Computers and Electronics in Agriculture 166 (2019) 104978

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Deep learning-based automatic recognition network of agricultural machinery images

Ziqiang Zhang[a,b,1], Hui Liu[a,*], Zhijun Meng[b], Jingping Chen[b]

[a] Information Engineering College, Capital Normal University, Beijing 100048, China
[b] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

ARTICLE INFO

Keywords:
Agricultural machinery images
Deep learning
Image recognition
Inception_v3 network
AMTNet

ABSTRACT

Due to the massive amount of data generated by the mobile Internet and the development of large-scale computing devices and technologies, the deep learning algorithm has experienced a breakthrough in terms of image recognition technology. Traditional image recognition requires the complex extraction of image features, whereas deep learning technology can automatically learn image features through multi-layer nonlinear transformation, which is especially proficient at extracting complex global features. An image annotation dataset containing the images of seven types of machines and six types of abnormal images was constructed in this study from the large number of machine images in the agricultural machinery operation supervisory service system. To improve the Inception_v3 network, a network called AMTNet was designed and trained for automatic recognition of agricultural machinery images. Under the same experimental conditions, AMTNet achieved recognition accuracies of 97.83% and 100% on validation sets Top_1 and Top_5, respectively, demonstrating better performance than the classic networks ResNet_50 and Inception_v3. To further test the performance of AMTNet, 200 images of each of the 13 types of machine images were selected as test sets. The average area under the curve and F1-score of the network for image recognition of various machines reached 92% and 96%, respectively. According to the test results, AMTNet shows good robustness to illumination, environmental changes, and small area occlusion, which meets the practical application requirements of intelligent supervision over agricultural machinery operation.

1. Introduction

In recent years, China's mode of agricultural production has undergone rapid transformation, with the degree of agricultural mechanization continuing to increase (Yang, 2011). In 2017, the Ministry of Agriculture issued the *Thirteenth Five-Year Plan for the Development of Agricultural Mechanization in China* (Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2017), proposing a strategic goal of achieving a comprehensive mechanization level of crops of more than 70% by 2020 and setting up approximately 500 demonstration counties to take the lead in realizing the mechanization of agricultural production. As the level of agricultural mechanization continues to grow, the demand for agricultural machinery clustering and intelligent management has promoted the integration of new information technology and agricultural machinery technology represented by the Internet of Things (Shaonong et al., 2015; Kang et al., 2018). Beijing Engineering Research Center for Intelligent Agricultural Machinery developed an agricultural machinery operation supervision service system (Yin et al., 2018) that consists of a vehicle-mounted monitoring terminal, Global Navigation Satellite System (GNSS) positioning sensor, machine operation monitoring sensor, vehicle-mounted waterproof camera, and machine recognition sensor. The location, operation status, speed, and image of the agricultural machinery are uploaded to the central server through the General Packet Radio Service (GPRS) network so that the supervisory department can check the real-time trajectory, historical trajectory, area, images, and other regulatory indicators of agricultural machinery operation through a Web browser.

The recognition of agricultural machinery is a daily task performed by supervising personnel. It is also an important index that governments use to issue subsidies for the use of agricultural machinery. At present, the agricultural machinery operation supervision service system has provided data management services for hundreds of thousands of users. Due to the quantity of daily real-time transmission of image data and the inefficiency, cost, subjectivity, and error rate of

[*] Corresponding author at: No. 56, North Road of Western 3rd-Ring, P.O. Box 801, Beijing 100048, China.
E-mail address: liuhui@cnu.edu.cn (H. Liu).
[1] Address: Beijing Nongke Mansion Room B-508, No. 11, ShuGuangHuaYuan Middle Road, Beijing 100097, China.

https://doi.org/10.1016/j.compag.2019.104978
Received 9 May 2019; Received in revised form 26 August 2019; Accepted 26 August 2019
Available online 10 September 2019
0168-1699/ © 2019 Elsevier B.V. All rights reserved.

