Mathematical Biosciences and Engineering

MBE, 20 (6): 10223–10243.
DOI: 10.3934/mbe.2023448
Received: 31 December 2022
Revised: 12 March 2023
Accepted: 19 March 2023
Published: 30 March 2023

http://www.aimspress.com/journal/MBE

Research article

A specific fine-grained identification model for plasma-treated rice growth using multiscale shortcut convolutional neural network

Wenzhuo Chen[1,#], Yuan Wang[1,#], Xiaojiang Tang[1], Pengfei Yan[1], Xin Liu[1], Lianfeng Lin[1], Guannan Shi[1], Eric Robert[3] and Feng Huang[2,3,4*]

[1] College of Information and Electrical Engineering, China Agricultural University, Beijing, 10083, China
[2] College of Science, China Agricultural University, Beijing 100083, China
[3] GREMI, UMR 7344, CNRS/Université d'Orléans, 45067 Orléans Cedex France
[4] LE STUDIUM Loire Valley Institute for Advanced Studies, Centre-Val de Loire region, France

[#]The same contribution

[*]**Correspondence:** Email: huangfeng@cau.edu.cn. Tel: +86-010-62737618.

Abstract: As an agricultural innovation, low-temperature plasma technology is an environmentally friendly green technology that increases crop quality and productivity. However, there is a lack of research on the identification of plasma-treated rice growth. Although traditional convolutional neural networks (CNN) can automatically share convolution kernels and extract features, the outputs are only suitable for entry-level categorization. Indeed, shortcuts from the bottom layers to fully connected layers can be established feasibly in order to utilize spatial and local information from the bottom layers, which contain small distinctions necessary for fine-grain identification. In this work, 5000 original images which contain the basic growth information of rice (including plasma treated rice and the control rice) at the tillering stage were collected. An efficient multiscale shortcut CNN (MSCNN) model utilizing key information and cross-layer features was proposed. The results show that MSCNN outperforms the mainstream models in terms of accuracy, recall, precision and F1 score with 92.64%, 90.87%, 92.88% and 92.69%, respectively. Finally, the ablation experiment, comparing the average precision of MSCNN with and without shortcuts, revealed that the MSCNN with three shortcuts achieved the best performance with the highest precision.

Keywords: multiscale shortcut convolutional neural network; fine-grained identification; plasma-

treated rice growth classification

1. Introduction

At present, the main methods of increasing crop production include chemical fertilizers, pesticides, auxin and so on. They were proved to have negative effects on our environment and human health, such as water quality deterioration, soil acidification, crop quality decline, food security problems and secondary pollution [1−3]. However, traditional scientific and technological methods, such as chemical and biological methods, cannot effectively solve the above problems, because the resistance of these compounds and both physical and chemical properties of food will produce unwanted changes. Based on environmentally friendly and food safety considerations, researchers continue to explore methods to increase production and improve quality. Low temperature plasma (LTP) technology was found to be capable of producing a huge number of ions, electrons, free radicals, ground and excited state molecules and so on, which readily react with the contacted materials. It was also a highly efficient and environmentally friendly agriculture technique. Moreover, numerous studies have demonstrated that the use of LTP technology in agriculture can improve quality and yield of crops [4−8]. To be specified, when seeds were exposed to plasma under atmospheric pressure, the formed active ingredients such as reactive oxygen species (ROS) and reactive nitrogen species (RNS) can disinfect seeds, disrupt dormancy and stimulate germination and growth. However, few experiments have been carried out to use plasma-treated seeds for field planting, that is, relevant experiments were mainly limited to laboratory research. Moreover, the treatment effects usually turned out to be different according to the types and parameters of plasma discharge. In order to identify the effect of different plasma discharge parameters on rice, in this paper the practical field planting and the growth data of plasma-treated rice were collected and graded according to the primary classification criterion, which include tiller number, plant height, leaf area, leaf length and chrominance.

In order to identify the plasma-treated rice, which exhibits similarities and subtle differences on both intra-class and inter-class, the primary characteristics of fine-grained visual categorization (FGVC) were analyzed. FGVC is a new important research area in computer vision that tries to differentiate sub-classes of objects in images with the same entry category [9]. Over the last few decades, experiments have been conducted with a wide range of approaches to basic image classification, which includes the machine learning algorithms such as decision tree [10], K nearest neighbors (KNN) [11], support vector machine (SVM) [12] and multi-layer perceptron (MLP) [13]. Moreover, manual-designed and single features such as color, local binary pattern (LBP) and histogram of oriented gradient (HOG) were commonly used for the classifier's inputs. Even if embedding methods stacked these independent features together, they could not adequately show the complex relationships among those features. It was not until AlexNet achieved a great success on ImageNet dataset which was perceived as a major breakthrough, and revealed the great advantages of deep learning models, the hierarchical CNN then began to play a dominant role in fundamental visual classification tasks [14−19].

Although the aforementioned machine learning and deep learning models stimulated the rapid growth of computer vision, the specific issue is that the basic classification tasks primarily focus on identifying entry-level images with clear visual distinctions, for example, the identification among cats, dogs and birds. In fact, complex FGVC tasks involving the distinction of closely related sub-classes

become more prevalent. For instance, in terms of common computer vision tasks, it is sufficient to locate a red automobile in heavy traffic. Whereas, for FGVC projects, they require the cooperation of human specialists and a network to offer detailed information so that the algorithm can determine the specific style, version, manufacturer, production age and category. Current studies have focused on FGVC objects, such as texture of feather [14,15], flowers[20-22], structure of aircraft [23−25], beak of birds [10,26], types of vehicle [27,28] and human [29,30]. All of these research have promoted FGVC to the forefront of computer vision. However, in some circumstances, the intra-class differences may be far greater than the inter-class differences, which resulted in a significant loss in identifying accuracy and robustness [31−33].

To improve the performance of FGVC on the growth identification and classification of plasma-treated rice, the MSCNN based on traditional CNN architecture using multiscale shortcuts was proposed for the extraction and fusion of rice image features. On the one hand, the mechanism of MSCNN benefited from the high resolution of the bottom-layer features (from stage A) which included information about location, texture, lighting, angle, pose, shape, articulation and color. On the other hand, it examined rich semantic properties of the top layers (Stage B) via repeated convolution operations. The information features were used in conjunction with the concept of objects classification. To be more specific, the pipelines (Stage A, Stage B and Stage C in Figure 1) of MSCNN enabled the realization of fast falling, sparsity and fully connection. Additionally, in order to form a multiscale structure from bottom to top and gain more complex distinctive information, the three shortcuts (S1, S2 and S3) from stage A to stage C followed the same direction. Simultaneously, grating was added to each shortcut to convert its output into one-dimensional vector, allowing comparatively coarser scale feature to incorporate with stage C. Ultimately, horizontal fusion of these features was applied through fully connection layer in Stage C to maintain coarser underlying features of fine-grained area in each input image, improving semantic representation of top layer outcomes. This improvement was verified to enable a more accurate and precise network.

In this paper, the data set of eleven groups of rice including with and without plasma treatment was constructed and categorized based on basic growth criterion. Then the data set was used to train and test MSCNN to obtain the effective identification results. Experimental results demonstrated that for the growth identification of rice treated by plasma, the proposed MSCNN was superior to other comparative classification algorithms in terms of accuracy, precision, F1 score and recall. Simultaneously, the ablation experiments were performed on MSCNN to verify the efficiency of the shortcuts and the optimal configuration of the proposed structure. The main contents and innovations of this article are as follows:

(1) 5,000 high-fidelity rice images at the tillering stage were taken as original data and were expanded to 10,000 images via data augmentation. At the same time, textual information includes the plasma parameters, tiller number, plant height, leaf area, leaf length and chrominance were also recorded.

