Fast Recognition System for Tree Images based on Caffe Platform and Deep Learning

Zongchen Li

Central South University of forestry science and technology <https://orcid.org/0000-0001-5422-5406>

Wenzhuo Zhang

Central South University of forestry science and technology

Guoxiong Zhou (✉ zhougx01@163.com)

Central South University of forestry science and technology <https://orcid.org/0000-0002-5142-4845>

Research

Keywords: Tree image recognition, image cleaning, Gabor filter, Dual-Task CNN

Posted Date: August 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-52291/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Fast Recognition System for Tree Images based on Caffe Platform and Deep Learning

ABSTRACT Aiming at the difficult problem of complex extraction for tree image in the existing complex background, we took tree species as the research object and proposed a fast recognition system solution for tree image based on Caffe platform and deep learning. In the research of deep learning algorithm based on Caffe framework, the improved Dual-Task CNN model (DCNN) is applied to train the image extractor and classifier to accomplish the dual tasks of image cleaning and tree classification. In addition, when compared with the traditional classification methods represented by Support Vector Machine (SVM) and Single-Task CNN model, Dual-Task CNN model demonstrates its superiority in classification performance. Then, in order for further improvement to the recognition accuracy for similar species, Gabor kernel was introduced to extract the features of frequency domain for images in different scales and directions, so as to enhance the texture features of leaf images and improve the recognition effect. The improved model was tested on the data sets of similar species. As demonstrated by the results, the improved deep Gabor convolutional neural network (GCNN) is advantageous in tree recognition and similar tree classification when compared with the Dual-Task CNN classification method. Finally, the recognition results of trees can be displayed on the application graphical interface as well. In the application graphical interface designed based on Ubuntu system, it is capable to perform such functions as quick reading of and search for picture files, snapshot, one-key recognition, one-key exit and so on.

KEYWORDS: Tree image recognition, image cleaning, Gabor filter, Dual-Task CNN

I. BACKGROUND

With more and more serious environmental pollution, trees are widely used in urban greening, and play a significant role in improving and protecting the environment. After careful selection and ingenious arrangement, all kinds of trees can play a major role in protecting environment, improving environment, beautifying environment and generating economic by-products^[1-2]. On the contrary, if the tree greening construction is not rationally planned, problems such as single variety, neglect of economic benefits, and blind construction will continue to be evident, resulting in the lack of sustainability of forestry development, the stagnant state of forestry economy, and the lack of long-term of forestry construction. These even leads to some unpredictable consequences^[3-4]. The analysis of tree species is an important theoretical basis for the development of forestry planning, providing a reference for landscape engineering design, management personnel and related researchers^[1-2]. From the creation of botany founded by Theophrastus to modern and enduring tree classification system, people have been improving the tree classification system based on the original tree classification system to provide better application for garden trees and guidance for forestry development. However, due to the current similarity among trees and the diversity of surrounding landscapes, the identification of tree species in forestry development has become a complex and professionally demanding work. The

cycle for training forestry professional is very long. It is difficult to attract high-end technical personnel. The traditional manual investigation methods is not sufficient to meet the requirements of modern forestry development because of its low efficiency, time consuming, and a fact that it is easy for workers to become fatigue. All these will affect the correctness of the results^[5-7]. Therefore, how to efficiently and quickly classify tree images has become an urgent problem to be solved.

Traditional machine learning often requires time-consuming and complex processing of tree image features, and it still requires a large amount of professional knowledge to manually design feature extraction methods and classifiers, and then identify and predict. When some features are not extracted during the extraction of artificial features, the useful information missed will not be recovered from the classification training, which will affect the detection results. While shallow layer learning models such as SVM, Boosting, and LR are superior to rule-based learning, their ability to express complex functions and generalization capability is not strong due to the finite samples and limited computational units. Shallow layer manual neural networks cannot identify tree species quickly and efficiently, because there are a large number of parameters, and there are defects such as over-fitting and long training time^[8]. CNN is a deep neural network with convolution structure^[9-11], which reduces the number of weights that need to be trained through weight

sharing and reduces the computational complexity of the network. Moreover, through the pooling operation, the network has a certain invariance to the local transformation of the input image, which improves the generalization ability of the network. The convolutional neural network can process complex big data through multi-layer nonlinear transformation, and replace the manual design features with automatic learning features, avoiding the error accumulation of manual design features. Through feature extraction and classifier joint optimization, the accuracy and recognition speed of tree classification are improved. In recent years, more and more researchers have begun to work on how to identify trees quickly and effectively. In order to better improve the convolutional neural network to make it more effective in tree classification applications, researchers continue to improve existing networks from various aspects such as network layer design, loss function design, activation function, and regular terms. Aivars Lorencs et al., based on multiple clustering attempts to design a multi-class classifier with better classification results through the collected multi-spectral image data near the tree top, but only achieved the goal of semi-automatic identification of species^[12]. Du Jianxin et al., used a new mobile center hypersphere classifier for classification based on the extraction of plant leaf shape features^[13]. Although the classification performance is improved compared with the two classifiers of 1-NN and K-NN, it is still greatly affected by the feature extraction process, and the classification performance needs to be improved. Hou Tong et al., used BP neural network combined with plant leaf feature shape to classify plants, and the average recognition rate reached 87.5%^[14]. However, manual neural networks require massive parameters, and have a long learning time. Stephen Gangwu et al., used the probabilistic neural network to automatically identify the leaves^[15]. Compared with the BP network, the classification speed is greatly improved, the process is simple and the noise tolerance is high, but this method has high requirements for the representativeness of the training samples and requires more storage space. Liu Jiandu et al., transformed the blade image into time-frequency domain image by wavelet transform, and then uses SVM classifier to classify the blade, which has better classification performance^[16-17]. Wang-Su Jeon et al., used the existing Flavia dataset to train traditional CNN model to identify plant leaves based on the TensorFlow framework. Compared with traditional machine learning, the CNN model has better classification performance in plant identification, but the tree species in the data set is limited and it also need to manually expand the data set for further testing^[18]. Sue Han Lee et al. used the CNN network to learn unsupervised feature representations for 44 different plant species. Experimental results using CNN models with different classifiers show consistency and superiority compared to existing technology solutions that rely on handcrafted features, but in the process of image

