



## OPEN ACCESS

## EDITED BY

Lun Hu,  
Chinese Academy of Sciences (CAS), China

## REVIEWED BY

Wei Peng,  
Kunming University of Science and  
Technology, China  
Xingyi Li,  
Northwestern Polytechnical University, China

## \*CORRESPONDENCE

Jian Su,  
✉ gzkjgq@126.com

RECEIVED 05 December 2025

REVISED 03 January 2026

ACCEPTED 12 January 2026

PUBLISHED 29 January 2026

## CITATION

Huang J, Su J and Cheng T (2026) IRMKD: an application of instance relation matrix in plant disease recognition.  
*Front. Bioinform.* 6:1761574.  
doi: 10.3389/fbinf.2026.1761574

## COPYRIGHT

© 2026 Huang, Su and Cheng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# IRMKD: an application of instance relation matrix in plant disease recognition

Jinqing Huang<sup>1</sup>, Jian Su<sup>2\*</sup> and Tengfei Cheng<sup>1</sup>

<sup>1</sup>School of Computer and Electronic Information, Guangxi University, Nanning, China, <sup>2</sup>Guangxi Science and Technology Evaluation Center, Nanning, China

**Background:** The recognition and prevention of plant diseases is very important to the growth process. At present, neural networks have achieved good results in plant disease identification, but the development of convolutional neural networks has brought a large number of network parameters and long recognition time, which greatly limits its application on devices that lack computing resources.

**Methods:** To solve this problem, We introduce a novel approach, dubbed instance-relation-matrix based knowledge distillation (IRMKD), that transfers mutual relations of data examples. For concrete realizations of IRMKD, we combine the correlation of the samples with the relationship between the characteristics of the instances and introducing multiple loss functions.

**Results:** Experimental results show that the proposed method improves educated student models with a significant margin. In particular, for traditional neural networks, our method significantly reduces memory usage and recognition time by an average of 92% and at the same time ensure that the recognition accuracy rate is above 93%, provides a new plant disease recognition method for devices with limited memory and computing resources.

**Conclusion:** IRMKD can significantly reduce the volume of the model and improve the recognition speed of the model on the premise of slightly reducing the accuracy of the verification set.

## KEYWORDS

convolutional neural network, deep learning, disease identification, Knowledge distillation, model compression

## Introduction

Crop diseases have seriously affected the world's agricultural economy and will cause severe damage to crop yields. Disease identification is the key to predicting agricultural yields, which is of great importance for economic stability and food security in the agricultural sector (Kuzuhara et al., 2020). With the development of deep learning technology, numerous structures or patterns of complex networks are being used to identify diseases. But the enormous computing complexity of these architectures has restricted their use in many downstream applications. In response to this situation, some researchers had proposed different methods of compression of models in recent years. Real et al. (2019) developed an image classifier that exceeds manual design, which makes the neural network model more compact. Deng et al. (2019) propose a compression technique which is not explored in the area of architecture taking into account the decomposition of the tensor. Zhang et al. (2018) jointly train a quantified DNN compatible with bits

and its related quantizer to obtain the effect of the compression model. There are also a number of methods to compress network models, including prunin and knowledge distillation (Deng et al., 2020).

As a typical type of model compression and acceleration, knowledge distillation can effectively train small student models from large teacher models (Gou et al., 2021). Knowledge distillation can be divided into the following categories: response-based knowledge distillation, feature-based knowledge distillation, and relation-based knowledge distillation.

**Response-based knowledge distillation:** Response-based knowledge distillation usually means that the student network responds to the neurons in the last output layer of the teacher model. Its main idea is to directly simulate the final prediction of the teacher model. In recent years, some scholars have further explored response-based knowledge to solve the problem of insufficient information when the ground truth tag is the conditional target (Meng et al., 2019).

**Feature-based knowledge distillation:** Deep neural network is good at learning multi-level feature representation. Specifically, feature-based knowledge from the middle layer is a good extension of response-based knowledge, especially for the training of thinner and deeper networks. Zhang et al. (2020) propose a new task-based feature distillation (TOFD) method, which is a convolution layer trained by task loss in a data-driven way. Chen et al. (2018) proposed a feature mapping-based knowledge extraction method called knowledge extract with feature maps (KDFM), which improves the efficiency of knowledge extraction by learning feature maps from the teacher network.

**Relationship-based knowledge distillation:** Both response-based and feature-based knowledge use the output of a specific layer in the teacher model, while relationship-based knowledge distillation further discusses the relationship between different layers or data samples on the basis of the above two methods. Lee and Song. (2019). propose a knowledge distillation method based on multi-head graphs. They explore the data relationship between any two feature graphs in a multi-attention network through graph knowledge. In order to explore the paired clues in the student network and the teacher network, Passalis use the student model to simulate the mutual information flow of the paired clues in the teacher model (Passalis et al., 2020).

At present, the compression method of the above model still has the problem of low compression rate or loss of model accuracy after compression. The common point of these methods is that in the process of knowledge distillation, they only pay attention to the consistency of the instance while ignoring the correlation between the samples (Peng et al., 2019). In fact, the correlation between samples is also very important for classification, because it directly reflects how teachers model the structure of different samples embedded in the feature space. Therefore, we propose a knowledge extraction method based on the relationship between examples. In addition to the widely used instance feature maps, our method also defines three new knowledge types: sample correlation, instance correlation and feature space transformations, and proposes an instance relation matrix (IRM) to model all types of knowledge.