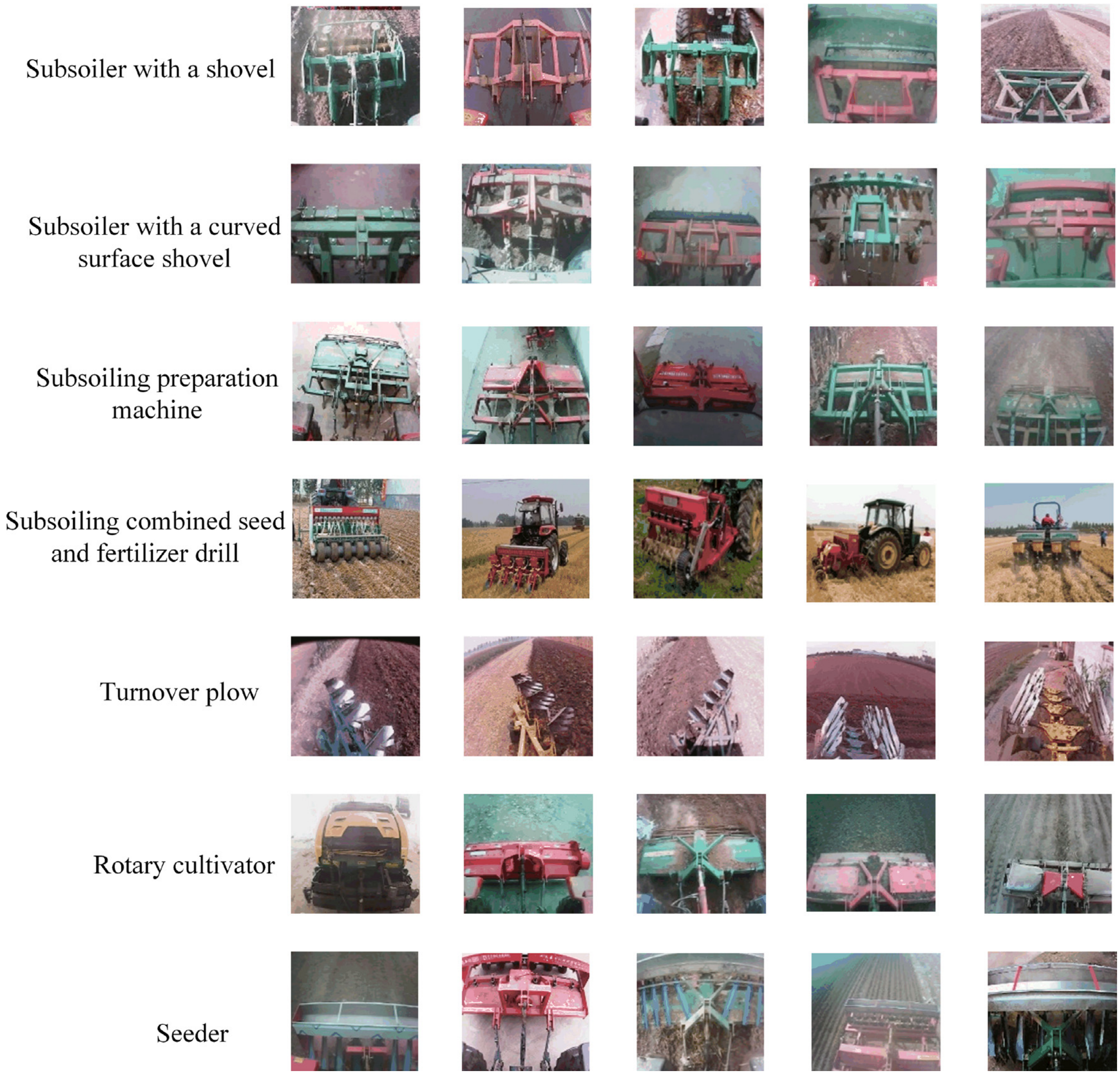

Subsoiler with a shovel
Subsoiler with a curved surface shovel
Subsoiling preparation machine
Subsoiling combined seed and fertilizer drill
Turnover plow
Rotary cultivator
Seeder

Fig. 1. Images of the seven types of agricultural machinery.

manual recognition, developing automatic recognition systems for agricultural machinery is of great significance.