preprocessing, the traditional manual methods to filter pictures are too inefficient^[19].

Based on the above references, this paper designs a rapid recognition system for tree image based on Caffe platform and deep learning, which can realize two tasks of image cleaning and tree classification, and the CNN model is the natural choice of multi-task problem because the learned convolution features can be shared by different advanced tasks^[20]. The improved multi-task Auto-Clean convolutional neural network model trains multi-label data sets with attribute values as numerical values, and obtains the joint features of the two task models^[21-23]. The extracted features are more discriminative, and can realize two tasks of image cleaning and tree classification, and the dual task model is less time-consuming and has better classification accuracy than the single task model. Firstly, we input the tree image into the neural network to determine whether the input image is a dirty one. If it is a clean image, the system starts to predict the tree category. Secondly, The trainable feature extractor and the trainable classifier in the CNN model minimize the loss function combined with the Auto and Clean tasks by forward propagation and backpropagation error algorithms, so that the extracted features can better reflect the essence of the image and the recognition model is better. At the same time, the loss function is used to drive the model to use the small batch random gradient descent method for self-weighted learning, so as to further adjust the task weight to obtain a pre-trained recognition model. Finally, we optimize the classifier through super parameter tuning to obtain a better recognition model. The experimental results show that the Dual-Task CNN model can identify the tree species accurately and in real time after simplified training. In addition, considering the influence of tree similarity on the recognition effect, we also introduce the Gabor layer on the improved Dual-Task CNN model, and propose a new re-improved Deep Gabor CNN model that can enhance the learning of local features of the blade to achieve a more effective distinction between different tree species^[24-27]. Finally, based on the Ubuntu system, this paper also designs the application graphical interface to realize the functions of fast reading and searching of picture files, quick photo taking, one-key recognition and one-button exit, which is convenient and intuitive.

II. DATA SET CONSTRUCTION AND PREPROCESSING

We took photos for trees and collected all kinds of image data from the six views of the trees to form an initial database, and then generated six models to identify the trees by pre-training and fine-tuning of the CaffeNet convolutional neural network. By comparing the classification performance of the recognition models obtained by training the six views, we select the tree blade image with higher classification accuracy to reconstruct the data set. Then, the data set is expanded by rotating, translation, zooming and horizontal flipping. Finally, the data is cleaned by script classification

and manual screening to form an experimental tree image data set.

FIGURE 1. Data Set Construction and Preprocessing

A. DATABASE CONSTRUCTION

1) INITIAL DATABASE CONSTRUCTION

At a botanical garden, we collected hundreds of thousands of trees in six different views in different environments and shapes for training and test sets during database construction. These six views include the entire plant, branches, flowers, fruit, leaves, and stem parts^[28].

Several pictures of the collected trees are shown in Figure 2. below.

FIGURE 2. Example Image of 6 Views of Trees in Training Set

The specific distribution of the training set and the test set in the acquired image is as follows.

TABLE I
TRAINING TEST SET DISTRIBUTION TABLE

View	Training	Testing
Branch	8588	2330
Entire	15233	5308
Stem	5468	500
Flower	13458	2500
Fruit	6588	1265
Leaf	19456	5544
Total	68791	17447

2) DATA SAMPLE SCREENING

We firstly pre-trains a model on the basis of CaffeNet convolutional neural network, and then uses the image fine-tuning strategy to transmit the learned identification information to get a better learning model, and realize the weight sharing of the two models of the two tasks. The formed joint features are further iteratively optimized by the BP algorithm, and finally an Auto-Clean CNN model achieving the dual tasks of experimental image cleaning and tree classification is designed. In the training set of the six tree views, we randomly selected 5000 images for the micro-training of the improved Dual-Task CNN network model, and decided the best view according to the classification performance of each model, and then use the best view as the data set for later model training optimization^[29].

The classification accuracy of the six views of the tree varies with the number of iterations is shown in the Figure.3 below.

