International Journal of Computational Intelligence Systems
Vol. **13(1)**, 2020, *pp.* 559–564
DOI: https://doi.org/10.2991/ijcis.d.200425.001; ISSN: 1875-6891; eISSN: 1875-6883
https://www.atlantis-press.com/journals/ijcis/

Research article

Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model

Zhenbo Li[1,2,3], Fei Li[1,2,3,*], Ling Zhu[1,2], Jun Yue[4]

[1]*College of Information and Electrical Engineering, China Agricultural University, No. 17 Tsinghua East Road, Haidian District, Beijing, 100083, China*
[2]*Computer Version Group, Key Laboratory of Agricultural Information Acquisition Technology, No. 17 Tsinghua East Road, Haidian District, Beijing, 100083, China*
[3]*Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, No. 17 Tsinghua East Road, Haidian District, Beijing, 100083, China*
[4]*College of Information and Electrical Engineering, LuDong University, 186 Hongqi W Road, Zhifu District, Yantai, 264025, China*

ARTICLE INFO

Article History
Received 18 Jul 2019
Accepted 04 Apr 2020

Keywords
Vegetable recognition and classification
Deep learning
VGG-nets
Framework of caffe

ABSTRACT

To improve the accuracy of automatic recognition and classification of vegetables, this paper presents a method of recognition and classification of vegetable image based on deep learning, using the open source deep learning framework of Caffe, the improved VGG network model was used to train the vegetable image data set. We propose to combine the output feature of the first two fully connected layers (VGG-M). The Batch Normalization layers are added to the VGG-M network to improve the convergence speed and accuracy of the network (VGG-M-BN). The experimental verification, this paper method in the test data set on the classification of recognition accuracy rate as high as 96.5%, compared with VGG network (92.1%) and AlexNet network (86.3%), the accuracy rate has been greatly improved. At the same time, increasing the Batch Normalization layers make the network convergence speed nearly tripled. Improve the generalization ability of the model by expanding the scale of the training data set. Using VGG-M-BN network to train different number of vegetable image data sets, the experimental results show that the recognition accuracy decreases as the number of data sets decreases. By contrasting the activation functions, it is verified that the Rectified Linear Unit (ReLU) activation function is better than the traditional Sigmoid and Tanh functions in VGG-M-BN networks. The paper also verifies that the classification accuracy of VGG-M-BN network is improved due to the increase of batch_size.

1. INTRODUCTION

In recent years, the development model of agriculture in China is changing from traditional agriculture to modern intelligent agriculture; the production of vegetables in agricultural products is also increasing. However, at present, vegetable picking, sorting and sales tasks still rely on manual completion, not only a large amount of labor force is consumed, but also the work efficiency is low, which seriously affecting the development of commercialization of vegetable products. The research of automatic recognition and classification [1–3] of vegetables provides important technical support to solve the above problems.

Scholar at home and broad have been devoted to fruit recognition and classification and significant progress has been made in this field. Pragati *et al.* [3] introduce new fruits recognition techniques. This combines four features analysis method shape, size and color, texture based method to increase accuracy of recognition. Proposed method used is nearest neighbor (NN) classification algorithm. Dubey *et al.* [4] extracted different state-of-art color and texture features and combined them to achieve more efficient and discriminative feature description. Multi-class support vector machine is used for the training and classification purpose. The experimental results show that this strategy combining multiple methods can achieve better recognition efficiency. Zhang *et al.* [5] proposed a method based on fitness-scaled chaotic artificial bee colony algorithm and feedforward neural network (FSCABC-FNN). The network extracted color, shape and texture as features. The method reached a rather good accuracy. In the meanwhile, Tao H W *et al.* [6] propose a method for recognition fruit and vegetable types based on color and texture features. Used a matching score fusion algorithm to fuse color and texture features, and finally, a NN classifier is used to realize fruit and vegetable recognition. As we can see, a majority of methods on fruit classification are traditional and old-fashion. Linear classifier and KNN classifier take great part in fruit classification and features are rare, which limits the development and accuracy of methods.