In many cases, traditional image processing technology is characterized by acquiring features of targets. There features include color, shape, and graining. Image processing systems then classify the images through methods such as artificial neural network (Roffman et al., 2018) and support vector machine (Thanh Noi and Kappas, 2018). Because of the great differences in color and brightness of images of different types of agricultural machinery and the great similarity between certain agricultural machinery, it is difficult and inefficient to extract image features using the traditional image recognition method. Therefore, there is a profound need for a simple, efficient, and accurate automatic agricultural machinery recognition method.

Due to the massive amount of data generated by the mobile Internet and the development of large-scale computing devices and technologies, the deep learning algorithm (LeCun et al., 2015) has seen a breakthrough in terms of image recognition technology. Instead of manually extracting features based on prior knowledge, the algorithm automatically learns image features through multi-layer nonlinear transformation under the data-driven design. The deep learning algorithm is especially proficient at extracting complex global features, with a strong robustness to the judgment of the transformation and rotation of objects.

Agricultural machinery recognition is quite different from flower recognition (Xia et al., 2017) and plant recognition (Lee et al., 2015). The main manifestations are as follows: (1) image data acquisition methods are different. Agricultural machinery images are mainly captured by the main vehicle terminal camera. Images are easily affected by illumination, temperature, humidity, wind, precipitation, and bumpy roads. Special preprocessing operations are needed for the collected agricultural machinery images; (2) the distribution structure of agricultural machinery image is more complex. A mature deep convolutional neural network can extract all the features of the image. However, the complex network structure, large number of parameters, and long training time are not conducive to the deployment and application of the network. Furthermore, there may be a certain degree of over-fitting, resulting in a slightly lower recognition accuracy (Dandan and Dongjian, 2019); (3) Image recognition technology based on deep learning requires a training dataset that covers a large number of annotated images. Current open-source datasets, such as cifar10, cifar100, and ImageNet, provide data regarding general objects, such as plants, cars, faces, and flowers, but there are no open-source data set of agricultural machinery. Therefore, it is necessary to perform studies on agricultural machinery recognition.

Currently, there are few studies on the recognition of agricultural machinery using convolutional neural networks, especially on the recognition of various types of agricultural machinery and abnormal images. Yang et al. used a convolutional neural network for subsoiler recognition (Yang et al., 2018), and its recognition rate reached 98.5%. However, due to the relatively simple processing of abnormal images, only a few types of agricultural machinery can be recognized by the network, and the accuracy tends to be less than 90%, which is insufficient for the requirements of practical applications of agricultural machinery operation supervision service systems.

In this study, an agricultural machinery image annotation dataset was constructed for 7 types of agricultural machinery images and 6 types of abnormal images. According to the actual needs of the supervisory system for automatic image recognition and the characteristics of agricultural machinery image, an AMTNet network was designed for the automatic recognition of agricultural machinery images. The feasibility and effectiveness of the AMTNet network were assessed by comparing it to ResNet_50 network and Inception_v3 network using the same validation set. To further test the performance of the AMTNet network, test sets were selected for verification experiments. The results showed the average values of AUC area under the curve and F1-score of AMTNet network to be 92% and 96% respectively, which does meet the requirements of practical applications of agricultural machinery operation supervision service system for the automatic recognition of agricultural machinery images.