(2) Using fine features on different scales to improve CNN by shortcuts, the practical goal to the fine-grained identification of the growth state of rice treated by plasma was achieved.

(3) Ablation study of MSCNN was conducted to determine the most effective network architecture for improving average precision and demonstrating the possible applicability of MSCNN.

2. Related works

To overcome the issues caused by the similarity of subclass image features, posture discrepancies and background interference, research has been carried out to solve the problems of traditional CNN networks. Normally, FGVC pipelines focus on locating and resembling critical region of interest (ROI). According to the supervision method of data set, there are two categories, namely strongly supervised and weakly supervised.

For the strong supervision method, apart from human expert annotation datasets (e.g., Stanford Dogs, Caltech-UCSD Birds), they also requires additional artificial markers, such as auxiliary bounding boxes and local coordinates information for specific locations, to help removing background noise and completing domain detection of foreground objects. Typical works are as follows. Zhang et al. [34] proposed a network that detected the less deformable parts and localizes other highly deformable parts with simple geometric alignment. Wei et al. [35] showed a novel Mask-CNN model without the fully connected layers basing on part of annotations, which can both locate the discriminate parts and generate weighted object by part of masks. Qi et al. [36] used two core modules as the selected module and representation module to exploit spatial relation and capture more discriminate details for FGVC. Zhang et al. [37] applied the original image as input data and used the generated visual description representing coarse-to-fine visual clues. However, these strongly supervised methods were still less accurate and insufficient when they encounter the extraction of discriminatory features and the location of significant regions, owing to substantial inter-class differences and minor intra-class differences.

Weakly supervised models aim to acquire local features without requiring further component annotation, that is, to classify fine-grained images solely based on the category label of image annotation. Various enhancements were made to modify the extraction function of traditional CNN to implement weakly supervision. Huang et al. [38] proposed a novel Part-Stacked CNN architecture by modeling subtle differences of object. Lee et al. [39] learned useful features directly from the raw representations of input data using CNN and gained the intuition of the chosen features based on a deconvolutional network (DN). Rohrbach et al. [40] introduced a hand-centric approach for fine-grained activity classification and detection and found that decomposition into attributes allowed sharing information across composites. Cai et al. [41] proposed a polynomial kernel based on predictor to capture higher-order statistics of convolutional activation for modeling interaction and extend polynomial predictor to integrate hierarchical activation via kernel fusion. Hu et al. [42] proposed a spatially weighted pooling (SWP) strategy and pooled the extracted features of DCNNs with the guidance of its learning masks, thus minimal modification was needed in terms of implementation. He et al. [16] presented a residual learning framework to ease the training of networks that are substantially deeper than those used previously and improved the accuracy from considerably increased depth.

Although weakly supervised networks are marginally worse than the strongly supervised ones, they are less expensive and more practical due to no need of extra local annotation. Moreover, the identification criterion of rice growth grade involves some parameters, such as the leaf area, tiller number, chrominance and so on, which are hardly labeled in images. Thus, weakly supervised method was selected to classify the growth grade of plasma-treated rice. The above mentioned weakly supervised networks normally consisted of numerous convolution layers, and even with the residual learning frame, these modules were still repeated for several times, which is not simple enough in terms of the network architecture. Moreover, for our project, prior to Softmax in the CNN architecture, the feature granularity was exceedingly coarse, and finer features from the bottom can help with distinguishing subtle discrimination that seldom can be conveyed backward. There is a lack of link,

that can take full use of the bottom convolution layers, which are more susceptible to fine-grained information (e.g., texture, direction and edge) through the output visualization results of each convolution layer module. Further, as finer information flows forward through the pipeline, it could be interpreted into semantic information in the top layer, such as the leaf area, tiller numbers and chrominance, which are used for the criterion of rice growth grade. Thus, the purpose of this research was to improve the traditional weakly supervised CNN architecture with simple shortcut connections and multiscale coarse feature description. The sections bellowed detail the architecture and optimization of the proposed framework procedure.

3. MSCNN

According to previous researches [43−46], features in the top layer are highly compressed, but they are too fine in space to accomplish precise positioning that led to a poor discrimination ability. In contrast, bottom layers in early stages of the model are sensitive to information like direction, texture and edge, but do not express sophisticated semantic representation. In an attempt to take advantage of the success of traditional CNN networks for object classification, this work concentrates on a comprehensive representation of both bottom and top layers for FGVC task of rice growth identification. The detailed architecture of MSCNN is divided into three components, that is, pipeline, multiscale shortcuts and main calculations as described below.

3.1. Pipeline

Based on CNN, MSCNN consists primarily of three stages, and three shortcuts were attached to the pipeline, as shown in Figure 1. Among these stages, stage A is used to help the input images falling fast into relatively small feature maps. Stage B adds dropout and local response normalization (LRN) to raise the sparsity of data and prevent overfitting. Additionally, shortcuts are used to preserve and grate the shallow finer features of each max-pool layer in stage A. Then, they were delivered to stage C, which implements multiscale-feature fusion using fully connected layers.

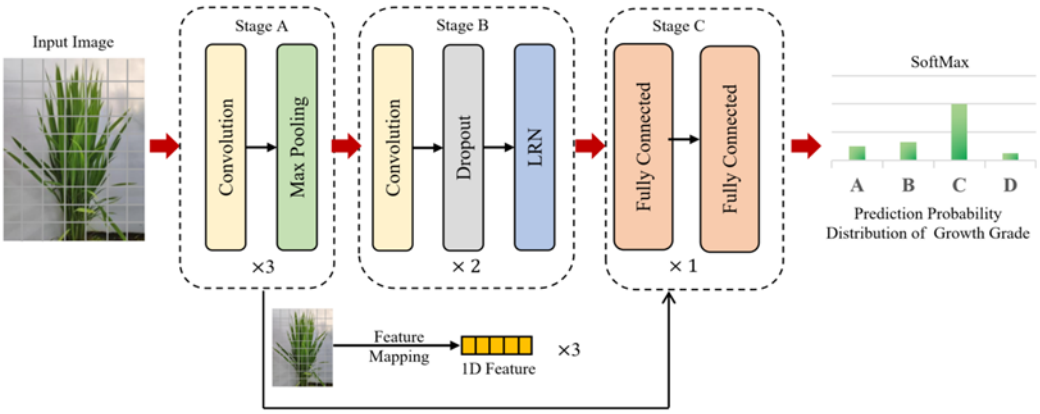

Figure 1. Pipeline of MSCNN.

In stage A, with the feeding of three-dimensional RGB images of rice (rice with and without plasma treatment) at the tillering stage, 5*5 convolution and 2*2 max-pool were stacked three times

to rapidly compress the feature map into 9*9*192. The modest convolution kernel size is used to expand the coarse degree of feature extraction and receptive field. However, in comparison to prior work on architecture uniformity, where 3*3 convolution filter throughout the whole network, the relatively large size in stage A introduces a large number of parameters. Thus, the pooling layer is introduced to solve this issue, as its primary job is down-sampling and dimension reduction. Likewise, for the FGVC task of rice growth identification, max pooling is selected to retain more information on texture detail and decrease convolution layer parameter errors caused by the deviation of the estimated mean.

Following that, LRN and dropout are attached following 3*3 and 2*2 convolution, which were repeated two times in stage B. Inspired by the discovery of genuine neurons, while highlighting its peak and constraining surrounding values to avoid neuron saturation, the LRN layer is used to suppress the output of the activation function laterally at the end of each convolution. Meanwhile, during the forward propagation, dropout also processes the activation value in a way of making certain neurons stop working with certain probability. These strategies increase the generalization and robustness due to the result of local features and data sparsity.