FIGURE A. Entire Image

FIGURE B. Stem Image

FIGURE C. Fruit Image

FIGURE D. Flower Image

FIGURE E. Branch Image

FIGURE F. Leaf Image

FIGURE 3. Comparison of Classification Accuracy of Trees 6 Views

It can be seen from Fig. 3(A) to Fig. 3(F) that after 5000 iterations, the recognition accuracy rates of the Fruit, Entire and Branch views are not more than 70%. The universality of the three views for the identification of tree species is generally not high, and it is difficult to identify in different scenarios accurately and effectively. In the six possible views of plants, bark and stems are important organs for trees so they are not suitable for identification. The use of flowers and fruits image does not have the universal significance of research for trees without flowers and fruits and cannot be the basis for the identification of trees. The leaves are the longest-lived organs in the various organs of the tree. The extraction process does not affect the normal growth of the trees, and the scanned images of the flowers and leaves have higher classification accuracy in 6 views, so we screened out all the tree leaf image sets for the latter training optimization.

3) DATA SET EXTENSION

When training deep convolution neural networks, the accuracy of the network can be improved and over-fitting can be prevented by using a large amount of training data. However, in the actual operation process, it is often difficult to obtain enough training samples, but the artificial expansion of the data can achieve similar effects, and can enhance the generalization ability and robustness of the training model. The data set used includes a total of 50

classes, and finally obtains 100,000 images by means of rotation, zooming and horizontal flipping^[30-31]. In this experiment, the data set is divided into two categories, 70% is the training set, and 30% the test set. The data is preprocessed by manual screening and script classification to obtain the tree image data set for training experiments, and finally 100,000 images of trees for training experiments were obtained.

FIGURE 4. Data Set Extension Example

B. DATA PREPROCESSING

1) SCRIPT CLASSIFICATION

Before using the convolutional neural network to classify the dataset, it is necessary to classify the messy tree image data in the database and store it in the folder corresponding to the tree type. We use scripts to implement the function of classifying trees into corresponding folders to facilitate the training, verification and testing of the later models. English, please get a native English-speaking colleague to carefully proofread your paper.

2) MANUAL SCREENING

A good data set is an important prerequisite for training a good model. Therefore, before inputting a tree image to a training model, it is necessary to clean the data set to ensure the authenticity and robustness of the training as much as possible. When there are fewer pictures in the current period, we mainly use manual cleaning to remove unqualified dirty pictures visible to human eyes, and whether human eyes can identify dirty pictures as the overall screening standard. The specific cleaning rules are as follows.

Blurred or Unrecognized pictures of tree species	Delete
Images with multiple distractions or species	Delete
Pictures with severe picture damage	Delete
Other pictures	Store in experimental data set

III. METHODS

RAPID RECOGNITION SYSTEM FOR TREE IMAGE BASED ON CAFFE PLATFORM AND DEEP LEARNING

A. DUAL-TASK CONVOLUTIONAL NEURAL NETWORK BASED ON CAFFE PLATFORM

Caffe has advanced deep learning algorithms and reference models to provide a good starting point for new research and applications, making a clean and modifiable framework available for many practitioners. The classification model trained by the Caffe team using magenet images is already very good and worth learning. When using our own data set to train, if the sample size is too small, the model has too low precision and takes a long time. However, using the parameters in the model that has been trained in Caffe as our initialization parameters reduces the requirements of training accuracy of the model on the data set, and improves the classification performance of the same data set training. Therefore, the Auto-Clean CNN model of this paper is based on CaffeNet's network framework and then uses our own dataset to train the model's feature extractor and classifier, and then the initialization parameters are continuously updated and optimized. After the entire fine-tuning process, we get our optimization model called Auto-Clean CNN^[32-33].

B. DESIGN OF DUAL-TASK CONVOLUTIONAL NEURAL NETWORK MODEL

The Auto-Clean convolutional neural network we designed based on the famous architecture proposed by Krizhevsky in 2012 mainly includes five convolutional layers, two connection layers, and a softmax loss layer, and among the layers also exist pooling, regularization, RELU activation functions and other operations. However, the network structure of the two models in Dual-Task CNN, which implements image cleansing and tree multi-class recognition, is not identical.

The first fully connected layer of each CNN model is represented as a feature vector of the input image, and the formed joint features are then input into two different specific softmax loss layers. These classification tasks collectively use the feature vectors output by the convolutional network and jointly update the network parameters during training. The structure of the improved Dual-Task CNN model is shown in Figure 5 below.

TABLE II
MANUAL CLEANING RULE TABLE

Special circumstances	Rules
-----------------------	-------

FIGURE 5. Dual-Task CNN Overall Structure

As can be seen from Figure 5, the input image with the attribute tag information (left side) is entered into the model, and each CNN will predict a binary attribute. First, the two CNN models in the Dual-Task CNN will form a joint feature between the last convolutional layer and the first fully connected layer through forward propagation after inputting the original image. Then, the formed joint features are input to the specific softmax loss layer and the multi-class softmax loss layer to complete the two tasks of tree classification and image cleaning. After learning the hidden layer parameters to achieve hard sharing and then iteratively optimizing the network through the back propagation algorithm, the risk of over-fitting is reduced, and the generalization ability of the model is also improved^[34]. For the Clean CNN network model, the training task is simpler than tree multi-class recognition, and image cleaning is a pre-processing of multi-level prediction. The last layer of the fully-connected layer can reduce a large number of errors in the training process and improve the generalization performance, but the fully-connected layer contains a large number of redundant parameters, only the fully-connected layer parameters can account for about 80% of the entire network parameters. This is not conducive to the rapid cleaning of the image, so in the clean CNN model we remove the last fully connected layer and directly connect the first fully connected layer to the softmax layer. Although the performance of the model is slightly reduced, the cleaning speed will be greatly improved.

