A CNN-LSVM MODEL FOR IMBALANCED IMAGES IDENTIFICATION OF WHEAT LEAF

T. Su, *S. Mu*, *A. Shi*, *Z. Cao*, *M. Dong*

Abstract: In order to improve the accuracy of convolutional neural networks (CNN) in imbalanced dataset classification, a novel hierarchical CNN-LSVM is proposed. Considering the imbalance in the number and spatial distribution of wheat leaf disease images, the improved local support vector machine (LSVM) replaces Softmax as the classifier of the model, and meanwhile a cost sensitive matrix is designed to assign the value for penalty factors in the optimized objective function of LSVM. It effectively improves the sensitivity of misclassification caused by imbalanced data. To verify the validity and practicability of CNN-LSVM, 6028 wheat leaf disease images containing 8 species are collected from planting bases in Shandong Agricultural University. Then the imbalanced and balanced standard image sets are generated by data augmentation and Borderline-Synthetic Minority Oversampling (Borderline-SMOTE). They have 36168 and 46176 images, respectively. The experimental results show that the average identification accuracies of the CNN-LSVM obtained on imbalanced and balanced standard datasets are 90.32% and 93.68%, respectively. And it starts to converge when the iteration times are close to 13000. CNN-LSVM has higher classification accuracy and lower iteration times, compared with CNN-Softmax, CNN-SVM, LSVM and support vector machine (SVM).

Key words: *convolutional neural networks, local support vector machines, deep learning, wheat leaf diseases, image identification*

1. Introduction

In recent years, CNN has made great breakthrough in face recognition, gait tracking and other fields [1]. Its application field is expanding constantly, and it has preliminary application in crop diseases and insect pests identification. Seyed et

---

*Tingting Su, Zhihao Cao, Mengping Dong; Shandong Agricultural University, College of information science and Engineering, Taian, Shandong, 271018, China, E-mail: m15065806906@163.com, czh@sdau.edu.cn, dongmengping@126.com

†Shaomin Mu – Corresponding author; Shandong Agricultural University, College of information science and Engineering, Shandong agricultural big data engineering laboratory, Taian, Shandong, 271018, China, E-mail: msm@sdau.edu.cn

‡Aiju Shi; Shandong Agricultural University, College of chemistry and materials science, Taian, Shandong, 271018, China, E-mail: saj31402@163.com

al. [2] combined the features of 5 kinds of wheat leaf disease images by utilizing Gabor filter and visual features, and then classified them by using a neural network system; Srdjan et al. [3] designed a plant disease recognition system based on CNN, which can distinguish 13 kinds of various normal and diseased leaves in the presence of complex background; Mads et al. [4] took 10413 weed images containing 22 different species as experimental samples and adopted CNN to identify the category of each sample under different dynamic conditions

Although the above CNN-based crop disease identification has achieved good results, CNN is unable to deal with imbalanced datasets better [5, 6]. The distributions of imbalanced datasets are generally seen as long-tailed i.e., most of the samples are densely distributed in a few categories while others only have a limited samples [7, 8]. Class imbalance is usually encountered in crop disease image data, mainly for the following 2 reasons: Firstly, because of the different incidence of different diseases, the number of images of each disease varies greatly. Secondly, because of the different extent and symptoms of the same disease and the impact of lighting, irregular appearance, water mist, acquisition device jitter, the density for the feature vector varies greatly in the spatial distribution.

In view of the above problems, a CNN-LSVM is proposed for the imbalanced images identification of wheat leaf disease in this paper. It effectively solves the problem of sample misidentification caused by class imbalanced, and improves the identification accuracy. The remainder of this paper is organized as follows. In Section 2, the imbalance of wheat leaf disease images and the architecture and improvement of CNN-LSVM are illustrated. Then the construction of the imbalanced and balanced standard image databases of wheat leaf diseases are presented. Section 3 presents the experimental results and discussion. Finally, conclusions are drawn in Section 4.

2. Materials and methodology

2.1 The imbalance of wheat leaf diseases images

The imbalance of wheat leaf disease image dataset attributes to the subjective factors such as temperature, humidity as well as regional distribution and objective factors.It is not advisable to collect specific categories in practice.And the imbalance of wheat leaf disease images is reflected in 2 aspects: the number of samples and the spatial distribution of feature vectors.

(1) Due to the variety of wheat and the randomness of growth environment in planting bases of Shandong Province, the images of wheat leaf diseases show obvious difference in quantity.

(2) The complexity of the feature vector distribution is another factor that leads to the deterioration of classifier performance. Because of the differences of many pathogenic mechanisms and the influence of uneven illumination, shadow occlusion and equipment shake during the image acquisition, the pixel distribution and signal intensity of the wheat image vary greatly. The complexity of the spatial distribution is generalized. It concretely includes

overlap, lack of representative samples, and small degree of separation between classes. The above situations are shown in Fig. 1(a) The imparity of quantity and (b) disequilibrium of quantity and spatial distribution.
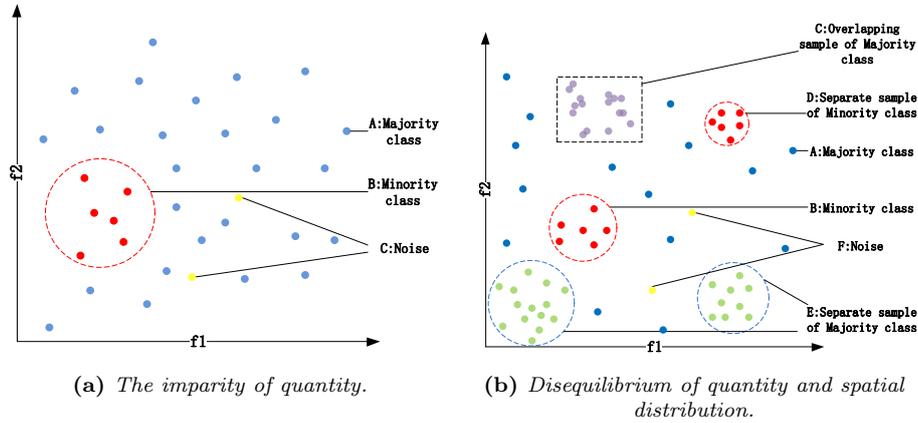

