IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 17, 2021, accepted August 28, 2021, date of publication August 30, 2021, date of current version September 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3109120

Research on Spider Sex Recognition From Images Based on Deep Learning

QIANJUN CHEN[1,2], YONGCHANG DING[2], CHANG LIU[2], JIE LIU[2,3], AND TINGTING HE[1,4,5,6]

[1]National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, Hubei 430079, China
[2]State Key Laboratory of Biocatalysis and Enzyme Engineering of China, College of Life Sciences, Hubei University, Wuhan, Hubei 430062, China
[3]Faculty of Resources and Environmental Sciences, Hubei University, Wuhan, Hubei 430062, China
[4]Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, Wuhan, Hubei 430079, China
[5]School of Computer, Central China Normal University, Wuhan, Hubei 430079, China
[6]National Language Resources Monitoring and Research Center for Network Media, Central China Normal University, Wuhan, Hubei 430079, China

Corresponding authors: Jie Liu (sparassidae@aliyun.com) and Tingting He (tthe@mail.ccnu.edu.cn)

ABSTRACT The rapid and accurate identification of spider sex is the first step in spider image recognition. The traditional artificial method used to identify the sex of mature spiders is mainly based on their genital structures (male palps or female epigynum) and highly dependent on the professional background of the identifiers. This article uses computer-based deep learning and transfer learning to identify the sex of spider, explores the design and application of convolutional neural networks in deep learning for spider sex recognition from images, and establishes a neural network model that displays excellent performance in experiments. In addition to optimizing the network model, we select appropriate hyperparameters to improve the accuracy of recognition and reduce the influence of human factors in the identification process. Through a comparison of multiple sets of experiments based on existing sample data collected in the laboratory, we find that the transfer learning method based on Xception can obtain better prediction accuracy than ResNet-152. After data augmentation, the optimization of a combined activation function and the fine-tuning of frozen layers, the prediction accuracy reaches 98.02%, and for an actual measurement of independent samples, the recognition accuracy reaches 92.38%. Therefore, the proposed method can basically replace manual identification and provide a reference for the artificial intelligence-based identification of spider species. Additionally, the model results indicate that male and female dimorphism may exist beyond the non-genital characteristics of spiders.

INDEX TERMS Deep learning, transfer learning, convolutional neural network, spider sex identification.

I. INTRODUCTION

There are nearly 50,000 species of spiders in the world and more than 5,000 species of spiders in China [1]. The identification of spider species mainly depends on the characteristics of the genitals (male palps or female epigynum); this process is time consuming, laborious, subjective, and highly dependent on spider classification experience. With the rapid development of the microbiology field, the importance of traditional taxonomy has been highly challenged. For example, there are few permanent taxonomy positions, and funding for taxonomy research is limited [2]. However, while traditional taxonomy faces many challenges, the rapid development of modern technology has led to new opportunities for developing new taxonomic methods, such as deep learning [3]; such methods have provided a new basis for spider taxonomy assessment [4]. In particular, the rapid development of deep learning in the past ten years has led to its use in image recognition in various fields [5], but research on spider image recognition has not yet been reported. The first step in spider species classification is identifying spider sex from images; this result directly determines the accuracy of subsequent species recognition tasks. Traditional methods for the artificial identification of the sex of mature spiders are mainly based on the genital structures (male palps or female epigynum) of the spiders [6]. In addition, the body color, pattern and shape is sexually dimorphic in some spider groups [7]. With 142 species, 2000 of which are endemic to East Asia, South Asia and Southeast Asia, *Pseudopoda* Jäger is the third largest genus in the family Sparassidae (World Spider Catalog 2021) [8]. These spiders are highly diversified in China, where 63 species have been reported. However, according to the results of laboratory investigations, there are at least 110 species in China. As a result, *Pseudopoda* is an ideal candidate for studying spider image recognition. In addition, although *Pseudopoda* spiders are typically

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

nocturnal, almost all species have similar body color (yellow generally) and spot patterns (fovea and radial furrows distinctly marked). It is difficult to manually distinguish males and females or identify species based on body coloration and spot patterns. However, multiple studies have found that the body color and spot patterns of spiders play important roles in attracting the opposite sex and increasing the success rate of courtship [9], [10]. Do body color and spot patterns of spiders of the genus *Pseudopoda* have similar functions, and can these features be identified by artificial intelligence? To date, no studies have addressed this topic.

In recent years, with the continuous improvement of deep learning technology, convolutional neural network (CNN) models have made considerable progress in the field of image recognition [11]–[14]. Each network has distinct characteristics, the recognition accuracy of networks has been continuously improved [15], and applications have been continuously refined in various fields, such as face recognition [16], medicine [17], [18], agriculture [19], [20] and others, with good results. This approach provides ideas and a research basis to study spider sex recognition. Since Alex Krizhevsky released AlexNet in 2012, many types of deep learning networks have been invented, such as VGGNet, GoogLeNet, Inception, ResNet, etc. The abstract reasoning ability of these networks has been continuously improved [21]. Additionally, computing frameworks are becoming increasingly mature. The currently popular computing frameworks include TensorFlow, Caffe, Theano, MXNet, Torch, and PyTorch [22], [23]. Among them, TensorFlow performs model training and testing based on large standardized data sets such as ImageNet; the prediction accuracy of TensorFlow is very high, and its generalization ability is very strong. According to the information on the official ImageNet website, the number of image samples in the dataset has reached 14197122, spanning 21841 categories; these images are manually labeled, thus providing sufficient samples to support learning and training in various models [24], [25]. However, the resolution of the images included in the standard database is low. In the process of machine learning, a model can only learn primary features from the training set, such as the outline or texture of an object; therefore, the standard set is not specific enough to meet the requirements in certain research areas. In terms of spider sex recognition, it is relatively easy to determine the shape of a spider's genitals, but features such as the back pattern are difficult to distinguish. Therefore, a learning model that can extract primary features while also learning minor features related to the target object is needed.

Transfer learning involves applying models trained with large datasets from source fields to data from target fields [26]. This approach is important for small-sample machine learning and can effectively alleviate the various problems caused by small sample sizes, such as overfitting and weak generalization ability, among others. Transfer learning in deep learning is widely used in small-sample learning, and the results are typically good [19], [27]–[30]. For example, Issam Dagher and Dany Barbara used networks such as VGG, ResNet, and Inception for transfer learning to solve problems related to face age estimation [31]. Li Miao and Wang Jingxian *et al.* applied a transfer learning method for crop disease recognition [32]. Ashraf Darwish and Dalia Ezzat *et al.* used transfer learning and integrated learning to identify corn disease problems [33].

In this paper, we performed spider sex image recognition based on 42 Chinese *Pseudopoda* species and a transferring learning method. We mainly addressed two questions: 1. Can image-based sex recognition be achieved for *Pseudopoda* spiders, and if so, how can the recognition accuracy be improved for small sample sets? 2. Do the non-genital features of *Pseudopoda* spiders, such as the body color and spot patterns, display sexual dimorphic trends and play important roles in image-based sex recognition?