C. CROSS ENTROPY LOSS FUNCTION

We use the learning method of forward propagation and backpropagation in convolutional neural networks. Firstly, we calculate the cross entropy loss function of Auto task and Clean task, and then minimize the loss function of combining the two tasks and update the weight through Auto-Clean convolution model to obtain a better identification model^[35-36].

The formula of the Cross Entropy Loss Function in the feature extraction process is described as follows:

We give N training picture sets and corresponding label sets $D = \{I_i, C_i, K\}_{i=1}^N$, where C_i indicates whether the picture is a dirty one and it is dirty when the input image cannot be classified, I_i represents the tree category to which the picture belongs, and K indicates the total number of trees. High-dimensional features $X_i \in R^{D_b \times 1}$ are extracted by

the above-mentioned trained Auto-Clean CNN model, which can be expressed as:

$$X_i = f(I_i, k, b, \beta, \gamma) \quad (1)$$

Where: $f(*)$ represents the nonlinear mapping of the input picture to the first fully connected layer feature, k and b represent the set of all convolution layer filters and offsets, β and γ represent the zooming and rotation of all bottom layers, and $\theta = (k, b, \beta, \gamma)$ represents a collection of all extracted feature parameters. The extracted feature X_i is shared in image clearing and multi-class prediction.

Assuming that $W^d \in R^{D_d \times D_d}$ and $b_k \in R^{D_d \times 1}$ are weight matrices and offset vectors in the fully connected layer, where D_d is the number of different trees in D , the resulting linear model is

$$y_i^d = W^d \cdot T_{X_i} + b^d \quad (2)$$

Where y_i^d output to the softmax layer is used to calculate the probability size of each tree category X_i in the training set.

$$\text{softmax}(y_i^d)_{K_n} = p(y_i^d = K_n | x_i) = \frac{\exp(y_{iK_n}^d)}{\sum_j \exp(y_{iK_j}^d)} \quad (3)$$

Where y_{ij}^d indicates the probability that X_i belongs to the j th tree, and the softmax(*) function maps the model's output y_i^d to the probability distribution of all tree images and the K_i image of the tree image. Finally, the result of the tree image can be obtained from

$$\hat{y}_i^d = \arg \max_{K_n} \text{softmax}(y_i^d)_{K_n} \quad (4)$$

Then the Cross Entropy Loss Function can be expressed as

$$L(I_i, y_i^d) = -\log(p(\hat{y}_i^d = y_i^d | I_i, \Phi, W^d, b^d)) \quad (5)$$

The loss function of the clean task is calculated in the same way as the auto task.

For the classification of tree images, image clearing and the connection of two fully connected layers, we use $W = \{W^d, W^c, W^f\}$ to represent its weight matrix to simplify the elimination of the offset coefficient. For a given training set D , the Auto-Clean convolution model minimizes the loss function combined by the two tasks. The loss function is

$$\Phi, W = \arg \min \sum_{i=1}^N \alpha_d L(I_i, y_i^d) + \alpha_c \sum_{i=1}^N L(I_i, y_i^c) \quad (6)$$

Where α_d and α_c , represent the weight of each task, respectively, and when α_d or α_c is zero, it is a single task model.

At the same time, the loss function drives the Auto-Clean CNN model to learn to extract feature parameters ϕ and

classification task weights W . During the test, we first test whether the input image is a clean one. If the image entered cannot be classified accordingly, the model will output the image as a dirty one. If the image is a clean one, the model will output the predicted tree category.

D. ADAPTIVE WEIGHT LEARNING

In the MTL-based convolutional neural network, how to set the weight of two tasks is also an important part of the research. In our previous work, we either treated equally the tasks we encountered or obtained weights through brute force search, and searching all weight combinations was time-consuming. We further train and optimize the model by automatically performing weight adjustments by using a mini-batch random gradient descent method in convolutional neural networks^[37].

Auto-Clean CNN automatically assigns α_d and α_c through the vector α_s . First, a softmax layer is added to the first fully connected layer, and $\omega_s \in R^{D_b \times 2}$ and $\varepsilon_s \in R^{2 \times 1}$, respectively, represent the weight matrix and offset vector corresponding to the weighted softmax layer. We have

$$\alpha_s = \text{soft max}(\omega_s T x_i + \varepsilon_s) \quad (7)$$

If $\alpha_d + \alpha_c = 1$, then $\alpha_s = [\alpha_d, \alpha_c]T$ is the dual task's adaptive loss rate. The formula is shown as follows:

$$\Theta, W, \alpha_s = \arg \min \sum_{i=1}^N \alpha_d L(I_i, y_i^d) + \alpha_c \sum_{i=1}^N L(I_i, y_i^c) \quad (8)$$

$$\alpha_d + \alpha_c = 1 \quad (9)$$

When the mini-batch random gradient descent method is used to solve the above optimization problem, the weight adjustment is aggregated in the batch sample, and the final model yields the setting weights of the two tasks through adaptive learning.

