

## Research Article

# A Sparse Deep Transfer Learning Model and Its Application for Smart Agriculture

**Zhikui Chen** <sup>1,2</sup> **Xu Zhang**<sup>1</sup> **Shi Chen**<sup>3</sup> and **Fangming Zhong**<sup>1</sup>

<sup>1</sup>The School of Software Technology, Dalian University of Technology, Dalian 116620, China

<sup>2</sup>The Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian 116620, China

<sup>3</sup>Emporia State University, Emporia, Kansas 66801, USA

Correspondence should be addressed to Zhikui Chen; zkchen@dlut.edu.cn

Received 10 March 2021; Accepted 9 June 2021; Published 23 June 2021

Academic Editor: Keping Yu

Copyright © 2021 Zhikui Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The introduction of deep transfer learning (DTL) further reduces the requirement of data and expert knowledge in various uses of applications, helping DNN-based models effectively reuse information. However, it often transfers all parameters from the source network that might be useful to the task. The redundant trainable parameters restrict DTL in low-computing-power devices and edge computing, while small effective networks with fewer parameters have difficulty transferring knowledge due to structural differences in design. For the challenge of how to transfer a simplified model from a complex network, in this paper, an algorithm is proposed to realize a sparse DTL, which only transfers and retains the most necessary structure to reduce the parameters of the final model. Sparse transfer hypothesis is introduced, in which a compressing strategy is designed to construct deep sparse networks that distill useful information in the auxiliary domain, improving the transfer efficiency. The proposed method is evaluated on representative datasets and applied for smart agriculture to train deep identification models that can effectively detect new pests using few data samples.

## 1. Introduction

Although they have many advantages in performance, deep neural network- (DNN-) based methods often require expert knowledge to label data samples for generating datasets in training. The heavy requirement of labeled data will result in significant training costs, which make it expensive for extension. The deep transfer learning (DTL) can reuse well-trained models for the identification task and transfer knowledge that is learned from laboratory data to help identify in-field data, which alleviates the dependency on labeled datasets to reduce the cost. However, DTL has still not changed the fact that a considerable number of parameters need to be calculated, because they transfer all parameters, and many of the trained DNNs are overparametric. For example, ResNet-18 is the network commonly used as the backbone of DNN-based image recognition, in which up to 11.2 M (million) parameters need to be trained during each epoch [1]. Among these parameters waiting to be transferred, the existence of redundant parameters which are irrelevant to

the target task would result in a large amount of unnecessary computation and memory cost. The redundancy in the network seriously affects the quality of transfer, as well as limiting the possibility for popularizing. Meanwhile, the methods of designing small efficient DNNs with fewer parameters will also face difficulty in inheriting knowledge from DTL, due to their unique network structure.

Agriculture is one of the most important basic industries, which covers a wide range of the world. Agricultural production faces many risks, of which pest and disease outbreaks are the most economic threats [2], and early identification of plants in the field is a crucial first step to detect and control the spread of diseases and pests [3]. Traditional in-field plant pest and disease identification methods rely on human's experience of manual observation and evaluation, with relatively low accuracy and efficiency in detection. With the development of intelligent agriculture, more technologies, i.e., remote sensing, IoT devices, computer vision [4], and unmanned aerial vehicle (UAV) [5], are providing new tools for in-field

plant pest and disease detection based on automated image recognition, which can help with large-scale early identification. At the same time, the data is growing. As the amount of data escalates, more identifications of plant pests and diseases by the aid of DNN are proposed. Images collected in the field are used to train deep networks, in which extracted features are used for recognition and classification. In recent years, DNN-based methods have been applied to the pest and disease identification of cash crops and grain crops [6–9].

For further agricultural extension, approaches with general applicability should be able to be used in mobile terminals, smartphones, and other small devices in the edge computing area. To this end, models must balance the performance with applicability and efficiency and adapt to limited processing power on the basis of ensuring detection accuracy. If only the most necessary parts would be transferred in DTL, it is possible to earn simplified models with lower device requirements for image recognition tasks, which could really reduce the computational cost and retain the advantage of inheriting knowledge by transfer learning. The newly proposed *lottery ticket hypothesis* (LTH) provides a theoretical possibility for this [10]. It finds representative sparse subnetworks by pruning the original network, which can be retrained to achieve equivalent or higher performance, but it uses fewer parameters (even only 5%-10% of the original's). It provides a new idea for plant pest and disease identification based on DTL: the network structure of the source domain is firstly pruned to obtain a sparse subnetwork with key knowledge. Then, the subnetwork, instead of the entire network in traditional DTL, will be transferred as the solution of the target task to achieve a sparse DTL. Thus, the requirement of expert knowledge and in-field data samples and the calculation of parameters can be reduced simultaneously.

As indicated above, in this paper, a sparse deep transfer learning method is proposed and applied to solve the problem of plant pest and disease identification based on image recognition. Firstly, a hypothesis is proposed that a transferable sparse subnetwork structure can be found and its portability can be verified. Then, the steps of the method are designed and used in DNN-based plant pest and disease identification, to seek and transfer an optimal sparse subnetwork to the target task to explore the application in practical problems. Finally, simulation experiments are carried out to show that the method can achieve an equivalent (or even higher) recognition accuracy with a more simplified network architecture and fewer parameters, while retaining the advantage of utilizing existing knowledge through transfer learning.

Thus, the main contributions can be concluded in two-fold aspects:

- (i) To relieve the lack of in-field labeled data and reduce the cost of collecting and labeling data samples by professionals for model training, a DTL-based method is designed, which can moderate the dependence of data in a plant pest and disease identification deep learning model

- (ii) To cope with the defect that the DTL-based method cannot reduce high computational complexity and high hardware requirements, a sparse transfer strategy is designed, which transfers the pruned network structure to reduce the parameters that need to be trained in the model, to simplify the network architecture, reduce the volume and computing cost of the model, and thereby provide the possibility of running the model on ordinary office computers, smartphones, and edge computing devices for better agricultural extension

The rest of this paper is organized as follows. Section 2 is the related works about DNN and traditional DTL-based plant identification and LTH. The proposed sparse transfer hypothesis and the sparse deep transfer learning strategy with its steps are given in Section 3. Section 4 proves the proposed sparse transfer hypothesis on benchmark datasets and verifies the performance of the proposed method on the real dataset. Finally, Section 5 concludes the whole work and gives future discussion.

