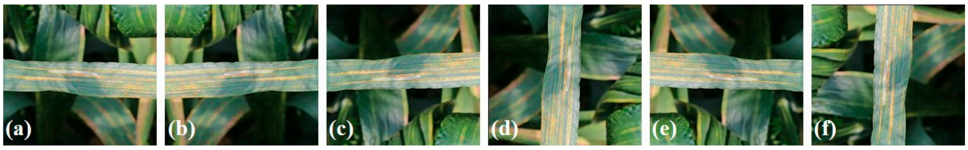


Figure 2. Geometric transformation of images in our Field-based Wheat Diseases Images (FWDI) dataset: (**a**) original image, (**b**) horizontal flip, (**c**) vertical flip and clockwise rotations of (**d**) 90°, (**e**) 180° and (**f**) 270°.

2.2. Convolutional Neural Networks (CNNs)

We selected seven representative CNNs for wheat disease identification: VGG-16, Inception-v3, ResNet-50, DenseNet-121, EfficientNet b6, ShuffleNet-v2 and MobileNetV3. These CNNs are described below.

2.2.1. VGG-16

VGG-16 is a CNN architecture with 16 convolutional layers [12]. It was used to win the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2014. A major contribution of VGG16 is that it improves AlexNet by replacing the convolutional layers with 3 × 3 filters of stride 1, and the max pooling layers with 2 × 2 filters of stride 2, and adopts the same convolution and max pooling filter sizes consistently throughout the whole architecture. It has two fully connected (FC) layers, each with 4096 output channels, followed by a final FC layer with the softmax operation to output probabilities of all classes.

2.2.2. Inception-v3

Inception-v3 is a CNN architecture from the Inception family [14] that contains inception modules originally used in Inception-v1 (also known as GoogLeNet) [7]. Like VGG, GoogLeNet achieved top results at ILSVRC 2014. Its key innovation is that the inception module computes multiple convolutions with filter sizes in parallel on the same input data, and concatenates their results as a single output. The underlying concept is to handle objects of various scales. The inception module also inserts extra 1 × 1 convolutions before the convolutions to reduce channel dimensions, allowing more-efficient computation. Based on previous versions, Inception-v3 added several new techniques, including the RMSProp optimizer, factorized 7 × 7 convolutions, BatchNorm in the auxiliary classifier, and label smoothing.

2.2.3. ResNet-50

ResNet, short for residual network, is a type of neural network that introduces residual learning [15]. It won first place at the ILSVRS 2015 classification competition and Common Objects in Context (COCO) 2015 competition in ImageNet detection, ImageNet localization, COCO detection and COCO segmentation. Instead of learning features, ResNet tries to learn residuals by adding shortcut connections (or skip connections). The skip connections in ResNet solve the problem of vanishing gradients in deep neural networks by allowing gradients to flow through these alternate shortcut paths. ResNet-50 is a variant of ResNet with 48 convolution layers along with 1 max pooling layer and 1 average pooling layer.

2.2.4. DenseNet-121

DenseNet-121 is a variant of a dense convolutional network (DenseNet) with 121 layers in the neural network. DenseNet is an architecture that focuses on making deep neural networks even deeper yet more efficient to train by using dense shortcut connections between layers [16]. Like ResNet50, DenseNet was developed to improve the decline in accuracy in visual object recognition caused by the vanishing gradient problem. To maximize information flow between the layers of the network, each layer is connected to all other deeper layers in the network. Unlike ResNet, DenseNet does not combine features through summation but instead uses concatenation.

2.2.5. EfficentNet-B6

EfficentNet-B6 is a lightweight and easy-to-deploy model in a new family of EfficientNet CNNs proposed by Google in 2019 [30]. It focuses on improving not only the accuracy but also the efficiency of models. EfficientNet starts from a small EfficientNet-B0 model obtained with grid search under a specific constraint, and uses a compound scaling to uniformly scale all dimensions of the model, i.e., depth, width and resolution. Compound scaling is motivated by the intuition that if the input image is bigger, then the network needs more layers to increase the receptive field and more channels to capture more finegrained patterns within the bigger image. EfficentNet-B6 is a variant that scales up from EfficientNet-B0 by ~8 times in terms of the number of parameters.

2.2.6. ShuffleNet-v2

ShuffleNet is an extremely computationally efficient CNN architecture designed for mobile devices with limited computing power. It uses pointwise group convolution and channel shuffle to reduce computational cost while maintaining accuracy [31]. ShuffleNet is optimized for direct (such as actual running speed or memory access) rather than indirect metrics to measure the network's computational complexity rather. Built upon ShuffleNet-v1, ShuffleNet-v2 introduces a new channel split operation and moves the channel shuffle operation further down the block.

2.2.7. MobileNetV3

MobileNet models are designed by Google for mobile and embedded vision applications [32]. They use depthwise separable convolutions to reduce the number of parameters, which results in lightweight deep neural networks. Specifically, they split the 3 × 3 convolutional layer of a conventional CNN into a 3 × 3 depthwise convolution and a 1 × 1 pointwise convolution. MobileNetV3, proposed in 2019 [33], is the latest generation of MobileNet. It is tuned to mobile phone central processing units (CPUs) by combining hardware-aware network architecture search (NAS) [34] complemented by the NetAdapt algorithm and novel architecture advances. The advances include new network designs such as the linear bottleneck and inverted residual structure in MobileNetV2, new efficient versions of nonlinearities practical for the mobile setting, and squeeze and excitation in MnasNet [30].