Finally, in stage C, the convolution operation with a 1*1 kernel size is applied as fully connection. It is used to map the learned distributed feature representation to the sample classification space, which corresponds to the predicted probability distributions for rice growth grade of A−D.

3.2. Multiscale shortcuts

In this research, the image difference among the rice (with and without plasma treatment) may not be obvious, resulting in a reduced precision for typical CNN classification. Thus, multiscale shortcut methodology is applied by fusing and reusing the features at component-level (①, ② and ③) and object-level (④), as shown in Figure 2. It is obvious that MSCNN is formed by attaching three shortcut connections (S1, S2 and S3) from different scales in Stage A to the pipeline of standard CNN.

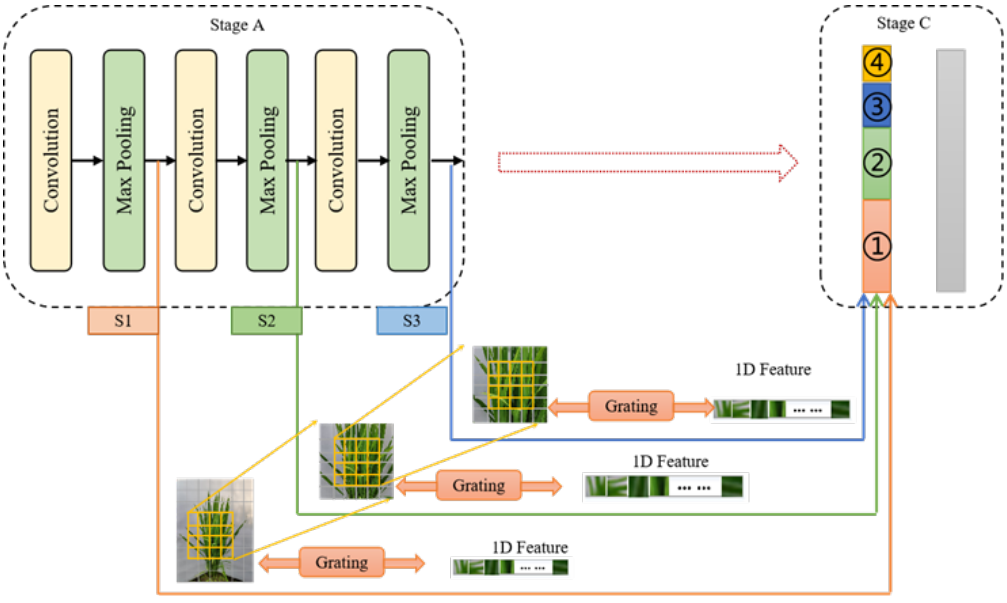

Figure 2. Shortcuts of MSCNN.

For forward propagation, due to the serial arrangement of network architecture, the bottom layers that reside in the fast fall blocks of Stage A, only learn shallow features such as edge, direction, shape, texture and color. These outputs were subsequently forwarded to the next operation of stage B. It is responsible for learning deep semantic features that contain global characteristics about the item at object-level to determine the growth grade and can locate attractive target region in order to gain a preliminary sense. In the meantime, grating is employed to convert the max-pooling layer outputs from three dimensions W*H*C (width, height and channels, respectively) to one-dimensional feature(1*1*WHC). In the final stage, the fully connected layer factionalized as a classifier, which provides an output of 1*1*4. Grating is needed to preprocess the 3D feature map in order to fuse coarse object-level information from stage B with the three component-level features of S1-S3 through the first fully connection layer in stage C.

While back propagating, optimization strategy is used to update the parameters based on the difference between ground truth and prediction received from stage C. Accordingly, the back-propagated differences of stage C are composed of four parts. Namely the differences between S1, S2, S3 and output of stage C, which will be propagated to respective layers. Then, they flow back until reach the initial layer to help update parameter and optimize the model configuration.

Because the multiscale shortcuts are only dedicated for feature preservation and grating, minor weight and input signal alterations have little effect on the model improvement process. Furthermore, unlike ResNet [16] and DenseNet [17], the spanning scale in Figure 2 is not fixed, allowing for customization shortcuts based on data requirements. Thus, the bidirectional propagation of multiscale distinctions shortcuts increases sample use of essential information in stage C by reserving and fusing all feature maps at all scales, allowing shortcut connections between shallow and deep layers to compensate for the lack of serial CNN perception.

3.3. Main calculations

In stage A, the input plasma rice images were divided into three dimensions, which are red, green and blue respectively. Then each dimension was processed as a feature map, convolution and max-pooling were used immediately to deconstruct it quickly. To be more specified, in order to strengthen and filter of the original intra-class features, convolution operations were introduced to extract different scales of features between image feature matrix and the convolution kernels, using the function as below:

$$C_{j,i}^{n} = f(\sum_{i=1}^{m} S_{j-1}^{n} * w_{j,i} + b_{j,i}), \tag{3.1}$$

where $C_{j,i}^{n}$ is the output of the i^{th}convolution kernel in the j^{th}convolution layer, $S_{j-1,i}^{n}$ is the i^{th} down-sample results in the $(j-1)^{th}$ down-sampling layer, $w_{j,i}$ is the weight matrix for the i^{th} convolution output in the $(j-1)^{th}$ down-sampling layer, $b_{j,i}$ is the corresponding bias, "*" is convolution operation and $f(\ )$ is the nonlinear activation function. Typically, the nonlinear activation function will be chosen from the Sigmoid, Tanh and ReLU. Due to the fact that the gradient of Sigmoid and Tanh changes gradually in the saturated zone, which is susceptible to the gradient fading and can critically decrease the model convergence rate. Here we use ReLU, as the networks activation function, it guarantees the sparsity of neuron links and declines the incidence of over-fitting.

Its formula is expressed as follows is

$$ReLU\left(C_{j,i}^{n}\right) = max\left(0, C_{j,i}^{n}\right) \tag{3.2}$$

It is worth noticing that the convolution operation significantly increases the computational costs and dimensionality. As a result, pooling layers were put fractionally behind to lighten the structure. The objective is to preserve certain immutability (rotation, translation and expansion), as well as to help lower the likelihood of overfitting, reserve critical information and booste robustness, (the ability to resist the distortion within a specified range). The mean-pooling, max-pooling and stochastic-pooling are the three commonly used pooling methods. Errors in feature extraction are mostly due to two factors. One is the neighborhood size which raises the variance of the estimated value, and the other is convolution layer parameter error which causes divergence from the calculated mean value. Generally speaking, the mean-pooling reduces the first error and preserves more background information of the image. Additionally, stochastic-pooling choose the pixel point value according to its probability which lead to a weak performance especially for FGVC. Due to the randomness, subtle discrimination information might not be kept. Whereas max-pooling selects the maximum value which can help lower the chance of the second error and help retain more texture information by obtaining local features. That is what FGVC applications require. So, we employ maximum pooling. Its formula is as follows:

$$h_{m,k} = maxa_{s_i,k}, s_i \in N_m, \tag{3.3}$$

where s_i is the pooling window size, $a_{s_i,k}$is the activation value in the k^{th} channel, m is the output feature map size after pooling, $h_{m,k}$ is the maximum value of each point in the k^{th} channel. Even after using noise-canceling and cutting, a small amount of noise interference remains. Consequently, a total of three fast fall blocks are applied to increase the robustness and capability of feature extraction.

4. Plasma seed treatment and data acquisition and preprocessing

4.1. Plasma seed treatment

In order to find optimal plasma parameters that can increase the rice production and quality significantly, eleven groups of rice were discussed here with and without plasma treatment. The group without plasma treatment is named as the control group (CK). There are three groups of plasma treated seeds with arc discharge (AD), radio frequency (RF) discharge and dielectric barrier discharge (DBD), respectively.