## 2. Related Works

With the development and popularization of image sensors in DNN-based plant pest and disease identification models, the use of the convolutional neural network (CNN) is becoming an important trend in agriculture. The pests and diseases can be detected and classified by insect individuals, lesions, and representative characteristic changes, which are usually manifested on the leaves of affected plants [11]. Thus, combined with the corresponding agricultural knowledge, images of healthy and diseased leaves can be used as the input of CNN to train the identification model. Methods have been applied to a variety of food and cash crops including but not limited to rice [12, 13], corn [14], tea [15–17], cannabis [18], and apple [19].

In the above deep CNN-based models, there exist two main problems:

- (1) *Data requirements*: the high cost of training models from scratch due to the lack of labeled data and the requirement of expert knowledge in labeling them
- (2) *Model size*: large-scale network architecture occupies much memory and resources, which is not suitable for low-computing-power devices. It increases the difficulty of storage and transmission, which limits the scope of application

For the contradiction between labeled data requirements and the lack of in-field data in (1), DTL is introduced. Mohanty et al. [20] combine DTL to improve efficiency in training a CNN model to identify 14 crop species and 26 diseases and achieve an accuracy of 99.35% with a hold-out test set. Ramcharan et al. [21] apply DTL to identify 2 types of pests and 3 diseases of cassava images taken in the field of Tanzania, inheriting the knowledge of image recognition from GoogLeNet\_Inception v3, and achieve an overall accuracy of 93% for unseen data using 11,670 original images.

Libo et al. [22] use DTL in real-time detection of cole diseases and pests to solve the unbalance classes and false positives generated in training. Thenmozhi and Reddy [23] design a crop pest classification method based on deep CNN and transfer learning. However, DTL usually fails to reduce the requirement of parameters, which results in too large models to deploy on low-computing-power hardware systems.

For (2), the network structure is optimized [24, 25], and particular structures of small efficient CNNs are usually designed with less computation and small volume that is easy to popularize in application. Rahman [26] proposes a two-stage small CNN architecture, which reduced the model size by 99% compared to VGG-16 while remaining an accuracy of 93.3%. Xing et al. [27] develop a weakly dense CNN model for citrus diseases and pest recognition, which is designed from the aspect of parameter efficiency and is simple enough for mobile devices. And regarding designing small efficient neural networks, some state-of-the-art memory-efficient CNN architectures, such as MobileNet [28] and SqueezeNet [29], are usually used as backbones or references. These simple efficient neural networks usually consume less power and take up less memory, which makes it easier to store and deploy them on low-power hardware systems. At the same time, by using fewer parameters, the models require less data for convergence and will be able to avoid dense computing. However, the biggest problem is that because of the distinction in network structure, it is difficult to combine DTL that reuses structures and parameters with these simple CNNs, for commonly used parameters' weights in DTL are usually based on dense networks such as VGG structure or ResNet structure.

To take advantage of DTL in building resource-efficient CNNs, researchers have made a series of efforts [30–32]. In 2019, Frankle and Carbin proposed the LTH [10], which can compress the model by finding representative sparse subnetworks to replace the original dense network. Because the subnetwork is retrainable while preserving the original performance, it may provide a theoretical possibility for generating simple efficient networks from the original dense network and then transfer it to the target task.

To sum up, in the face of these challenges, in this paper, the LTH is modified by using it in DTL to generate transferable sparse structures. Therefore, it transfers only the most necessary knowledge while reducing the volume of the network, to realize the sparse deep transfer learning.

### 3. Methods

The LTH states that for a feedforward DNN, there is an implicit optimal sparse subnetwork structure which is retrainable to achieve the same accuracy as the dense network within the number of original's iterations. It can succeed in finding the subnetwork to retain knowledge and ability from large-scale datasets such as ImageNet in visual recognition tasks [33]. The possibility that whether the subnetwork is able to transfer in the discussion of LTH has been raised, but whether the transfer can always be realized between tasks has so far been inconclusive.

In this section, on the basis of the studies above, a sparse transfer method named WLTs-SDTL is proposed. It transfers only the most important part of the original network, and LTH is modified for generating sparse subnetworks in DTL. The method is then applied in plant pest and disease identification based on image recognition.

#### 3.1. Sparse Transfer Hypothesis for WLT-Nets

**3.1.1. Reviewing the LTH.** The particular subnetwork is generated when randomly initialized, and to seek for it is like to find a “winning lottery ticket” in the original network. We named it as a WLT-net. The retrainable WLT-net is able to be found by unstructured pruning according to the following conditions:

Consider a feedforward network whose loss function  $\ell = \xi(x)$  and initial parameters are defined by  $\lambda_i$ . After training and optimizing, the network achieves the minimum validation of  $\ell$  with the accuracy rate  $\eta\%$  when the number of iterations is  $n$ . The exists WLT-net  $\xi(x; M(\lambda_i))$  from  $\xi(x)$  with iteration  $n' \leq n$  and accuracy rate  $\eta' \geq \eta\%$ , in which  $M(\lambda_i) = m \odot \lambda_i$  is the Mask function; a mask  $m \in \{0, 1\}^{|\lambda|}$  is used for determining and marking which weights will be retained after pruning.

Then, give the definition of the domain and task in DTL: the source domain is denoted as  $D_S$  with task  $T_S$  for providing transferable knowledge and the target domain is denoted as  $D_T$  with task  $T_T$ , in which  $D = \{\chi, P(X)\}$ ,  $X = \{x_1, x_2, \dots, x_n\} \in \chi$  and  $T = \{y, f(x)\}$ ;  $y$  is the label space and  $f(x)$  can be regarded as the nonlinear loss function of DNN to map  $x$  to  $y$ .

Thus, combined with DTL, the sparse transfer hypothesis for WLT-nets is proposed in Assumption 1. The proof procedure shows that when meeting the conditions, a task can be regarded as generated by LTH from a larger dense DNN, while retaining the most necessary backbone architecture and knowledge. Conversely, the proposed hypothesis can be used in designing a transferable WLT-net from an existing arbitrary dense network  $T_S$ , to inherit whose knowledge. The required ability is then able to transfer from  $D_S$  to  $D_T$  through a small efficient WLT-net that is isomorphic to  $T_T$ , which completes the process of a WLT-net-based sparse DTL.

**Assumption 1.** Sparse transfer hypothesis for WLT-nets.

For the task  $T$  waiting to be solved, we modeled it as DNN. The network can be regarded as  $T_T$  in  $D_T$ , where  $T_T = \xi(x; M(\lambda_i))$ .