(a) *The imparity of quantity.*　　**(b)** *Disequilibrium of quantity and spatial distribution.*

Fig. 1 *Imbalance of wheat leaf disease images. In (a) the red circles represent the minority class, the blue ones represent the majority class, and the yellow ones represent the noise; besides, in (b) the purple circles represent the overlapping sample of majority class, while the green ones represent the separate sample of Majority class.*

The red and blue circles respectively represent the minority and the majority samples in Fig. 1(a) and 1(b), and both are imbalanced. There is only quantitative imbalance in Fig. 1(a). It is obvious that the number of blue samples is significantly higher than that of the red ones. The samples in Fig. 1(b) are not only imbalanced in quantity, but also in spatial distribution: there exist overlapping samples and separated items (sub clustering). Lacking of representative samples makes it difficult for them to be concerned and studied by CNN, which leads to the misleading of model growth.

2.2　The structure of CNN-LSVM

For the identification of imbalanced wheat leaf disease images, the CNN-LSVM is designed. Aiming at the imbalanced data problem, the optimization objective function of LSVM is improved, and the cost sensitive matrix is designed to define the values of the penalty factor in LSVM. The overall framework of CNN-LSVM is illustrated. It can be divided into 2 major phases: Feature Extraction and Classification in Fig. 2.

2.2.1　Feature extraction

The size of input images is (256×256) pixels. The batch size is 100 images. Feature Extraction are performed simultaneously on GPU_1 and GPU_2. The main body of
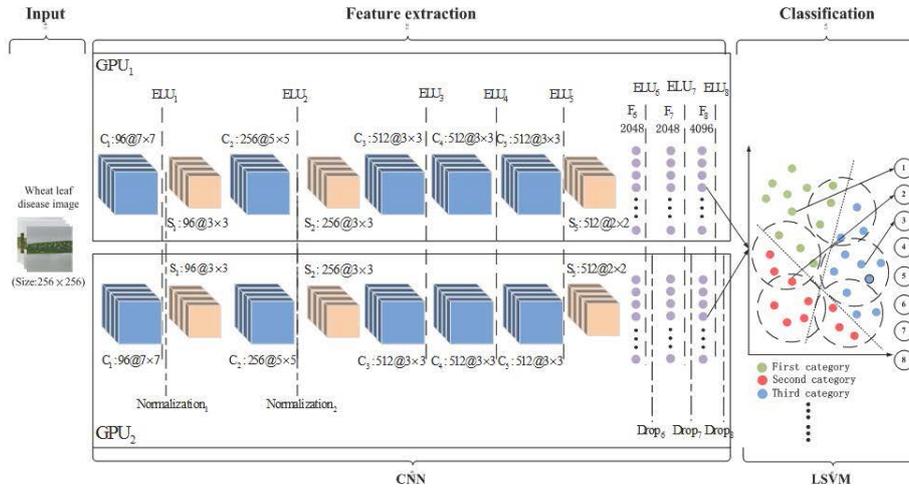
Fig. 2 *The structure of CNN-LSVM.*

Feature Extraction is CNN, which is responsible for feature extraction of wheat leaf disease images. It can be divided into 11 main hidden layers: 5 convolutional layers $(C_1, C_2, C_3, C_4, C_5)$, 3 subsampling layers (S_1, S_2, S_5) and 3 fully connected layers (F_6, F_7, F_8). In order to mimic the link sparsity of the biological nervous system and avoid the vanishing gradient as well as overfitting, ELU, normalized layer and Dropout are inserted among the main hidden layer as the function layers.

(1) Convolutional layer   The convolutional layer simulates the feature extraction process of primary cells with local receptive fields and weight sharing. It is not desirable that the convolutional network is either too deep or too shallow, and five convolution layers is adopted experientially. The convolutional layer structure is shown in Tab. I.

As can be seen from Tab. I, with the deeper of the convolutional layers, the kernel size becomes smaller while the kernel sum gradually increases from 100 to 1024. It aims at better extracting the fine-grained features in the deep level of the images. Stride is the moving step of the filter in horizontal and vertical directions. Padding is the number of 0 supplemented at the edge of the filter in the horizontal and vertical directions. It is designed to ensure that the width of the image is divisible by the size of the filter. Lr_weight and Lr_bias represent learning rates of weights and offsets, respectively, and they affect the rate of parameter optimization in the optimization process. Decay_weight and Decay_bias respectively represent the attenuation rates of weights and offsets, which can reduce overfitting. Momentum coefficient is equivalent to the damping term, it reduces the sensitivity of the network to the local details of the error surface, and effectively suppresses the local optimum, so as to reduce the learning oscillation trend and improve the convergence process of the model.