The concept of deep learning [7–11] originates from artificial neural networks. Deep learning forms more abstract high-level features through the combination of low-level features, which to find the distributed features of the data. In 2012, Hinton's team, known

*Corresponding author. Email: leefly072@126.com

as the "father of deep learning," won the ImageNet image classification [12–15] contest using deep learning methods. The classification accuracy was over 10% higher than the second was. The result was a huge shock in the field of computer vision; it has set off a craze for deep learning in academia. The deep learning network mainly includes Stacked AutoEncoder, Restricted Boltzmann Machine, Deep Belief Network and Convolutional Neural Network. Among them, the convolution neural network (CNN) [16–20] has the most significant effect in the image recognition task. What is more, CNN is applied to fruit and vegetables classification as well, which takes a great effect. For instance, Ashutosh *et al.* [21] experimented on food or nonfood classification and food recognition utilizing a pre-trained GoogleNet model. The results showed that the accuracy of food and nonfood classification was 99.2%, while the accuracy of food classification was only 83.6%. Li *et al.* [22] proposed an integrated convolutional neural network method to solve the problem of food type recognition in refrigerator. Effectively enhance the dominant role of color in object recognition. Improved the problem of low accuracy due to occlusion and Angle change, the average accuracy was 92.2%. Therefore, in order to further improve the recognition and classification accuracy of vegetables and speed up the convergence rate of the network, this paper proposes to use the improved VGG convolutional neural network model to achieve the recognition and classification of vegetables.

In this paper, based on the improved VGG [23] network model, the method of recognition and classifi-cation of vegetable images was established, and 10 kinds of vegetable image data sets were built, which were divided into training set and test set. Combining the output feature of the first two full-connected layers to improve the recognition and classification accuracy of the vegetable images, add the BN layers to the network to improve the convergence speed. Using Caffe [24,25] of the open source deep learning framework, the vegetable image dataset was trained with the improved VGG network model and tested on the test set.

2. THE MAIN TEXT

The structure of the paper is as follows: Section 3 introduces our work in image database, the use of improved VGG model. Section 4 presents the experiments and our results step by step. Finally, Section 5 draws conclusions and future tasks.

3. INSTRUMENTS AND METHODS

The model data and methods used in the study are described in detail in the following sections.

3.1. Vegetable Image Dataset Preparing and Preprocessing

We have been to build the vegetable image data set and expanded to 48,000 images. Among them, the training set accounted for 80% (38,400) and the test set accounted for 20% (9,600). The images were obtained from the image database of ImageNet (50%), website crawler (30%) and shoot by myself (20%). This paper was trained ten categories of vegetables, they are broccoli, pumpkin, cauliflower, mushrooms and cucumber, Chinese cabbage, tomato, eggplant, garden pepper and carrots. Because the size of each image in the vegetable image dataset is different, the image size is normalized to 224*224. The vegetable image dataset used in this paper has the same number of images per category of vegetables (1200). The number of images is relatively small; therefore, the vegetable image data set is rotated at 90, 180 and 270 degrees by method of data expansion, trained by this method, the image data set have been enlarged by 4 times, the rotation effects of vegetable images are shown in Figure 1.

In Figure 1, each category of vegetables from left to right is the image of original, rotating 90°, 180° and 270°. Each image size is different in vegetable image data set, in order to neatly arrange the image, the image is processed into the same size in Figure 1 (80*80). The data expansion method is suitable for training and testing images. In the training phase, the increase data can produce additional training samples, thereby reducing the impact of overfitting. In the test phase, the increase in data helps to improve the classification accuracy rate.