2. Materials and methods

2.1. Image classification

The agricultural machinery operation supervision service system contains tens of millions of images related to different types of machines, mixed with various abnormal images. In this study, agricultural machinery was divided into seven types: the subsoiler with a shovel, subsoiler with a curved surface shovel, subsoiling preparation machine, subsoiling combined seed and fertilizer drill, turnover plow, rotary cultivator, and seeder (Fig. 1). According to the quality, shooting angle, and shadowing of the images, the abnormal images were divided into six types: black and white, blurred, pointing to the sky, road photographed, object photographed, and occlusion (Fig. 2). The image dataset established in this study contained a total of 125,000 images, and the dataset was divided into two mutually exclusive sets using the “hold-out” method (Dwork et al., 2015); these sets were called the training set and the validation set. The training set was used to train the network, which included a total of 100,000 images, and the validation set was used to verify the accuracy of the network, which contained 25,000 images. Table 1 shows the distribution of image data about the agricultural machinery.

2.2. Image preprocessing of agricultural machinery

Compared with the image data in other research fields, the operating environment of agricultural machinery is relatively harsh; therefore, the quality of the images obtained is comparatively poor, the backgrounds of the images are usually complex, and the shooting angle varies. Image preprocessing is an important approach to remove damaged images, eliminate the influence of noise such as differences in the background, color, and size of the images, reduce the calculated amount of network training, improve the efficiency of the algorithm, and make the network more accurate. In view of the problems related to agricultural machinery images, a compiled algorithm was used to process scripts in batches and carry out image preprocessing in four ways: image clipping, image color adjustment, motion blur elimination, and image noise reduction.

Table 1
Distribution of datasets of the agricultural machinery images.

No.	Image type	No. of images in the training set	No. of images in the validation set
0	Black and white	5000	1250
1	Blurred	2000	500
2	Pointing to the sky	1000	250
3	Road photographed	1500	375
4	Object photographed	3500	875
5	Occlusion	2000	500
6	Subsoiler with a shovel	15,000	3750
7	Subsoiler with a curved surface shovel	10,000	2500
8	Subsoiling preparation machine	15,000	3750
9	Subsoiling combined seed and fertilizer drill	10,000	2500
10	Turnover plow	10,000	2500
11	Rotary cultivator	15,000	3750
12	Seeder	10,000	2500
–	Total	100,000	25,000

2.2.1. Image cropping

The original images were taken by different devices monitoring agricultural machinery in different regions, and therefore the image size varied greatly. Since the neural network involves a fixed input node, it is necessary to unify the image size before inputting the pixels of the image into the neural network. In this study, the bilinear interpolation method (Gribbon et al., 2003) was used to crop the images to 64 × 64 pixels per inch. The core idea of the bilinear interpolation method is to perform a linear interpolation calculation in both the X axis and the Y axis. Unlike the nearest neighbor interpolation, which is relatively coarse, and the bicubic interpolation, which involves a large computational cost, the bilinear interpolation method has a stable processing effect and a small computational complexity.

2.2.2. Image color adjustment

The agricultural machinery image recognition network can be trained to be affected as little as possible by irrelevant factors by adjusting the hue, brightness, and contrast of the images. In this study, the image preprocessing functions provided by TensorFlow API (i.e., the hue function, the brightness function, and the contrast function) were used to process the original images of the agricultural machinery. Through image preprocessing, the image of a subsoiler with a shovel with poor hue (Fig. 3a), the image of a rotary cultivator with high local brightness (Fig. 3c), and the image of a seeder with low contrast (Fig. 3e) were all restored to clear images, as shown in (Fig. 3b, d, and f, respectively).