II. SAMPLE MATERIALS, METHODS AND PROCEDURES

A. SPIDER MICRO GRAPHICS DATA SET

The spider samples used in this study were stored at the Center for Behavioural Ecology and Evolution (CBEE; College of Life Sciences, Hubei University, Wuhan, China). These samples contain 3,133 habitus photos for 30 *Pseudopoda* species (Table 1). We randomly took photos in dorsal view, ventral view or both views for each spider. All photos were taken with a Leica DFC450 digital camera attached to a Leica M205C stereomicroscope, with 10–20 photographs

TABLE 1. Distribution of spider sample data.

Species	Male	Female
P. anguilliformis	40	39
P. bibulba	42	42
P. bicruris	42	41
P. breviducta	42	42
P. cangschana	221	242
P. coenobium	42	0
P. confusa	40	40
P. contraria	0	40
P. daliensis	42	42
P. digitata	0	40
P. emei	42	42
P. gibberosa	40	40
P. interposita	42	42
P. kunmingensis	40	40
P. lushanensis	42	40
P. mediana	40	80
P. namkhan	42	56
P. nyingchiensis	42	0
P. peronata	0	40
P. physematosa	0	42
P. rivicola	42	42
P. robusta	40	40
P. roganda	40	40
P. saetosa	40	40
P. semiannulata	40	42
P. semilunata	42	42
P. sicyoidea	42	42
P. signata	40	42
P. sinapophysis	67	57
P. songi	266	256
Totals	1500	1633

taken in different focal planes and combined using image stacking software (Leica LAS). The captured TIFF files were converted into JPEG format through Python, the file size was drastically reduced while maintaining a resolution of 2560×1920, and a standard data set was established in JPEG format. The image annotation result is shown in Figure 1.

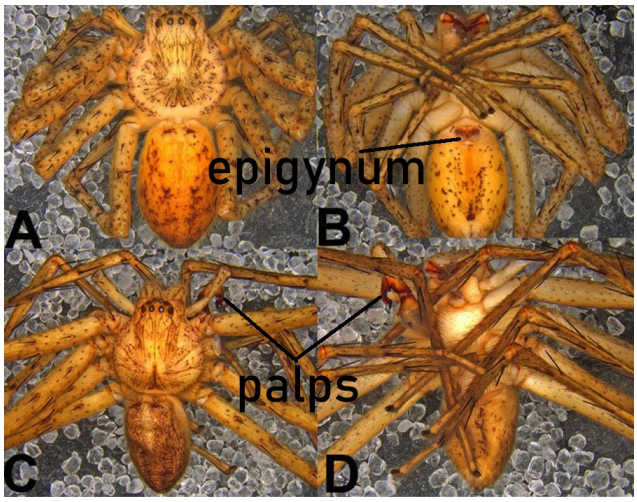


FIGURE 1. Example of spider micro graphics. *Pseudopoda contraria* Jäger & Vedel, 2007: A. Overall photo of female spider, dorsal view; B. Overall photo of female spider, ventral view; C. Photo of male spider, dorsal view; D. Photo of male spider, ventral view.

In this study, the samples of 30 *Pseudopoda* species are divided into a model training set, a validation set, and a test set, and samples for the other 12 *Pseudopoda* species are used as generalization test set A (Test set A is used in a supplementary experiment to verify the reliability of the machine learning model); this test set included 800 samples (Table 2). In addition, to verify whether nongenital structural features, such as the back color and pattern of spiders, display female and male dimorphism, 328 pictures without any genital structure information are manually selected as generalization test set B from set A. According to the general practice of model training, after randomly shuffling the samples, the 30 species sets are divided into training set and validation set at a ratio of 3 to 1; additionally, 4/5 of the validation set is used for model validation, and 1/5 is used as the model test set. The final training set contains 2350 samples, the validation set contains 626 samples, and the test set contains 157 samples. The data in Table 1 and Table 2 show that the samples are balanced [34], [35].

B. TRANSFER LEARNING METHOD

A transfer learning method for image recognition with TensorFlow is used, and the base model is trained based on ImageNet. The model is designed with five flows: a data augmentation flow, a data preprocessing flow, a general feature extraction flow, a domain feature extraction flow and a label prediction flow (Figure 2). The general feature extraction flow adopts the structures and parameters of the basic network, this part does not need to be trained, so there is no backward propagation. In the other hand, in the domain feature extraction flow and label prediction flow such as the spider field,the parameters of these layers need to be retrained, so there are forward and backward propagation, the parameters are adjusted through backward propagation. Among them, the domain feature extraction flow and the label prediction flow are redesigned compared to the traditional flows. To retain the contributions of subtle features in the domain feature extraction flow, the ReLU activation function is modified to an ELU function. In the label prediction flow, the feature output of the convolutional layer is obtained by global average pooling, and a dropout layer is added before the fully connected layer to prevent overfitting; the dropout rate is set to 0.2, that is, 20% of neurons are randomly discarded [33], [36].

TABLE 2. Data distribution for generalization test set A.

Species	Male	Female
P. songi	40	68
P. sp1	40	12
P. sp3	40	40
P. sp4	0	40
P. sp5	0	40
P. sp6	40	40
P. sp7	40	40
P. sp9	40	40
P. sp11	40	40
P. sp15	0	40
P. sp16	0	40
P. sp19	40	40
Totals	320	480

After several groups of experiments, Xception is finally selected as the base model for transfer learning, and the model architecture is shown in Figure 3 and Figure 4.

C. OPTIMIZATION AND PARAMETER SELECTION OF LEARNING MODEL

1) BASIC MODEL SELECTION

Based on our sample size and data characteristics, ResNet-152 and Xception are selected as the basic candidate models for the transfer learning network, and one base model is selected based on the final experimental results [37], [38].

2) DATA RESOLUTION SELECTION

The resolution of input samples has a considerable influence on the prediction accuracy of a model. The default resolution of Resnet-152 is 224×224, and that of Xception is 299×299. The default resolution of the original model is low. Thus, to study the impact of the resolution on the model accuracy, this study designs three groups of experiments with image resolutions of 299×299, 800×600 and 1600×1200.