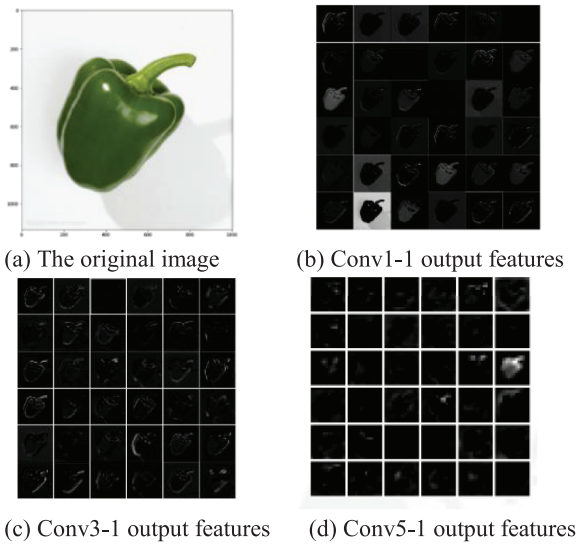
3.2. Improvement of VGG Model

In order to improve the classification accuracy of vegetables in VGG-16 network, fusing fc6 (fully connected layer) and fc7 of network output features, we call the improved network VGG-M. The VGG-M network fused fc6 and fc7 layers with double dimensions, resulting the network convergence is slow when the model is trained. To solve this problem, five BN layers were added,

Figure 1 Part of the vegetable data expansion.

Table 1 Three kinds of network structure comparison.

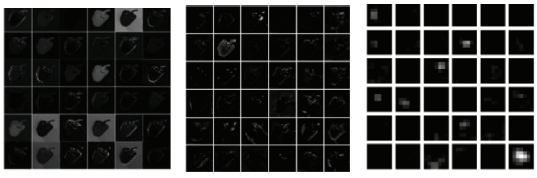
VGG-16	VGG-M	VGG-M-BN
Input (224*224 RGB Image)	Input (224*224 RGB Image)	Input (224*224 RGB Image)
conv1-1-64	conv1-1-64	conv1-1-64
conv1_2-64	conv1_2-64	Batch Normalization1
pool1	pool1	conv1_2-64
		pool1
conv2_1-128	conv2_1-128	conv2_1-128
conv2_2-128	conv2_2-128	Batch Normalization1
pool2	pool2	conv2_2-128
		pool2
conv3_1-256	conv3_1-256	conv3_1-256
conv3_2-256	conv3_2-256	Batch Normalization3
conv3_3-256	conv3_3-256	conv3_2-256
pool3	pool3	conv3_3-256
		pool3
conv4_1-512	conv4_1-512	conv4_1-512
conv4_2-512	conv4_2-512	Batch Normalization4
conv4_3-512	conv4_3-512	conv4_2-512
pool4	pool4	conv4_3-512
		pool4
conv5_1-512	conv5_1-512	conv5_1-512
conv5_2-512	conv5_2-512	Batch Normalization5
conv5_3-512	conv5_3-512	conv5_2-512
pool5	pool5	conv5_3-512
		Pool5
Fc6	Fc6	Fc6
Fc7	Fc7	Fc7
Fc-1000	Fc6,7	Fc6,7
	Fc-10	Fc-10
SoftMax	SoftMax	SoftMax



(a) The original image (b) Conv1-1 output features

(c) Conv3-1 output features (d) Conv5-1 output features

Figure 2 Output features of partially convolutional layers.



(a) pool1 (b) pool3 (c) pool5

Figure 3 Output features of partially pooling layers.

and the network was called VGG-M-BN. The improved VGG network model and the traditional VGG network model are shown in Table 1.

In Table 1, VGG-M has 17 layers (except the input layer), there is one more fully connected layer than the traditional VGG-16 network. VGG-M has 22 layers (except the input layer), There are 13 convolutional layers, 5 BN layers (new), 4 fully connected layers (Fc6, 7 layer added) and SoftMax layer. The convolutional layer and the pooling layer appear alternately.

The main function of the convolutional layer is feature extraction, which use the 3*3 conv. layers throughout the whole net can decrease the number of parameters when compared to AlexNet model. The incorporation of 1*1 conv. layers can increase the nonlinearity of the decision function [23]. We will take garden pepper as an example, visualizing the output features of convolutional layers (only show the first 36) as shown in Figure 2.

Figure 2 shows that the contour feature of the image near the input layer is clear and close to the shape of the original image. When the number of layers increases, the features gradually change into blur and abstract, it can be seen that the various layers of the model represent different levels of abstraction.