Name	C_1	C_2	C_3	C_4	C_5
Kernal_Size	7×7	5×5	3×3	3×3	3×3
Kernal_Sum	100	256	512	768	1024
Stride_Col	2	2	1	1	1
Stride_Row	2	2	1	1	1
Padding_Col	0	1	1	1	0
Padding_Row	0	1	1	1	0
Lr_weight	0	1	1	1	1
Lr_bias	0	2	2	2	2
Decay_weight	0.005	1.000	1.000	1.000	1.000
Decay_bias	0.005	0.005	0.005	0.005	0.005
Momentum coefficient	0.9	0.9	0.9	0.9	0.9
Output_Sum	100	256	512	768	1024

Tab. I *Architecture of convolutional layer.*

(2) Subsampling layer In order to avoid the loss of critical information and suppress overfitting during the progress of dimensionality reduction of inputted feature maps, the max pooling and mean pooling are applied to extract abundan features. The subsampling layer structure is shown in Tab. II. The filter sizes of

Name	S_1	S_2	S_5
Type	Mean pooling	Max pooling	Mean pooling
Kernal_Size	3×3	3×3	3×3
Kernal_Sum	2048	2048	4096
Stride_Col	1	1	1
Stride_Row	1	1	1
Padding_Col	0	0	0
Padding_Row	0	0	0
Lr_weight	1	5	5
Lr_bias	2	10	10
Momentum coefficient	0.9	0.9	0.9
Output_Sum	2048	2048	4096

Tab. II *Architecture of Subsampling layer.*

max pooling and mean pooling are equal to their step size. The input feature maps are firstly partitioned into non-overlapping matrices. Each matrix then outputs the maximum and average signal value, thus selecting representative features in the local receptive field, reducing the dimension of the intermediate expression layer and getting the feature after dimensionality reduction better transformation invariance.

(3) Fully connected layer The fully connected layer is used to reassemble the local features extracted by convolution layers and subsampling layers into a

complete graph by weight matrix, which realizes the rasterization of 2 dimensional feature maps. Due to the size 256×256 of input image is larger, 3 fully connected layers are constructed and the convolution kernels in per fully connected layer are increased for nonlinear mapping ability of the network. The fully connected layer structure is shown in Tab. III. In particular, LSVM is used as classifier to output identification results and other information at the back of F_8.

Name	F_6	F_7	F_8
Type	Inner-product	Inner-product	Inner-product
Kernal_Size	1×1	1×1	1×1
Stride_Col	3	3	1
Stride_Row	3	3	1
Padding_Col	0	0	0
Padding_Row	0	0	0
Momentum coefficient	0.9	0.9	0.9
Output_Sum	100	256	1024

Tab. III *Architecture of fully connected layer.*

(4) Exponential linear unit nonlinearity   Due to the interference of sunlight, fog, dust and other factors, the intensity range of pixel signal is very large. When the signal value is too large or too small, the gradient of the Sigmoid and the Tanh in the traditional CNN changes slowly and becomes closer to zero in the soft saturated region. It is easy to cause vanishing gradient. Recently, ReLU has attracted much attention for alleviating the above vanishing gradient. But with the increase of network depth, training samples and iteration, some samples will sink into hard saturation region, and "neuron death" will occur, which seriously affects the model convergence.

In order to solve the above problems, accelerate the training speed of the network, ensure the sparsity of neurons link and avoid overfitting, exponential linear unit (ELU) is used as the nonlinear activation function in this paper, and the initial value of α is set at 0.25 and then adjusted by self-learning. Compared with Sigmoid, Tanh and ReLU, its convergence speed is faster and identification accuracy is higher.

(5) Normalized layers   As noted above, owing to the influence of sunlight, water mist, dust and other factors, the range of the signal intensity in gathered images is extremely wide. For example, an image with both reflections and shadows has a larger range of pixels. If this kind of images are used as input directly, it will deteriorate model growth. The role of signals with large value range in the model learning will be too large, and the role of signals with small value range will be too small. Moreover, the range of the function domain is limited, and the input data need to be mapped into this domain.

To solve the above problems, the local response normalization is used in the 2 Normalized layers Normalization$_1$ and Normalization$_2$. LRN generates horizontal

local inhibition. It enhances the generalization of CNN by creating the competition among neurons in a local region [9]. As a branch of LRN, the channel internal normalization is utilized and its scope is extended in the independent channel. The received signal is normalized as Eq. (1).

$$y_{p,q}^{(i)} = x_{p,q}^{(i)}/(k + (\alpha/n) \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (x_{p,q}^{(j)})^2)^\beta,$$ (1)

where $x_{p,q}^{(i)}$ presents the output of the signal corresponding to the p, q position; and α, β respectively denote the scaling factor and exponential term; and n represents the local size of the normalized range. The variables α, β and n are initialized to 0.0001, 0.75 and 5 respectively, following [1].

2.2.2 Classification

LSVM can make full advantage of the local information of the samples and utilize the adjacent samples to participate in the classification, which effectively reduce the complexity of the model. Hence LSVM can make full use of local information to complete classification task. Therefore, LSVM is the main body in classification, its task is to identify wheat leaf disease images according to the image features extracted by CNN. $\sigma(\bar{x}, x)$ is introduced to measure the similarity between test samples and training samples, and the optimized objective function of LSVM is seen in Eq. (2).

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C \sum_{\{i=1|y_i=+1\}}^{n} \sigma(\bar{x}, x)\xi_i$$
$$s.t.y_i((w \cdot \phi(x_i) + b) \geq 1 - \xi_i$$ (2)
$$\xi_i \geq 0, i = 1, 2, \dots, n$$
$$\xi \in [0, 1],$$

where $\mathbf{w}$ is the normal vector of optimal classification hyperplane, C is a penalty parameter used to measure penalties for misclassification of samples. ξ_i is a slack variable that indicates the non conformity extent of the constraint $y_i((w \cdot \phi(x_i)+b) \geq 1 - \xi_i$. ξ_i is the vector of x_i in feature space.

In order to improve identification accuracy of LSVM on imbalanced datasets, in this paper, multiple penalty factors $\sum_{m=1}^{8} \sum_{n=1}^{8} C(m, n)$ are introduced in the cost function to improve the sensitivity of LSVM to the misclassification of imbalanced datasets Eq. (3).