E. SUPER PARAMETERS TUNING

1) PARAMETER CONFIGURATION OF TRAIN.PROTOTXT FILE

The parameter configuration file defines super-parameters such as the learning rate multiple of the weight parameter and the offset parameter, the convolution kernel size, the step size, the down sampling parameter, the full connection layer parameter and the classification accuracy rate, designs the object needed to be optimized by the model, train the network and test the network.

2) PARAMETER CONFIGURATION AND OPTIMIZATION OF SOLVER.PROTOTXT FILE

This parameter configuration file defines some super parameters that need to be set during the network model training process, such as the basic learning rate, the weight attenuation coefficient, the number of iterations, the momentum, and whether to use the GPU or the CPU. Since the loss function is not absolutely a convex function, it may be only a local optimum value found according to the CNN model described above, and optimization to it is also needed.

The file calculates the loss in the forward process and the gradient in the reverse process, and alternately calls the forward and backward algorithms to iteratively update the parameters. The process is as follows:

Step 1: Call the train.prototxt configuration file, design the objects that need to be optimized, the training network for learning and the test network for evaluation;

Step 2: Update parameters by forward and backward iterations;

Step 3: Regularly evaluate the test network and set the test cycle ;

Step 4: display the state of the model and solver during the optimization process.

F. COMPARISON BETWEEN CNN MODEL AND EXISTING METHODS

By reviewing some data and previous experiments, we found that traditional learning methods such as SVM and BP neural network are not effective enough to identify complex tree images in comparison with CNN models. Even if there is a over-fitting in CNN models, CNN models' experimental results are better than these traditional learning algorithms. In the improved research of CNN model, the Dual-Task convolutional neural network we designed includes two model structures of image cleaning and tree recognition compared with the ordinary single-task CNN network, and it can extract more representative features based on the correlation between the two tasks. It not only improves the accuracy of tree identification and saves labor costs, but also reduces the burden on trainers and promotes the long-term development of tree classification.

Among them, the recognition accuracy of various tree classification methods is shown in Table 3 below.

TABLE III
COMPARISON OF CLASSIFICATION PERFORMANCE OF EXISTING METHODS

Existing Models	Classification accuracy of test set (%)
BP neural network ^[15]	87.5%
SVM ^[38]	85.12%~90.625%
Single-Task CNN	91.1%
Dual-Task CNN	95.6%

1) COMPARISON OF RECOGNITION ACCURACY OF IMPROVED CNN MODEL

Figure 6 shows the average classification accuracy of the Dual-Task CNN model and Single-Task CNN model. It can be seen from the graph that the Dual-Task CNN model has higher average classification accuracy than the Single-Task CNN model. Figure 7 shows the variation of the classification accuracy of 25 species under the Dual-Task and Single-Task models with the number of iterations. When the number of iterations is 80000, the classification accuracy of the Dual-Task CNN model is increased from 91.1% in the single task to 95.6%, which is a significant improvement.

FIGURE 6. Average classification accuracy of dual-task and single-task

FIGURE 7. Classification accuracy of Single-Task CNN with iterations

FIGURE 8. Classification accuracy of Dual-task CNN with iterations

2) COMPARISON OF RECOGNITION SPEED OF IMPROVED CNN MODEL

Figure 8. shows the recognition time of different types of trees under the dual-task and single-task models. It can be seen that compared with the single-task model, the dual-task model not only improves the classification accuracy, but also

shortens the recognition time of various trees by half. The unique design of the Auto-Clean model making the addition of an image cleaning task do not increase the recognition time of the model. On the contrary, the joint features formed by the two related tasks promote the learning of tree classification, which greatly reduces the recognition time of trees, thus the accuracy and feasibility of the dual-task classification based on the Auto-Clean CNN model designed are demonstrated.

FIGURE 9. Various tree identification times under the two-task and single-task models

3) COMPARISON OF CLASSIFICATION PERFORMANCE WITH OR WITHOUT DATA SET PREPROCESSING

The following two pictures show the comparison of the classification accuracies between the dual-task and single-task models without dataset pre-processing. It can be seen from the four pictures that high quality training data is necessary. Comparing the classification accuracy with or without pre-processing, the classification performance of the model recognition time and recognition accuracy after data pre-processing is better, and the classification accuracy of the single-task model after pre-processing is approximately equal to that of the dual-task model without pre-processing, and there is even a tendency to exceed.

FIGURE 10. Classification accuracy of Single-Task CNN without iterations

FIGURE 11. Classification accuracy of Dual-task CNN without iterations

G. DESIGN OF RE-IMPROVED DEEP GABOR CNN MODEL

1) INFLUENCE OF SPECIES SIMILARITY ON CLASSIFICATION PERFORMANCE OF CNN MODEL

From the Single-Task CNN to the Dual-Task CNN model, the accuracy of image recognition of 25 types of trees have been greatly improved. However, by analyzing the classification results of various tree species, it is found that the improvement of the classification effect of the Dual-Task CNN model compared with the ordinary CNN network is mainly concentrated on the leaf species with obvious characteristics, and for the identification of similar tree species, the effect of the Dual-Task CNN model is still unsatisfactory.