2.3. Comparison and Evaluation

2.3.1. Training Strategies of CNNs

The purpose of training is to find CNN parameter values that minimize the differences between predictions and ground truth labels on a training dataset [22]. The parameters optimized include kernels in convolution layers and weights in fully connected layers and other layers, e.g., batch normalization. In this study, we compared three training strategies (Figure 3): (1) training from scratch, (2) transfer learning with fixed feature extraction, and (3) transfer learning with retuning.

Training a CNN from scratch (Figure 3a) is a process to train the whole network on the target dataset (i.e., the FWDI dataset) with all its parameters randomly initialized. In contrast, transfer learning, which includes fixed feature extraction and retuning, initializes the network parameters (usually those in the layers before the fully connected layers) using a model pretrained on a large dataset, (i.e., the PlantVillage dataset in our case). Fixed feature extraction (Figure 3b) fixes the values of the parameters initialized from the pretrained model, using them as a feature extractor, and replaces the fully connected (FC) layers of the pretrained model with new FC parameters trained on the target dataset. Retuning (Figure 3c) does not only train the FC layers, but also retunes all the parameters initialized from the pretrained model. Note that in all the training strategies, we replace the fully connected layers with global average pooling, which generates one feature map for each corresponding target category and feeds the resulting vector directly into the softmax layer [35]. The use of global average pooling can reduce computational complexity and the number of model parameters, thus reducing overfitting as a structural regularizer.

Our codes are written in Python and all the CNNs were implemented with Keras [36] running on top of TensorFlow. We run our codes on a computer with Tesla V100, 16 GB RAM and Ubuntu 16.04. To ensure comparability between the different experiments, the hyperparameters for all CNNs used in this study are chosen over a validation set randomly selected from the training sets. Hyperparameters are the variables that determine the network structure or how the network is trained. We set the initial learning rate as 0.001 and use ReduceLROnPlateau to reduce the learning rate when a metric has stopped improving.

Adam [37], an adaptive learning rate optimization algorithm, is employed to minimize the loss. We set the batch size as 32 to train the models (i.e., 32 samples are used to compute the loss and update the network parameters at one time). The maximum number of iterations is set to 50 according to trial-and-error experiments.

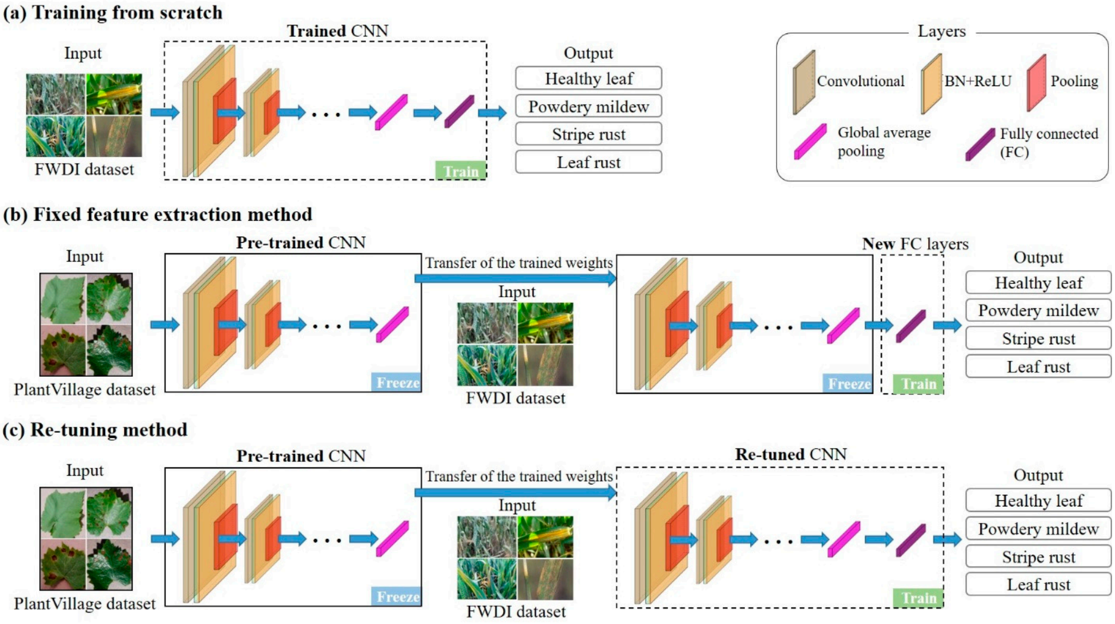


Figure 3. Training strategies of convolutional neural network (CNN) used in this study.

2.3.2. Model Assessment

For each deep model, we report *accuracy*, *memory* and *time* to measure the performance of CNNs. *Accuracy* is defined as the ratio of correctly classified labels (*match*) to the total number of images (*N*) in the test set:

$$Accuracy = \frac{\sum_i^N match_i}{N}, \quad (1)$$

where *i* is the *i*th target image. In addition, a confusion matrix is used to report the models' accuracy for each class (i.e., powdery mildew, stripe rust, leaf rust and healthy leaf).