3) DATA AUGMENTATION SELECTION

This study uses five augmentation methods, namely, random flipping, random rotation, random crop- ping, random scaling and random correction of contrast, and the related parameters are randomly designed according to the habits shown in Table 3. In addition, two sets of experiments are performed

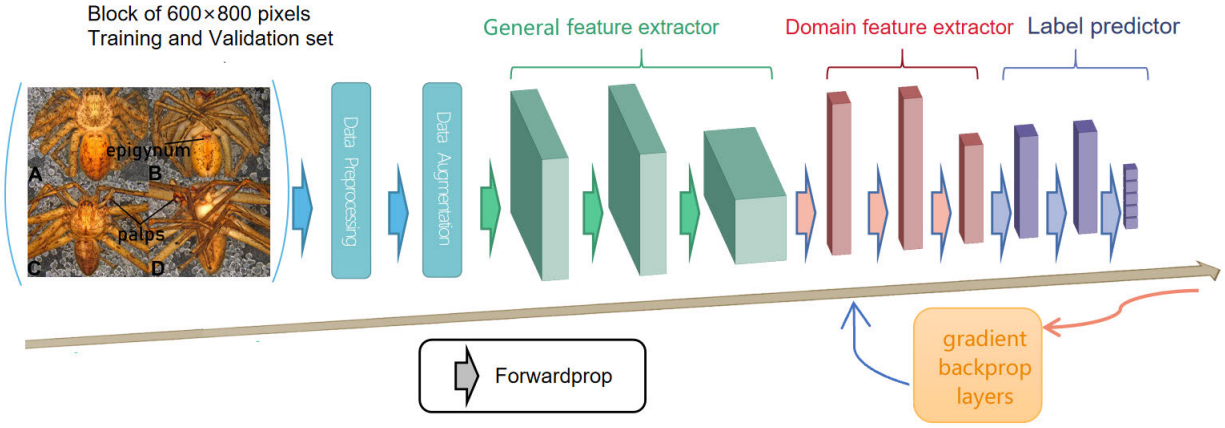


FIGURE 2. Transfer learning model and flow.

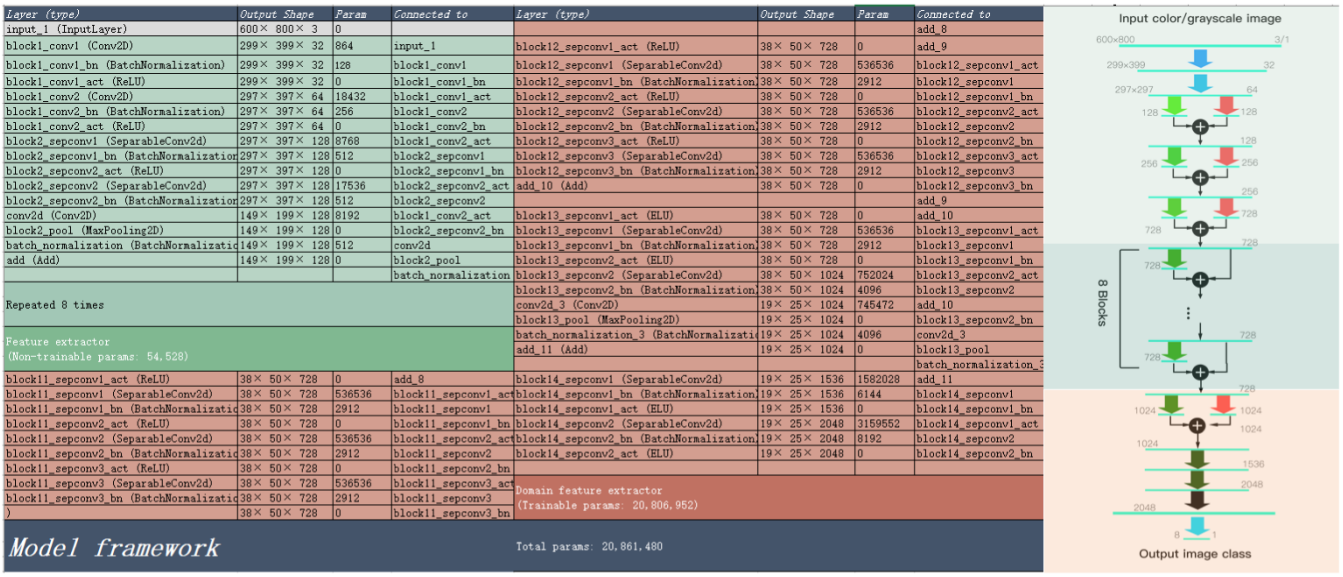


FIGURE 3. Network architecture of the transfer learning model based on Xception.

TABLE 3. Parameter settings for data augmentation processing.

Method	Parameter	Value
RandomFlip	Mode	horizontal_and_vertical
RandomRotation	Factor	0.2
RandomZoom	height_factor	0.8
RandomContrast	Factor	(0.5,0.5)
RandomCrop	width,height	Half of origin size

for the random cropping problem: random cropping and removing random cropping [39].

4) SELECTION OF FROZEN LAYERS IN THE GENERAL FEATURE EXTRACTION FLOW

The parameters of the general feature extraction flow are obtained by ImageNet training, and they reflect the general rule of the network model for sample feature extraction. The Xception network has 134 layers grouped into 14 blocks [38]; according to the characteristics of this network, four sets of experiments are designed to compare the effect of the depth of the general feature extraction flows in the network on the accuracy of the model. In these experiments, the first 66, 86, 96 and 126 network layers are frozen.

5) ACTIVATION FUNCTION OPTIMIZATION

The activation function affects the output results of each layer and has a direct impact on the merit of the final prediction results. Some reports have indicated that ELU and ReLU activation functions perform well in various machine learning domains [38], but ELU provides a wider excitation boundary than ReLU. The Xception network model uses a ReLU activation function, and in this paper, four sets of experiments involving the activation function in domain feature extraction are performed: 1) ReLU is used as the default, 2) only the

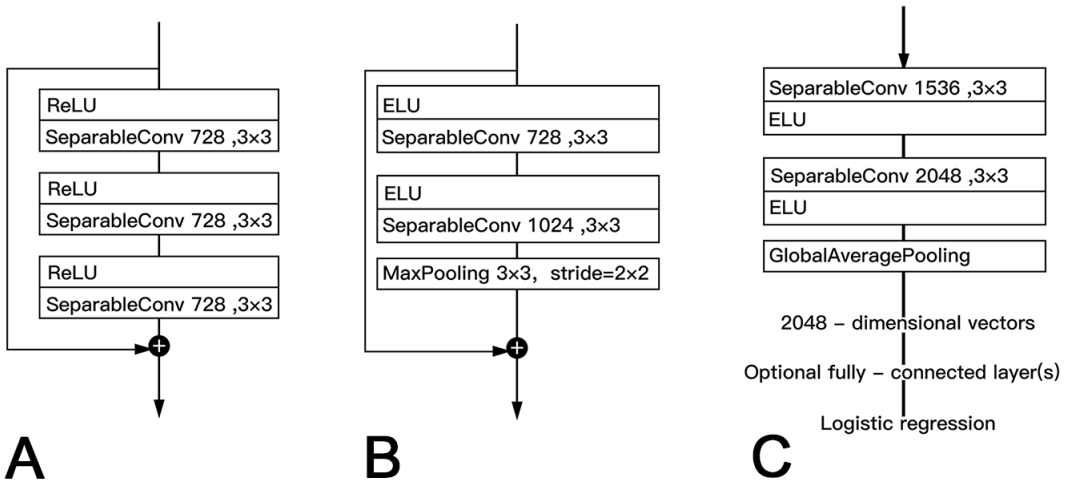


FIGURE 4. Detail of some layers of the transfer learning model and the corresponding flows: A. Repeated steps in the Xception model; B. The activation function is changed to ELU in Block 13; C. The activation function is changed to ELU in Block 14.