According to the reverse reasoning of LT hypothesis,  $\exists T_S$ , the corresponding dense network in  $D_S$ , which makes  $T_T$  to be the WLT-net of  $T_S$ , only if  $T_T$  satisfies the following conditions simultaneously.

When the loss function  $\ell$  achieves minimum validation (compared with  $T_S$ ):

- (1) The number of iterations is  $n_T \leq n_S$
- (2) The percentage of accuracy rate is  $\eta_T \geq \eta_S$

- (3)  $\xi(x; M(\lambda_i))$  is able to be obtained from  $T_S$  by unstructured pruning using the Mask function  $M(\lambda_i)$

$$M(\lambda_i) = m \odot \lambda_i, \text{ and mask } m \in \{0, 1\}^{|\lambda_i|}.$$

Now that  $T_T$  is viewed as the WLT-net generated from  $T_S$ , which is proven that it can keep the performance in LT hypothesis, it provides a shortcut to transfer knowledge through sparse structures.

When meeting the conditions, a small efficient network can be regarded as the sparse part of a larger dense DNN. In this way, transferable knowledge can be sought from an isomorphic structure in  $D_S$  for reusing in  $D_T$ , to realize the sparse DTL.

The WLT-net in LTH has been proven to retain the original network's performance with only 5%-10% of parameters left. In particular, when  $\lambda_i$  is used for initializing retraining, the performance is better than that of random initialization. It shows that (1) the main functions of the network are retained in the most important parameters and (2) when the original dense network is initialized, knowledge and skills are simultaneously initialized into the particular WLT-net. Since WLT-net is retrainable, these knowledge and skills are able to transfer within a sparse network backbone as long as the network  $T_T$  in the target domain has an identical structure.

**3.2. WLT-Net-Based Sparse Deep Transfer Learning.** In this section, the specific methods and steps of WLTs-SDTL are proposed. The process is illustrated in Figure 1.

Concretely, the steps of implementation are as follows:

**Step 1.** Locate and identify  $D_S$  in DTL. The structure of  $T_S$  will determine the subsequent sparse structure.

- (i) When  $D_S$  is representative in the corresponding research field, such as ImageNet, which consists of a vast scale of generic data samples, a well-trained  $T_S$  using a high-performance hardware device could always be promising
- (ii) It should be reasonable for choosing a DNN, which has been trained and used to solve certain problems in the practical application, to be the  $T_S$
- (iii) It is also able to train a new DNN from scratch as  $T_S$ , if the task is too unique to seek references. Then, the proposed method will compress it with an acceptable additional computation cost to get a simple efficient network that is easy to promote and deploy

**Step 2.** Prepare to seek WLT-nets in the DNN of  $T_S$ .

- (i) Based on the proposed *sparse transfer hypothesis for WLT-nets*, an iterative pruning algorithm will be used. The weight of initial parameters in  $T_S$  is denoted by  $\lambda_i$ , while after  $n$  time iterations,  $\lambda_n$  denotes the weight when the loss function achieves minimum validation. In addition,  $\lambda_\alpha (\alpha < n)$  is recorded, preparing for *late reset*, a skill to speed up convergence and improve accuracy

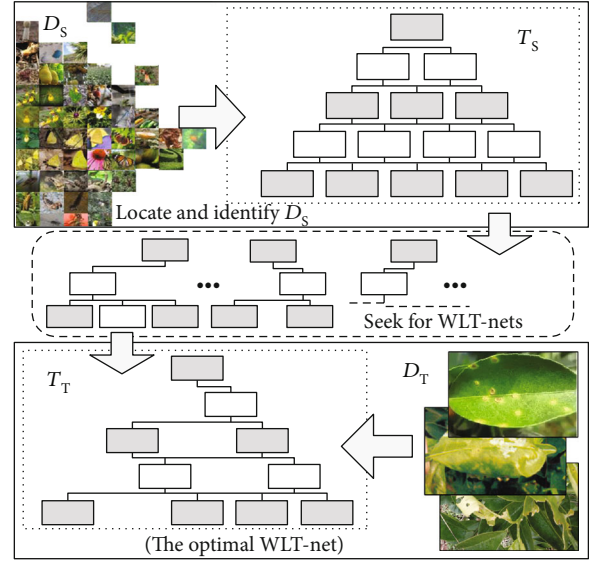


FIGURE 1: Process of the WLTs-SDTL method.

- (ii)  $\mu$  is defined as the rating standard score to measure the contribution of parameters' weights in DNN, whose definition is

$$\mu(\lambda_i) = \max \left( 0, \frac{\lambda_i \lambda_n}{|\lambda_i|} \right). \quad (1)$$

**Step 3.** Generate WLT-nets by unstructured pruning.

- (i) Rank  $\mu$  by score. The Mask function  $M(\lambda_i) = m \odot \lambda_i$  is introduced, using a mask  $m \in \{0, 1\}^{|\lambda_i|}$  to determine whether a parameter should be remained
- (ii) Set a pruning ratio  $p$  for each epoch. Then, in the layers of DNN, for the weights of parameters whose scores are in the top  $p\%$ , use the mask  $m = 1$  to label them on behalf of retaining. Oppositely, the residual  $(100 - p)\%$  parameters will be pruned, whose masks  $m$  are set to 0. The value of  $p$  can be defined for each layer separately

**Step 4.** Parameters are processed according to the value of the mask  $m$  obtained in step 3.

- (i) For parameters whose mask  $m = 1$ , reset their weights to the recorded  $\lambda_\alpha$  in step 2
- (ii) For parameters whose mask  $m = 0$ , prune them. Their weights will be frozen in the subsequent training of DNN, which results in a sparse network architecture
- (iii) Different from the original LTH, the variation trend of weight is considered. The weights of the parameters being pruned are frozen at 0 only when they tend to 0. When the variation trend of weights in



the training is moving away from 0, they are frozen at  $\lambda_i$

*Step 5.* Repeat step 2 to step 4, until the optimal transferable WLT-net  $\xi$  is obtained.

*Step 6.* Use  $\xi$  as  $T_T$  in the target domain  $D_T$  to realize the sparse DTL based on the optimal WLT-net. In the fine-tuned training dataset of  $D_T$ , initialize the DNN with  $\lambda_\alpha$ , while the frozen parameters are not trained.