Memory (in megabytes, MB) is the GPU memory occupied by model parameters:

$$Memory = \frac{N_{parm} \times 4}{1024 \times 1024}, \quad (2)$$

where N_{parm} is the number of model parameters, which is multiplied by 4 bytes, the size of each parameter represented by a float variable. Model parameters are weights in the network that are learned during training, mainly comprised of the weights and biases that need to be learned in the convolution layers and FC layers, and the mean and deviation parameters that need to be learned in the batch normalization.

Time (in milliseconds, ms) represents the processing time of a network to return a result for one image through forward propagation of the network.

3. Results

3.1. Accuracy of CNNs Trained by Different Training Sets and Strategies

As shown in Figure 4, the accuracies of CNNs trained on the augmented training set were generally higher than those trained using the original training set. This indicates that image augmentation of training sets can improve the accuracy of CNNs when the training data size is small. Augmentation of training sets had a greater impact on training accuracy than validation accuracy, as expected. For the validation data, augmentation showed the greatest benefit for models trained from scratch, highlighting the importance of large training sets for CNNs and the value of data augmentation for overcoming dataset size limitations.

Figure 4 also shows that the training strategy affects the accuracy of CNNs. CNNs trained from scratch had relatively low accuracy for early iterations compared to those of the other two training strategies, indicating slow convergence. CNNs trained by fixed feature extraction converged more rapidly than the other approaches, but their accuracies were generally the lowest. In contrast, the accuracies of CNNs using retuning were generally higher than those using the other two training strategies.

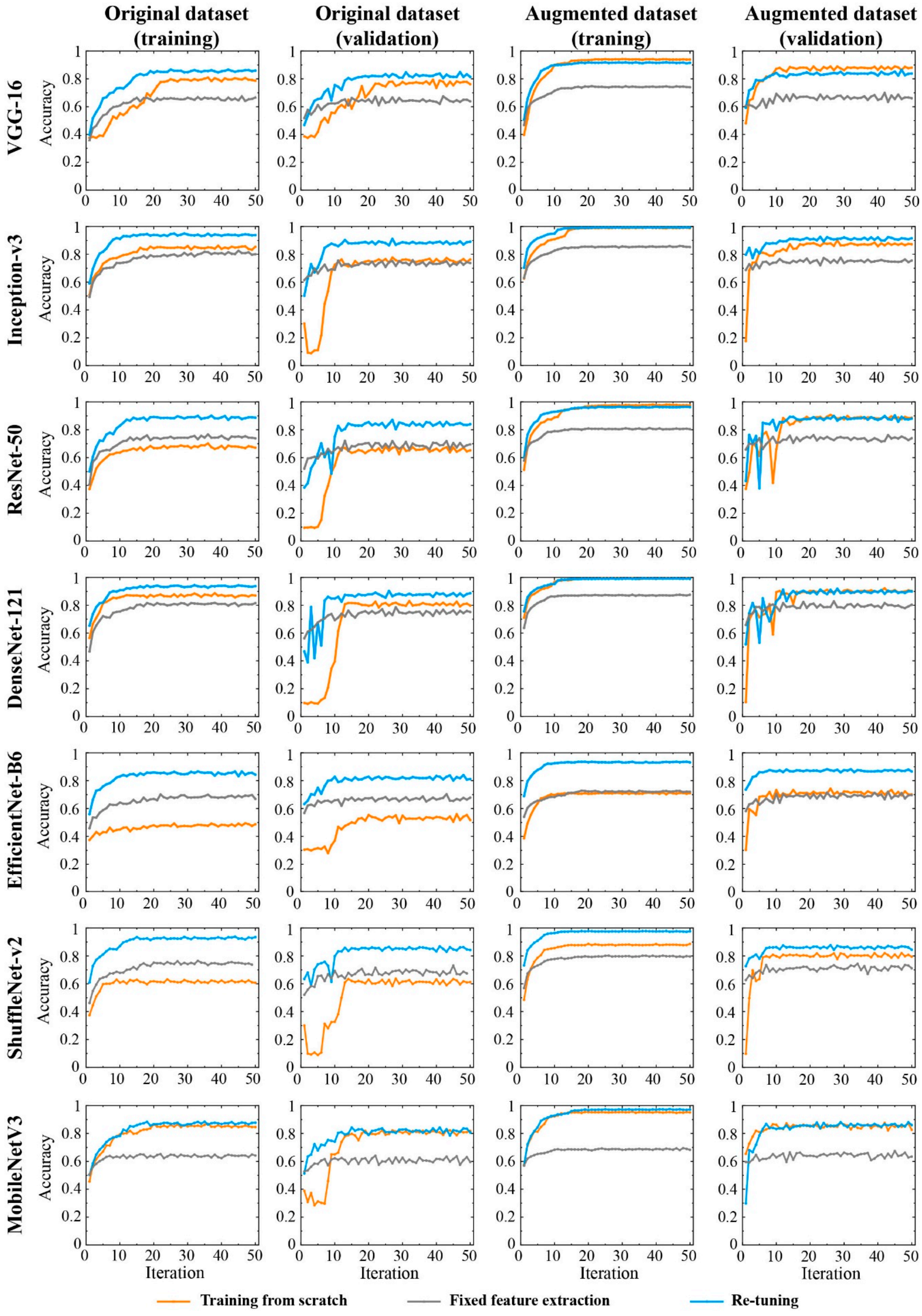
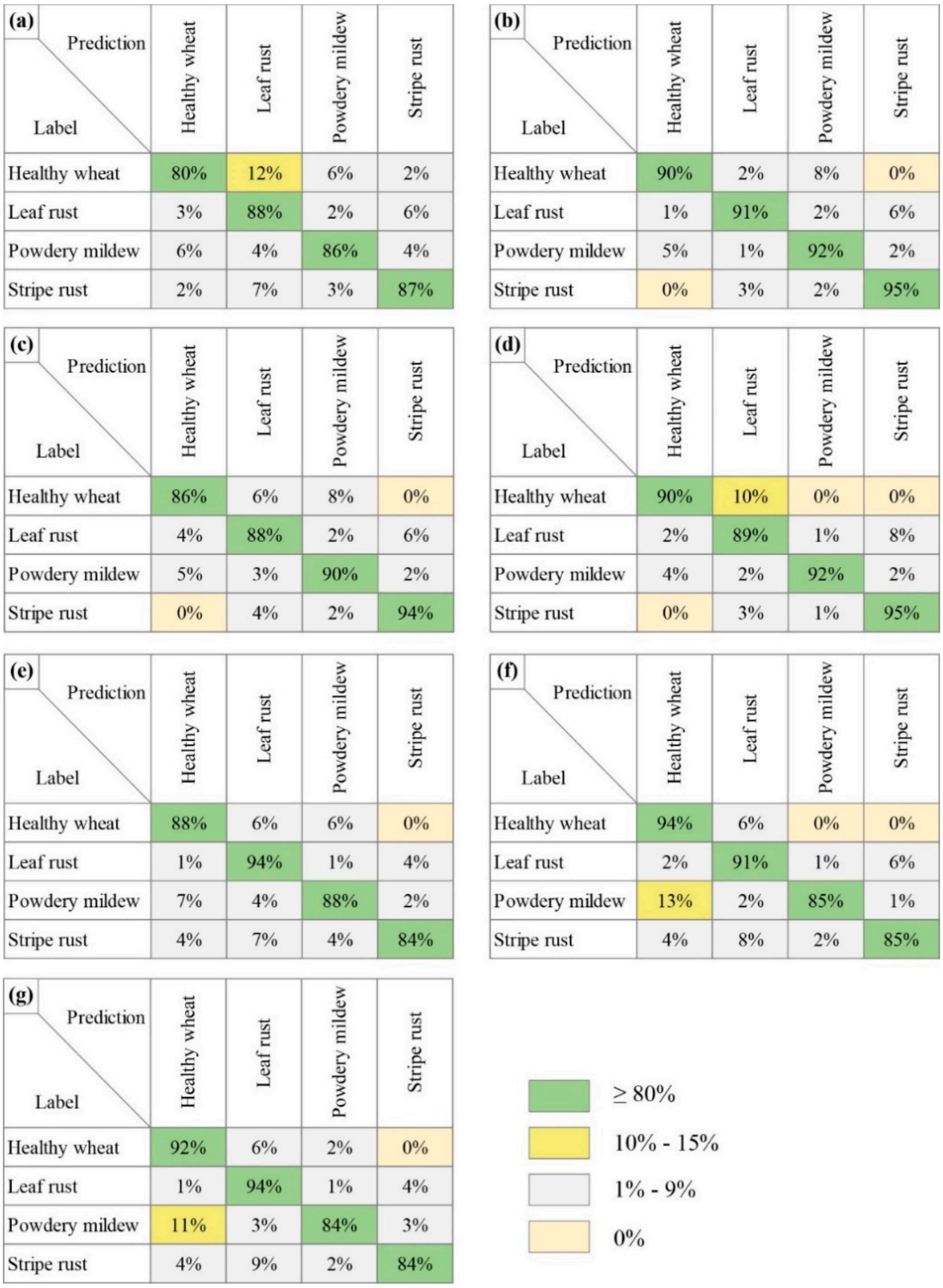