TABLE 4. Evaluation values for the basic training models in selection experiments.

Experiments	V1	V2	V3	V4	V5	V6	V7	V8
ResNet-152	0.9823	1	0.9747	0.9983	0.3401	0.0068	0.0236	0.3334
Xception	0.9802	0.9826	0.9739	0.9738	0.1211	0.0667	0.0001	0.0544

Block 14 activation function is modified to an ELU function, 3) the Block 13 and Block 14 activation functions are modified to ELU functions, and 4) the activation function in Block 14 is removed [40].

D. GENERALIZED PRACTICAL TEST EXPERIMENTS

To verify the network model and assess the female and male dimorphism of the nongenital characteristics of spiders, two sets of experiments are performed. With optimally trained network models, prediction experiments based on generalization test set A and generalization test set B are performed.

E. MODEL EVALUATION STANDARD

The model is assessed based on 8 metrics, named V1, V2, V3, V4, V5, V6, V7 and V8, which are defined as follows.

$$accuracy = \frac{quantity\ of\ \ correctly\ identified}{quantity\ of\ \ total} \tag{1}$$

$$V1 = Max(validation\ accuracy\ over\ 200\ epochs) \tag{2}$$

$$V2 = Max(training\ accuracy\ over\ 200\ epochs) \tag{3}$$

$$V3 = Mean\left(\begin{array}{c} top\ 20\ of\ \ validation\ accuracy \\ over\ 200\ epochs \end{array}\right) \tag{4}$$

$$V4 = Mean\left(\begin{array}{c} training\ accuracy\ corresponding \\ to\ the\ top\ 20\ of\ \ the\ validation \\ accuracy\ over\ 200\ epochs \end{array}\right) \tag{5}$$

$$V5 = Mean\left(\begin{array}{c} validation\ loss\ corresponding \\ to\ the\ top\ 20\ of\ \ the\ validation \\ accuracy\ over\ 200\ epochs \end{array}\right) \tag{6}$$

$$V6 = Mean\left(\begin{array}{c} training\ loss\ corresponding \\ to\ the\ top\ 20\ of\ \ the\ validation \\ accuracy\ over\ 200\ epochs \end{array}\right) \tag{7}$$

$$V7 = Abs(V3 - V4) \tag{8}$$

$$V8 = Abs(V5 - V6) \tag{9}$$

V1 reflects the highest accuracy that the model can achieve for the validation set during the whole training process, V2 reflects the highest accuracy achieved for the training set, and the difference between V1 and V2 reflects the degree of fitting of the model. V3 and V4 reflect the stability of the model prediction accuracy, V5 and V6 reflect the cross-entropy loss of the model, and V7 and V8 provide intuitive feedback regarding the fit of the model [41].

III. RESULTS

A. BASE MODEL SELECTION

In the base model selection experiments, 200 training epochs were considered for the two groups of experiments, and the evaluation metrics V1-V8 are shown in Table 4. Additionally, the accuracy and loss value curves are plotted in Figure 5. The data show that the values of V1, V2, V3, V4, V5, V7, and V8 for ResNet-152 are generally higher than those for Xception; notably, only V6 is lower for ResNet-152 than for Xception. Overall, the accuracy of ResNet-152 is higher, but ResNet-152 appears to be overfitting the results. Based on Figure 5, the loss distribution of the validation set is not smooth enough and has a tendency to be overestimated. Thus, based on previous deep network model research [15], in this study, the Xception network is used as the base neural

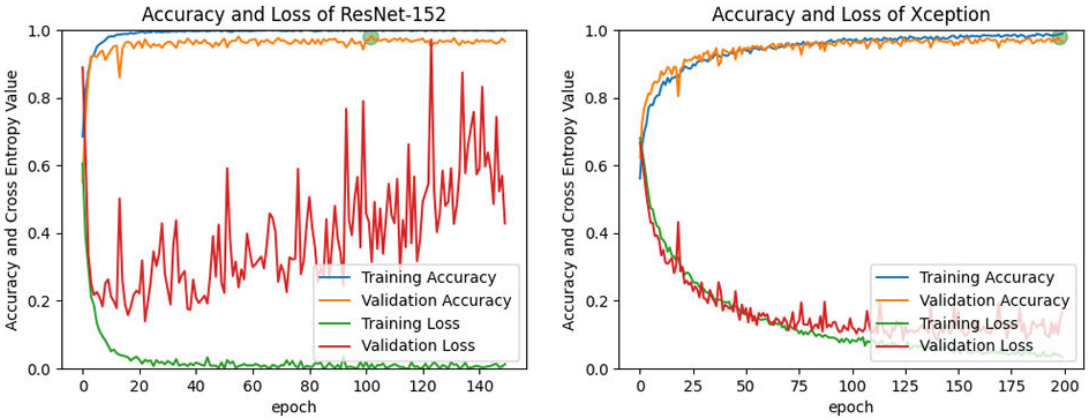


FIGURE 5. Curves of model accuracy and loss values in the base model selection experiments.

TABLE 5. Evaluation values of the sample input resolution selection experiments.

Experiments	V1	V2	V3	V4	V5	V6	V7	V8
299×299	0.8611	0.8502	0.8515	0.8272	0.3990	0.3736	0.0244	0.0253
800×600	0.9557	0.9243	0.9458	0.9000	0.1585	0.2364	0.0458	0.0779
1920×1600	overflow	overflow	overflow	overflow	overflow	overflow	overflow	overflow

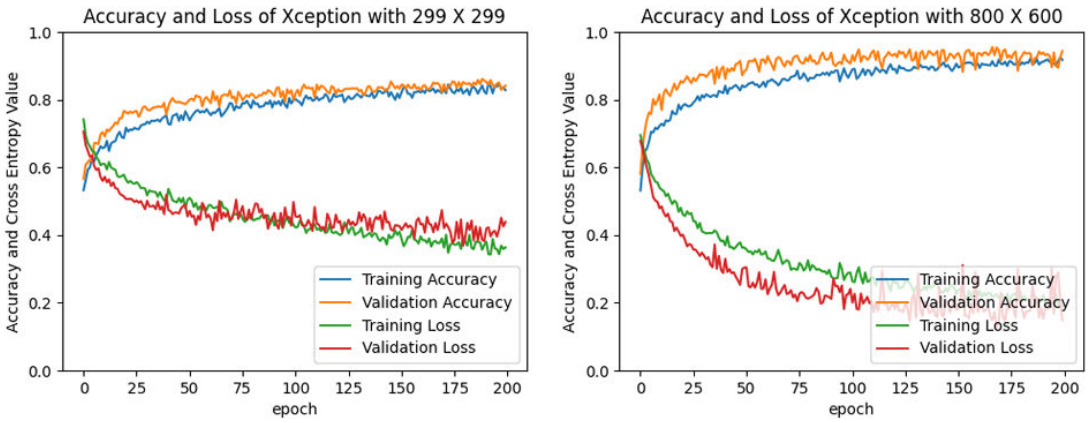


FIGURE 6. Curves of the model accuracy and loss values based on selection experiments with different sample input resolutions.

network model, and subsequent experiments and conclusions are based on Xception.