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C \sum_{i=1}^{n} \sigma(\bar{x}, x)\xi_i + \sum_{m=1}^{8} \sum_{n=1}^{8} C(m, n)\xi_i$$
$$s.t.y_i((w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, m \neq n$$ (3)
$$m = 1, 2, \dots, n$$
$$n = 1, 2, \dots, n.$$

To solve the above 2 programming problems, Lagrange multipliers α_i and β_i are introduced to construct Lagrange functions shown in Eq. (4).

$$L(w, b, \alpha, \beta) = \min_{w,b,\xi} \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} + C \sum_{i=1}^{n} \sigma(\bar{x}, x)\xi_i + \sum_{m=1}^{8} \sum_{n=1}^{8} C(m, n)\xi_i \qquad (4)$$

$$- \sum_{i=1}^{y} \alpha_i(y_i((w \cdot \phi(x_i) + b) - 1 + \xi_i)) - \sum_{i=1}^{N} \beta_i \xi_i.$$

The dual problem, equivalent to the Eq. (3), can be obtained by using formula Eq. (4), that is, the following 2 convex programming problems Eq. (5).

$$\min \frac{1}{2} \sum_{m=1}^{8} \sum_{n=1}^{8} \alpha_m \alpha_n y_m y_n - \sum_{n=1}^{N} \alpha_i$$

$$s.t. y_i \alpha_i = 0$$

$$0 \leq \alpha_i \leq C(m, n), m \neq n \qquad (5)$$

$$m = 1, 2, \ldots, 8$$

$$n = 1, 2, \ldots, 8.$$

The optimal solution vector $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_i^*)$ of the above optimization problem can be found, α_i and β_i then the decision function $f(x)$ of LSVM can be expressed as Eq. (6).

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \alpha_i^* K * (\bar{x}, x) + b\right). \qquad (6)$$

The penalty factors can be defined through the cost sensitive matrix of Tab. IV. In Tab. IV, the horizontal items represent the correct identification and the vertical items represent the actual identification. $C(m, n)$ represents a penalty factor when

P / T	1	2	3	4	5	6	7	8
1	0	$C(1,2)$	$C(1,3)$	$C(1,4)$	$C(1,5)$	$C(1,6)$	$C(1,7)$	$C(1,8)$
2	$C(2,1)$	0	$C(2,3)$	$C(2,4)$	$C(2,5)$	$C(2,6)$	$C(2,7)$	$C(2,8)$
3	$C(3,1)$	$C(3,2)$	0	$C(3,4)$	$C(3,5)$	$C(3,6)$	$C(3,7)$	$C(3,8)$
4	$C(4,1)$	$C(4,2)$	$C(4,3)$	0	$C(4,5)$	$C(4,6)$	$C(4,7)$	$C(4,8)$
5	$C(5,1)$	$C(5,2)$	$C(5,3)$	$C(5,4)$	0	$C(5,6)$	$C(5,7)$	$C(5,8)$
6	$C(6,1)$	$C(6,2)$	$C(6,3)$	$C(6,4)$	$C(6,5)$	0	$C(6,7)$	$C(6,8)$
7	$C(7,1)$	$C(7,2)$	$C(7,3)$	$C(7,4)$	$C(7,5)$	$C(7,6)$	0	$C(7,8)$
8	$C(8,1)$	$C(8,2)$	$C(8,3)$	$C(8,4)$	$C(8,5)$	$C(8,6)$	$C(8,7)$	0

Tab. IV *Cost sensitive matrix. T stands True, and P stands Predict. Note: 1 – Normal leaf; 2 – Mechanical damage leaf; 3 – Powdery mildew; 4 – Bacterial leaf streak; 5 – Cochliobolusheterostrophus; 6 – Stripe rust; 7 – Leaf rust; 8 – Bacterial leaf blight.*

a sample of real class j is wrongly divided into i class. Akbani et al. [10] proposed a method that uses the numbers of two sets of samples to set penalty factor for each other. That is: $C_+ = \text{Maj\_num}$, $C_- = \text{Min\_num}$. Maj_num and Min_num represent the number of minority class samples and majority class samples, respectively. However, the way to define the penalty parameters only by the number of samples has some limitations. It is due to that the imbalance of data is reflected not only in the difference of the samples number, but also in the complexity of the samples distribution in feature space. The above assignment method takes into account the previous condition only, without regard to the latter. Therefore, a new penalty factor assignment method is proposed in this paper, which takes into account both the imbalance of the number and the spatial distribution of wheat leaf disease images.

Fig. 3 shows the concepts of the hyper sphere, the sample center and the mean distance between centers of the two classes of samples in Algorithm 1.

Algorithm 1 Framework of the penalty factor assignment.

Require:

Minority sample $\{\text{Min}_i | 1 \leq i \leq \text{Min\_num}\}$;

Majority sample $\{\text{Maj}_i | 1 \leq i \leq \text{Maj\_num}\}$

1: The minority samples Min_i and the majority samples Maj_i are clustered respectively to find the central points $\bar{x}_{\text{Min}}$ and $\bar{x}_{\text{Maj}}$

$$\bar{x}_{\text{Min}} = \frac{\sum_{i=1}^{\text{Min\_num}} \text{Min}_i}{\text{Min\_num}},$$

$$\bar{x}_{\text{Maj}} = \frac{\sum_{i=1}^{\text{Maj\_num}} \text{Maj}_i}{\text{Maj\_num}},$$

where Min_numandMaj_num are the samples number of the two kinds of samples respectively.