Figure 4. Comparison of training accuracy and validation accuracy of CNNs developed by the original FWDI and augmented datasets using different training strategies (i.e., training from scratch, fixed feature extraction and retuning).

3.2. Wheat Disease Diagnosis

Since the CNNs trained on the augmented dataset using retuning provided the best results, we explore their accuracy for classification of the test dataset in Figure 5. Inception-v3 had the highest class-based recall accuracy of 90–95% for healthy wheat and the three diseases (Figure 5b), while VGG-16 had the lowest recall accuracy of 80–88% (Figure 5a). The lowest recall class-based accuracy for all the methods was 80%.



(a)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	80%	12%	6%	2%
Leaf rust	3%	88%	2%	6%
Powdery mildew	6%	4%	86%	4%
Stripe rust	2%	7%	3%	87%

(b)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	90%	2%	8%	0%
Leaf rust	1%	91%	2%	6%
Powdery mildew	5%	1%	92%	2%
Stripe rust	0%	3%	2%	95%

(c)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	86%	6%	8%	0%
Leaf rust	4%	88%	2%	6%
Powdery mildew	5%	3%	90%	2%
Stripe rust	0%	4%	2%	94%

(d)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	90%	10%	0%	0%
Leaf rust	2%	89%	1%	8%
Powdery mildew	4%	2%	92%	2%
Stripe rust	0%	3%	1%	95%

(e)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	88%	6%	6%	0%
Leaf rust	1%	94%	1%	4%
Powdery mildew	7%	4%	88%	2%
Stripe rust	4%	7%	4%	84%