In this paper, combining the plant village disease data set (Hughes and Salathé, 2015) and the complex background data set provided by the Guangxi Academy of Agricultural Sciences,

a lightweight convolutional neural network compression method based on knowledge distillation (Hinton et al., 2015) is proposed. The test results in the real environment show that our method can significantly reduce the memory usage of the model while maintaining or slightly reducing the accuracy of the model. In addition, the method we propose is versatile. Whether deploying the model on a cloud server or a local device, this method can improve the recognition speed of the model while reducing memory usage and training overhead. Our main contributions can be summarized in the following three areas:

1. For the first time, we combine the four kinds of knowledge of sample correlation and instance feature, instance relationship and cross-layer feature space transformation to carry out knowledge distillation.
2. For the first time, the concept of instance relation matrix (IRM) is proposed, and the instance relation matrix and its transformation were used to model all types of knowledge. The instance relationship matrix can be represented by the data structure of the three-dimensional array  $IRM[i][j][k]$ , where  $i$  and  $j$  represent the Euclidean distance between the  $i$ th feature map and the  $j$ th feature map, and  $k$  represents the same the  $k$ th sample in the batch.
3. Introducing multiple loss functions to supervise the training of the student network is used to help students learn different kinds of knowledge stored in IRMs, and then obtain the final loss function  $L_{MTK}$  by weighting, and then prove the superiority of the method through experimental results.

## Materials and methods

### Data preprocessing

The train set and validation set used in our experiments are based on the Plant Village dataset (Hughes and Salathé, 2015). It contains 82,161 pictures of plant leaves of varying sizes from 24 plants in 55 classes. The data set contains images with clean background and congested background, as shown in Figure 1. Clean background images consist of isolated leaves with uniform backgrounds, while cluttered background images comprise partial or full images of plants taken in a natural background. The number of images in each class ranges from 43 to 6,359. This data set is divided into three different sets. PlantLeaf1 contains 18 classes which contain pictures with a cluttered background. None of the images in this dataset contains laboratory-conditioned images. PlantLeaf2 contains 11 classes, which constitute both clean and cluttered images. Clean background images were used in the training, while cluttered background images were used in the testing of this dataset. PlantLeaf3 consists of 16 classes of 11 plants. These classes contain both clean and cluttered images, whereas the number of images per class varied from 892 to 5,507. This dataset consists of 10 classes of 10 different crop species and 6 classes of tomato plants infected by different diseases. The number of classes and frames for each PlantLeaf data set is detailed in Table 1.

To generalize the model and ensure a robust model, these image datasets were augmented using different data augmentation processes, such as flipping, random crops, rotations, shifts, and



**FIGURE 1**  
Partial presentation of plant village dataset.

a combination of these techniques. Data augmentation aims to prevent overfitting by training the model to large data created artificially model.

## Overview of four knowledge distillation method

In this section, a structured disease identification method and a lightweight neural network reduction method are proposed. The overall design of this study is shown in Figure 2.

Knowledge distillation is first proposed in (Weng and Preneel, 2011) for model compression. The key idea of knowledge distillation is that the soft probability of trained teachers' network output contains not only class labels, but also more information about data points. For example, if multiple categories of high probability areas signed to an image, it may mean that the image must be located near the decision boundary between these categories. Therefore, forcing students to imitate these probabilities should enable students' network to absorb some knowledge that teachers have found in the information outside the training label itself.

In the learning process of knowledge distillation (Shorten and Khoshgoftaar), the student model is trained by imitating the output of the teacher model in the same sample. In the traditional **Softmax** classifier, given any input image, the model generates a vector  $S^t(x) = [S_1^t(x), S_2^t(x), \dots, S_K^t(x)]$  where  $S^t(x)$  represents the score corresponding to the  $k$ th disease. We use **Softmax** as the classifier at the end of the neural network to convert the output  $S^t(x)$  of the neural network into probability Distribution  $p^t(x)$ , as shown in Formula 1.

$$p_k^t(x) = \frac{e^{s_k^t(x)}}{\sum_j e^{s_j^t(x)}} \quad (1)$$

Hinton et al. (2015) proposed that the output of a well-trained teacher model would be infinitely close to the real output of One-Hot coding, which causes useful inter class information to be ignored in the training process, and directly lead to the unsatisfactory training effect of the student model. Therefore, it is necessary to use the temperature scale to "soften" these probabilities, as shown in Formula 2.

$$\tilde{p}_k^t(x) = \frac{e^{s_k^t(x)/T}}{\sum_j e^{s_j^t(x)/T}} \quad (2)$$

where  $T > 1$  is an adjustable super parameter. By adding the parameter  $T$  to classifier, students would similarly produce a softer classification probability distribution  $\tilde{p}^s(x)$ , thus preserving the probability relationship between different categories of samples. Compared with the traditional One-Hot coding hard tag as the training target, because the soft target output by the teacher model after the Soft-Softmax classifier well retains the probability relationship between different categories of samples, it usually brings better performance. The loss function of students is a linear combination of the typical cross entropy loss function  $L_{cls}$  and the loss function  $L_{KD}$  in the process of knowledge distillation (Kim and Rush, 2016), as shown in Formula 3–5.