This way, important skills and information can be inherited from  $D_S$  through the transfer of a sparse WLT-net architecture. The requirement of data samples and computing parameters in the DNN training are both reduced.

Compared with the original LTH, in the proposed WLTs-SDTL method, we have made optimizations and improvements in the following three aspects:

- (i) The LTH is extended and modified for generating sparse networks in DTL. By regarding  $T_T$  as a WLT-net of  $D_S$ , the correlation between  $D_S$  and  $D_T$  is established, enabling knowledge to be transferred between tasks. When following the idea in the original LTH,  $T_T$  should be initialized using a  $\lambda_i'$  in  $D_T$ , not  $\lambda_\alpha$ , so that the knowledge obtained from  $D_S$  will be lost. In the proposed WLTs-SDTL, since we regard  $T_T$  as a part generated from  $T_S$ , it can then be initialized naturally with  $\lambda_\alpha$  conforming to the LTH, while retaining the performance. Thus, the proposed method can achieve the knowledge transfer using sparse WLT-net effectively
- (ii) There is a more reasonable standard to evaluate parameters in pruning. In the pruning process of the original LTH, the rating standard score  $\mu$  in Mask function  $M$  which evaluates the contribution of parameters' weights is defined as  $\mu = |\lambda_i|$ . Clearly, it failed to consider the sign change of parameters in training, i.e., across the zero axis. As shown in formula (1), the proposed  $\mu$  emphasizes the role of signs, which leads to a correct expression of the trend in weight changing. Comparative experiments prove that it can improve the final performance of the sparse network
- (iii) Further consider the influence of the trend in weight changing during training on the freeze and reset of parameters. In the original LTH, after pruning, the parameters will be reset at  $\lambda_i$  ( $m = 1$ ) or frozen at 0 ( $m = 0$ ). Subsequently, the *late reset* [33] is proposed to use the recorded weights after a period of training iterations for the reset, to make the convergence faster and the final accuracy higher. In the proposed WLTs-SDTL method, we also adopt it that  $\lambda_\alpha$  ( $\alpha < n$ ) is used for resetting parameters instead of  $\lambda_i$  in step 4.

Furthermore, since the *late reset* is effective, considering the frozen ones, the weights of pruned parameters will be frozen at 0 to avoid subsequent training, but why 0?

A reasonable explanation is that these weights contribute less to the network and are not important. However, if it

really does not matter, these weights could be set to any value, instead of a particular 0, without affecting the network's performance. In experiments that freeze these parameters to  $\lambda_i$ , it shows that, similar to *late reset*, the validity might depend on whether a specific value can reflect the changing trend of weights in training to some extent. When freezing at 0 correctly, it is equivalent to letting weights whose trends are getting closer to 0 reach to their final value in advance. Thus, in the paper, a parameter will be frozen at 0 only if its trend tends to be 0. When the trend is away from 0, freeze the parameter at its value in  $\lambda_i$  to reduce the impact.

## 4. Experiment

In this section, experiments are designed to verify the hypothesis and evaluate the performance of the proposed method in actual solutions. Firstly, the proposed *sparse transfer hypothesis for WLT-nets* and *WLTs-SDTL* are verified on the benchmark datasets. Then, the WLTs-SDTL method is used to design a detection model based on image recognition and applied to actual solutions of plant pest and disease identification using open-source lab datasets. Finally, a small scale of real datasets that we collected in Chongqing, China, is used in training a model that realized the citrus greening disease (Haunglongbing) identification. The experiments of training models are run on the server which contains 2 Inter Xeon Silver 4110 8-core CPUs and 2 NVIDIA Tesla M60 GPUs (128 G), while some validations are able to run on ordinary office computers since the sparse models are used.

**4.1. Verification of WLT-Net-Based Sparse DTL.** The sparse transfer hypothesis for WLT-nets is verified on the benchmark datasets, i.e., CIFAR-10 and SmallNORB. Specifically, define a DNN-based task  $T_S$  on the  $D_S$  using CIFAR-10, and the proposed method is used to find transferable sparse WLT-net, which will be used as  $T_T$  in DTL. The accuracy and computational load of the model after transfer will be compared with those of the fully connected dense DNN structure trained on SmallNORB, the dataset of  $D_T$ . Through the above approaches, it validates whether the modified hypothesis is able to realize a sparse DTL to reduce the parameters while maintaining the accuracy in the proposed WLTs-SDTL method.

**4.1.1. Datasets.** Identification of plant pests and diseases can be modeled as a multiclassification task based on image recognition. Therefore, two of the classical datasets, CIFAR-10 [34] and SmallNORB [35], are chosen for designing a simple experiment to evaluate the feasibility of the hypothesis, which are widely used in identifying ubiquitous objects and regarded as the benchmark to validate various models. The properties of the datasets are shown in Table 1. Since the channel and image size of  $T_S$  are different from those of  $D_S$ , channel conversion and 4-pixel padding are applied at the training time.

**4.1.2. Settings.** About the structure of the network, which contains both the original fully connected dense DNN and the sparse WLT-net generated from it, ResNet-18 is chosen for the backbone. As the classical deep residual network's

TABLE 1: Properties of experimental parameters.

Dataset	CIFAR-10	SmallNORB
Class	10	5
Train	50,000	40,000
Test	10,000	10,000
Image size	$32 \times 32 \times 3$	$28 \times 28 \times 1$
Domain	$D_S$	$D_T$
Parameter settings	SGD [ $5e^{-3}$ , $1e^{-3}$ , $1e^{-4}$ ] with momentum 0.9, weight decay $1e^{-4}$ , and learning rate = 0.01	

18-layer version (with 17 convolution layers and 1 fully connected layer, 11.2 M parameters to train), it is also used in the original LTH, so the same configuration is set for the experiments. Settings of experimental parameters are shown in Table 1 too. The pruning rate in finding WLT-net is set to 20% with batch size 128, the maximum number of iterations 30,000, and at most 50 epochs in each iteration. 10 rounds of iterative pruning are performed to find the optimal WLT-net. During the operation, only the convolution layers will be pruned. When retraining the transferred sparse network in  $T_T$  for optimizing, by convention, weights of the convolution layers are frozen and only the fully connected layer is fine-tuned.

**4.1.3. Verify the Sparse Transfer Hypothesis for WLT-Nets.** Firstly, experiments are designed to validate the performance of WLT-net compared with the original dense network on  $D_S$ . After training in the source domain CIFAR-10, the original dense ResNet-18 achieved an average accuracy of 89.43% in the test dataset, with 11,173,962 parameters used.