B. DATA RESOLUTION SELECTION

With a single Nvidia Tesla V100 GPU and a 32 GB video memory card, the resolution of 1920 × 1600 in the experiments directly led to memory overflow, and the learning model could not be trained. Therefore, the corresponding group of experiments was abandoned. The remaining two groups of experiments were performed, and V1-V8 values were calculated; the results are shown in Table 5. The accuracy and loss value curves are plotted in Figure 6. The figures show that the duration and parameter fits for training at a resolution of 800 × 600 were greater than those at a resolution of 299 × 299; however, the fine features of the original images were much better preserved, thereby improving the accuracy of the model. In the term of V3 values, for images with an 800 × 600 resolution, the model can reach 0.9458; for a 299 × 299 resolution, the V3 value of the model reaches only 0.8515. Thus, the prediction accuracy of the former is improved by 11.07%, and the model that uses images with an 800 × 600 resolution has obvious advantages, especially considering the results in Figure 6.

C. DATA AUGMENTATION SELECTION

In this case, the V1-V8 values for the two groups of experiments were obtained. The experimental results are shown in Table 6, the curves of accuracy and loss values are shown in Figure 5. In the case of no random cropping, the prediction accuracy for the validation set can reach 0.96 after

TABLE 6. Evaluation values of the randomly cropping for data augmentation experiments.

Experiments	V1	V2	V3	V4	V5	V6	V7	V8
With random cropping	0.9557	0.9243	0.9458	0.9000	0.1585	0.2364	0.0458	0.0779
Without random cropping	0.9786	0.9894	0.9736	0.9812	0.1001	0.0506	0.0076	0.0501

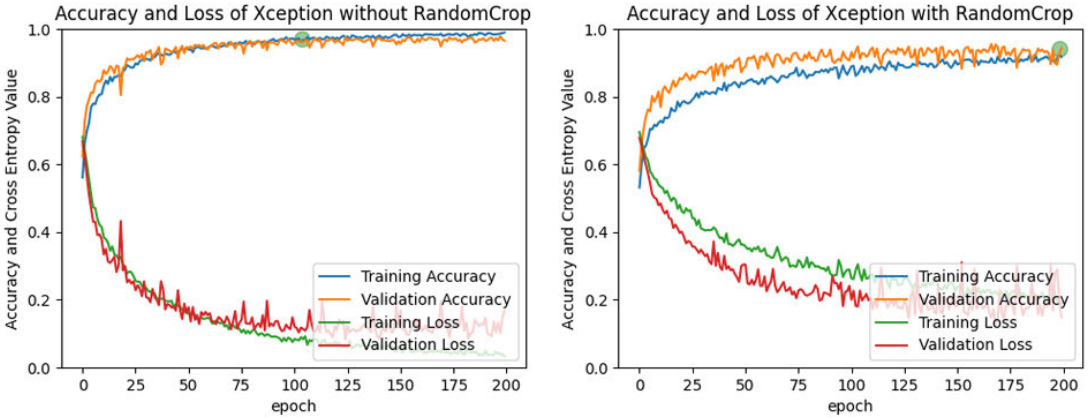


FIGURE 7. Curves of the model accuracy and loss values based on random cropping in the data augmentation experiments.

TABLE 7. Evaluation values from the selection experiments based on general feature extraction with frozen layers.

Experiments	V1	V2	V3	V4	V5	V6	V7	V8
66 frozen layers	0.9786	0.9923	0.9769	0.9838	0.1122	0.0428	0.0069	0.0694
86 frozen layers	0.9786	0.9881	0.9739	0.9790	0.1225	0.0566	0.0051	0.0659
96 frozen layers	0.9802	0.9826	0.9739	0.9738	0.1211	0.0667	0.0001	0.0544
126 frozen layers	0.9405	0.9098	0.9283	0.8982	0.2491	0.2593	0.0301	0.0102

104 epochs of training, and with random cropping, the accuracy reaches 0.93 after 200 epochs of training. Based on both Table 6 and Figure 7, the prediction accuracy of the model is greatly reduced after adding random cropping, possibly because spiders are small, and some subtle features that play an important role in sex discrimination may be cropped. Thus, the random cropping of data led to the loss of these subtle features. Consequently, the random cropping augmentation method is not considered further.

D. SELECTION OF FROZEN LAYERS IN THE GENERAL FEATURE EXTRACTION FLOW

The four groups of models in the experimental design were trained for 200 epochs, and the V1-V8 values were obtained, as shown in Table 7. The accuracy and loss value curves are plotted in Figure 8. The experimental results indicate that the accuracy for the training set is generally higher than that for the validation set after 125 epochs of training when the number of frozen layers is 66; additionally, an overfitting phenomenon appears. The prediction accuracy is lower overall and the cross-entropy loss slowly decreases when the number of frozen layers is set to 126 layers. The results for 86 and 96 frozen layers are similar, but the V1 value in the latter case is slightly higher and V7 is smaller; thus, model performance is best when 96 layers are frozen. Subsequently, 96 frozen layers were used in all other experiments involving the spider sex recognition model.