In this paper, the pooling layer using max pooling operation, this action seeking maximum value in each field, regardless of where is the maximum value in the region, the value of the Max pooling operation is the same. Therefore, the operation can realize translation, rotation and scale invariance, which provides a strong robustness for the model. The output features of the pooling layers are visualized, as shown in Figure 3.

The input of the pooling layer comes from the last convolutional layer, through max pooling operation, the number of network structure parameters is reduced while retaining the main features. Preventing the occurrence of the overfitting, it also improves the generalization ability of network.

The activation function of the VGG-M-BN network uses the Rectified Linear Units (ReLUs). The unsaturated nonlinear characteristics of the function is effective to alleviate the problem that the traditional Sigmoid and Tanh activation functions are easy to produce gradient disappear. Moreover, when using the ReLU function, the output does not tend to saturate as the input increases gradually, it can solve the problem of gradient disappearing. And when training, ReLU can speed up the network training.

The BN layer mainly solves the problem that the middle layer data distribution changes during the training process, to prevent the gradient from disappearing or exploding and speeding up network convergence. The specific BN operation is to normalize the activation value of each neuron in the hidden layer.

$$\hat{x}^{(k)} = \frac{x^{(k)} - E\left[x^{(k)}\right]}{\sqrt{Var\left[x^{(k)}\right]}} \tag{1}$$

$x^{(k)}$ represents the activation value of a neuron, $E\left[x^{(k)}\right]$ refers to the average of each batch of training data neurons $x^{(k)}$, $Var\left[x^{(k)}\right]$ represents the variance of each batch of training data neuron $x^{(k)}$. After this transformation, the data has a normal distribution of the mean of 0 and the variance of 1, in order to increase the value of the

derivative, increase the liquidity of reverse communication information and accelerate the training convergence rate. However, this will lead to a decline in Internet expression. In order to prevent this, each neuron adds two adjustment parameters γ and β, as shown in formula (2).

$$y^{(k)} = \gamma^{(k)}\hat{x}^{(k)} + \beta^{(k)} \tag{2}$$

Among them, the calculation of γ is shown in formula (3).

$$\gamma^{(k)} = \sqrt{Var\left[x^{(k)}\right]} \tag{3}$$

The calculation of β is shown in formula (3).

$$\beta^{(k)} = E\left[x^{(k)}\right] \tag{4}$$

By training and learning parameters γ and β, it can restore the original network to learn the features distribution, to keep the model expressive. Using BN on the convolutional layer, a policy of similar weight sharing is used; the whole feature map is treated as a neuron to reduce the number of parameters. In the network structure of this paper, after adding BN layer, the learning rate is increased, the dropout layer is removed and the network generalization ability are improved.

The last of the fully connected layer is a SoftMax layer with 1000 outputs, according to the number of packets of the vegetable image data set, the number of output layer classification is changed from the original 1000 to 5. The traditional gradient descent method is used to train the parameters of the network. The whole learning process in the network gradually reduces the learning rate, to improve the learning speed of the initial network while ensuring that the model can make the loss function reliable convergence.

4. EXPERIMENTS AND DISCUSSION

Our experimental environment is used the operating system Windows10, Deep learning framework Caffe[1], CUDA8.0, cuDNN5.1 and Python3.5, using GPU to calculate, the video card is GTX1080 and 8G memory.

In this paper, the vegetable image dataset was trained with the VGG-M-BN network. To verify the effect of batch_size on classification accuracy. Setting batch_size to 1, 2, 4 and 8, respectively.

In Table 2, we can find that the accuracy increases with the increase of batch_size. The larger the batch_size, the more accurate the reduction direction, the smaller the training shock, in the meanwhile, increased batch_size can improve memory utilization efficiency, and speed up processing. As can be seen from Figure 4, the batch_size setting not only has a great impact on accuracy, but also affects the convergence speed of the network. When batch_size is 8, the network converges fastest.

Table 2 batch_size experimental results.

Batch_size	1	2	4	8
Top1 accuracy (%)	36.5	86.6	92.4	96.5

[1] http://caffe.berkeleyvision.org/

The ReLU used in this paper is compared with the traditional Sigmoid and the Tanh function. The experimental results are shown in Table 3.