2: The sums of Euclidean distances $\text{sum}_{i=1}^{\text{Min\_num}} \| x_{\text{Min}} - \bar{x}_{\text{Min}} \|^2$ and $\text{sum}_{i=1}^{\text{Min\_num}} \| x_{\text{Min}} - \bar{x}_{\text{Min}} \|^2$ between the samples and the sample centers of two categories are obtained, Min_numandMaj_num by dividing the number of samples per class, and the mean distances Min_AvgandMaj_Avg of two classes of samples are obtained.

$$\text{Maj\_avg} = \frac{\sum_{i=1}^{\text{Maj\_num}} \| x_{\text{Maj}} - \bar{x}_{\text{Maj}} \|^2}{\text{Maj\_num}},$$

$$\text{Min\_avg} = \frac{\sum_{i=1}^{\text{Min\_num}} \| x_{\text{Min}} - \bar{x}_{\text{Min}} \|^2}{\text{Min\_num}},$$

The mean distances approximate the radius of a hypersphere in a high-dimensional space containing two classes of samples;

3: Finally, the ratio of the samples number to the mean of the center distance is used to assign values to penalty factors:

$$C_+ = \frac{\text{Min\_num}}{\text{Min\_avg}},$$

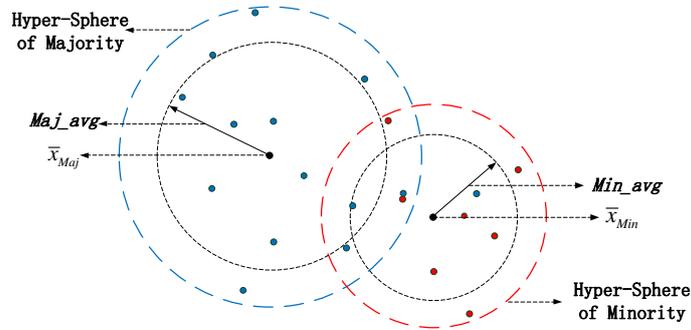$$C_+ = \frac{\text{Maj\_num}}{\text{Maj\_avg}}.$$

Fig. 3 *The hyper spheres, sample centers and mean distance between centers of two types of samples. Blue circle – Hyper-Sphere of Majority; red circle – Hyper-Sphere of Minority.*

2.3 Data description

As there are no public large-scale image sets of wheat leaf disease, therefore, high-fidelity images are collected from wheat planting bases in our experiments. Then the original images are processed by data augmentation and Borderline-SMOTE to respectively construct the imbalanced and balanced standard image sets, which improves the size and hierarchy of samples.

2.3.1 Image acquisition

From the wheat planting bases of Shandong Province, 6028 images containing 6 common wheat leaf diseases, normal leaves and mechanical damage leaves were collected, taken with a Canon EOS700D 18-135. The image format is JPEG with a resolution of 6088×3166, and each image is a 24-bit color bitmap. The samples, numbers and proportions of each disease are shown in Fig. 4 and Tab. V.

Name	Normal leaf	Mechanical damage leaf	Powdery mildew	Bacterial leaf streak	Cochliobolus hetero-strophus	Stripe rust	Leaf rust	Bacterial leaf blight
Number	1161	1037	482	362	1146	331	1061	493
Prop. [%]	18.5	17.2	8.0	6.0	19.0	5.5	17.6	8.2

Tab. V *Number and proportion of each wheat disease image in original image set.*

(a) *Normal leaf (a)*   **(b)** *Mechanical damage leaf (a)*   **(c)** *Powdery mildew (a)*   **(d)** *Bacterial leaf streak (a)*

(e) *Normal leaf (b)*   **(f)** *Mechanical damage leaf (b)*   **(g)** *Powdery mildew (b)*   **(h)** *Bacterial leaf streak (b)*

(i) *Cochliobolusheterostrophus (a)*   **(j)** *Stripe rust (a)*   **(k)** *Leaf rust (a)*   **(l)** *Bacterial leaf blight (a)*

(m) *Cochliobolusheterostrophus (b)*   **(n)** *Stripe rust (b)*   **(o)** *Leaf rust (b)*   **(p)** *Bacterial leaf blight (b)*
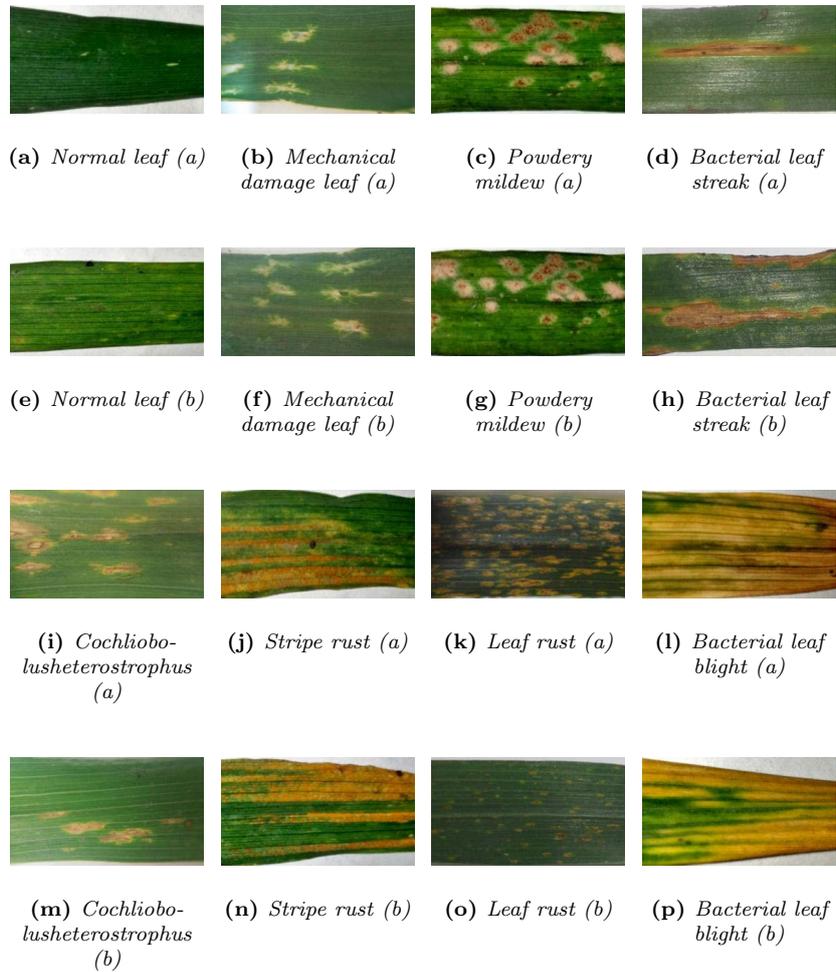
Fig. 4 *Samples of wheat disease images.*

2.3.2   Data augmentation and oversampling

CNN self-learning relies on iterative training on a large-scale image set. However, if the amount of data is too small, it is prone to overfitting, which makes the training error very small and while that of testing very large [11]. In order to increase the size and diversity of wheat leaf disease image set, 5 ways in Tab. VIII are adopted to implement image set augmentation.