E. ACTIVATION FUNCTION OPTIMIZATION

The four groups of experiments involved training for 200 epochs, and the V1-V8 values were obtained, as shown in Table 8. The accuracy and loss value curves are in Figure 9. Good prediction accuracy for the four groups of experiments. The V1 value of the default ReLU experiment reached 0.9786, and the V3 value reached 0.9736. In the experiment in which the activation function of Block 14 was modified to an ELU function, the V1 value reached 0.9802, and the V3 value reached 0.9739. In the group experiment which the activation functions of Block 13 and Block 14 modified to ELU functions, the V1 value reached 0.9618, and the V3 value reached 0.9521. In the experiment in which the activation function of Block 14 was removed, the V1 value reached 0.9791, and the V3 value reached 0.9747. The accuracies achieved for the training set and the validation set in these four experiments were compared. In the group experiment with the ReLU activation function, the accuracy of the training set was higher than that for validation set after 200 epochs of training, and overfitting occurred. In contrast, in the experiment with the ReLU activation function combined with the ELU function, the accuracy for the training set was similar to that for the validation set after 200 epochs of training; additionally, an overfitting state was not reached, indicating the model can continue learning after 200 epochs. This potentially increase in ability may increase model accuracy. Based on

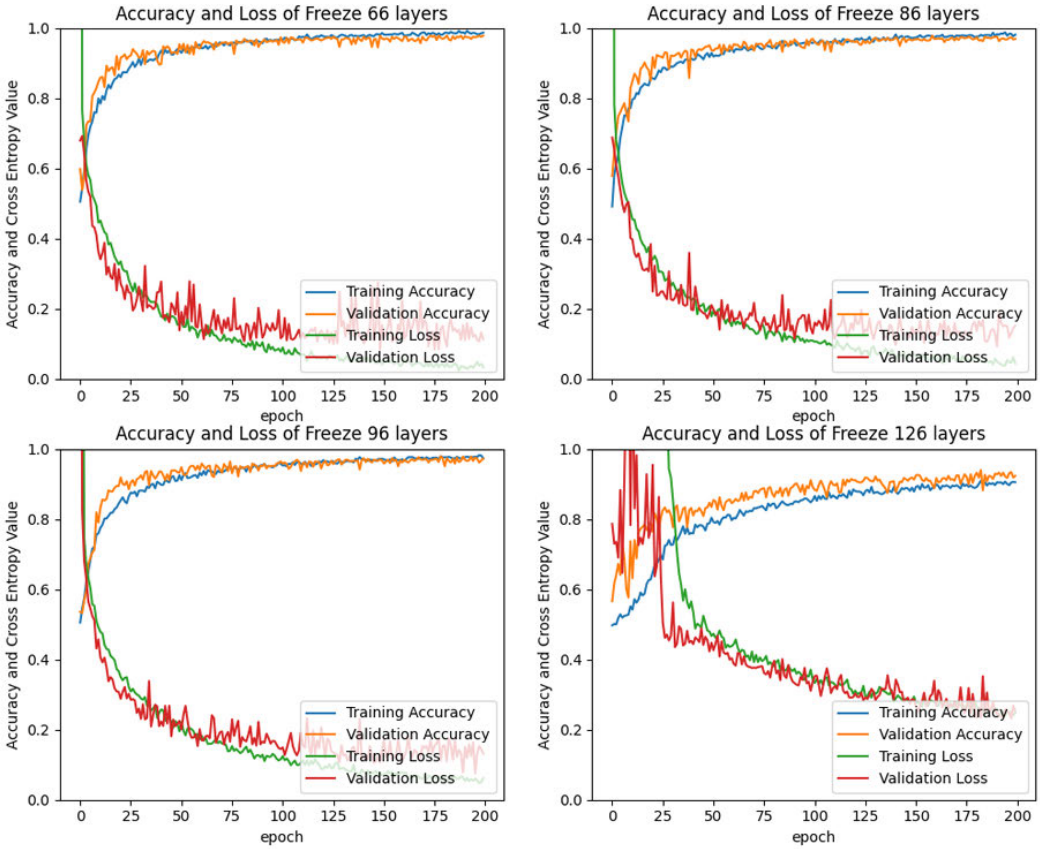


FIGURE 8. Curves of the model accuracy and loss values for the experiments involving general feature extraction with frozen layers.

TABLE 8. Evaluation values for the activation function optimization experiments.

Experiments	V1	V2	V3	V4	V5	V6	V7	V8
ReLU	0.9786	0.9894	0.9736	0.9812	0.1001	0.0506	0.0076	0.0501
Block14 ELU	0.9802	0.9826	0.9739	0.9738	0.1211	0.0667	0.0001	0.0544
Block13-14 ELU	0.9618	0.9455	0.9521	0.9298	0.1669	0.1833	0.0224	0.0164
Without an activation function in Block 14	0.9791	0.9753	0.9747	0.9669	0.1225	0.0901	0.0078	0.0324

a comparison of the second and third experiments, the second experiment displayed a faster gradient decrease and higher accuracy than the third experiment. Finally, based on a comparison of the second and fourth experiments, the V1 value of the second was higher than that of the fourth. In summary, the second group of experiments corresponded to the best experimental results when the ReLU function was changed to an ELU function in the Block 14 of Xception.

TABLE 9. Learning model structure and parameters.

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 600, 800, 3)]	0
sequential (Sequential)	(None, 600, 800, 3)	0
tf.math.truediv (TFOpLambda)	(None, 600, 800, 3)	0
tf.math.subtract (TFOpLambda)	(None, 600, 800, 3)	0
xception (Functional)	(None, 19, 25, 2048)	20861480
global_average_pooling2d(GlobalPooling2D)	(None, 2048)	0
dropout (Dropout)	(None, 2048)	0
dense (Dense)	(None, 2)	4098

F. GENERALIZED PRACTICAL TEST EXPERIMENTS

After the experiment and optimization process, the final number of parameters in the deep neural network was 20,865,578, of which 10,020,434 need to be learned during training in this study. The model structure is shown in Table 9. According to the results, after 179 epochs, the prediction accuracy reaches a maximum value of 0.9802; the corresponding model and parameters are used in predictions based on the real test set A and real test set B. For test set A, 61 samples were incorrectly predicted, and the prediction accuracy reached 92.38%. For test set B, 19 samples were incorrectly predicted, and the prediction accuracy reached 94.21%. Based on the results of the two sets of generalization experiments, machine learning can be effectively used in spider sex image recognition tasks, and the existence of male and female dimorphism phenomenon in relation to nongenital features is tentatively verified.