FIGURE A. Banana tree

FIGURE B. Ginkgo tree

FIGURE 12. Tree species with distinct characteristics

FIGURE A. Osmanthus tree

FIGURE B. Mango tree

FIGURE 13. Tree species with similar characteristics

The accuracy of the above tree species image recognition is shown in the table below. By analyzing the recognition

accuracy of 25 categories and the similarity with other species in the dataset, we can also conclude that the classification accuracy of Dual-Task CNN compared to the Single-Task CNN model is mainly reflected in species with low similarity. The higher the species similarity, the lower the classification accuracy of the class.

TABLE IV
THE CLASSIFICATION PERFORMANCE OF DUAL-TASK CNN VARIES WITH THE SIMILARITY OF TREE SPECIES

Tree Species	Single-Task CNN(%)	Dual-Task CNN(%)
Banana tree	91.7%	97.0%
Ginkgo tree	92.0%	97.5%
Osmanthus tree	89.0%	90.0%
Mango tree	84.0%	84.7%

In the existing spatial information extraction technology, the Gabor filter can enhance the robustness of feature learning for different scales and directions, especially for images with spatial transformation, so in order to improve the classification effect of DTY model on similar species, a new deep Gabor Convolutional Neural Network (GCNN) is proposed. We introduce the Gabor filter to generate the Gabor feature as the input of the Dual-Task CNN, thus enhancing the extraction of the edge and texture features of the blade image, and achieving the purpose of improving the recognition rate of similar tree species. Moreover, the re-improved GCNN is easy to implement and can be compatible with the Caffe deep learning framework used in this experiment.

2) GABOR FILTERS

In image processing, the Gabor function is a linear filter for edge extraction, which is very suitable for the expression and separation of image texture. In the spatial domain, a two-dimensional Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave and consists of a real part and an imaginary part. The real part of the Gabor filter can be expressed as the following formula.

$$g(x, y, \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right) \quad (10)$$

where the values of x and y are:

$$x' = x \cos \theta + y \sin \theta \quad (11)$$

$$y' = -x \sin \theta + y \cos \theta \quad (12)$$

From a wide range of experiments, we have found that the value of the wavelength is specified in pixels, and it is usually greater than or equal to 2, but cannot be greater than one-fifth of the size of the input image. What we usually use is that its reciprocal frequency is greater than or equal to 0.2. The parameter of direction (θ) specifies the direction of the parallel stripe of the Gabor function, which takes values from 0 to 360 degrees. The phase offset (φ) ranges from -180 degrees to 180 degrees. Among them, 0 and 180 degrees respectively correspond to the center-symmetric center-on function and the center-off function, and -90 degrees and 90

degrees correspond to the antisymmetric function. The aspect ratio (γ) refers to the spatial aspect ratio, which determines the shape of the Gabor function. When $\gamma = 1$, the shape is round, and when $\gamma < 1$, the shape is elongated with the direction of the parallel stripe.

We use the Gabor filter to replace the conventional first layer convolutional layer, and the extracted feature vector is input into the CNN model for classifier training. In previous experiments, a convolutional layer was usually used as the first layer. However, experiments show that the Gabor filter improves the results for the feature extractor. The improved depth Gabor CNN model can effectively enhance the recognition of similar species.

IV. DISSCUSION

A. THE CONSTRUCTION OF HARDWARE AND SOFTWARE ENVIRONMENT

This study uses a desktop computer configured with Intel i5 GPU, 8GB of running memory, and ROG STRIX-GTX1060 GPU. The experimental software environment is Ubuntu 16.04. 1 Desktop TLS operating system, and Caffe is installed as a framework for deep learning. The data set used contained 15,000 tree leaf images preprocessed by manual screening and script classification and expanded to 100,000 images finally, and the training set accounted for 80% and the test set 20%.

B. RAPID RECOGNITION OF TREE IMAGES BASED ON GCNN

1) STEPS DESCRIPTION

Experiments were carried out on the data sets by using CaffeNet CNN. In order to obtain better experimental results and speed up the experimental training, we can obtain better classification results by parameter tuning of pre-trained model in the network. The experimental steps are as follows:

Step 1: Clean and classify pictures, eliminate blurred pictures and other pictures with more distracting items, and generate listing files. Due to the small quantity of pictures at the beginning, we can use the manual classification method to pre-process the obvious pictures, such as removing blurred, multi-interference and multi-tree pictures, and then transfer them to the model for secondary cleaning;

Step 2: Convert images into lmdb format that is easily processed by the Caffe framework and calculate their mean files to speed up and enhance accuracy of training and testing;

Step 3: When designing the model to be used for training, use the training model that comes with the Caffe framework, and modify the layer of the model according to our needs;

Step 4: Configure the parameters in the train.prototxt and solver.prototxt files, and alternately call the forward and backward algorithms to iteratively update the parameters to avoid local optimization of non-convex functions, and train the dataset to generate a caffemodel file;

Step 5: Adjust the deploy.prototxt file and convert the mean file to a mean.npy file;

Step 6: Call Caffe's python interface, make the above modifications based on Caffe's own Caffenet model and use our dataset to train the classification model, and then use the classification model to identify the test image.

V. RESULTS

First, the following 50 CNN models were trained in 50 tree species. It can be seen from Figure 14 that the improved Dual-Task CNN model has a higher classification accuracy than the Single-Task CNN model, and the re-improved GCNN model has better classification accuracy for the data set.