Table 3 shows the accuracy of ReLU activation function is significantly higher than that of traditional Sigmoid and Tanh function. Figure 5 is a graph of the Top1 accuracy with the change of iteration times when the VGG-M-BN network uses three activation functions respectively. We can see that the convergence rate of ReLU activation function is obviously faster than Sigmoid and Tanh function.

In order to verify the relationship between the accuracy rate of vegetable image classification and the number of image data set, a total of 24000, 12000, 6000, 3000 and 1500 images were randomly selected and trained on the VGG-M-BN network from 48000 vegetable image data set. Among them, the training data set accounted for 80%, test data set accounted for 20%.

As we can see from Table 4, when the number of vegetable image data sets is decreasing, the accuracy of Top1 is decreasing. When the

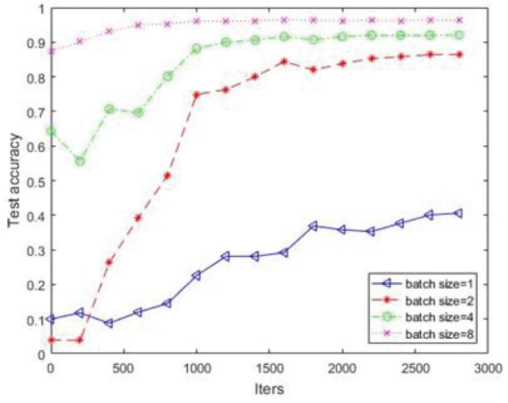


Figure 4 The accuracy of different batch_size.

Table 3 Activation function experiment results.

Activation Function	Sigmoid	Tanhn	ReLU
Top1 accuracy (%)	27.4	70.3	**96.5**

Legend: Bold indicates the highest accuracy rate in the comparison experiment results.

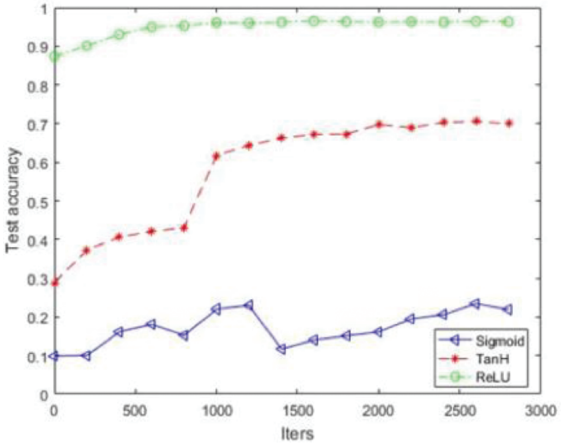


Figure 5 The accuracy of different activation functions.

number of data sets was less than 6000, the accuracy of Top1 was less than 90%. The accuracy rate dropped by nearly 10 percentage points compared with the dataset of 48000. The experiment verifies that the number of image data sets can affect the accuracy of recognition and classification. The accuracy of different number of dataset is shown in Figure 6.

The experimental results of VGG-M, VGG-M-BN network and traditional VGG-16 and AlexNet networks are shown in Table 5 and the accuracy curve of Top-1 accuracy is shown in Figure 7.

In Table 5, The Top1 accuracy of VGG-M and VGG-M-BN network is about 10 % higher than that of traditional AlexNet and VGG network. Although the accuracy of VGG-M network is improved, the convergence rate of the network is slow, in order to solve this problem, the BN layer was added to the VGG-M network.

As can be seen from Figure 7, VGG-M iterated 1500 times, and the network converges. While the VGG-M-BN network has converged at 500 iterations. After increasing the BN layer, the convergence rate is improved. Moreover, the convergence speed of the VGG-M-BN network is faster than that of the traditional VGG and AlexNet networks. Experimental results verify the effectiveness of the proposed method.