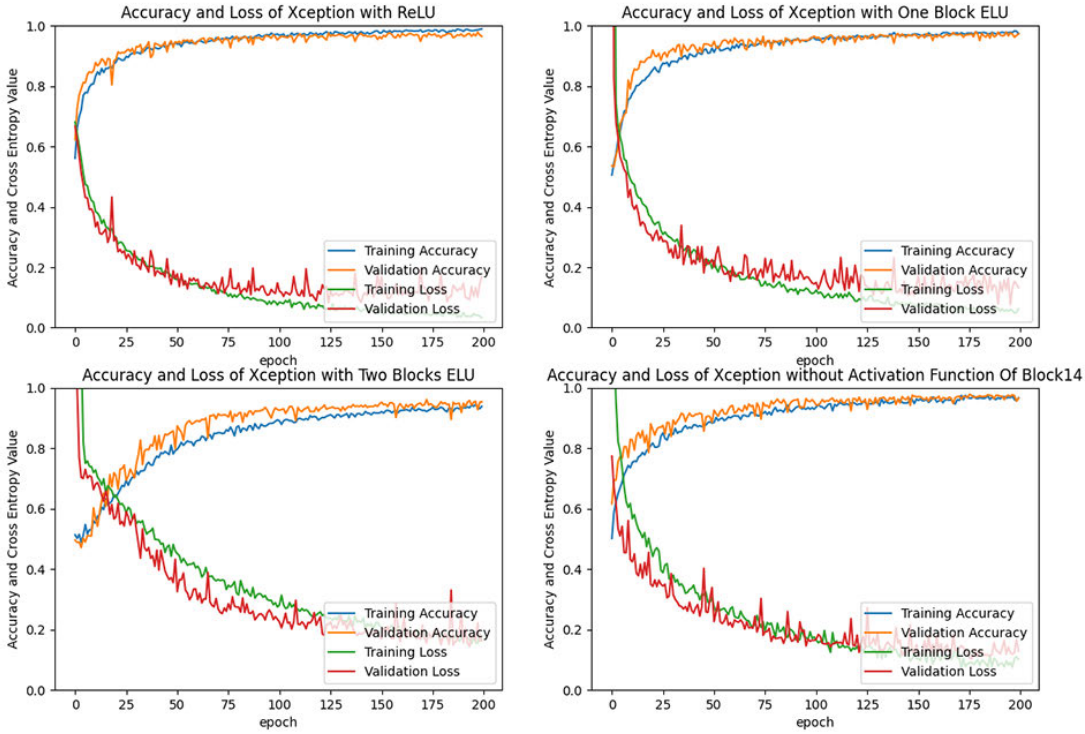


FIGURE 9. Curves of the model accuracy and loss values for the activation function optimization experiments.

IV. DISCUSSION

In this study, 42 Chinese *Pseudopoda* species belonging to the Sparassidae family were considered in spider sex recognition based on convolutional neural networks, and through model tests, data augmentation and model optimization, the prediction accuracy for the validation set reached 98.02%; additionally, the generalization accuracy for independent samples reached 92.38%. Thus, the proposed method can replace manual identification and provide a reference for future spider species image recognition problems.

Deep learning models are generally classified into several categories, and popular models include VGG, ResNet, GoogLeNet, and Inception, among others. According to Alfredo Canziani *et al.*, ResNet-152, Inception-V3, and Inception-V4 are highly suitable as base models for image recognition with transfer learning. Xception is based on the improved Inception-V3 network, and it outperforms Inception-V3 and Inception-V4 in some aspects. Our results show that the Xception model is slightly better than ResNet-152 in spider sex recognition from images, which may be related to the characteristics of the samples. The results of this study confirm that deep learning problems with small samples can be effectively solved using transfer learning. For small-sample training sets, transfer learning can take advantage of the existing knowledge obtained through training with large general datasets, and the unique abilities of the model can be applied in new domains. For example, this approach was applied for sex recognition from images of spiders of the genus *Pseudopoda* in this study and could be applied in future studies of spider species identification from images.

Data augmentation is achieved by randomly transforming the existing samples using certain rules; this process is analogous to the randomness of taking pictures in a natural environment and increases the number of samples in the study set. This approach can prevent the occurrence of overfitting due to the effects of a small training set and thus improve the generalization ability of the network. Currently, the common data augmentation methods include random flipping, random rotation, random cropping, random scaling, boundary enhancement, random deletion, randomly blending, and random contrast correction. It has been confirmed that data augmentation directly affects the learning ability and training accuracy of models. In this study, in addition to four common data augmentation methods (random flipping, random rotation, random scaling and random contrast correction), we focus on the effects of random cropping and varying the resolution of input images on model performance in spider sex recognition, which requires a fine scale. In addition to genitalia and other features, the dorsal pattern, color, morphology, etc. of spiders is also related to sex, and random cropping leads to the loss of important features.

Data preprocessing techniques also influence whether model learning can be successful, as reflected by excellent learning ability and high training accuracy. For example, in this study, the photos are uniformly processed into

JPEG format, which greatly reduces the capacity of the images, thus reducing the hardware demands of the network model; considering the existing hardware capabilities, a high resolution needs to be used as much as possible to retain subtle features of a sample in addition to the general outline or textural features of an object, thus improving the prediction accuracy of the model.

In this study, the number of frozen layers of the general feature extraction flow is experimented, and the results show that the number of frozen layers of the general feature extraction flow for transfer learning is neither more nor less, and it is necessary to combine specific target samples and experiment to find the appropriate number of frozen layers. The activation function of domain feature extraction flow is redesigned, ReLU and ELU activations are combined, and Dropout layer is added in label prediction flow to prevent model overfitting, etc. All these optimizations can have a great impact on the prediction ability of the model. However, in some experiments, the accuracy achieved for the validation set may be higher than that for the training set, possibly due to the randomness of sample segmentation, small sample sizes and the dropout of some neurons. Moreover, removing the activation function from Block 14 does not improve the accuracy of the model, potentially due to the use of transfer learning and a small sample set.

In addition, the manual identification of spider sex relies mainly on the genital structures of mature spiders (male palps or female epigynum). However, we found that the studied computer was able to identify the sex of spiders when it was not possible to do so manually based on pictures, and the generalization accuracy was as high as 94.21%. This result suggests that nongenital features such as the body color and the spot patterns of *Pseudopoda* spiders, may be dimorphic at the genus level. This phenomenon is currently unrecognizable by humans. Although male and female dimorphism in body morphology is typical in some specific taxa, such as *Nephila* species in the family Araneidae (the female is almost five times larger than the male), it has rarely been reported in nocturnal spiders such as Sparassidae and Lycosidae. However, it must be noted that the sample size in this study was relatively small, and the generalization test set contained only 328 images of 12 species; therefore, this conclusion needs to be further confirmed by expanding the sample. In addition, how the computer specifically identifies spiders based on their body color and dorsal pattern for sex recognition needs to be further confirmed using deep neural network interpretation, and this result is only a preliminary conclusion.

In conclusion, this study proposes a deep learning–transfer learning model based on Xception, and the training model can be used to solve sex recognition problems for spiders after optimization. This approach provides a reference for future studies of spider species recognition from images.

REFERENCES

[1] World Spider Catalog. (2021). *World Spider Catalog. Version 22.0. Natural History Museum Bern*. Accessed: Mar. 3, 2021. [Online]. Available: http://wsc.nmbe.ch

[2] M. C. Orr, R. R. Ferrari, A. C. Hughes, J. Chen, J. S. Ascher, Y.-H. Yan, P. H. Williams, X. Zhou, M. Bai, A. Rudoy, F. Zhang, K.-P. Ma, and C.-D. Zhu, "Taxonomy must engage with new technologies and evolve to face future challenges," *Nature Ecol. Evol.*, vol. 5, no. 1, pp. 3–4, Jan. 2021.

[3] S. Strauß, "From big data to deep learning: A leap towards strong AI or 'intelligentia obscura?'" *Big Data Cogn. Comput.*, vol. 2, no. 3, p. 16, 2018.