(f)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	94%	6%	0%	0%
Leaf rust	2%	91%	1%	6%
Powdery mildew	13%	2%	85%	1%
Stripe rust	4%	8%	2%	85%

(g)

Label \ Prediction	Healthy wheat	Leaf rust	Powdery mildew	Stripe rust
Healthy wheat	92%	6%	2%	0%
Leaf rust	1%	94%	1%	4%
Powdery mildew	11%	3%	84%	3%
Stripe rust	4%	9%	2%	84%

Figure 5. Confusion matrix analysis of different CNNs based on the augmented training set and retuning with 50 iterations. CNNs include (**a**) VGG-16, (**b**) Inception-v3, (**c**) ResNet-50, (**d**) DenseNet-121, (**e**) EfficentNet-B6, (**f**) ShuffleNet-v2 and (**g**) MobileNetV3.

Three typical causes of recognition errors were identified: disease in different stages, small-sized targets and environmental effects (Figure 6). As shown in Figure 6a, the powdery mildew in this image was in the late stage of disease development, and the associated white mold layer had developed into white-yellow spots, leading to misclassification as leaf rust. Alternatively, if the leaf or the spots on the leaf were small in the image, the disease was easily misclassified (Figure 6b). White spots due to residual water stains on healthy leaves were identified incorrectly as powdery mildew (Figure 6c).

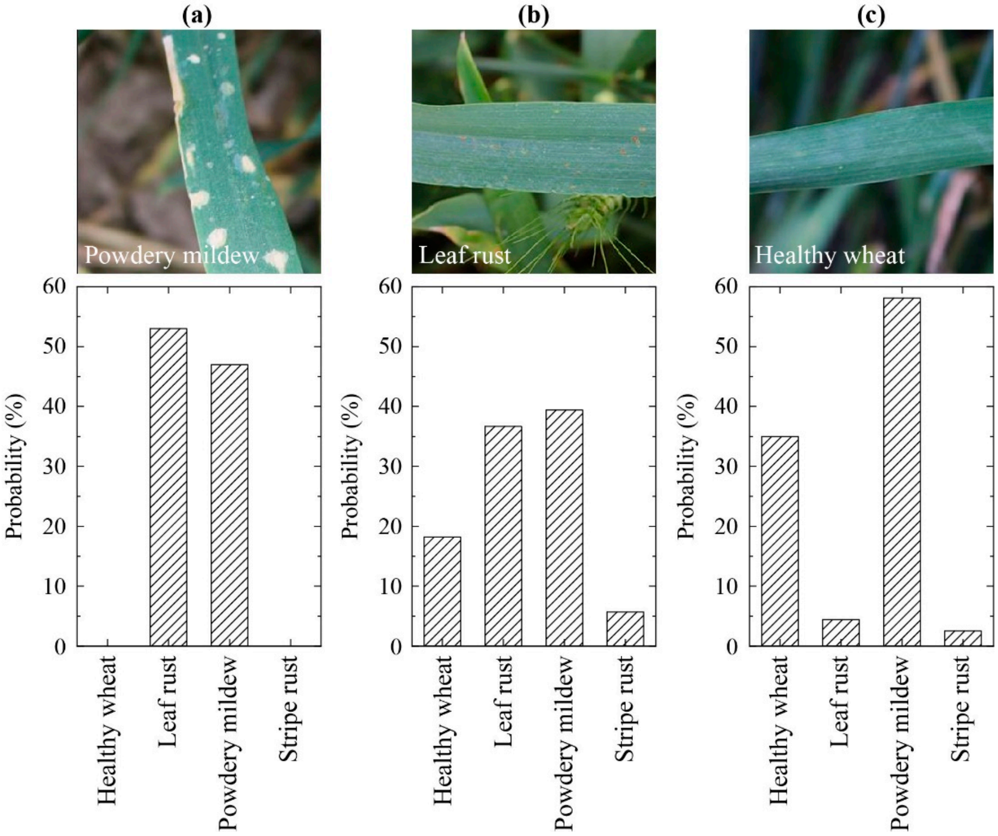


Figure 6. Examples of incorrectly identified images of wheat diseases: (**a**) powdery mildew, (**b**) leaf rust and (**c**) healthy wheat. The graphs below each image indicate the class-based classification probabilities for the specific image shown.

3.3. Comparative Evaluation of CNNs

Figure 7 shows the comparison of CNNs based on retuning in terms of accuracy, memory and processing time for wheat disease diagnosis. For all CNNs, the GPU memory occupied by model parameters (*memory*) was less than 100 MB. Because the first FC layer was replaced by global average pooling for all the models, the number of parameters of the FC layer was reduced compared to the originally proposed models. ResNet50 required the largest *memory* (98 MB), followed by Inception-v3 (83.2 MB), due to the large number of channels in their convolutional operations. Although DenseNet-121 is a deep network, its parameters occupied only 30.87 MB of GPU memory because 1 × 1 convolution was used to reduce the number of feature channels. MobileNetV3, with deep detachable convolution, had the lowest *memory* (11.95 MB), making it a lightweight network, but at the cost of a relatively low accuracy (88.08%). The highest accuracy was obtained by Inception-v3 (92.53%), followed by DenseNet-121 (91.92%) and ResNet50 (90.3%). While VGG-16 had the lowest accuracy, its operation time was the shortest, with an average time of 4.64 ms per image. Given that VGG-16 had only 16 layers, training from the input layer to the output layer was fast. Conversely, DenseNet-121, with 121 layers, needed more time to identify each image (20.31 ms).