$$L = \alpha L_{cls} + (1 - \alpha) L_{KD} \quad (3)$$

$$L_{KD} = -T^2 \sum_k \tilde{p}_k^t(x) \log \tilde{p}_k^t(x) \quad (4)$$

$$L_{cls} = - \sum_k q_k \log \tilde{p}_k^t(x) \quad (5)$$

where  $T$  and  $\alpha$  are adjustable super parameters. The common choices are  $T \in 3, 4, 5$  and  $\alpha \in [0.5, 0.9]$ ,  $q_k(x)$  is the real label of the sample.

TABLE 1 Generated database for training and validation.

Class name	Plant common name	Disease common name	Disease scientific name	Images (number)
C01	Apple	Apple scab	<i>Venturia inaequalis</i>	630
C02	Apple	Apple black rot	<i>Botryosphaeria obtusa</i>	621
C03	Apple	Cedar apple rust	<i>Gymnosporangium juniperi-virginianae</i>	275
C04	Apple	Healthy	—	1,645
C05	Blueberry	Healthy	—	1,502
C06	Cherry	Healthy	—	854
C07	Cherry	Powdery mildew	<i>Podosphaera spp</i>	1,052
C08	Corn	Cercospora leaf spot	<i>Cercospora zeae-maydis</i>	513
C09	Corn	Common rust	<i>Puccinia sorghi</i>	1,192
C10	Corn	Northern leaf blight	<i>Exserohilum turcicum</i>	985
C11	Corn	Healthy	—	1,162
C12	Grape	Black rot	<i>Guignardia bidwellii</i>	1,180
C13	Grape	Black measles	Pc	1,383
C14	Grape	Isariopsis leaf spot	<i>Pseudocercospora vitis</i>	1,076
C15	Grape	Healthy	—	423
C16	Orange	Huanglongbing	<i>Candidatus liberibacter</i>	5,507
C17	Peach	Bacterial spot	<i>Xanthomonas campestris</i>	2,297
C18	Peach	Healthy		360
C19	Pepper	Bacterial spot	<i>Xanthomonas campestris</i>	997
C20	Pepper	Healthy	—	1,478
C21	Potato	Early blight	<i>Alternaria solani</i>	1,000
C22	Potato	Late blight	<i>Phytophthora infestans</i>	1,000
C23	Potato	Healthy	—	152
C24	Raspberry	Healthy	—	371
C25	Soybean	Healthy	—	5,090
C26	Squash	Powdery mildew	<i>Erysiphe cichoracearum</i>	1835
C27	Strawberry	Leaf scorch	<i>Diplocarpon earlianum</i>	1,109
C28	Strawberry	Healthy	—	456
C29	Tomato	Bacterial spot	<i>Xanthomonas campestris</i>	2,127
C30	Tomato	Early blight	<i>Alternaria solani</i>	1,000
C31	Tomato	Late blight	<i>Phytophthora infestans</i>	1909
C32	Tomato	Leaf mold	<i>Fulvia fulva</i>	952
C33	Tomato	Septoria leaf spot	<i>Septoria lycopersici</i>	1771

(Continued on the following page)

TABLE 1 (Continued) Generated database for training and validation.

Class name	Plant common name	Disease common name	Disease scientific name	Images (number)
C34	Tomato	Spider mites	Tetranychus urticae	1,676
C35	Tomato	Target spot	Corynespora cassiicola	1,404
C36	Tomato	Tomato mosaic virus	Tomato mosaic virus	373
C37	Tomato	Yellow leaf curl virus	Begomovirus	5,357
C38	Tomato	Healthy	—	1,591
Total				54,305