5. CONCLUSION

In order to improve the accuracy of recognition and classification of vegetables, based on the improved VGG network model, this paper proposes a target recognition and classification method. The main contributions include: Self-built ten kinds of vegetable image data sets. The images were obtained from the image database of ImageNet, website crawler and shoot by myself. The image data extension method is used to reduce overfitting in the learning process. At the same time, the image is pretreated and normalized. In this paper, the recognition and classification methods of vegetable images of VGG-M and VGG-M-BN are presented. On the basis of the traditional VGG network, VGG-M combines the output features of the first two fully connected layers, the accuracy was 95.8%. The accuracy of VGG-M-BN network after adding BN layer reached 96.5%, in the meanwhile, the convergence rate of the network has accelerated. This paper is also verified the influence of the number of data sets, the size of batch_size and the different activation functions on the recognition and classification accuracy.

Although the recognition and classification of vegetable image data set has achieved some results, compared with the traditional AlexNet and VGG network Top1 accuracy has improved, but there are still deficiencies can be improved from the following aspects:

1. Improving vegetable image data set. There are fewer kinds of vegetables used in this paper, the variety of vegetables can be added on this basis, to make it contain all the vegetables we see everyday.
2. Adopt ensemble learning methods. On the basis of this paper, ensemble learning methods can be used, combining the characteristics and advantages of different models to improve the accuracy of recognition and classification of vegetable images.

Table 4 | Different number of experimental results.

Index	48000	24000	12000	6000	3000	1500
Top1 accuracy (%)	**96.5**	94.1	92.5	90.4	88.5	84.4

Legend: Bold indicates the highest accuracy rate in the comparison experiment results.

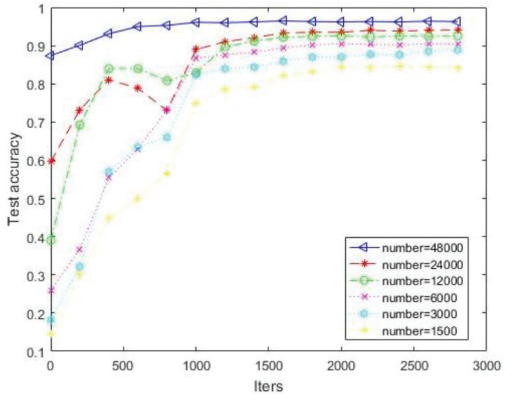


Figure 6 | The accuracy of different number of dataset.

Table 5 | Comparison of network experiment results.

Index	AlexNet	VGG	VGG-M	VGG-M-BN
Top1 accuracy (%)	86.3	92.1	95.8	**96.5**

Legend: Bold indicates the highest accuracy rate in the comparison experiment results.

CONFLICT OF INTEREST

No conflict of interest has been declared by the authors.

AUTHORS' CONTRIBUTIONS

Zhenbo Li: contributed to the conception of the study; Ling Zhu: performed the experiment and performed the data analyses and wrote the manuscript; Fei Li: contributed significantly to analysis and manuscript preparation; Jun Yue: helped perform the analysis with constructive discussions.

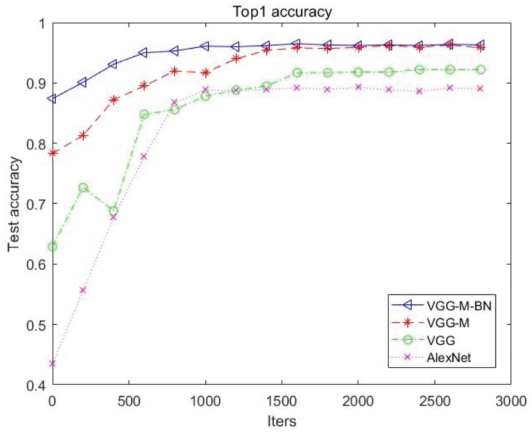


Figure 7 | Top1 accuracy.

ACKNOWLEDGMENTS

This work was supported by the Hebei Province School Science and Technology Cooperation Development Fund Project (Grant No. 18047405D, Grant No. 201805510811480), the International Science and Technology Cooperation Program of China (Grant No. 2015DFA00530), and the Key Research and Development Plan Project of Shandong Province (Grant No. 2016CYJS03A02).