The experimental results are shown in Figure 2, to display the relationship among the pruning level, number of remaining parameters, and average accuracy. When only 10.7% of the original parameters are retained after pruning (1,212,145 used), an average accuracy of 89.24% is still able to be achieved. As illustrated in the figure, it is proven that the pruning method in the proposed sparse transfer hypothesis for WLT-nets can guarantee the accuracy and reduce parameters while generating sparse subnetworks.

**4.1.4. Verify WLTs-SDTL.** Then, on the basis of the above analysis, to validate whether the sparse WLT-net can inherit knowledge from  $T_S$ , more experiments are designed, and the effect of initialization on the results is compared. As for SmallNORB on  $D_T$ , when training a dense ResNet-18 from scratch, the average accuracy achieves 89.9%. The optimal WLT-net generated in each round under different pruning levels is transferred to  $T_T$ , respectively, and the accuracy is compared. Since the influence of trend in weight change during training has been further considered in this paper, to better compare and prove the effectiveness of the proposed method, four kinds of initialization approaches are used for the sparse network separately: (a) original LTH (using  $\lambda_i$ ), (b) random initialization, (c) late reset method (using  $\lambda_\alpha$ ), and (d) proposed WLTs-SDTL. The performance of the orig-

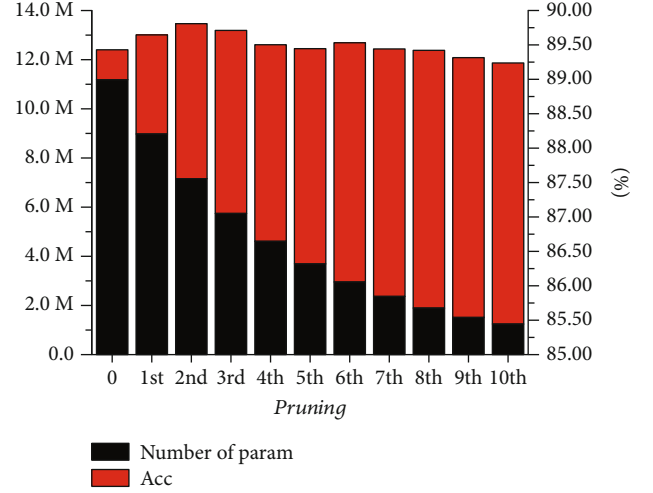


FIGURE 2: Experimental results: verify the sparse transfer hypothesis for WLT-nets.

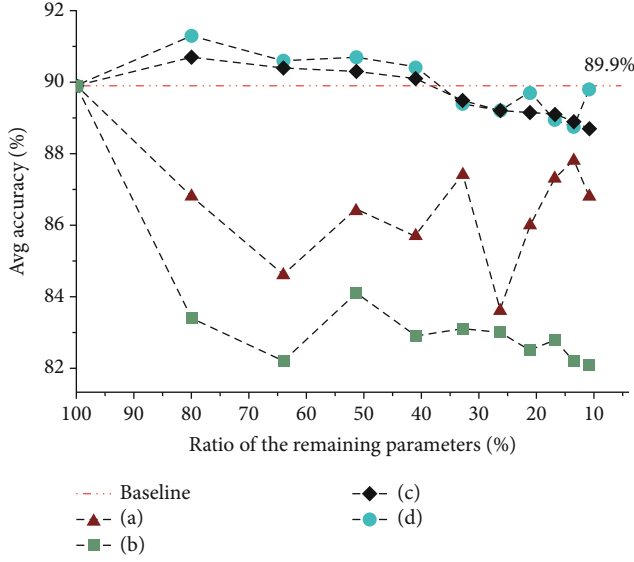
inal dense network before pruning is used as the baseline. The experimental results are shown in Figure 3.

As the experimental data shows, a better performance can be obtained than that of training dense DNN directly when proper pruning is carried out. It proves that the WLTs-SDTL can transfer the necessary ability from  $T_S$  in  $D_S$  to  $T_T$  in  $D_T$ . The sparse transferable structure is able to greatly save the cost of training parameters, while the overall accuracy can be kept, which realizes a sparse DTL. Compared with the initialization approaches (a), (b), and (c), the proposed WLTs-SDTL is more effective on the whole. At the same time, according to the results, when running a deep-level pruning on the original model, in which only 10% of the parameters remains (1,212,145 compared with the original 11,173,962), the precision loss is acceptable sometimes, which corresponds to the original LTH. Thus, when the target task  $T_T$  can accept a slight performance loss in exchange for generalization ability, it offers the possibility of using DNN-based deep computing methods on devices with low computational power such as mobile terminals, smartphones, or edge computing devices.

In summary, experiments on benchmark datasets have verified the feasibility of the proposed *sparse transfer hypothesis for WLT-nets* and *WLTs-SDTL*.

**4.2. Identification of Pests and Diseases Based on WLTs-SDTL.** In this section, the proposed WLTs-SDTL is used to train a sparse network from a dense detection model based on image recognition and applied in actual solutions of plant pest and disease identification. Specifically, the common diseases of tomato leaves are identified, inheriting the ability from ImageNet and using open-source lab datasets for  $T_T$ 's fine-tuning training.

**4.2.1. Datasets.** The ResNet-18 network pretrained on ImageNet is used as  $D_S$  to provide the necessary knowledge from weight of parameters. As for the  $T_T$  in identifying pests and diseases on tomato leaves, the PlantVillage dataset is chosen. The PlantVillage [36] is an open-source image dataset of



Param. remains with best Acc				
80%	64%	51.2%	40.96%	32.77%
91.3% (d)	90.6% (d)	90.7% (d)	90.4% (d)	89.5% (c)
26.21%	20.97%	16.77%	13.42%	10.74%
89.2% (c,d)	89.7% (d)	89.1% (c)	88.9% (c)	89.8% (d)

FIGURE 3: Experimental results: using WLTs-SDTL on benchmark datasets.

leaves from 14 crops, which contains 26 categories of plant pests and diseases and corresponding healthy leaves.

Since the samples of different categories in the original dataset are uneven, the crop tomato with sufficient samples is selected, in which categories with fewer samples and images with poor quality are eliminated. Then, data enhancement methods such as horizontal flip are used to adjust the sample size of each category to the same. Finally, a total of 8 categories of pests and diseases/1 healthy leaf are defined; meanwhile, the image size is adjusted to  $64 \times 64$  uniformly. The specific properties of the dataset are shown in Table 2.