Table 1 shows the parameters of rice groups with and without plasma treatment and their corresponding average tiller number. For group AD (No.2-5), the processing power is 455 W, the discharge area is 30 mm^2, the reaction medium is air with the flow rate of 1.5 L/min and the processing time is 0.6, 1.2, 1.8 and 2.4 s, respectively. The processing time in the group RF (No.6-8) was 60, 120 and 180 s, with argon of 80 pa , the RF power of 60 W and the discharge area of 8 mm^2. In the group DBD (No.9-11), the processing power was 45, 72 and 92 W, respectively, with the discharge area of 13 mm^2, processing duration of 30 s, frequency of discharge voltage of 9.5 kHz and using air as the diacharge gas.

After being treated with different plasma parameters listed in Table 1, the rice seeds were planted in the same location with the same environment (in Taizhou city, Jiangsu province, from May to November). During the rice growing period, tiller number is critical to identify its growth grade, and plays a significant role in rice yield. Thus, when these groups of rice reached the tillering stage, the number of tillers were collected. In Table 1, the data shows that group CK has the minimum tillering number (19), less than all the other plasma treated groups which show the obvious improvement by plasma treatment. Among these groups, No. 3 (in the group AD) has the maximum tillering number (34), indicating appropriate plasma parameters can significantly increase the tillering number and boosting growth vitality. Moreover, further experimental result also showed that No.3 (in group AD) has the highest yield.

Table 1. Parameter settings of plasma treatment and corresponding average tiller number.

Group	Number	Power/W	Discharge area/mm^2	Processing duration/s	Reaction medium	Average tiller number
CK	1	-	-	-	Air	19
AD	2	455	30	0.6	Air	29
	3	455	30	1.2	Air	34
	4	455	30	1.8	Air	29
	5	455	30	2.4	Air	27
RF	6	60	8	60	Argon	28
	7	60	8	120	Argon	25
	8	60	8	180	Argon	25
DBD	9	45	13	30	Air	25
	10	72	13	30	Air	24
	11	92	13	30	Air	27

4.2. Data acquisition

In our investigation, during the tillering stage of the rice, 5,000 high-fidelity rice images were obtained from top and side view for the 11 groups in Table 1. The image format was JPEG and each one was a 24-bit color bitmap. In addition, the plant height, leaf length and area, tiller numbers and chrominance of the 11 groups of rice were also measured. As seen in Figure 3, from left to right, these images show the growth status of group CK, DBD, RF and AD, which correspond to group No.1, 10, 6 and 4 in Table 1, respectively. Obviously, it shows that the rice of No.10 (in group AD) has the highest plant height and the most tillers.

According to the standard of rice growth grade issued by China Agriculture Press, it includes five types of parameters for rice growth (tiller number, plant height, leaf area, leaf length and chrominance) and the characteristics of each type are divided into four grades of A-D shown in Table 2. Then, each type of data measured in our experiment can correspond to the four different grades in Table 2. For example, the rice of No.10 (in group A) has the tiller number of 34, plant height of 45 cm and leaf area of 39 cm^2. Combined with the standard in Table 2, the rice of No.10 represents the growth grade D for tiller number, grade C for plant height and grade C for leaf area, respectively.

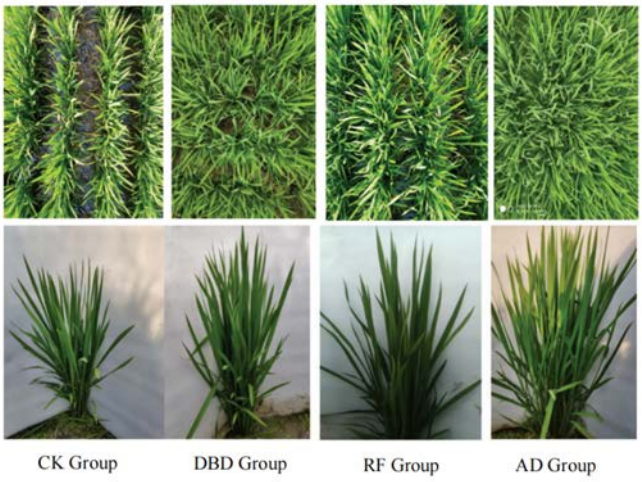

Figure 3. Top and side view of rice images.

Table 2. Rice growth grade.

Grade	A	B	C	D
Number of tillers	10-16	17-23	24-30	31-37
Plant height/cm	30-35	36-41	42-47	48-53
Leaf area/cm^2	15-25	26-36	37-47	48-60
Leaf length/cm	16-23	24-31	32-39	40-50
Chrominance	Weak	Average	Strong	Ultra-Strong

4.3. Image augmentation

Appropriately labeled samples can alleviate over-fitting in the model training stage [47]. In order to make the data more comprehensible and the MSCNN network more robust, the data is preprocessed before training. Data augmentation operations is the common method of data preprocessing, which includes vertical deformation, elastic distort, oblique quadrangle, rotation and blur, color filtering, noise addition, PCA jittering and scaling blur [31, 48-51]. The previous studies have shown that these augmentation strategies are satisfactory. In our experiment, the operations of rotation, oblique quadrangle, vertical deformation and elastic distort were randomly implemented for data preprocessing shown in Figure 4.

Figure 4. Image augmentation operations.

Table 3. Rice growth grade.

Grade	A	B	C	D
Number of original images	885	1336	1790	989
Number of images after augmentation	1770	2672	3580	1978
Proportion	17.70%	26.72%	35.80%	19.78%

Our field planting results have showed that the rice of No.10 has the best overall growth state. However, according to Table 2, this group acquires grade D, C and C in terms of tillering number, plant height and leaf area, respectively. In order to give a comprehensive growth grade for each group of rice, the weights of the five parameters in Table 1 were assigned with 0.5, 0.3, 0.2, 0 and 0, respectively. Then, according to the calculated overall results after weighted, the rice images in our dataset were labeled with four grades of A (weak), B (normal), C (good) and D (outstanding). After the weights were set, the overall growth grade of No.10 was modified to be D. Table 3 shows the numbers of original and augment images of rice with the overall four growth grades. It shows the original images with grade A of 885, grade B of 1336, grade C of 1790 and grade D of 989. Then these original images were augmented randomly with the four operations in Figure 4 to obtain the doubled number of original images. The dataset after augment will be used to train and test MSCNN model.

5. Computational experiments

In this section, we conducted computational experiments on our dataset, in order to compare the proposed MSCNN with the mainstream CNNs on our defined FGVC task of plasma-treated rice growth. The experimental setup and selected evaluation indicators for our dataset were conducted first. Then, sensitivity analysis was performed on critical parameters (optimizer training techniques, learning rate and batch size) to fine-tune the MSCNN architecture and assist in choosing the most effective hyper parameters for MSCNN. Likewise, analyses were given based on the comparison MSCNN with the main CNN networks. Lastly, the ablation study was conducted to show the effectiveness and feasibility of adding shortcuts in the MSCNN network.

5.1. Computing environment

Due to the enormous number of iterations and high data throughput, the training procedure is time consuming. Subsequently, a powerful graphics processing unit (GPU) is crucial. In this study, all computational experiments were conducted on NVIDIA GeForce GTX 1050 Ti GPU. In addition, it used Intel (R), Core (TM) i7-9700K (3.00GHz) processor with 32GB memory. The operating system was Windows 10 (64-bit). The neural network model was constructed using TensorFlow1.14, an integrated deep learning computing framework developed by Google Brain project. In order to achieve faster graphical computation and less storage cost, CUDA toolkit 10.0 was also executed. The specific integration environment of the model is the interactive language development platform Spyder under Anaconda 3.6, which used Python as the programming language.