On the one hand, the achievements of the model can be further improved by data enhancement. The generalization ability of the model can be improved by increasing the number of images via rotation and translation and so on, while the robustness of the model can be further improved by images added noise [12,13]. On the other hand, it expands the range of data sets, which makes the features mined

by the network own more invariance to operations such as rotation and scaling. In addition, it enlarges the gap between the majority class and the minority class in quantity, making the image imbalance more obvious and thus better test the actual performance of CNN-LSVM in dealing with imbalanced image set.

Image augmentation enabled every wheat leaf disease images to produce six corresponding enhanced images respectively, and finally the imbalanced standard image set (ISIS) containing 36168 images is obtained. The number and proportion of each kind of images in ISIS are shown in Tab. VI. Later, CNN-LSVM and other algorithms are used to classify the imbalanced and balanced wheat leaf disease image sets to test its actual performance in different environments. In this paper, the above ISIS are oversampled by Borderline-SMOTE to generate balanced standard image set $BSIS$. The balanced standard image set $BSIS$ contains 46176 images.
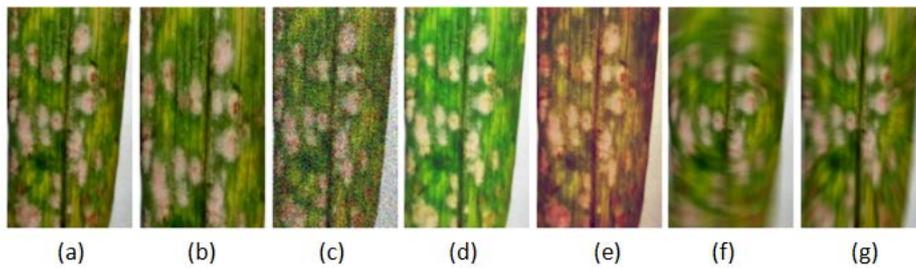

Fig. 5 *Images processed by data augmentation. (a) – Original image; (b) – Random crop; (c) – Noise addition; (d) – Color Jittering; (e) – PCA Jittering; (f) – Rotation blur; (g) – Scaling blur.*

Name	Normal leaf	Mechanical damage leaf	Powdery mildew	Bacterial leaf streak	Cochliobolus hetero-strophus	Stripe rust	Leaf rust	Bacterial leaf blight
Number	6696	6222	2892	2172	6876	1986	6366	2958
Prop. [%]	18.5	17.2	8.0	6.0	19.0	5.5	17.6	8.2

Tab. VI *Number and proportion of each kind of images in ISIS.*

Name	Normal leaf	Mechanical damage leaf	Powdery mildew	Bacterial leaf streak	Cochliobolus hetero-strophus	Stripe rust	Leaf rust	Bacterial leaf blight
Number	6696	6222	5784	4344	6876	3972	6366	5916
Prop. [%]	14.5	13.5	12.5	9.4	14.9	8.6	13.8	12.8

Tab. VII *Number and proportion of each kind of images in BSIS.*

Name	Detail operations
Random crop	Randomly select Crop Center on the original image of the (256×256) size. Then use the (224×224) size window to capture it. Finally extend the screenshot to (256×256) so as to unify the image input size.
Noise addition	A 30 % Gauss noise is added to the original image, with an offset of 0.2 and a standard deviation of 0.3.
Color jittering	The original image hue, saturation and brightness are increased by 20 %, the contrast is increased by 30 %, the sharpness is decreased by 10 %.
PCA jittering	Firstly, the mean and standard deviations of 3 color channels of R, G and B are calculated, and the input data of the network is normalized. Then, the dimension is reduced through PCA in the RGB color space of the training set pixel, and the 3 principal direction vectors p_1, p_2, p_3 and 3 features $\lambda_1, \lambda_2, \lambda_3$ of the RGB space are obtained. Then add $[p_1, p_2, p_3][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^{\mathrm{T}}$ to each pixel $I_{xy} = [I_{Rxy}, I_{Gxy}, I_{Bxy}]^{\mathrm{T}}$ of the image. α_1 is a random variable satisfying the mean of 0 and the difference of 0.1.
Radial blur	The original image is radially blurred by rotation blur and scaling blur respectively. The rotating fuzzy unit is 10, and the scaling fuzzy unit is 30. They are used to simulate the effects of quickly rotating or moving cameras to achieve radiation likefunctions.

Tab. VIII *Data augmentation.*

3.   Results and discussion

3.1   Experimental environment

Training a deep CNN on the large scale images through a large number of iterations requires the use of high performance GPUs. The model of the GPUs used in our experiment is NVIDIA GeForce GTX 1080.

Besides, the computer model is Lenovo Idea Centre Y900, the processor is Intel Core i7 6700K $4.0GHz/L38M$, and the memory size is 16GB. Furthermore, the operating system is Ubuntu 14.04.4 64 bits, and the Caffe, an open source convolutional architecture for fast feature embedding exploited by the Berkeley Vision and Learning Center $BVLC$, is applied to implement the model. The Python is utilized as the programming language to adapt to the core of Caffe.