**4.2.2. Settings.** Deeper-level pruning is carried out. 15 rounds of iterative pruning are performed to find the optimal WLT-net, with at least 3.6% of the parameters being retained (406,495 compared with 11,173,962 in the original dense net). Other experimental settings are the same as them in the previous section.

**4.2.3. Identify Pests and Diseases of Tomato Leaves.** The experimental results are shown in Figure 4. The identification model using dense ResNet-18 can achieve 96.44% accuracy after training in  $D_T$ , and a series of sparse networks with different volumes are obtained, respectively, using WLTs-SDTL.

The highest accuracy is up to 97.69% in pruning level 5, when 67% of the parameters are removed. By and large, WLTs-SDTL can guarantee the accuracy in plant identification while reducing parameter computation. When pruning properly, the accuracy can be higher than that of the original

TABLE 2: Properties of the PlantVillage dataset.

Category	Sample set	Training set
Tomato healthy	1,592	1,500
Tomato bacterial spot	2,127	1,500
Tomato early blight	1,000	1,500
Tomato late blight	1,910	1,500
Tomato Septoria leaf spot	1,771	1,500
Tomato spider mites	1,653	1,500
Tomato mosaic virus	373	Unused
Tomato leaf mold	952	1,500
Tomato target spot	1,404	1,500
Tomato TYLCV	5,357	1,500

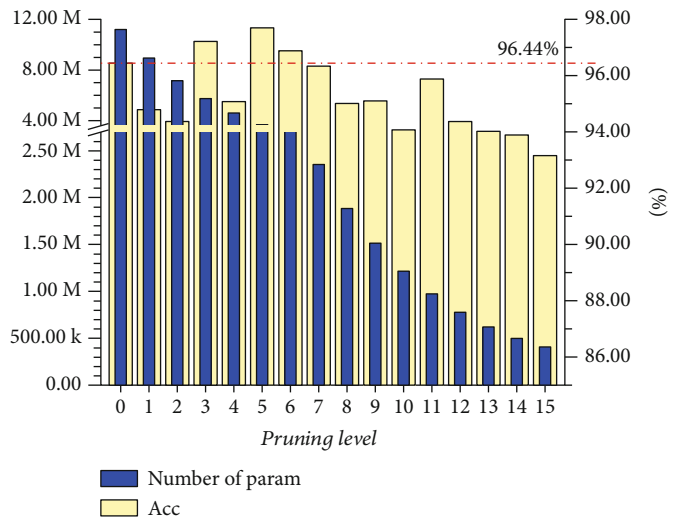


FIGURE 4: Experimental results of WLTs-SDTL on the PlantVillage dataset.

dense net. If the optimal performance is required, fine-grained pruning can be gradually carried out between levels near the best accuracy, which is between 30% and 50% in this set of experiments. For example, we can set the pruning rate to 10% or less in each round, to find a balance between performance and volume of the model.

When only 3.6% of the parameters are retained, the sparse network is still able to achieve an accuracy of 93.16%, with no other than 406,495 parameters to be trained. Considering that the accuracy is acceptable in daily identification tasks, it verifies that the proposed WLTs-SDTL can generate a small efficient network suitable for mobile terminals or edge computing devices with low computational power in the practical application of pest and disease identification.

**4.3. Using WLTs-SDTL for Real Collected Data Identification.** In this section, a small scale of datasets collected in Chongqing, China, is used in training the identification model to detect the citrus greening disease (Haunglongbing), combined with lab data. Citrus greening (Haunglongbing) is a

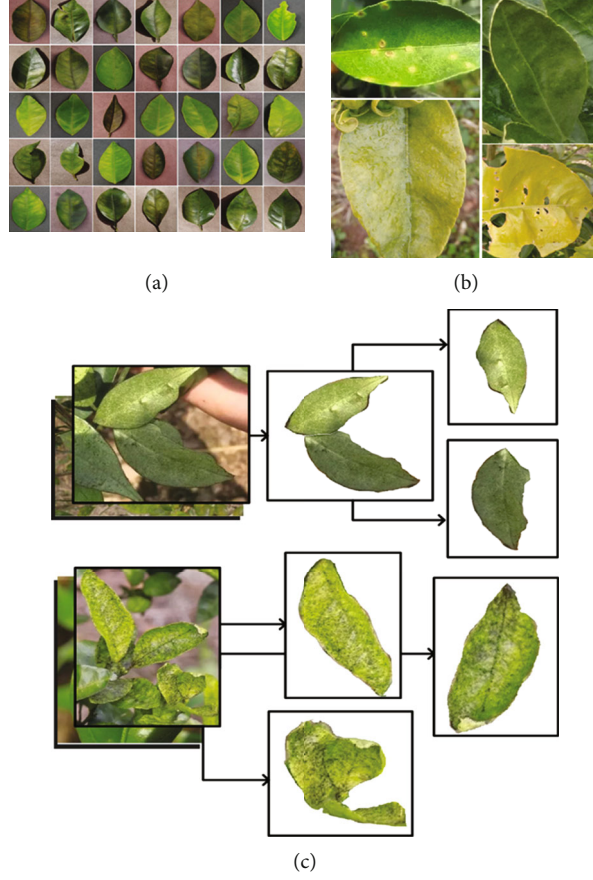


FIGURE 5: The composition of the dataset: (a) samples in PlantVillage; (b) collected samples; (c) process of clipping.

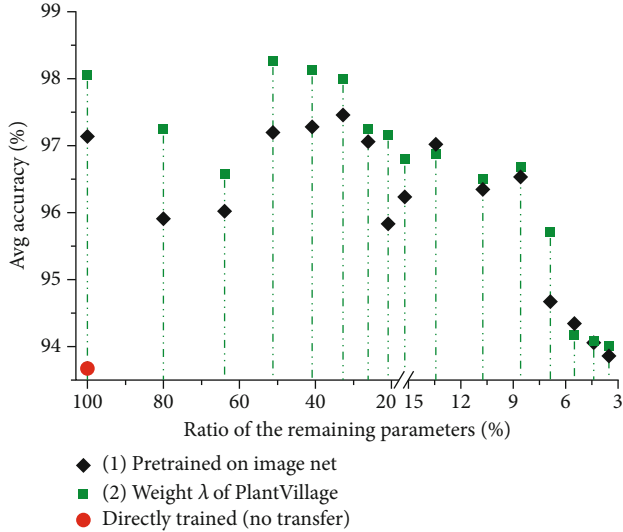


FIGURE 6: Experimental results of using WLTs-SDTL in identifying Haunglongbing.

devastating disease in the world citrus production, which seriously restricts the development of the citrus industry [37]. The etiolation of leaves can be used for its in-field identification.