5.2. Evaluation indicators

Plasma-treated rice growth identification is a novel application field of FGVC. In the FGVC tasks,

there are various evaluation indicators, including receiver operating characteristic (ROC), area under curve (AUC), precision recall curve (PRC), confusion matrix, accuracy, precision, recall and F1-score. Among these indicators, AUC, ROC and PRC are so comprehensive and it is difficult to observe the designed model with single indicator readily and intuitively. Subsequently, accuracy, precision, recall and F1 score are considered for the effectiveness verification of the MSCNN. The four evaluation indicators are defined as follows.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}, \tag{5.1}$$

$$Average\ Precision = 1/N \sum_{i=1}^{N} \frac{TP_i}{TP_i+FP_i}, \tag{5.2}$$

$$Average\ Recall = 1/N \sum_{i=1}^{N} \frac{TP_i}{TP_i+FN_i}, \tag{5.3}$$

$$F1\ Score = 1/N \sum_{i=1}^{N} \frac{2TP_i{}^2}{TP_i(2TP_i+FP_i+FN_i)}, \tag{5.4}$$

where TP_i, TN_i, FP_i and FN_i are the number of samples that belong to true positive (TP), true negative (TN), false positive (FP) and false negative (FN) of the i^{th} category respectively. N is the total category number of testing samples in the test dataset. Due to the task in this study is the multiple classification, the calculated average values of these indicators would be used for evaluation.

5.3. Parameter settings and sensitivity analysis

Prior to training the network, it is critical to configure the parameters. According to the followed experimental results, the following section discusses the training super parameter set in detail, as shown in Table 4.

Table 4. Parameter settings of MSCNN.

Super parameter	Option/Value
Training strategy	RMSProp
Learning rate	0.001
Momentum unit	0.95
Minimum sample input batch	64
Decay rate	0.1
Decay step	10000
Maximum iteration number	20000

(1) Different optimized training strategies

Gradient-based optimization training strategy RMSProp was selected for MSCNN at the training stage according to the result shown in Figure 5 for the following reasons. To determine the ideal model training technique, four practical and popular strategies including stochastic gradient descent (SGD), Momentum, root mean square prop (RMSProp) and Adam are compared. As shown in Figure 5, these four training strategy curves are generally parallel throughout the training process, and the precision increases with the epoch number. All the optimal precision values are obtained near the 100th epoch.

As can be seen, the precision by SGD is remarkably the lowest. This is due to noise introduced by randomly selecting the gradient descent. This randomness may presumably lead to an ambiguous direction of weight update, trapping the gradient descent process at the saddle point. Although the Momentum method outperforms SGD, its update direction is completely reliant on the gradient calculated by the current batch and is therefore extremely unstable. The curves of RMSProp and Adam are similar. The former improves Momentum strategy by storing information of the prior gradients, and the learning step is gradient-dependent, whereas it does not have correction variables. Whereas Adam improved RMSProp by explicitly incorporating the estimation of the first-order moment (exponentially weighted), which could smooth the gradient. Nevertheless, Figure 5 demonstrates that the precision and speed of RMSProp is slightly better than Adam. This may be due to the addition of three shortcuts that speed up proportion changes in the gradient during backward propagation and alleviate the significant deviation generated by second-order moment estimation in the initial stage of RMSProp. As a result, the optimized training strategies of RMSProp is used in this study.

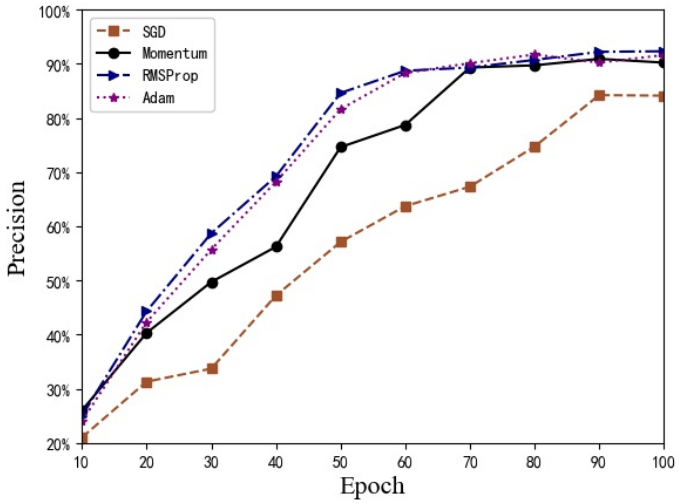

Figure 5. Effects of MSCNN with different training strategies on plasma-treated rice growth dataset.

(2) Different learning rate and momentum unit

After the RMSProp was selected as the training strategy of MSCNN, the following experiments are performed to confirm the appropriate learning rate (LR) and momentum (M). LR is used to regulate the rate of gradient descent. Excessive LR may result in a loss decrease or divergence. Whereas small LR can slow down the converge speed. Meanwhile, the presence of M strengthens the parameters change when current gradient direction is the same as proceeding, otherwise retards the change. Figure 6 shows the precision curves with the combinations of LR = 0.01 and 0.001 and M = 0.9 and 0.95, respectively. It is self-evident that when M equals 0.95, the overall precision is more than 90%. When the LR is 0.0001, the curves imply more instantly convergence to discover the optimal point than LR = 0.001. This is because with a small LR (0.0001) the gradient decreases accompanied with oscillation near the extreme point, and its combination with the large momentum of 0.95 can accelerate the convergence. According to the comparison of the curve trend shown in Figure 6, it shows that the precision of MSCNN with the combination of LR = 0.0001 and M=0.95 is the highest with steadier trend.

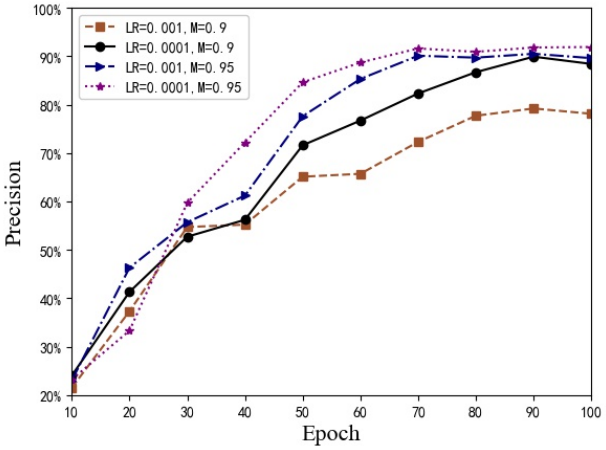

Figure 6. Effects of MSCNN with different learning rate and momentum unit on plasma-treated rice growth dataset.

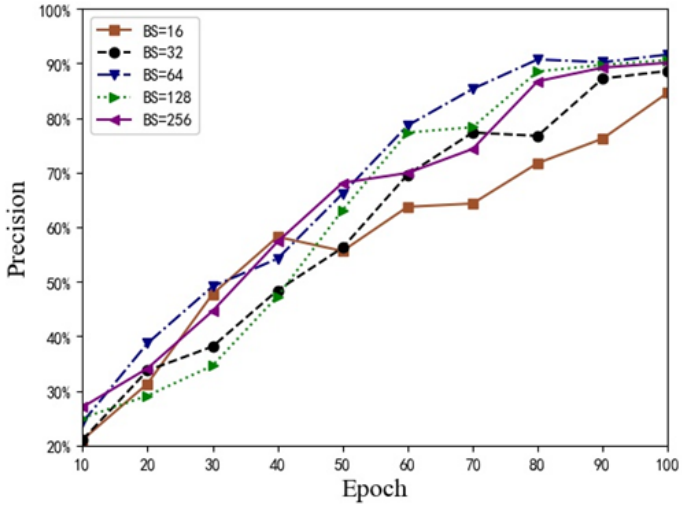

Figure 7. Effects of MSCNN with different learning rate and momentum unit on plasma-treatedr rice growth dataset.