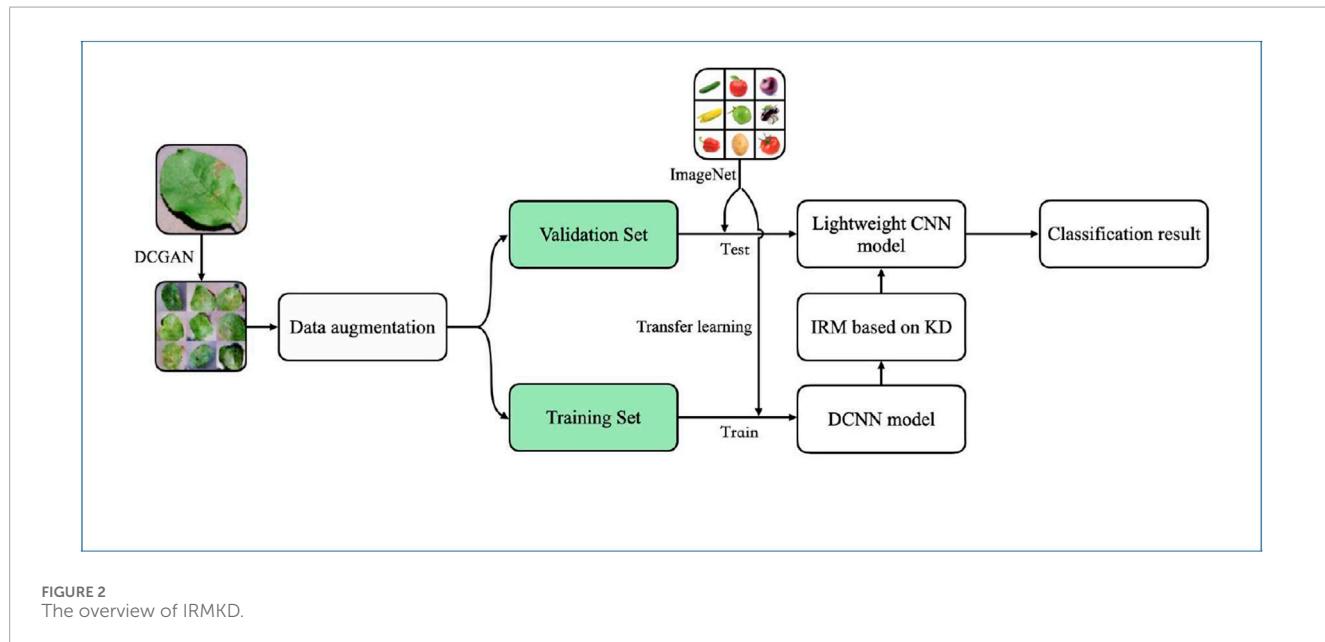


TABLE 2 Accuracy of different knowledge distillation methods under different network structures.

Teacher net.	Student net.	Baseline	KD	AT	SP	LIRM	$L_{MTK}$	TNA
VGG16 (304.0 MB)	MobileNet (20.8 MB)	91.57%	91.20%	92.27%	93.86%	92.58%	<b>93.96%</b>	95.85%
AlexNet (356.6 MB)	MobileNet (20.8 MB)	91.45%	90.89%	93.37%	93.64%	92.46%	<b>93.87%</b>	94.46%
GoogleNet (36.6 MB)	MobileNet (20.8 MB)	91.48%	91.14%	91.98%	92.45%	92.67%	<b>93.32%</b>	93.28%
ResNet (179.8 MB)	MobileNet (20.8 MB)	91.53%	90.17%	91.67%	91.23%	92.08%	<b>93.35%</b>	93.40%

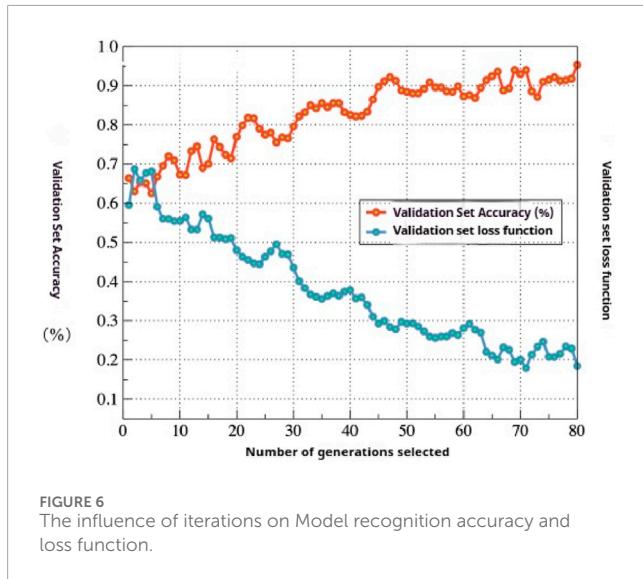
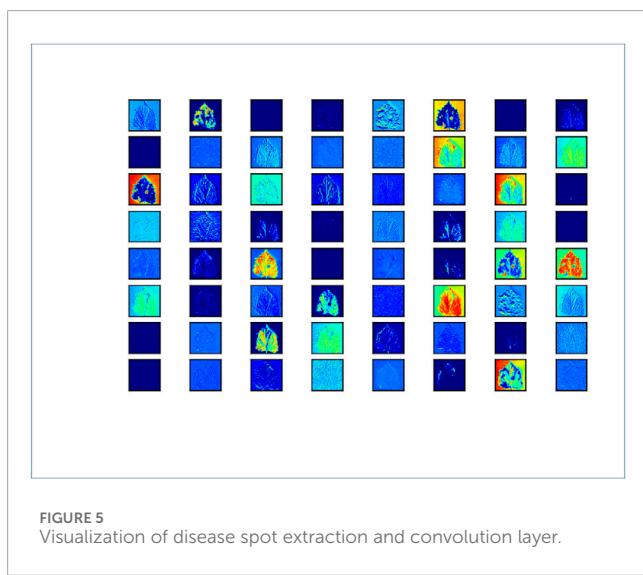
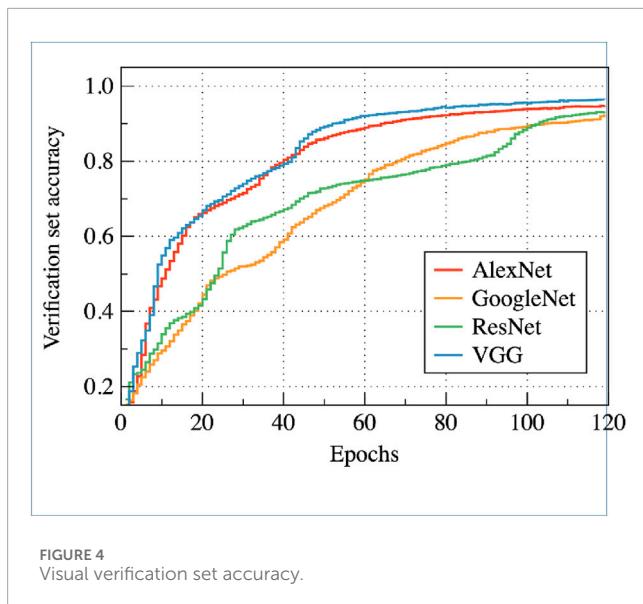
Bold values indicate the highest accuracy.