3.2   Experimental results and analysis

3.2.1   Experiment A: CNN-LSVM, CNN-Softmax and CNN-SVM

In order to validate the performance of CNN-LSVM, it is compared with CNN-Softmax and CNN-SVM. The 3 algorithms perform 180000 iterative training on

ISIS and BSIS respectively. Then the training models are used to identify the test sets. Additionally the feature visualization of some filters in CNN-LSVM are shown in Fig. 6. And there latively obvious dynamic change process of identification accuracies are shown in Fig. 7 and 8.
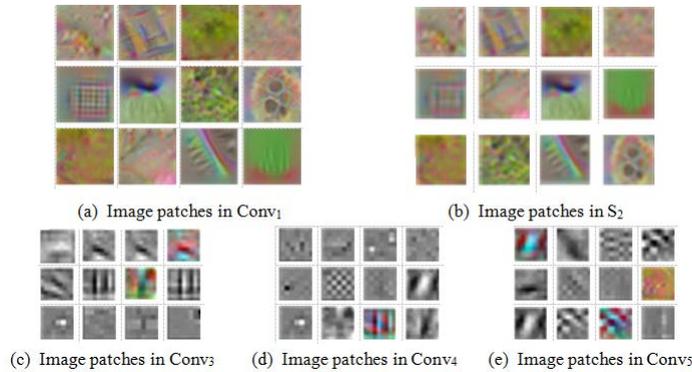

(a) Image patches in $Conv_1$ (b) Image patches in S_2

(c) Image patches in $Conv_3$ (d) Image patches in $Conv_4$ (e) Image patches in $Conv_5$

Fig. 6 *The filters of some main layers in CNN-LSVM visualized as small images patches.*
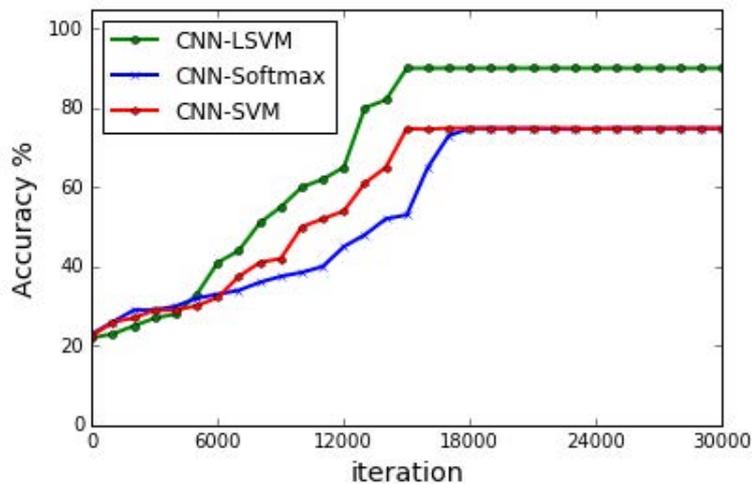
Fig. 7 *Test accuracy of CNN-LSVM, CNN-Softmax and CNN-SVM in ISIS.*

Fig. 7 indicates that when dealing with imbalanced datasets, the average identification accuracy of the common CNN-Softmax and CNN-SVM models is around 74.87 %, which is significantly lower than the average identification accuracy of CNN-LSVM 90.32 %. In addition, the CNN-LSVM began to converge when the iterations reached about 13000 times. But CNN-Softmax and CNN-SVM needs more than 15000 iterations to achieve stable identification accuracy. The experimental results and process show that CNN-LSVM has higher identification accuracy and learning rate for imbalanced image set.
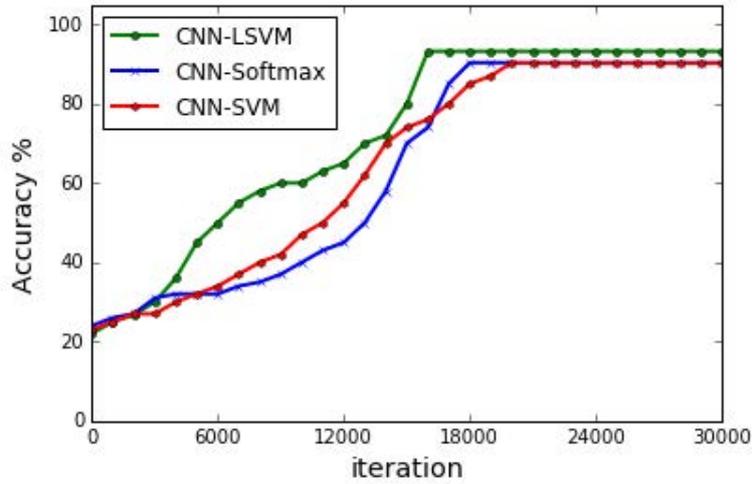
Fig. 8 *Test accuracy of CNN-LSVM, CNN-Softmax and CNN-SVM in BSIS.*

Fig. 8 indicates that the identification accuracy of CNN-LSVM is higher than that of CNN-Softmax and CNN-SVM when dealing with a balanced image set, but the gap is not obvious. In terms of converge rate, CNN-LSVM still owns a faster learning rate than CNN-Softmax and CNN-SVM.

3.2.2 Experiment B: CNN-LSVM, LSVM and SVM

In order to validate the advantages of CNN in automatic extraction of image features and the effect of LSVM to alleviate the imbalanced data problem, CNN-LSVM is compared with LSVM and SVM. To get the training model, the 3 algorithms perform 100000 iterative training on ISIS and BSIS respectively. The processing objects of KNN and SVM are the 49-dimensional feature vectors extracted from the wheat leaf disease images. In addition, the radial basis function RBF is adopted as the kernel function in SVM. The penalty parameter C, gamma, and slack variable are initialized to 10, 0.02 and 0.001, respectively. The concrete feature extraction information and identification accuracies of each compared algorithm on ISIS and BSIS are respectively shown in Tab. IX and X.