FIGURE A. Classification accuracy of the tree varies with the number of iterations

FIGURE B. Test loss of the tree varies with the number of iterations

FIGURE 14. Comparison of classification performance of 50 tree species

The DCNN and GCNN models obtained by training two types of similarly similar species are shown in Figure 15 and Figure 16. When the number of iterations reaches 4000, the classification accuracy of the GCNN model is increased from 87% to 98% compared with the DCNN model. From the ROC curve, the GCNN model is also better than the DCNN model for the classification performance of similar species.

FIGURE A. Classification accuracy of the tree varies with the number of iterations

FIGURE B. Test loss of the tree varies with the number of iterations

FIGURE 15. Tree species with similar characteristics

FIGURE 16. Comparison chart of ROC curve

VI. CONCLUSIONS

Aiming at the problem of complex image extraction in existing complex backgrounds, we propose an Auto-Clean convolutional neural network model that can achieve image

cleaning and multi-class prediction of trees, and improve the generalization and classification ability of the model. The joint feature shares through parallel training and learning of related tasks, and then the BP algorithm and the mini-batch random gradient descent algorithm are used to optimize the model. In order to avoid local optimization of non-convex functions, the super parameters in the model are selected and tuned, and the training model is further trained. Finally, upon the above improvements and experiments, the training precision of the Auto-Clean dual-task model and the ordinary single-task model can reach 96.0% and 91.7%, and the recognition time is as low as 2s, achieving the goal of efficiently and accurately identifying the number of types.

Then, in order to reduce the interference of species similarity on the classification performance of Dual-Task CNN model and further improve the classification performance of the tree image recognition system, we propose a re-improved deep Gabor convolutional neural network model called GCNN. The Gabor filter is used to enhance the extraction of edge and texture features of the tree image, and the extracted features are input into the CNN model to achieve the goal of enhancing the recognition accuracy of similar species. The experimental results also show that the re-improved deep Gabor convolutional neural network is more advantageous in distinguishing similar trees.

At present, recognition work of tree image is also interfered by many factors such as species type, species similarity, and complexity of the surrounding environment. However, with the further research and improvement of convolutional neural networks, it is believed that there will be more models with excellent performance. It can be applied to the classification and identification of trees to make efforts for the improvement and protection of the ecological environment.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation in China (Grant Nos. 61703441)

REFERENCES

- [1] Tyrväinen, Liisa, et al. "Benefits and uses of urban forests and trees." Urban forests and trees. Springer, Berlin, Heidelberg, 2005, pp. 81-114.
- [2] Yu, Z. H. A. N. G. "Urban Garden Planning and Design and Ornamental Plant Protection." Journal of Landscape Research, vol.10, no.4, pp. 159-162, Aug. 2018.
- [3] Feng Ma, "Analysis of Forestry Ecological Sustainable Development Measures." Seed science & technology, vol.38, no. 11, pp. 23-24, Dec. 2018.
- [4] Lindenmayer, David B., Chris R. Margules, and Daniel B. Botkin. "Indicators of biodiversity for ecologically sustainable forest management." Conservation biology, vol.14, no.4, pp.941-950, Aug. 2000.