## Relational knowledge distillation based on IRM

As shown in Figure 3, for multiple DNN layers of the teacher model, an matrix is constructed, where is the number of DNN layers selected, and each element in the matrix represents the Euclidean distance between two characteristic graphs with corresponding subscripts. The matrix provides sufficient and general information about the characteristic distribution, so that the

extracted knowledge can guide student networks with different structures. At present, most teacher-student frameworks based on knowledge distillation rely on strong constraints at the instance level.

At the same time, the correlation among multiple samples is also valuable for knowledge distillation. Using the sample correlation, the student model can better learn the relationship between different samples. Therefore, in the process of knowledge distillation, this paper takes  $n$  samples as the input of neural network at the same time.



Because each sample will generate a  $m \times m$  matrix, we will finally obtain a  $n \times m \times m$  three-dimensional matrix, which is  $IRM$ .

Let the input data set of the network be  $X = \{x_1, x_2, \dots, x_n\}$ ,  $n$  is the sample number of Mini-batch,  $C_i^T(x_i)$  and  $C_i^S(x_i)$  are the output characteristic graphs of the  $i$ th sample in the  $j$  layer of teacher and student models respectively,  $f_{a,h}^T(x_i)$  an mapping  $f_{a,h}^S(x_i)$  represents the Euclidean distances of the output characteristic graphs of the  $g$ th and  $h$ th layers in the teacher student model. Let  $F_i^T$  and  $F_i^S$  represents the set of distances between the  $j$ th layer and each layer in the  $m$ th layer in the teacher and student models, respectively via Formulas 6.1, 6.2:

$$F_j^T(x_i) = \text{matrix} (f_{j,1}^T(x_i), f_{j,2}^T(x_i), \dots, f_{j,m}^T(x_i)) \quad (6.1)$$

$$F_j^S(x_i) = \text{matrix} (f_{j,1}^S(x_i), f_{j,2}^S(x_i), \dots, f_{j,m}^S(x_i)) \quad (6.2)$$

At the same time, the mapping functions are introduced as follows (Formulas 7.1, 7.2):

$$\theta: F \rightarrow G \in W^{mxm} \quad (7.1)$$

$$\varphi: G \rightarrow H \in E^{nxmxm} \quad (7.2)$$

where  $G$  is the distance matrix of  $m \times m$  and  $H$  is the three-dimensional matrix of  $n \times m \times m$ , each element in  $H$  represents the distance between the characteristic graph of layer  $C_g$  and layer  $C_h$  in the input  $i$ th sample.

The  $IRM$  formula can be expressed as follows (Formula 8):

$$L_{IRM} = \Psi(\varphi^T(\chi), \varphi^S(\chi)) \quad (8)$$

where  $\Psi$  is the loss function as Formula 9:

$$\Psi(H^T, H^S) = \sum_{\substack{i \in [1, n] \\ j, k \in [1, m]}} \|h_{i,j,k}^T - h_{i,j,k}^S\|_2^2 \quad (9)$$

In order to avoid too strict constraints, cross layer feature space transformation is introduced as the third type of knowledge, and an  $IRM$  transformation is proposed to model the knowledge. The feature space transformation is a more relaxed description than the dense fitting of teacher's case features in the middle layer. By combining  $IRM$  with  $IRM$  transformation, this method has more general, moderate and sufficient knowledge than the existing methods. Finally, two loss functions for  $IRM$  and  $IRM$  transformation are designed and optimized to improve the performance of the student model. Firstly, the mapping function is defined as follows (Formula 10):