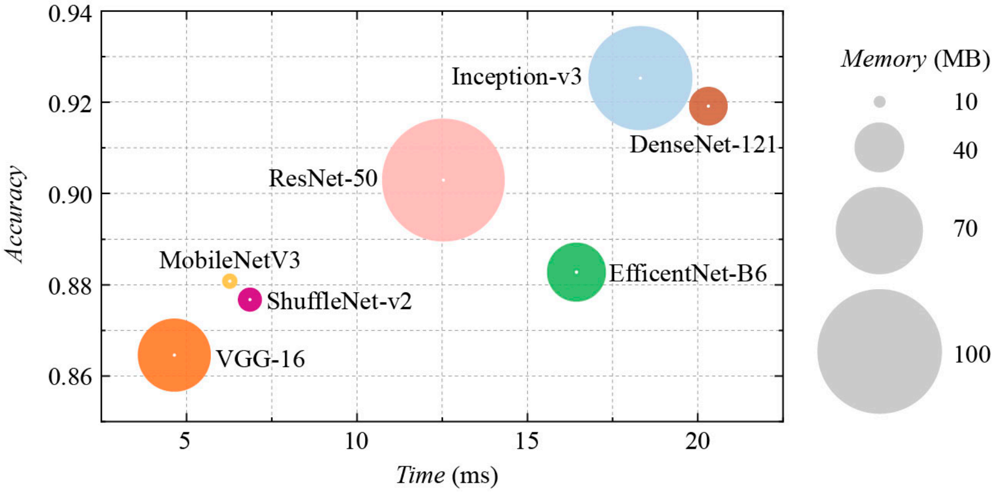


Figure 7. Performance of different models for wheat disease diagnosis. *Accuracy* is the ratio of correctly classified labels to the total number of images in the test. *Memory* is the GPU memory occupied by model parameters. *Time* represents the processing time of a network needed to return a result for one image. See Section 2.3.2 for details.

4. Discussion

In this study, we assessed seven CNNs based on three different training strategies for wheat leaf disease diagnosis. According to our results, the potential factors influencing the performance of CNNs in wheat disease identification include the network architecture of the CNN model, the training strategy and the input dataset. Moreover, the limitations and prospects of the CNN model applied in precision agriculture are highlighted, with the aim of enhancing fast and accurate crop disease diagnosis.

4.1. Influencing Factors of CNNs Applied in Crop Diseases Diagnosis

The main factor influencing the performance of different CNNs in crop disease diagnosis is the network architecture. Generally, deep CNN models have higher accuracy than shallow networks [16]. For example, Too et al. [13] found that DenseNet-121 (with 121 layers) and ResNets (with 50, 101 and 152 layers) had test accuracies over 99%, while the VGG net with 16 layers achieved 81.83% for plant disease identification. Similarly, our results show the lowest accuracy for VGG-16 (86%) and relatively high accuracy for deeper networks, such as ResNet50 (90%) and DenseNet-121 (92%) (Figure 7). In contrast, Wang et al. [38] found that VGG-16 performed the best, with the highest accuracy of 90.4% on the PlantVillage dataset, outperforming VGG-19, Inception-v3 and ResNet50. Fuentes et al. [39] also found the best performance for VGG-16 in tomato plant disease and pest diagnosis rather than the deeper networks. These results indicate that more complex and deeper network do not necessarily yield greater accuracy on different tasks.

The performance of CNNs is also affected by the training strategy used. Our results indicate that retuning improved the accuracy and reduced overfitting, a finding consistent with previous studies [23,24,39]. Compared to the other two training strategies, CNNs trained from scratch had slow convergence. This is because all parameters were randomly initialized and optimized from scratch without any prior knowledge. Conversely, the CNNs trained by fixed feature extraction converged more rapidly but had the lowest accuracy. This is because only the FC layer was trained after the transfer, making the network easier to train. However, because the other parameters were directly transferred from the PlantVillage dataset, they were not optimized for the specific features of the FWDI. Therefore, retuning, which is comprised of learning general characteristics from a large dataset and then modifying the pretrained network to adapt to the characteristics of the small dataset, is recommended for tasks with limited training samples.

As the basic source of information, the input dataset is crucial for training CNNs. First, the dataset size is important, as discussed above [40], and consequently data augmentation is often used to increase the number of images in the training set [41]. While the number of training images for each wheat disease was unequal (Table 1), the classification accuracy of different diseases did not depend on the relatively small disparity in image number (Figure 5) but was mainly affected by the enlarged dataset size after data augmentation (Figure 4). By providing geometrically transformed replicates of the training images, augmentation improves the accuracy of CNNs by providing a larger and more general training dataset (Figure 4), an observation that has also been noted in previous studies [40,42]. Second, image quality can interfere with classification results. In particular, images collected in field conditions can be affected by environmental factors such as complex backgrounds, unstable lighting, image blur due to movement or lack of focus, and the presence of water drops or marks, all of which might lead to misclassification (e.g., Figure 6c). Third, confusing representation of disease symptoms is also a potential factor causing uncertainty in crop disease diagnosis. Two typical causes of recognition errors observed in this study were diseases in different stages of development and symptoms that had a small size (Figure 6a,b). In addition, the simultaneous occurrence of more than one disease and different diseases that have similar symptoms can also weaken the robustness of CNN models [40]. Therefore, annotated datasets of large size and rich variety will always be required for crop disease diagnosis.