2.2.3. Elimination of motion blur

Images of agricultural machinery are taken by vehicle-mounted cameras during the operation of the agricultural machinery, so motion blur can often be found in the images. The so-called motion blur refers to image blur caused by the relative motion between the camera and the subject. From the technical point of view, there are currently three ways to process blurred images: image super-resolution reconstruction, image enhancement, and image restoration. Image restoration

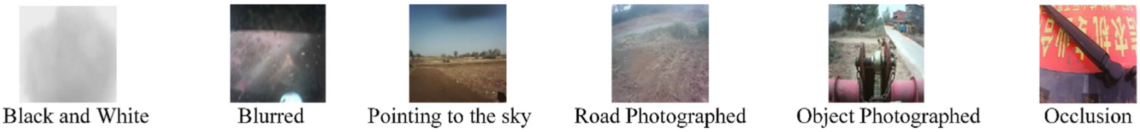

Black and White Blurred Pointing to the sky Road Photographed Object Photographed Occlusion

Fig. 2. The six types of abnormal images.

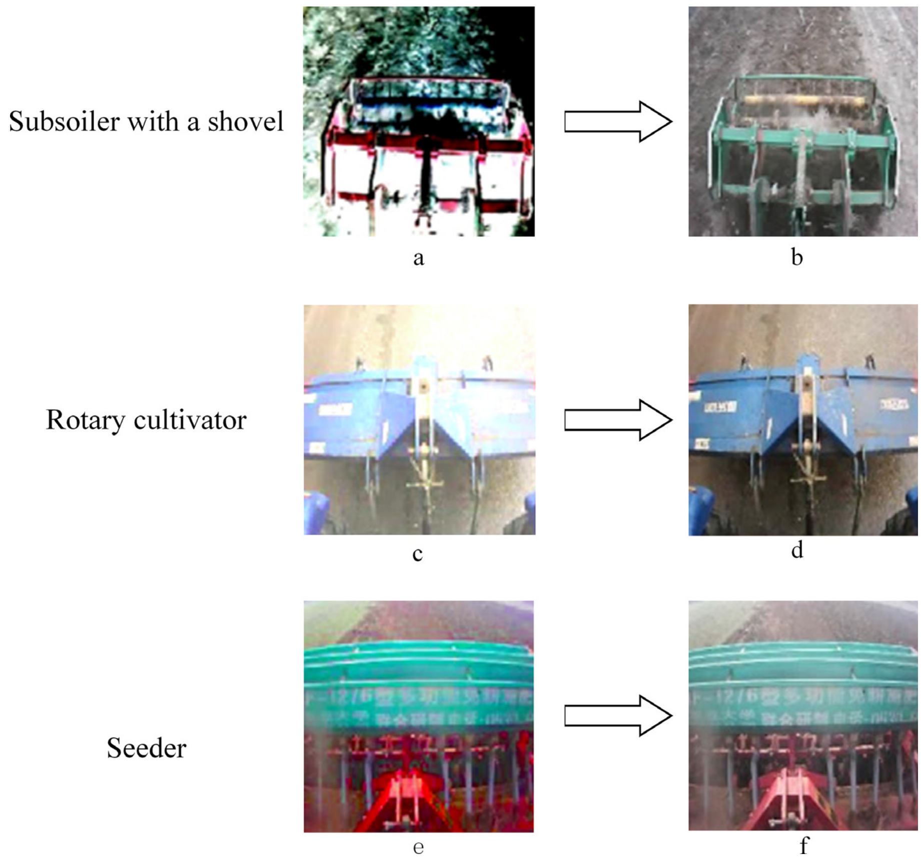


Fig. 3. Agricultural machinery images before and after color adjustment.