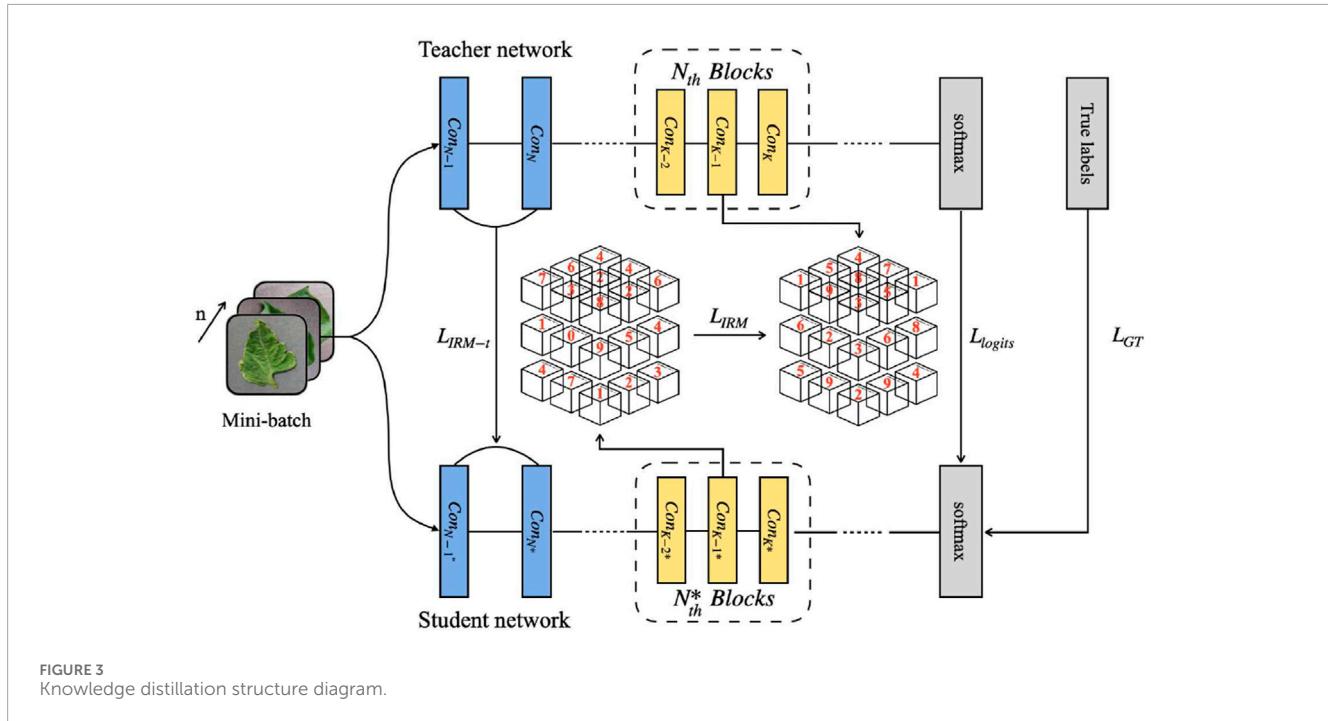
$$\Phi: \chi \rightarrow D \in R^{n \times n} \quad (10)$$

where  $D$  is a two-dimensional matrix of  $n \times n$ , let each element of  $D_i^T$  and  $D_i^S$  represents the distance of the output characteristic graph of the  $x_i$  sample and the  $x_{i'}$  sample input in the teacher and student models at the  $j$ th layer respectively.  $L_{IRM-t}$  extracted the transformation knowledge of feature space by calculating the difference change of the feature graph between two layers in the network model,  $D_g^T(\chi) - D_h^T(\chi)$  is the amount of knowledge flow information from layer  $g$  to layer  $h$  in the teacher network, then  $IRM-t$  formula can be defined as follows (Formula 11):

$$L_{IRM-t} = \| (D_g^T(\chi) - D_h^T(\chi)) - (D_g^S(\chi) - D_h^S(\chi)) \|_2^2 \quad (11)$$

TABLE 3 Comparison with other research methods.

Model	Original images			Pre-processed images		
	Accuracy (%)	Parameters (M)	Time (S/Epoch)	Accuracy (%)	Parameters (M)	Time (S/Epoch)
GRNNs	97.27	165.35	6,714	96.54	165.24	6,102
PNNs	98.06	247.52	7,039	96.17	248.11	6,421
RBF	96.06	214.89	6,273	97.36	215.36	5,926
PCA and BP	95.44	156.16	5,452	98.26	153.23	4,857
IRMKD	93.45	19.8	1,206	94.67	20.2	1,087

FIGURE 3  
Knowledge distillation structure diagram.

$L_{\text{logits}}$  formula represents the *Softmax* loss of teachers' network output and students' network output as [Formula 12](#):

$$L_{\text{logits}} = \|Y^T - Y^S\|_2^2 \quad (12)$$

Finally, we define an  $L_{\text{MTK}}$  loss function is used to train student network, which is based on  $\text{IRM-}t$  transform loss ( $L_{\text{IRM-}t}$ ),  $\text{IRM}$  loss ( $L_{\text{IRM}}$ ) and *Softmax* loss ( $L_{\text{logits}}, L_{\text{GT}}$ ) as follows [Formula 13](#).

$$L_{\text{MTK}} = \alpha L_{\text{IRM-}t} + (1 - \alpha) L_{\text{IRM}} + \beta L_{\text{logits}} + \gamma L_{\text{GT}} \quad (13)$$

where  $L_{\text{GT}}$  is the loss function between the real tag and the student network output, and  $\alpha, \beta, \gamma$  are the super parameters. Using MTK loss can optimize the student network and obtains three types of knowledge from the teacher network.