**4.3.1. Datasets.** We have collected 1,266 images from Chongqing, China, as well as from Internet and monographs, in which 238 samples are Haunglongbing. In the process of photographing samples in the field, there may be more than one leaf in a photo, and more training samples can be obtained through clipping, as illustrated in Figure 5(c). Haunglongbing can also be identified by fruit; however, due to the lack of samples, images of diseased fruit are screened out. After the clipping, filtering, and image intensification operations, 1,500 samples are generated for training. Meanwhile, in the category orange haunglongbing of the PlantVillage dataset used in Section 4.2.1, which contains 5,507 images of leaves, 5,000 images are randomly selected (Figure 5(a)). Finally, the training dataset of 6,500 samples is created.

**4.3.2. Settings.** 15 rounds of deep-level iterative pruning are performed, while the other experimental settings are the same as them in the previous section. About the initial weight used for late reset, two kinds of options are chosen: (1) the weight of ResNet-18 was pretrained on ImageNet, which has been used as  $D_S$  in Section 4.2.1 and (2) the final weight  $\lambda_B$  of identification task  $T_T$  in Section 4.2.3 has been trained on PlantVillage datasets. The contrast experiment is designed to compare whether the proposed method can inherit better ability from a more *similar domain*.



**4.3.3. Identifying Haunglongbing.** After training on the original dense network, when performing a traditional intensive DTL (no parameters pruned), the initial weight (1) achieves 97.14% accuracy while the final weight (2) achieves 98.06%. Then, the proposed WLTs-SDTL is used, respectively; the relationship of accuracy and remaining parameters is shown in Figure 6. The highest accuracy is not even able to achieve 94% when training a dense ResNet-18 from scratch, and pruning the network will lead to the loss of overall accuracy; its performance on simplified WLT-net can be no longer discussed in this section (because the performance will not be better than that using DTL).

The experimental results show that, compared with training directly, DTL can achieve better initial performance in identifying citrus Haunglongbing disease under the help of collected data. And regarding the proposed sparse DTL, the following are obtained:

- (1) As the parameters decrease, the overall accuracy declines. However, it is still within an acceptable range and higher than no-transfer dense network. When the weight from a similar identification task is used in (2) shown in Section 4.3.2, a higher initial accuracy can be obtained, and the trend of accuracy is more stable in subsequent sparsification than (1) shown in Section 4.3.2. Thus, a similar task, which has been well trained for identifying other pests and diseases, might be a better initial choice in practical application
- (2) Although fewer parameters are used, the model can achieve higher accuracy in some pruning levels: in (1) shown in Section 4.3.2, among the 3rd to the 5th levels, the average accuracy, respectively, achieves 97.20%, 97.28%, and 97.46% when 51.2%, 40.96%, and 32.77% of the parameters remain higher than the original 97.14%. And in (2) shown in Section 4.3.2, in the 3rd and 4th pruning level, it achieves 98.26% and 98.13% compared with the original 98.06%, while it is 97.99% in the 5th level. The reason is that sparsification reduces the redundancy of parameters, and the negative feedback of low-contribution parameters is inhibited
- (3) Fine-grained pruning in the optimal range can be proceeded for the best performance. And in the experimental results of this paper, note that the range is always between 50% and 30%. Therefore, we speculate that in the proposed WLTs-SDTL, the priority could be given to these pruning levels when models' performance is preferred. And when the volume of the model needs to be compressed as much for widely deploying, it shows that the sparse model can use about 10% of the original parameters (accuracy 96.67% when using 8.59%) to maintain an acceptable performance close to that of the original. The limiting small model with only 3.6% of the original parameters is also taken into account, whose accuracy is able to achieve 94.01%, still higher than that of the dense

net without transfer, and is more likely to be used for low-computing-power devices or edge computing devices

To sum up, when the proposed WLTs-SDTL is used in an actual solution of identifying diseases of plants, the sparsification of the network can be realized through pruning to save the computational overhead of parameters while maintaining or even improving performance. Thus, the balance between performance and model size can be dynamically adjusted, and the deployment possibility of low-computational-power equipment is provided.

## 5. Conclusion

In this paper, a sparse deep transfer learning model is proposed. The method is aimed at modeling the identification of plant pest and disease with limited collected data in the field, and a sparse DTL strategy is designed to transfer only the most important architecture and optimize models' size.

Specifically, (1) the sparse transfer hypothesis is proven, which succeeds in modifying LTH to reduce the parameter computation in DTL by generating sparse transferable WLT-nets. (2) The sparse transfer method named WLTs-SDTL is formally proposed, in which the compressing strategy is designed to construct a deep sparse network, distills useful information from the auxiliary domains, and improves the transfer efficiency. (3) The proposed method is applied to detect pests and diseases with few data samples in training deep identification models. The hypothesis is verified by the benchmark dataset; meanwhile, the proposed method is evaluated on the representative datasets.

Experimental results show that when the proposed method is used in actual solutions, the sparsification of the network can save the cost of computing parameters while maintaining or sometimes improving the performance, thereby dynamically adjusting the balance between the model's accuracy and size, providing the deployment possibility in low-computational-power devices.

Moreover, the sparse strategy can be promoted in identifying new pests and diseases of the plant with few data and even widely used in other tasks based on image recognition and lack of data. On that occasion, depending on specific tasks, it is supposed to wisely choose the suitable network architecture and balance the accuracy and volume of models.

In the future, the proposed method will be studied on more other domains to overcome the scarcity of data and the redundancy of model parameters, improving the effectiveness of sparse deep transfer learning.

## Data Availability

The open-source datasets used in this paper, such as IMGnet, are freely available in various deep learning frameworks such as PyTorch and TensorFlow.

## Conflicts of Interest

The authors declare that they have no conflict of interest.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 61672123 and 62076047) and the Fundamental Research Funds for the Central Universities (Nos. DUT20LAB136 and DUT20TD107).