REFERENCES

[1] N.-Q. Pham, T.-S. Nguyen, J. Niehues, *et al.*, Very deep self-attention networks for end-to-end speech recognition, arXiv preprint arXiv: 1904.13377, 2019.
[2] L. Zhu, Z. Li, C. Li, *et al.*, High-performance vegetable classification from images based on AlexNet deep learning model, Int. J. Agric. Biol. Eng. 11 (2018), 217–223.
[3] S. Ciptohadijoyo, W.S. Litananda, M. Rivai, *et al.*, Electronic nose based on partition column integrated with gas sensor for fruit identification and classification, Comput. Electr. Agric. 121 (2016), 429–435.
[4] P. Ninawe, S. Pandey, A completion on fruit recognition system using K-nearest neighbors algorithm, Int. J. Adv. Res. Comput. Eng. Technol. 3 (2014), 2352–2356.
[5] S.R. Dubey, S. Jalal, Fruit and vegetable recognition by fusing color and texture features of the image using matching learning, Int. J. Appl. Pattern Recognit. 2 (2015), 160–181.
[6] Y. Zhang, S. Wang, G. Ji, P. Phillips, Fruit classification using computer vision and feedforward neural network, J. Food Eng. 143 (2014), 167–177.
[7] H.W. Tao, L. Zhao, J. Xi, *et al.*, Fruits and vegetables recognition based on color and texture features, Trans. Chin. Soc. Agric. Eng. 30 (2014), 305–311.
[8] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex, Physiol. 160 (1962), 106–154.
[9] S.H. Lee, C.S. Chan, S.J. Mayo, *et al.*, How deep learning extracts and learns leaf features for plant classification, Pattern Recognit. 71 (2017), 1–13.
[10] P. Wang, W. Li, S. Liu, *et al.*, Large-scale isolated gesture recognition using convolutional neural networks, in International Conference on Pattern Recognition, IEEE, Cancun, Mexico, 2016, pp. 7–12.
[11] X.W. Gao, R. Hui, Z. Tian, Classification of CT brain images based on deep learning networks, Comput. Methods Prog. Biomed. 138 (2017), 49–56.
[12] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[13] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science. 7 (2006), 504–507.
[14] G. Hinton, A practical guide to training restricted Boltzmann machines, Momentum. 9 (2010), 926–947.
[15] G.E. Hinton, S. Osindero, Y.W. The, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006), 1527–1554.
[16] C. Hentschel, T.P. Wiradarma, H. Sack, Fine tuning CNNS with scarce training data — adapting imagenet to art epoch classification, in IEEE International Conference on Image Processing, IEEE, Phoenix, AZ, USA, 2016, pp. 3693–3697.
[17] V. Ferrari, M. Guillaumin, Large-scale knowledge transfer for object localization in ImageNet, in IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Providence, RI, USA, 2012, pp. 3202–3209.
[18] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015), 436–444.
[19] X. Zeng, L.I. Jie, Time-frequency image recognition based on convolutional neural network, Machinery & Electronics, 34 (2016), 25–29.
[20] T. Zhou, An image recognition model based on improved convolutional neural network, J. Comput. Theor. Nanosci. 13 (2016), 4223–4229.
[21] M. Alotaibi, A. Mahmood, Improved Gait recognition based on specialized deep convolutional neural networks, in Applied Imagery Pattern Recognition Workshop, IEEE, Washington, DC, USA, 2015, pp. 1–7.
[22] Z. Kaixuan, H. Dongjian, Recognition of individual dairy cattle based on convolutional neural networks, Trans. Chin. Soc. Agric. Eng. 31 (2015), 181–187.
[23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014. https://arxiv.org/abs/1409.1556
[24] A. Singla, L. Yuan, T. Ebrahimi, Food/non-food image classification and food categorization using pre-trained GoogLeNet model, in International Workshop on Multimedia Assisted Dietary Management, ACM, Amsterdam, The Netherlands, 2016, pp. 3–11.
[25] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv: 1502.03167, 2015. https://arxiv.org/abs/1502.03167