## Experiment results and discussion

The hardware environment of this experiment includes Intel i9-10900x (3.20ghz) 10 core 20 thread CPU, NVIDIA geforce RTX2080ti 11 GB\*2 server. The software environment is Windows10 64 bit system, CUDA 9.0, cudnn 7.0, PyCharm 2018.2. The front end and back end of the experimental framework for training model are keras and tensorflow, respectively.

In this paper,  $128 \times 128$  three channel RGB images are used to train 200 epochs. The size of batchis 512, the initial learning rate is 0.1, the momentum is 0.9, the weight depth is  $10^{-4}$ , and the number of classes is, the random seed is 2. The experimental results of different knowledge distillation methods under different network structures are shown in [Table 2](#).

In our experiment, VGG16, AlexNet, GoogleNet and ResNetis used as the teacher's network structure, and MobileNet is used as the student's network structure. The teacher column represents

the accuracy of the teacher model. First, a teacher model is trained on plant village with 4 different neural network. After 120 iterations, the highest accuracy of 4 teacher model reaches 95.85%. The column “baseline” indicates the accuracy of the basic student model. The same sample is used to train the student model on MobileNet and the model parameter is only 28.0 MB. Under the same conditions, the accuracy of the model is 91.57%. Figure 4 shows the accuracy change curve of the verification set during the training of four different teacher models.

The KD column represents the accuracy of the training results after the distillation of basic knowledge. From the table, we can see that the accuracy of the model after the distillation of knowledge has improved compared with the baseline column. AT and SP are the model accuracy of other knowledge distillation methods,  $L_{IRM}$  and  $L_{IRM-t}$  are the model accuracy of the proposed method. It can be seen that using the knowledge distillation method based on the instance relation matrix to transfer the knowledge from the pretrained teacher model VGG16 to the untrained student model MobileNet, the accuracy of the model had been significantly improved, with the highest accuracy of 93.60%. The network structures of teachers and students used are VGG16 (304.0 MB) and MobileNet (28.0M), and the accuracy is only 2.25% different from that of the teacher model. Moreover, the accuracy of the model with multiple loss functions is higher than that with only IRM loss, which proves that the IRM transformation loss ( $L_{IRM-t}$ ) is useful. Figure 5 shows the visualizations of the activation effects of different convolutional layers after IRMKD distillation. As can be seen from the figure, the model has effectively learned to extract and activate the disease spots on the sample leaves.

To evaluate the performance of the IRMKD method in real-world scenarios, this paper tested the MobileNet model trained using this method on the mango powdery mildew dataset provided by the Plant Protection Research Institute of the Guangxi Academy of Agricultural Sciences. The changes in accuracy and loss function during the model training process are illustrated in Figure 6. From the figure, it can be observed that during the training process, the loss function decreases, and simultaneously, the prediction accuracy on the test set shows an overall increasing trend. Moreover, the model converges rapidly, achieving a good convergence state after 50 iterations, the highest accuracy reached is 95.54%.

This experiment compares VGG16, AlexNet, GoogleNet and Resnet four main neural network structures, and  $L_{MTK}$  model's verification set accuracy, parameter size in three-channel RGB image of  $128 \times 128$  sizes. The network with higher verification set accuracy and model parameter quantity is taken as teacher model. Otherwise, it is taken as a student model. It can be seen from the table that MobileNet had the least network parameters, and its accuracy is slightly lower than VGG16 and Resnet. Therefore, this paper chose MobileNet as the student model in knowledge distillation. The results show that VGG16 can make the model the highest accuracy in the search space. At the same time, the average accuracy of  $L_{MTK}$  model optimized by knowledge distillation method is 97.62%, which is slightly improved compared with the other four network structures. The distributed MobileNet model has better performance in memory and average recognition time than other networks. The

average recognition time is shortened to 0.218 s, and the model size compression is only 19.83MB, in general, the model after knowledge distillation has higher recognition performance, and can meet the requirements of different application scenarios in recognition time and disk occupation. This proves the effectiveness and feasibility of this method.

In order to compare the performance between our proposed method and other plant disease recognition models, we compared the four methods: generalized regression networks (GRNNs) (Wang et al., 2012a), probabilistic neural networks (PNNs), radial basis function (RBF) (Wang et al., 2012b), BP network with PAC (Wang et al., 2012c). The experimental results are shown in Table 3. The test results show that GRNN and PNN have the highest driving accuracy, 97.27%, 98.06% respectively. It can be seen that these four methods have higher recognition accuracy. The accuracy of RBF neural network, PCA and BP network are 96.06% and 95.44%, respectively, which is slightly lower than the previous four methods. Furthermore, although our method is not as accurate as the 6 methods above, it is clearly ahead of the other methods in terms of model parameters and training speed. In addition, the experimental results also show that the accuracy of the model after pre-processing is improved compared to the model without pre-processing, with an average increase of 0.54%.

Finally, the accuracy of verification set and the change of IRM loss value are visualized. It can be seen that within 100 iterations, with the increase of the number of iterations, the accuracy of verification set increases from about 65% to about 92%, and the loss value of IRM decreases from about 0.6 to about 0.2. The model effect is significantly improved.