(3) Different batch size

In order to acquire the optimal batch size (BS) and thus increase memory use and training efficiency, the performances of MSCNN utilizing different BSs during the training were compared, shown in Figure 7. During fine-tuning operation in back propagation, the network first averages the loss attained from each instance in each batch and then calculates the gradient based on the model output. Consequently, BS determines the gradient smoothing degree between adjacent batches. Typically, despite of batch normalization, when BS is small, the difference between adjacent batches is quite large, resulting in more severe gradient oscillation and divergence. On the opposite, when BS is large, the subtle difference makes it easy to fall into a local minimum. Thus, it is critical to recognize the balance value [52–54]. As shown in Figure 7, it reveals that with the rise of BS, precision goes up rapidly and the oscillation amplitude declines in the beginning. However, it is worth noticing that when BS is 128 or more, the precision curve tends to be irregular and the oscillation amplitude increases.

That is because large batch size introduces less noise that makes it preferable for a sharp minimize but increases the time required to get the same accuracy value. Finally, according to the higher precision value and faster convergence, the BS of 64 was selected for the training of MSCNN on rice images.

5.4. Identification results

The comparison of MSCNN and the five mainstream models,(i.e., AlexNet, MobileNet, ResNet, VGG-16 and VGG-19) [15,16,19,55] are conducted with the parameters in Table 4 and the training of 100 epochs.

As seen in Figure 8, the proposed MSCNN achieves a competitive performance that outperforms the others on each evaluation indicator. It has the greatest accuracy, recall, precision and F1 score of 92.64%, 90.87%, 92.88% and 92.69%, respectively. Thus, due to its novel structure, multiscale shortcuts in MSCNN allow for the integration of the higher layers' outputs in the first stage A without requiring frequent intermediary and transfer operations, resulting in a significantly superior self-optimization. Thus, compared with the similar connection in ResNet [15], the residual connection enables the network more sensitive to the fluctuation of weights and data, but it also brings a high possibility of over-fitting. As the best model, MSCNN surpass ResNet by 11.76%, 4.74%, 0.24% and 4.52% in terms of each indicator in turn. Moreover, this comparison result proves that it is feasible to improve the network performance through multiscale shortcuts. Additionally, as typical simple and deep networks, VGG-16 and VGG-19 [55] are only slightly better than AlexNet [19], confirming once again the importance of depth in visual representations. However, the loss of shallow features for the upper layers and identical kernel size throughout the whole convolution operation in VGG-16 and VGG-19 [55] are perhaps the common reasons for their weaker performance on FGVC tasks. Moreover, MSCNN outperforms AlexNet [19] and MobileNet [15], which are representatives of pure network representation and cannot cater for the stringent requirements of FGVC. Because their bottom-up, feed-forward architectures are incapable of fusing features from different layers for identification. In addition to the detailed comparison with the above models, the difference between MSCNN and the more advanced transformer model was also analyzed. First of all, transformers use multi-head self-attention mechanism and reduce the number of parameters of network. However, it lacks some of inductive biases inherent to CNN, such as translation equivalence and locality. Therefore, it needs a huge size of data to compliment the pre-train process in order to acquire the inductive bias. However, due to our dataset only has 5000 original pictures, transformer is not the best choice for dealing with our task of growth identification of plasma treated rice.

Moreover, Figure 9 shows the confusion matrix of the MSCNN. The abscissa represents the prediction classification of our plasma-treated rice growth, while the ordinate represents their ground truth classification. The color bar on the right indicates the degree of accuracy in growth grade. The number in each square represents the correct prediction percentage of each grade corresponding ground truth grade. For instance, 0.984 in the top left corner indicates that all 1770 images of Grade A are put into the model for training, among which 1742 images are identified as grade A, accounting for 98.4% of the total. As illustrated in Figure 9, the values in the four squares on the diagonal are significantly higher than the other squares, indicating that the MSCNN network accounts for the correct identification at the four grades, showing the effectiveness and feasibility of our model on the growth grade classification of plasma treated rice.

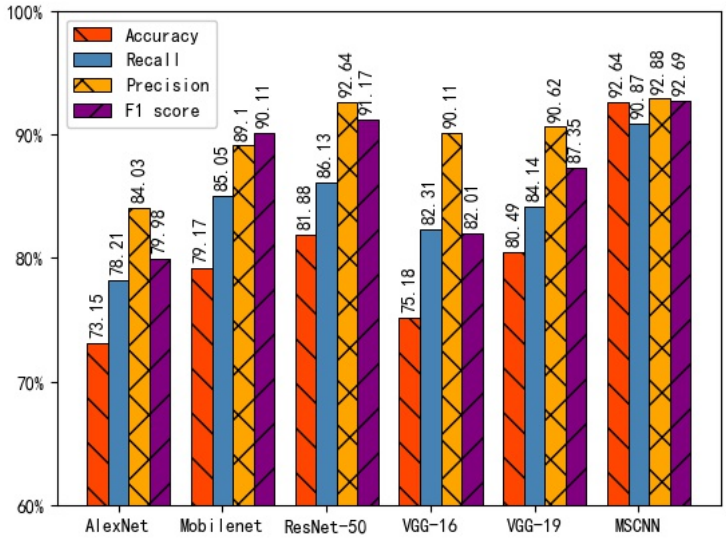

Figure 8. Comparison of models on plasma-treated rice growth dataset.

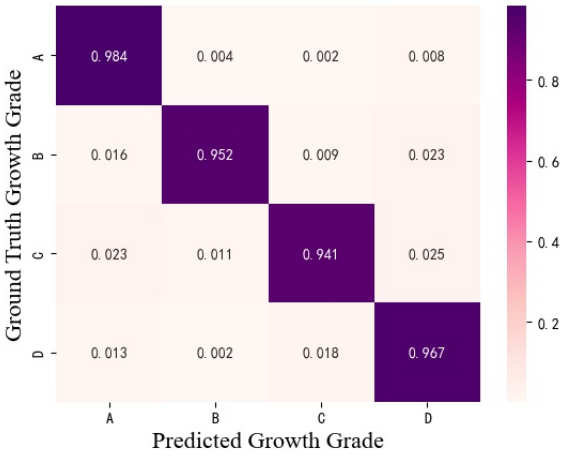

Figure 9. MSCNN confusion matrix on plasma-treated rice growth dataset.

5.5. Ablation study

To further study the effectiveness of multiscale shortcuts in MSCNN, the ablation experiment of the MSCNN with the combinations of different shortcuts were conducted. As shown in Table 5, the MSCNN with three shortcuts (S1+S2+S3) reaches the highest average precision (92.88%). This result can be explicable in terms of the MSCNN architecture. To be more specific, in the forward direction, the first fully connection layer in stage C fuses the feature maps consists of four components (i.e., the component-level feature maps come from S1, S2, S3 and object-level output feature of stage B). Then, the back-propagated errors from stage C to stage A flow through these shortcuts to improve the model performance. Hence, the coarse-scale representations and local-specific discriminations can be fully accounted for simultaneously through all the three shortcuts to provide multiple-perspectives. In addition, the MSCNN with two shortcuts (S1+S2, S1+S3, S2+S3) achieves the average precision of 89.06% and with only one of the three shortcuts (S1, S2, S3) has the lowest average precision (82.88%).