[4] R. O. Sinnott, D. Yang, X. Ding, and Z. Ye, "Poisonous spider recognition through deep learning," in *Proc. Australas. Comput. Sci. Week Multiconf.*, Feb. 2020, pp. 1–7.

[5] M. Zahangir Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. Shamima Nasrin, B. C Van Esesn, A. A S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*. [Online]. Available: http://arxiv.org/abs/1803.01164

[6] D. Song, "Classification of spiders," *Sichuan J. Zool.*, vol. 2, pp. 37–41, Mar. 1985.

[7] S. M. Van Belleghem, R. Papa, H. Ortiz-Zuazaga, F. Hendrickx, C. D. Jiggins, W. Owen McMillan, and B. A. Counterman, "Patternize: An R package for quantifying colour pattern variation," *Methods Ecol. Evol.*, vol. 9, no. 2, pp. 390–398, Feb. 2018.

[8] H. Zhang, P. Jäger, and J. Liu, "Establishing a new species group of Pseudopoda Jäger, 2000 with the description of two new species (araneae, sparassidae)," *ZooKeys*, vol. 879, p. 91–115, Oct. 2019.

[9] M. L. Lim, M. F. Land, and D. Li, "Sex-specific UV and fluorescence signals in jumping spiders," *Science*, vol. 315, no. 5811, p. 481, 2007.

[10] J. Li, Z. Zhang, F. Liu, Q. Liu, W. Gan, J. Chen, M. L. M. Lim, and D. Li, "UVB-based mate-choice cues used by females of the jumping spider *Phintella vittata*," *Current Biol.*, vol. 18, no. 9, pp. 699–703, May 2008.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[12] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*. [Online]. Available: http://arxiv.org/abs/1408.5882

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[14] N. Ketkar, "Convolutional neural networks," in *Deep Learning With Python*. Berlin, Germany: Springer, 2017, pp. 63–78, doi: 10.1007/978-1-4842-2766-4_5.

[15] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," 2016, *arXiv:1605.07678*. [Online]. Available: http://arxiv.org/abs/1605.07678

[16] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, p. 342, Jan. 2020.

[17] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. V. D. Laak, B. V. Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[18] S. Hassantabar, N. Stefano, V. Ghanakota, A. Ferrari, G. N. Nicola, R. Bruno, I. R. Marino, K. Hamidouche, and N. K. Jha, "CovidDeep: SARS-CoV-2/COVID-19 test based on wearable medical sensors and efficient neural networks," 2020, *arXiv:2007.10497*. [Online]. Available: http://arxiv.org/abs/2007.10497

[19] S. Gu, H. Gao, and X. Hang, "Research progress on image recognition technology of crop pests and diseases based on deep learning," *Trans. Chin. Soc. Agricult. Machinery*, vol. 50, no. S1, pp. 313–317, 2019.

[20] A. L Chandra, S. Vikas Desai, W. Guo, and V. N Balasubramanian, "Computer vision with deep learning for plant phenotyping in agriculture: A survey," 2020, *arXiv:2006.11391*. [Online]. Available: http://arxiv.org/abs/2006.11391

[21] A. Santoro, F. Hill, D. Barrett, A. Morcos, and T. Lillicrap, "Measuring abstract reasoning in neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4477–4486.

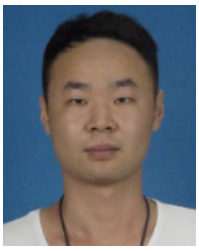
[22] A. F. Gad, "TensorFlow recognition application," in *Practical Computer Vision Applications Using Deep Learning With CNNs*. Berkeley, CA, USA: Apress, 2018, doi: 10.1007/978-1-4842-4167-7_6.

[23] F. Chollet. (2015). *Keras: Deep learning library for theano and Tensorflow*. [Online]. Available: https://keras.io/k

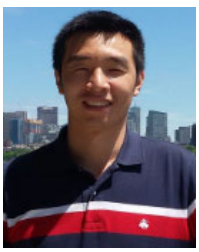
[24] ImageNet. (2021). *ImageNet*. Accessed: Apr. 5, 2021. [Online]. Available: http://image-net.org/
[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, C. Alexander Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[26] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[27] Q. Qian, Y. Qin, Y. Wang, and F. Liu, "A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis," *Measurement*, vol. 178, Jun. 2021, Art. no. 109352.
[28] S. Feng, H. Zhou, and H. Dong, "Application of deep transfer learning to predicting crystal structures of inorganic substances," *Comput. Mater. Sci.*, vol. 195, Jul. 2021, Art. no. 110476.
[29] Y. Cao, M. Jia, P. Ding, and Y. Ding, "Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network," *Measurement*, vol. 178, Jun. 2021, Art. no. 109287.
[30] Z. Zheng, H. Qi, L. Zhuang, and Z. Zhang, "Automated rail surface crack analytics using deep data-driven models and transfer learning," *Sustain. Cities Soc.*, vol. 70, Jul. 2021, Art. no. 102898.
[31] I. Dagher and D. Barbara, "Facial age estimation using pre-trained CNN and transfer learning," *Multimedia Tools Appl.*, vol. 80, pp. 20369–20380, Mar. 2021.
[32] L. Miao, W. Jingxian, L. Hualong, H. Zelin, Y. XuanJiang, H. Xiaoping, Z. Weihui, Z. Jian, and F. Sisi, "Method for identifying crop disease based on CNN and transfer learning," *Smart Agricult.*, vol. 1, no. 3, pp. 46–55, 2019.
[33] A. Darwish, D. Ezzat, and A. E. Hassanien, "An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis," *Swarm Evol. Comput.*, vol. 52, Feb. 2020, Art. no. 100616.
[34] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 39, no. 2, pp. 539–550, Apr. 2009.
[35] G. E. Batista, R. C. Prati, and M. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, 2004.
[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[38] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[39] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
[40] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2015, *arXiv:1511.07289*. [Online]. Available: http://arxiv.org/abs/1511.07289
[41] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*. [Online]. Available: http://arxiv.org/abs/2010.16061

QIANJUN CHEN is currently pursuing the Ph.D. degree with the National Engineering Research Center for E-Learning, Central China Normal University, China. His research interests include information retrieval, data mining, and machine learning.

YONGCHANG DING is currently pursuing the master's degree with the College of Life Sciences, Hubei University, China. His research interests include machine learning and ecology.



CHANG LIU is currently pursuing the master's degree with the College of Life Sciences, Hubei University, China. His research interests include machine learning, ecology, and software.



JIE LIU is currently a Full Professor with the College of Life Sciences, Hubei University, China. He has published 37 papers and 26 of which have been included in SCI journals. His research interests include ecology and spider taxonomy.

TINGTING HE is currently a Full Professor and the Dean of the School of Computer Science, Central China Normal University, China. Her research interests include natural language processing, sentiment analysis, information retrieval, machine learning, and deep learning.