	CNN-LSVM [%]	LSVM [%]	SVM [%]
ISIS	90.32	71.25	54.37
BSIS	93.68	78.25	73.75

Tab. IX *Identification results of CNN-LSVMHL, LSVM and SVM on ISIS and BSIS.*

As can be seen from Tab. IX, the identification accuracy of CNN-LSVM is significantly higher than those of LSVM and SVM in the identification experiments of ISIS and BSIS. This is because CNN can automatically extract and combine the

Feature Category	Parameter	Dimension Number	Dimension Total
Feature Category	H, S and V of the first moment in HSV color space.	3	6
	H', S' and V' of the second moment in HSV color space.	3	
Texture features	Energy (E), entropy (H), inertia moment (I) and Relevance (C) of Gray Level Co-occurrence Matrix on the 0 degree in RGB and HSV color space.	8	32
	Energy (E), entropy (H), inertiamoment (I) and Relevance (C) of Gray Level Co-occurrence Matrix on the 45 degreein RGB and HSV color space.	8	
	Energy (E), entropy (H), inertia moment(I) and Relevance (C) of Gray Level Co-occurrence Matrix on the 90degree in RGB and HSV color space.	8	
	Energy (E), entropy (H), inertia moment (I) and Relevance (C) of GrayLevel Co-occurrence Matrix on the 135 degree in RGB and HSVcolor space.	8	
Texture features	Hu invariant moments: m(1),m(2),m(3),m(4),m(5),m(6),m(7).	7	11
	Measure of disease spot. (S)	1	
	Perimeter of disease spot. (L)	1	
	Roundness of disease spot. (C)	1	

Tab. X *Feature extraction of wheat leaf disease images.*

abstract features in the deep level of images. They are not only abundant in number, but also better reveal the true distribution of pixels. The number of artificially extracted features is limited and they are usually located in the shallow level of the image and thus the representation ability is not strong, which leads to the lower identification accuracy. The identification accuracy of LSVM in identification of ISIS is also obviously higher than that of SVM, which is due to that the penalty factor and the assignment mechanism are added to LSVM, the algorithm becomes more sensitive to the number, proportion and spatial distribution of imbalanced samples. The penalty factors can correct the trend of the model in time when there is an error caused by the imbalance phenomenon. Although the accuracy of LSVM in identification of BSIS is higher than that of SVM, the superiority is not significant. This is because BSIS does not have the imbalance problem and the 2 algorithms are equal in this respect, but LSVM can make better use of the local information of samples to obtain higher identification accuracy.

4. Conclusions

In order to improve the low identification accuracy of imbalanced wheat leaf diseases images by CNN, LSVM is introduced as classifier to generate CNN-LSVM in this paper. By adding the penalty factors in the optimization objective function corresponding to the imbalance of the data set, the model is more sensitive

to the misidentification caused by imbalance. Penalty factors are assigned by cost sensitive matrix. The cost sensitive matrix comprehensively takes into account the imbalance between the number and spatial distribution of wheat leaf disease images.

The experimental results show that CNN-LSVM has higher classification accuracy and faster model convergence rate when identifying imbalanced and balanced wheat leaf disease images. Especially in the classification experiments of imbalanced datasets, the improvement of accuracy is more obvious.

Acknowledgement

References

[1] KRIZHEVSKY A., SUTSKEVER I., HINTON G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, pp. 1096–1105, doi: 10.1016/j.protcy.2014.09.007.

[2] MOUSAVI S.A., HANIFELOO Z., SUMARI P., et al. Enhancing the diagnosis of corn pests using gabor wavelet features and SVM classification, 2016.

[3] SLADOJEVIC S., ARSENOVIC M., ANDERLA A., et al. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence & Neuroscience*. 2016, 2016(6), pp. 1–11, doi: 10.1155%2F2016%2F3289801.

[4] DYRMANN M., KARSTOFT H., MIDTIBY H.S. Plant species classification using deep convolutional neural network.*Biosystems Engineering*. 2016, 151, pp. 72–80, doi: 10.1016/j.biosystemseng.2016.08.024.

[5] RAJ V., MAGG S., WERMTER S. Towards effective classification of imbalanced data with convolutional neural networks. *IAPR Workshop on Artificial Neural Networks in Pattern Recognition. Springer*, Cham, 2016, pp. 150–162.

[6] AL-HADDAD L., MORRIS C.W., Boddy L. Training radial basis function neural networks: Effects of training set size and imbalanced training sets. *Journal of Microbiological Methods*, 2000, 43(1), pp. 33–44, doi: 10.1016/S0167-7012(00)00202-5.

[7] BOUVRIE J. Notes on Convolutional Neural Networks. *MIT CBCL Tech Report*, Cambridge, MA. 2006, doi: 10.1016/j.protcy.2014.09.007.

[8] RADIVOJAC P., CHAWLA N.V., DUNKER A.K., OBRADOVIC Z. Classification and Knowledge Discovery in Protein Databases. *Journal of Biomedical Informatics*. 2004, 37, pp. 224–239, doi: 10.1016/j.jbi.2004.07.008.

[9] DAHL G.E., SAINATH T.N., HINTON G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2013, pp. 8609–8613, doi: 10.1109/ICASSP.2013.6639346.

[10] AKBANI R., KWEK S., JAPKOWICZ N. Applying support vector machines to imbalanced datasets. *European conference on machine learning*, Springer, Berlin, Heidelberg, 2004, pp. 39–50, doi: 10.1007/978-3-540-30115-8_7.

[11] HU J., LU J., TAN Y.P. Deep transfer metric learning. *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2015, pp. 325–333, doi: 10.1109/CVPR.2015.7298629.

[12] WANG S.H., LV Y.D., SUI Y. et al. Alcoholism Detection by Data Augmentation and Convolutional Neural Network with Stochastic Pooling. *Journal of Medical Systems*, 2018, pp. 1–11, doi: 10.1007/s10916-017-0845-x.

[13] ZHANG Y.D., DONG Z., CHEN X., et al. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimedia Tools and Applications*, 2017, doi: 10.1007/s11042-017-5243-3.