- [5] Cronquist, Arthur, and Armen Leonovich Takhtadzhian. An integrated system of classification of flowering plants. Columbia University Press, 1981.
- [6] Li, Apg. "An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II." *Botanical journal of the Linnean Society*, vol.141, no.4, pp.399-436, 2003..
- [7] Angiosperm Phylogeny Group. "An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III." *Botanical Journal of the Linnean Society*, vol.161, no.2, pp.105-121, 2009.
- [8] Liang Chang, Xiao-Ming Deng, Ming-Quan Zhou, Zhong-Ke Wu, Ye Yuan, Shuo Yuan, Hong-An Yuan. "Convolutional neural networks in image understanding," *Acta Automatica Sinica*, vol.42, no.9, pp.1300-1312, Sep. 2016
- [9] Fei-Yan Zhou, Linpeng Jin, Jun Dong. "Review of Convolutional Neural NetworkChinese," *Chinese Journal of Computers*, vol.40, no.6, pp.1229-1251, Jun. 2017.
- [10] Weibo Liu, et al. "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp.11-26, 2017.
- [11] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview," *Neural networks*, vol.61, pp.85-117, 2015.
- [12] Dinuls R, Erins G, Lorencs A, et al. "Tree species identification in mixed Baltic forest using LiDAR and multispectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.5, no.2, pp. 594-603, Feb. 2012.
- [13] Du J X, Wang X F, Zhang G J. "Leaf shape based plant species recognition," *Applied mathematics and computation*, vol.185, no.2, pp.883-893, Feb. 2007.
- [14] Tong Hou, Lihong Yao, Jiangming Qi. "Research on Plant Identification Based on Leaf Shape Characteristics," *Hunan Agricultural Sciences*, vol.39, no.4, pp.123-125, Apr. 2009 .
- [15] Wu, Stephen Gang, et al. "A leaf recognition algorithm for plant classification using probabilistic neural network," 2007 IEEE international symposium on signal processing and information technology. IEEE, 2007.
- [16] Liu J, Zhang S, Deng S. "A method of plant classification based on wavelet transforms and support vector machines," *International Conference on Intelligent Computing*. Springer, Berlin, Heidelberg, 2009, pp. 253-260.
- [17] Priya, C. Arun, T. Balasaravanan, and Antony Selvadoss Thanamani. "An efficient leaf recognition algorithm for plant classification using support vector machine," *International conference on pattern recognition, informatics and medical engineering (PRIME-2012)*. IEEE, 2012.
- [18] Lee, Sue Han, et al. "Deep-plant: Plant identification with convolutional neural networks." 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015.
- [19] Kang, Le, et al. "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks." 2015 IEEE international conference on image processing (ICIP). IEEE, 2015.
- [20] Korpusik, Mandy, and James Glass. "Convolutional neural networks and multitask strategies for semantic mapping of natural language input to a structured database." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp.6174-6178, 2018.
- [21] Sindagi, Vishwanath A., and Vishal M. Patel. "Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting." 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2017.
- [22] Guo, Michelle, et al. "Dynamic task prioritization for multitask learning." *Proceedings of the European Conference on Computer Vision (ECCV)*, pp.270-287, 2018.
- [23] Liu, Xiaoshuang, et al. "Driver Fatigue Detection Using Multitask Cascaded Convolutional Networks." *International Conference on Intelligence Science*. Springer, Cham, pp.143-152, 2017.
- [24] Sarwar, Syed Shakib, Priyadarshini Panda, and Kaushik Roy. "Gabor filter assisted energy efficient fast learning convolutional neural networks." 2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED). IEEE, pp.1-6, 2017.
- [25] Yao, Hu, et al. "Gabor feature based convolutional neural network for object recognition in natural scene." 2016 3rd International Conference on Information Science and Control Engineering (ICISCE). IEEE, pp.386-390, 2016.
- [26] David, Emilian, Paul Ungureanu, and Liviu Goras. "On the feature extraction performances of CNN gabor-type filters in texture recognition applications." 2006 10th International Workshop on Cellular Neural Networks and Their Applications. IEEE, pp.1-6, 2006.
- [27] Calderón, Andrés, Sergio Roa, and Jorge Victorino. "Handwritten digit recognition using convolutional neural networks and gabor filters." *Proc. Int. Congr. Comput. Intell*, 2003.
- [28] Reyes A K, Caicedo J C, Camargo J E. "Fine-tuning Deep Convolutional Networks for Plant Recognition," *CLEF (Working Notes)*, 2015.
- [29] Sladojevic, Srdjan, et al. "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational intelligence and neuroscience*, 2016.
- [30] Zhou, Hao, et al. "Faster R-CNN for Marine Organism Detection and Recognition Using Data Augmentation,"

Proceedings of the International Conference on Video and Image Processing. ACM, 2017.

- [31] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks," European conference on computer vision. Springer, Cham, 2014.
- [32] Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding," Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.
- [33] Yanai, Keiji, and Yoshiyuki Kawano. "Food image recognition using deep convolutional network with pre-training and fine-tuning," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2015.
- [34] Han, Song, et al. "Learning both weights and connections for efficient neural network," Advances in neural information processing systems, 2015.
- [35] Panchapagesan, Sankaran, et al. "Multi-Task Learning and Weighted Cross-Entropy for DNN-Based Keyword Spotting," Interspeech. 2016.
- [36] Collobert, Ronan, and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning." Proceedings of the 25th international conference on Machine learning. ACM, 2008.
- [37] Chun-Yun Zhang, Peng-Da Qin, Yi-Long Yin. "Self-adaptation multi-gram weight learning strategy for sentence representation based on convolutional neural network", Computer Science, vol.44, no.1, pp.66-64, Jan. 2017.
- [38] Yang Yang. "Research of tree species identification based on the theory of wavelet transform and SVM." Northeast Forestry University, 2017.

Figures

Figure 1

Data Set Construction and Preprocessing

Figure 2

Example Image of 6 Views of Trees in Training Set

FIGURE A. Entire Image

FIGURE B. Stem Image

FIGURE C. Fruit Image

FIGURE D. Flower Image

FIGURE E. Branch Image

FIGURE F. Leaf Image

Figure 3

Comparison of Classification Accuracy of Trees 6 Views

Figure 4

Data Set Extension Example

Figure 5

Dual-Task CNN Overall Structure

Figure 6

Average classification accuracy of dual-task and single-task

Figure 7

Classification accuracy of Single-Task CNN with iterations

Figure 8

Classification accuracy of Dual-task CNN with iterations

Figure 9

Various tree identification times under the two-task and single-task models

Figure 10

Classification accuracy of Single-Task CNN without iterations

Figure 11

Classification accuracy of Dual-task CNN without iterations

FIGURE A. Banana tree

FIGURE B. Ginkgo tree

Figure 12

Tree species with distinct characteristics

FIGURE A. Osmanthus tree

FIGURE B. Mango tree

Figure 13

Tree species with similar characteristics

FIGURE A. Classification accuracy of the tree varies with the number of iterations

FIGURE B. Test loss of the tree varies with the number of iterations

Figure 14

Comparison of classification performance of 50 tree species

FIGURE A. Classification accuracy of the tree varies with the number of iterations

FIGURE B. Test loss of the tree varies with the number of iterations

Figure 15

Tree species with similar characteristics

Figure 16

Comparison chart of ROC curve