4.2. Limitations and Prospects for Precision Agriculture

For precision agriculture, the diagnosis of crop disease severity is of great significance for precise disease prevention and control. However, previous studies have mainly focused on classifying disease type, rather than identifying the severity of each type of disease. Therefore, in future work, it will be necessary to collect a large number of images of different severities for each of the various diseases to assess the capabilities of CNNs to detect crop disease severity.

This study used RGB images for wheat disease identification. In addition to RGB cameras, more-informative sensors (e.g., multispectral or hyperspectral sensors) can be effectively used for image acquisition of crop disease [4,43,44]. Generally, multi- and hyperspectral sensors are mounted on unmanned aerial vehicle (UAV) and can obtain more spectral information of crop disease and cover larger areas than can be accomplished with proximal measurement [45,46]. Therefore, it is worth further exploring the performance and potential of CNNs and different training strategies on crop disease diagnosis.

Lightweight CNNs on mobile devices could be a breakthrough technology for assisting farmers in agricultural management, and therefore should be improved for practical applications. The ability to analyze larger images with CNNs would also be an important development. The size of input images for existing networks is usually no more than 256 × 256 pixels, and this may not resolve subtle symptoms, which can be important for early disease detection [47,48]. In addition, model accuracy should be improved. While the lightweight CNNs (e.g., ShuffleNet V2 and MobileNetV3) had high classification speed and occupied low GPU memory, their accuracies were relatively low in our experiments (Figure 7). Furthermore, given that the GPU performance of mobile devices is inferior to that of GPUs on computers, classification speed will be even slower when lightweight CNNs are deployed to mobile devices. Hence, the tradeoffs among accuracy, time and memory should be considered in model design.

5. Conclusions

We assessed the performance of seven state-of-art CNNs (VGG-16, Inception-v3, ResNet-50, DenseNet-121, EfficentNet-B6, ShuffleNet-v2 and MobileNetV3) in wheat fungal disease identification. Our analysis, undertaken using the public PlantVillage dataset and our FWDI dataset, quantified the impact of training strategies and other potential factors on the detection accuracy of CNNs. Considering training strategy effects, the results show that retuning increased the overall accuracies compared to training from scratch and fixed feature extraction.

Network architecture is a basic characteristic of different CNNs and directly affects their performance. The results demonstrated the importance of balancing recognition accuracy, operation time and model parameters in the model design for different tasks. In addition to CNN model architecture, image quality and capture conditions (such as complex backgrounds, unstable lighting and image blurring) might lead to misclassification, especially in field conditions. Moreover, symptom representation of leaf disease is a potential factor causing uncertainty in crop disease diagnosis. If the size of spots on a leaf is too small in the image, the disease will likely be classified with low confidence or misclassified. To benefit from the great potential of CNNs, annotated datasets of large size and rich variety remain a crucial element for crop disease diagnosis, and robust CNNs on mobile devices are desired for practical applications.

Author Contributions: Conceptualization, X.Y. and J.J.; methodology, H.L. and J.J.; software, H.L.; validation, J.J. and C.H.; formal analysis, J.J.; investigation, H.L. and J.J.; writing—original draft preparation, J.J., H.L. and X.Y.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Civil advance space technology research project (D040104), the National Natural Science Foundation of China (31971780), the Key Projects (Advanced Technology) of Jiangsu Province (BE 2019383), the National Key Research and Development Program of China (2021YFE0194800) and Collaborative Innovation Center for Modern Crop Production co-sponsored by Province and Ministry (CIC-MCP).

Acknowledgments: We thank Timothy A. Warner for his help with manuscript review.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Curtis, B.; Rajaram, S.; Macpherson, H. *Bread Wheat: Improvement and Production*; Food and Agriculture Organization of the United Nations (FAO): Rome, Italy, 2002.
2. Figueroa, M.; Hammond-Kosack, K.E.; Solomon, P.S. A review of wheat diseases-a field perspective. *Mol. Plant Pathol.* **2018**, *19*, 1523–1536. [CrossRef] [PubMed]
3. Wang, X.F.; Huang, D.S. A Novel Density-Based Clustering Framework by Using Level Set Method. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1515–1531. [CrossRef]
4. Khosrokhani, M.; Nasr, A.H. Applications of the Remote Sensing Technology to Detect and Monitor the Rust Disease in the Wheat–a Literature Review. *Geocarto Int.* **2022**, 1–27. [CrossRef]
5. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef]
6. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Processing Syst.* **2012**, *25*, 1097–1105. [CrossRef]
7. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]
8. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.
9. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [CrossRef]
10. Krizhevsky, A. One weird trick for parallelizing convolutional neural networks. *arXiv* **2014**, arXiv:1404.5997. [CrossRef]
11. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; Lecun, Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. *arXiv* **2013**, arXiv:1312.6229. [CrossRef]
12. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]
13. Too, E.C.; Yujian, L.; Njuki, S.; Yingchun, L. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **2019**, *161*, 272–279. [CrossRef]
14. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826. [CrossRef]