It may be due to that only one or two-scale feature preserves insufficient coarser information after grating compared with more shortcuts. Table 5 shows that the more shortcuts, the better the model performance, which proves the effectiveness of multiscale shortcuts in MSCNN.

Table 5. Precision of the combinations of shortcuts in MSCNN.

MSCNN model with different shortcuts	Precision	Average value of precision
With three shortcuts：S1+S2+S3	**92.88%**	**92.88%**
With two shortcuts：S1+S2	88.63%	
With two shortcuts：S1+S3	89.57%	89.06%
With two shortcuts：S2+S3	88.98%	
With one shortcut：S1	81.67%	82.88%
With one shortcut：S2	83.42%	
With one shortcut：S3	83.54%	
With no shortcut	84.18%	84.18%

6. Conclusion

In this paper, a dataset of plasma-treated rice growth images was constructed with the textual characteristics including tiller number, plant height and leaf areas. Then for the proposed MSCNN model, three shortcuts were attached to the traditional CNN structure to improve the performance by discriminating the differences among different scales and the grating layers were attached to each shortcut to create a one-dimensional feature vector. After that, all four-scale feature maps were concatenate horizontally to utilize the feature fusion in both component and object level. Consequently, the model improved the utilization rate of key information with more use of the hidden information to compensate for the serial CNN perception capability. Simultaneously, ReLU was employed as the activation function, and LRN and Dropout were added to the main neural pipeline as functional auxiliary layers to minimize gradient dispersion and overfitting. Compared with other mainstream models, the proposed MSCNN model has a simpler architecture but achieves the best identification performance according to the four evaluation indicators of accuracy, recall, precision and F1 score, which were 92.64%, 90.87%, 92.88% and 92.69%, respectively. Lastly the ablation study showed the best performance was acquired when the MSCNN with three shortcuts.

Acknowledgments

This work is supported by the National Science Fund of China with Grant No.11675261 and S&T Program of Hebei with Grant No.22375411D.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. A. M. Khaneghah, L. M. Martins, A. M. Von Hertwig, R. Bertoldo, A. S. Sant'Ana, Deoxynivalenol and its masked forms: Characteristics, incidence, control and fate during wheat

and wheat-based products processing - A review, *Trends Food Sci. Technol*., **71** (2018), 13−24. https://doi.org/10.1016/j.tifs.2017.10.012

2. N.S. Poluxeni, M. Sotirios, K. Chrysanthi, S. Panagiotis, H. Luc, Chemical pesticides and human health: the urgent need for a new concept in agriculture, *Front. Public Health*, **4** (2016) 148−148. https://doi.org/10.3389/fpubh.2016.00148
3. X. Lei, R. Qiu, Evaluation of food security in China based entropy TOPSIS model and the diagnosis of its obstacle factors, *J. China Agric. Univ.*, **27** (2022), 1−14. https://doi.org/10.11841/j.issn.1007-4333.2022.12.01
4. Y. T. Hui, D. C. Wang, Y. You, C. Y. Shao, C. S. Zhong, H. D. Wang, Effect of low temperature plasma treatment on biological characteristics and yield components of wheat seeds (Triticum aestivum L.), *Plasma Chem. Plasma Process.*, **40** (2020), 1555−1570. https://doi.org/10.1007/s11090-020-10104-z
5. H. Liu, Y. H. Zhang, H. Yin, W. X. Wang, X. M. Zhao, Y. G. Du, Alginate oligosaccharides enhanced triticum aestivum L. tolerance to drought stress, *Plant Physiol. Biochem.*, **62** (2013), 33−40. https://doi.org/10.1016/j.plaphy.2012.10.012
6. B. Šerá, P. r Špatenka, M. l Šerý, N. Vrchotová, I. a Hrušková, Influence of plasma treatment on wheat and oat germination and early growth, *IEEE Trans. Plasma Sci.*, **38** (2010), 2963−2968. https://doi.org/10.1109/TPS.2010.2060728
7. R. Thirumdas, A. Kothakota, U. Annapure, K. Siliveru, R. Blundell, R. Gatt, et al., Plasma activated water (PAW) Chemistry, physico-chemical properties, applications in food and agriculture, *Trends Food Sci. Technol*., **77** (2018), 21−31. https://doi.org/10.1016/j.tifs.2018.05.007
8. L. Tonks, Oscillations in ionized gases, *Plasma and Oscillations*, Elsevier, 1961, 122−139. https://doi.org/10.1016/B978-1-4831-9913-9.50014-5
9. B. Zhao, J. S. Feng, X. Wu, S. C. Yan, A survey on deep learning-based fine-grained object classification and semantic segmentation, *Int. J. Autom. Comput*., **14** (2017), 119−135. https://doi.org/10.1007/s11633-017-1053-3
10. A. Srivastava, E. Han, V. Kumar, V. Singh, Parallel formulations of decision-tree classification algorithms, *High Performance Data Mining*, Springer, Boston, 1999, 237−261. https://doi.org/10.1007/0-306-47011-X_2
11. G. D. Guo, H. Wang, D. Bell, Y. X. Bi, KNN model-based approach in classification, in *OTM Confederated International Conferences CoopIS, DOA, and ODBASE*, (2003), 986−996. https://doi.org/10.1007/978-3-540-39964-3_62
12. A. Tharwat, A. E. Hassanien, B. E. Elnaghi, A BA-based algorithm for parameter optimization of Support Vector Machine, *Pattern Recognit. Lett.*, **93** (2017), 13−22. https://doi.org/10.1016/j.patrec.2016.10.007
13. N. Coskun, T. Yildirim, The effects of training algorithms in MLP network on image classification, in *Proceedings of the International Joint Conference on Neural Networks*, (2003), 1223−1226.
14. J. Deng, J. Krause, F. F. Li, Fine-grained crowdsourcing for fine-grained recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2013), 580−587. https://doi.org/10.1109/CVPR.2013.81
15. E. Gavves, B. Fernando, C. G. Snoek, A. W. Smeulders, T. Tuytelaars, Fine-grained categorization by alignments, in *Proceedings of the IEEE International Conference on Computer Vision*, (2013), 1713−1720. https://doi.org/10.1109/ICCV.2013.215

16. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770−778. https://doi.org/10.1109/CVPR.2016.90
17. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 4700−4708. https://doi.org/10.1109/cvpr.2017.243
18. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *Proceedings of the 32nd International Conference on Machine Learning*, (2015), 448−456.
19. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, **60** (2017), 84-90. https://doi.org/10.1145/3065386
20. S. Jin, H. X. Yao, X. S. Sun, S. C. Zhou, L. Zhang, X. S. Hua, Deep saliency hashing for fine-grained retrieval, *IEEE Trans. Image Process.*, **29** (2020), 5336−5351. https://doi.org/10.1109/TIP.2020.2971105
21. Y. Jing, W. Wang, L. Wang, T. N. Tan, Learning aligned image-text representations using graph attentive relational net-work, *IEEE Trans. Image Process.*, **30** (2021), 1840−1852. https://doi.org/10.1109/TIP.2020.3048627
22. L. L. Zhang, J. Liu, M. N. Luo, X. J. Chang, Q. H. Zheng, Deep semisupervised zero-shot learning with maximum mean discrepancy, *Neural Comput.*, **30** (2018), 1426−1447. https://doi.org/10.1162/neco_a_01071
23. K. Liu, D. Liu, L. Li, N. Yan, H. Q. Li, Semantics-to-signal scalable image compression with learned revertible representations, *Int. J. Comput. Vis.*, **129** (2021), 2605−2621. https://doi.org/10.1007/s11263-021-01491-7
24. L. Qi, X. Q. Lu, X. L. Li, Exploiting spatial relation for fine-grained image classification, *Pattern Recognit.*, **91** (2019), 47−55. https://doi.org/10.1016/j.patcog.2019.02.007
25. L. Wang, K. He, X. Feng, X. T. Ma, Multilayer feature fusion with parallel convolutional block for fine-grained image classification, *Appl. Intell.*, **52** (2022), 2872−2883. https://doi.org/10.1007/s10489-021-02573-2
26. M. Srinivas, Y. Y. Lin, H. Y. M. Liao, Deep dictionary learning for fine-grained image classification, in *2017 IEEE International Conference on Image Processing*, (2017), 835−839. https://doi.org/10.1109/ICIP.2017.8296398
27. L. Liao, R. M. Hu, J. Xiao, Q. Wang, J. Xiao, J. Chen, Exploiting effects of parts in fine-grained categorization of vehicles, in *2015 IEEE international conference on image processing*, (2015), 745−749. https://doi.org/10.1109/ICIP.2015.7350898
28. K. Wang, M. Z. Liu, YOLOv3-MTis A YOLOv3 using multi-target tracking for vehicle visual detection, *Appl. Intell.*, **52** (2022), 2070−2091. https://doi.org/10.1007/s10489-021-02491-3
29. S. M. Pan, W. Q. Feng, Y. W. Chong, Attribute-guided global and part-level identity network for person re-identification, *Int. J. Pattern Recognit. Artif. Intell.*, **36** (2022), 2250011. https://doi.org/S0218001422500112
30. C. Wang, J. Y. Sun, S. W. Ma, Y. Q. Lu, W. Liu, Multi-stream network for human-object interaction detection, *Int. J. Pattern Recognit. Artif. Intell.*, **35** (2021), 2150025. https://doi.org/10.1142/S0218001421500257

31. Z. Q. Lin, S. M. Mu, F. Huang, K. A. Mateen, M. J. Wang, W. L. Gao, et al., A unified matrix-based convolutional neural network for fine-grained image classification of wheat leaf diseases, *IEEE Access*, **7** (2019), 11570−11590. https://doi.org/10.1109/ACCESS.2019.2891739
32. Z. Q. Lin, S. M. Mu, A. J. Shi, C. Pang, X. X. Sun, A novel method of maize leaf disease image identification based on a multichannel convolutional neural network, *Trans. ASABE*, **61** (2018), 1461−1474. https://doi.org/10.13031/trans.12440
33. H. Lu, Z. G. Cao, Y. Xiao, Z. W. Fang, Y. J. Zhu, Fine-grained maize cultivar identification using filter-specific convolutional activations, in *2016 IEEE International Conference on Image Processing*, (2016), 3718−3722. https://doi.org/10.1109/ICIP.2016.7533054
34. X. P. Zhang, H. K. Xiong, W. G. Zhou, Q. Tian, Fused one-vs-all features with semantic alignments for fine-grained visual categorization, *IEEE Trans. Image Process.*, **25** (2015), 878−892. https://doi.org/10.1109/TIP.2015.2509425
35. X. S. Wei, C. W. Xie, J. X. Wu, C. H. Shen, Mask-CNN is Localizing parts and selecting descriptors for fine-grained bird species categorization, *Pattern Recognit.*, **76** (2018), 704−714. https://doi.org/10.1016/j.patcog.2017.10.002
36. L. Qi, X. Q. Lu, X. L. Li, Exploiting spatial relation for fine-grained image classification, *Pattern Recognit.*, **91** (2019), 47−55. https://doi.org/10.1016/j.patcog.2019.02.007
37. Y. Zhang, X. S. Wei, J. X. Wu, J. F. Cai, J. B. Lu, V. A. Nguyen, et al., Weakly supervised fine-grained categorization with part-based image representation, *IEEE Trans. Image Process.*, **25** (2016), 1713−1725. https://doi.org/10.1109/TIP.2016.2531289
38. S. L. Huang, Z. Xu, D. C. Tao, Y. Zhang, Part-Stacked CNN for fine-grained visual categorization, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 1173−1182. https://doi.org/10.1109/CVPR.2016.132
39. S. H. Lee, C. S. Chan, S. J. Mayo, P. Remagnino, How deep learning extracts and learns leaf features for plant classification, *Pattern Recognit.*, **71** (2017), 1−13. https://doi.org/10.1016/j.patcog.2017.05.015
40. M. Rohrbach, A. Rohrbach, M. Regneri, S. Amin, M. Andriluka, M. Pinkal, et al., Recognizing fine-grained and composite activities using hand-centric features and script data, *Int. J. Comput. Vision.*, **119** (2016), 346−373. https://doi.org/10.1007/s11263-015-0851-8
41. S. Cai, W. Zuo, Z. Lei, Higher-order integration of hierarchical convolutional activations for fine-grained visual categorization, in *Proceedings of the IEEE International Conference on Computer Vision*, (2017), 511−520.
42. Q. Hu, H. Wang, T. Li, C. Shen, Deep CNNs with spatially weighted pooling for fine-grained car recognition, *IEEE. Intell Transp.*, **91** (2019), 47−55. https://doi.org/10.1016/j.patcog.2019.02.007
43. P. J. Burt, E. H. Adelson, *Readings in computer vision*, Elsevier, Piscataway, 1987, 671−679. https://doi.org/10.1016/B978-0-08-051581-6.50065-9
44. C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, *IEEE Trans. Pattern Anal. Mach. Intell.*, **35** (2012), 1915−1929. https://doi.org/10.1109/TPAMI.2012.231
45. B. Hariharan, P. Arbeláez, R. Girshick, J. Malik, Hyper columns for object segmentation and fine-grained localization, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2015), 447−456. https://doi.org/10.1109/CVPR.2015.7298642
46. J. Weber, J. Malik, Robust computation of optical flow in a multi-scale differential framework, *Int. J. Comput. Vis.*, **14** (1995), 67−81. https://doi.org/10.1007/BF01421489

47. H. L. Zheng, J. L. Fu, T. Mei, J.B . Luo, Learning multi-attention convolutional neural network for fine-grained image recognition, in *Proceedings of the IEEE international conference on computer vision*, (2017), 5209−5217. https://doi.org/10.1109/ICCV.2017.557
48. T. Y. Lin, A. RoyChowdhury, S. Maji, Bilinear CNN models for fine-grained visual recognition, in *Proceedings of the IEEE International Conference on Computer Vision*, (2015), 1449−1457. https://doi.org/10.1109/ICCV.2015.170
49. A. Fawzi, H. Samulowitz, D. Turaga, P. Frossard, Adaptive data augmentation for image classification, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2013), 580−587. https://doi.org/10.1109/ICIP.2016.7533048
50. R. Dellana, K. Roy, Data augmentation in CNN-based periocular authentication, in *2016 6th International Conference on Information Communication and Management*, (2016), 141−145. https://doi.org/10.1109/INFOCOMAN.2016.7784231
51. J. Johnson, A. Karpathy, F. F. Li, Densecap is Fully convolutional localization networks for dense captioning, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 4565−4574. https://doi.org/10.1109/CVPR.2016.494
52. N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, P. T. P. Tang, On large-batch training for deep learning is generalization gap and sharp minima, (2016). https://doi.org/10.48550/arXiv.1609.04836
53. H. Li, Z. Xu, G. Taylor, T. Goldstein, Visualizing the loss landscape of neural nets, in *32nd Conference on Neural Information Processing Systems*, **31** (2018).
54. P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, et al., Accurate, large minibatch SGD is Training ImageNet in 1 hour, (2017). https://doi.org/10.48550/arXiv.1706.02677
55. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *Proceedings of International Conference on Learning Representations*, (2015). https://doi.org/10.48550/arXiv.1409.1556

AIMS Press