## References

- [1] N. Shazeer, A. Mirhoseini, K. Maziarz et al., "Outrageously large neural networks: the sparsely-gated mixture-of-experts layer," 2017, <https://arxiv.org/abs/1701.06538/>.
- [2] J. Zhang, Y. Huang, R. Pu et al., "Monitoring plant diseases and pests through remote sensing technology: a review," *Computers and Electronics in Agriculture*, vol. 165, article 104943, 2019.
- [3] A. S. Gurle, S. N. Barathe, R. S. Gangule, S. D. Jagtap, and T. Patankar, "Survey paper on tomato crop disease detection and pest management," *International Journal of Applied Evolutionary Computation*, vol. 10, no. 3, pp. 10–18, 2019.
- [4] Y. Li, J. Zhou, X. Zheng, J. Tian, and Y. Y. Tang, "Robust sub space clustering with independent and piecewise identically distributed noise modeling," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 8720–8729, Long Beach, CA, USA, 2019.
- [5] D. Gao, Q. Sun, B. Hu, and S. Zhang, "A framework for agricultural pest and disease monitoring based on Internet-of-things and unmanned aerial vehicles," *Sensors*, vol. 20, no. 5, article 1487, 2020.
- [6] E. Mwembaze and G. Owomugisha, "Machine learning for plant disease incidence and severity measurements from leaf images," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 158–163, Anaheim, CA, USA, 2016.
- [7] 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, article 2022, 2017.
- [8] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299–315, 2017.
- [9] A. A. Alfarisy, Q. Chen, and M. Guo, "Deep learning based classification for paddy pests & diseases recognition," in *Proceedings of 2018 International Conference on Mathematics and Artificial Intelligence*, pp. 21–25, Chengdu, China, 2018.
- [10] J. Frankle and M. Carbin, "The lottery ticket hypothesis: finding sparse, trainable neural networks," 2019, <https://arxiv.org/abs/1803.03635/>.
- [11] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
- [12] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks," *Neurocomputing*, vol. 267, pp. 378–384, 2017.
- [13] D. Li, R. Wang, C. Xie et al., "A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network," *Sensors*, vol. 20, no. 3, p. 578, 2020.
- [14] P. Bhatt, S. Sarangi, A. Shivhare, D. Singh, and S. Pappula, "Identification of diseases in corn leaves using convolutional neural networks and boosting," *ICPRAM*, 2019, pp. 894–899, SciTePress, 2019.
- [15] X. Sun, Y. X. ShaominMu, Z. Cao, and T. Su, "Image recognition of tea leaf diseases based on convolutional neural network," in *2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pp. 304–309, Jinan, China, 2018, 10.1109/SPAC46244.2018.8965555.
- [16] G. Hu, X. Yang, Y. Zhang, and M. Wan, "Identification of tea leaf diseases by using an improved deep convolutional neural network," *Sustainable Computing: Informatics and Systems*, vol. 24, article 100353, 2019.
- [17] J. Chen, Q. Liu, and L. Gao, "Visual tea leaf disease recognition using a convolutional neural network model," *Symmetry*, vol. 11, no. 3, p. 343, 2019.
- [18] K. P. Ferentinos, M. Barda, and D. Damer, "An image-based deep learning model for cannabis diseases, nutrient deficiencies and pests identification," in *Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence, EPIA 2019*, Vila Real, Portugal, 2019.
- [19] K. Zhang, Y. Guo, X. Wang, J. Yuan, and Q. Ding, "Multiple feature reweight DenseNet for image classification," *IEEE Access*, vol. 7, pp. 9872–9880, 2019.
- [20] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," 2016, <https://arxiv.org/pdf/1604.03169/>.
- [21] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Frontiers in Plant Science*, vol. 8, p. 1852, 2017.
- [22] Z. Libo, H. Tian, G. Chunyun, and M. Elhoseny, "Real-time detection of cole diseases and insect pests in wireless sensor networks," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 3, pp. 3513–3524, 2019.
- [23] K. Thenmozhi and U. S. Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Computers and Electronics in Agriculture*, vol. 164, article 104906, 2019.
- [24] Y. Li, J. Zhou, J. Tian, X. Zheng, and Y. Y. Tang, "Weighted error entropy based information theoretic learning for robust subspace representation," *IEEE Transactions on Neural Networks and Learning Systems (T-NNLS)*, pp. 1–15, 2021.
- [25] Y. Li, R. Liang, W. Wei, W. Wang, J. Zhou, and X. Li, "Temporal pyramid network with spatial-temporal attention for pedestrian trajectory prediction," *IEEE Transactions on Network Science and Engineering (T-NSE)*, 2021.
- [26] C. R. Rahman, "Identification and recognition of rice diseases and pests using deep convolutional neural networks," 2018, <https://arxiv.org/abs/1812.01043>.
- [27] S. Xing, M. Lee, and K. K. Lee, "Citrus pests and diseases recognition model using weakly dense connected convolution network," *Sensors*, vol. 19, no. 14, article 3195, 2019.
- [28] G. Andrew, "MobileNets: efficient convolutional neural networks for mobile vision applications," 2017, <https://arxiv.org/abs/1704.04861/>.
- [29] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1mb model size," 2016, <https://arxiv.org/abs/1602.07360/>.
- [30] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient transfer learning," 2016, <https://arxiv.org/abs/1611.06440/>.

- [31] J. Liu, Y. Wang, and Q. Yu, "Sparse deep transfer learning for convolutional neural network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31no. 1, pp. 2245–2251, San Francisco, California, USA, 2017.
- [32] C. Reinhold and M. Roisenberg, "Filter pruning for efficient transfer learning in deep convolutional neural networks," in *18th International Conference on Artificial Intelligence and Soft Computing*, pp. 191–202, Zakopane, Poland, 2019.
- [33] J. Frankle, G. K. Dziugaite, D. M. Roy, and M. Carbin, "The lottery ticket hypothesis at scale," 2019, <https://arxiv.org/abs/1903.01611v1/>.
- [34] R. C. Çalik and M. F. Demirci, "Cifar-10 image classification with convolutional neural networks for embedded systems," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1-2, Aqaba, Jordan, 2018.
- [35] Y. LeCun, F. J. Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, vol. 2, pp. II–104, Washington, DC, USA, 2004.
- [36] D. P. Hughes and M. Salath'e, "An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing," 2015, <https://arxiv.org/abs/1511.08060/>.
- [37] Y. Lan, Z. Huang, X. Deng et al., "Comparison of machine learning methods for citrus greening detection on UAV multi-spectral images," *Computers and Electronics in Agriculture*, vol. 171, article 105234, 2020.