15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [CrossRef]
16. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [CrossRef]
17. Lu, J.; Hu, J.; Zhao, G.; Mei, F.; Zhang, C. An in-field automatic wheat disease diagnosis system. *Comput. Electron. Agric.* **2017**, *142*, 369–379. [CrossRef]
18. Picon, A.; Alvarez-Gila, A.; Seitz, M.; Ortiz-Barredo, A.; Echazarra, J.; Johannes, A. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* **2019**, *161*, 280–290. [CrossRef]
19. Bao, W.; Yang, X.; Liang, D.; Hu, G.; Yang, X. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* **2021**, *189*, 106367. [CrossRef]
20. Panigrahi, S.; Nanda, A.; Swarnkar, T. A Survey on Transfer Learning. In *Proceedings of the Intelligent and Cloud Computing*; Springer: Singapore, 2021; pp. 781–789.
21. Pan, S.J.; Qiang, Y. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
22. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]
23. Ramcharan, A.; Baranowski, K.; McCloskey, P.; Ahmed, B.; Legg, J.; Hughes, D.P. Deep Learning for Image-Based Cassava Disease Detection. *Front. Plant Sci.* **2017**, *8*. [CrossRef]
24. Jiang, Z.; Dong, Z.; Jiang, W.; Yang, Y. Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Comput. Electron. Agric.* **2021**, *186*, 106184. [CrossRef]
25. Amara, J.; Bouaziz, B.; Algergawy, A. A deep learning-based approach for banana leaf diseases classification. In *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband*; Gesellschaft für Informatik e.V.: Stuttgart, Germany, 2017.
26. Brahimi, M.; Boukhalfa, K.; Moussaoui, A. Deep learning for tomato diseases: Classification and symptoms visualization. *Appl. Artif. Intell.* **2017**, *31*, 299–315. [CrossRef]
27. Cowger, C.; Miranda, L.; Griffey, C.; Hall, M.; Murphy, J.; Maxwell, J.; Sharma, I. *Wheat Powdery Mildew*; CABI: Oxfordshire, UK, 2012; pp. 84–119.
28. Wan, A.M.; Chen, X.M.; He, Z. Wheat stripe rust in China. *Aust. J. Agric. Res.* **2007**, *58*, 605–619. [CrossRef]
29. Samborski, D. Wheat leaf rust. In *Diseases, Distribution, Epidemiology, and Control*; Elsevier: Amsterdam, The Netherlands, 1985; pp. 39–59.
30. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114. [CrossRef]
31. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856. [CrossRef]
32. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
33. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324. [CrossRef]
34. Zoph, B.; Le, Q.V. Neural architecture search with reinforcement learning. *arXiv* **2016**, arXiv:1611.01578.
35. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
36. Chollet, F. Keras: The Python Deep Learning Library. *Astrophys. Source Code Libr.* **2018**, ascl:1806.1022. Available online: https://ui.adsabs.harvard.edu/abs/2018ascl.soft06022C/abstract (accessed on 16 June 2022).
37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
38. Wang, G.; Sun, Y.; Wang, J. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Comput. Intell. Neurosci.* **2017**, *2017*, 2917536. [CrossRef] [PubMed]
39. Fuentes, A.F.; Yoon, S.; Lee, J.; Park, D.S. High-Performance Deep Neural Network-Based Tomato Plant Diseases and Pests Diagnosis System With Refinement Filter Bank. *Front. Plant Sci.* **2018**, *9*, 1162. [CrossRef] [PubMed]
40. Barbedo, J.G.A. Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* **2018**, *172*, 84–91. [CrossRef]
41. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. *ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design*; Springer: Cham, Switzerland, 2018.
42. Liu, B.; Zhang, Y.; He, D.; Li, Y. Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks. *Symmetry* **2018**, *10*, 11. [CrossRef]
43. Mustafa, G.; Zheng, H.; Khan, I.H.; Tian, L.; Jia, H.; Li, G.; Cheng, T.; Tian, Y.; Cao, W.; Zhu, Y. Hyperspectral Reflectance Proxies to Diagnose In-Field Fusarium Head Blight in Wheat with Machine Learning. *Remote Sens.* **2022**, *14*, 2784. [CrossRef]
44. Su, J.; Liu, C.; Chen, W.-H. UAV Multispectral Remote Sensing for Yellow Rust Mapping: Opportunities and Challenges. *Unmanned Aer. Syst. Precis. Agric.* **2022**, 107–122. [CrossRef]
45. León-Rueda, W.A.; León, C.; Caro, S.G.; Ramírez-Gil, J.G. Identification of diseases and physiological disorders in potato via multispectral drone imagery using machine learning tools. *Trop. Plant Pathol.* **2022**, *47*, 152–167. [CrossRef]

46. Zhang, S.; Li, X.; Ba, Y.; Lyu, X.; Zhang, M.; Li, M. Banana Fusarium Wilt Disease Detection by Supervised and Unsupervised Methods from UAV-Based Multispectral Imagery. *Remote Sens.* **2022**, *14*, 1231. [CrossRef]
47. Khaled, A.Y.; Abd Aziz, S.; Bejo, S.K.; Nawi, N.M.; Seman, I.A.; Onwude, D.I. Early detection of diseases in plant tissue using spectroscopy – applications and limitations. *Appl. Spectrosc. Rev.* **2018**, *53*, 36–64. [CrossRef]
48. Liu, J.; Wang, X. Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. *Plant Methods* **2020**, *16*, 83. [CrossRef] [PubMed]