## Conclusion

Aiming at the problem of redundancy and low accuracy of traditional plant disease identification methods, this paper proposes a structured compression method based on knowledge distillation. Compared with the classic convolutional neural network model, this method has advantages in performance. Our method can slightly improve the accuracy rate, while greatly reducing the amount of parameters, shortening the recognition time, so that the model can meet higher real-time requirements. This article compares and analyzes the performance of different models. The main experimental results and conclusions obtained are as follows:

1. We compare the performance of different knowledge distillation methods on the Plant Village dataset in 4 network structures of VGG, AlexNet, GoogleNet and ResNet. The experimental results show that the highest accuracy of IRMKD is 93.96% when VGG is used as the teacher model.
2. Compared with other latest knowledge distillation methods, the results show that the Distilled-MobileNet model can slightly improve the accuracy rate while significantly reducing the parameter amount and memory of the model, and speeding up the model recognition speed.
3. In this paper, the model trained by IRMKD method is compared with other state of the art plant disease recognition methods. The experimental results show that IRMKD can significantly reduce the volume of the model and improve

the recognition speed of the model on the premise of slightly reducing the accuracy of the verification set.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

## Author contributions

JH: Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft. JS: Funding acquisition, Investigation, Project administration, Resources, Writing – review and editing. TC: Methodology, Project administration, Supervision, Validation, Writing – review and editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the 2024 Autonomous Region-level College Student Innovation and Entrepreneurship Training Program (Project No. 202410593001S).

## References

Chen, W. C., Chang, C. C., and Lee, C. R. (2018). “Knowledge distillation with feature maps for image classification,” in *Asian conference on computer vision*, 200–215.

Deng, C., Sun, F., Qian, X., Lin, J., Wang, Z., and Yuan, B. (2019). “Tie: energy-efficient tensor train-based inference engine for deep neural network,” in *Proceedings of the 46th international symposium on computer architecture*, 264–278.

Deng, L., Li, G., Han, S., Shi, L., and Xie, Y. (2020). Model compression and hardware acceleration for neural networks: a comprehensive survey. *Proc. IEEE* 108 (4), 485–532. doi:10.1109/jproc.2020.2976475

Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021). Knowledge distillation: a survey. *Int. J. Comput. Vis.* 129 (6), 1789–1819. doi:10.1007/s11263-021-01453-z

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv Preprint arXiv:1503.02531*. doi:10.48550/arXiv.1503.02531

Hughes, D., and Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv Preprint arXiv:1511.08060*. Available online at: <https://api.semanticscholar.org/CorpusID:11897587>.

Kim, Y., and Rush, A. M. (2016). Sequence-level knowledge distillation. *arXiv preprint arXiv:1606.07947*

Kuzuhara, H., Takimoto, H., Sato, Y., and Kanagawa, A. (2020). “Insect pest detection and identification method based on deep learning for realizing a pest control system,” in *2020 59th annual conference of the Society of instrument and control engineers of Japan (SICE)* (IEEE), 709–714.

Lee, S., and Song, B. C. (2019). Graph-based knowledge distillation by multi-head attention network. *arXiv Preprint arXiv:1907.02226*. Available online at: <https://api.semanticscholar.org/CorpusID:195847947>.

Meng, Z., Li, J., Zhao, Y., and Gong, Y. (2019). “Conditional teacher-student learning,” in *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (IEEE), 6445–6449.

Passalis, N., Tzelepi, M., and Tefas, A. (2020). “Heterogeneous knowledge distillation using information flow modeling,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2339–2348.

Peng, B., Jin, X., Li, D., Zhao, S., Wu, Y., and Liu, J. (2019). “Correlation congruence for knowledge distillation,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 5007–5016.

Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. (2019). Regularized evolution for image classifier architecture search. *Proc. AAAI Conf. Artif. Intell.* 33, 4780–4789. doi:10.1609/aaai.v33i01.33014780

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6 (1), 1–48. doi:10.1186/s40537-019-0197-0

Wang, H., Li, G., Ma, Z., and Li, X. (2012a). “Application of neural networks to image recognition of plant diseases,” in *2012 international conference on systems and informatics (ICSAI2012)* (IEEE), 2159–2164.

Wang, H., Li, G., Ma, Z., and Li, X. (2012b). “Image recognition of plant diseases based on principal component analysis and neural networks,” in *2012 8th international conference on natural computation (IEEE)*, 246–251.

Wang, H., Li, G., Ma, Z., and Li, X. (2012c). “Image recognition of plant diseases based on backpropagation networks,” in *2012 5th international congress on image and signal processing (IEEE)*, 894–900.

Weng, L., and Peneel, B. (2011). “A secure perceptual hash algorithm for image content authentication,” in *IFIP international conference on communications and multimedia security*, 108–121.

Zhang, D., Yang, J., Ye, D., and Hua, G. (2018). “Lq-nets: learned quantization for highly accurate and compact deep neural networks,” in *Proceedings of the European conference on computer vision* (Munich, Germany: Springer Science), 365–382.

Zhang, L., Shi, Y., Shi, Z., Ma, K., and Bao, C. (2020). Task-oriented feature distillation. *Adv. Neural Inf. Process. Syst.* 33. doi:10.5555/3495724.3496961

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.