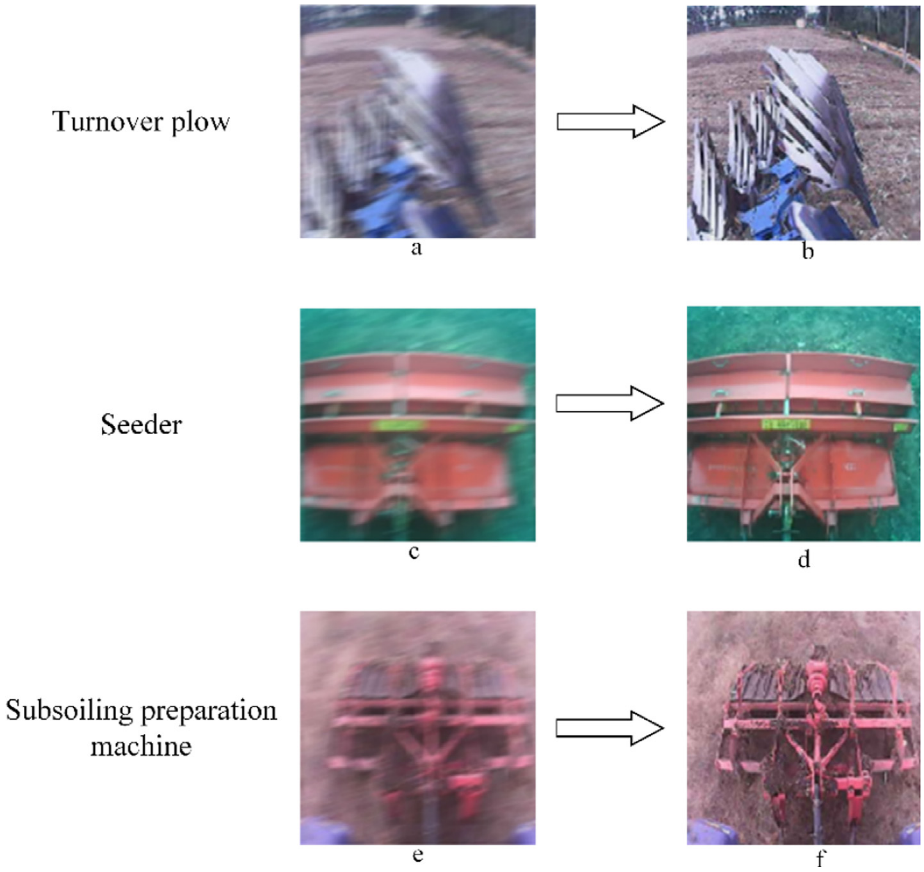


Fig. 4. Agricultural machinery images before and after the elimination of motion blur.

establishes a degradation network based on the prior knowledge of image degradation, based on which various inverse degradation processing algorithms are used to recover the images gradually, which improves the image quality. In this study, the Wiener filtering algorithm, an image restoration method (Buades et al., 2005), was used to preprocess the images with motion blur (i.e., Fig. 4a, c, and e) into clear images of the agricultural machinery (i.e., Fig. 4b, d, and f).

2.2.4. Image denoising

The noise of an image is often pixels or pixel blocks that cause a strong visual effect. In agricultural machinery images, noise is introduced during the process of image acquisition due to the influence of the sensor's material properties, electronic components, circuit structure, working environment, and the way of manual installation. This noise includes dark current noise, thermal noise caused by resistance, photon noise, and photoresponse non-uniform noise. Additionally, in the image signal transmission process, digital images will also be polluted by various noise during their transmission due to imperfections in the transmission medium and recording equipment. Moreover, in some steps of image processing, noise will be introduced into the images when the input images are not as high-quality as expected. The images of agricultural machinery are usually affected by salt-and-pepper noise, which is currently processed by the non-linear filter using methods such as the anomaly detection method (Qiao et al., 2002), the median filter method (Esakkirajan et al., 2011), and the pseudo-median filter method (Liu et al., 2010). In this study, the median filter method was used to sort the pixels of the local area according to their gray levels, and the median of the gray level in this area was selected as the gray value of the current pixel. As shown in Fig. 5, the original noise in the image of the seeder (Fig. 5a) and those in the image of the subsoiler with a shovel (Fig. 5d) were reduced after being processed by the 5 × 5 median filter (Fig. 5b and e). Clear images of the agricultural machinery were obtained through the processing of the 3 × 3 median filter (Fig. 5c and f). With these two filtering denoising processes, unnecessary error can be avoided because the useful information in the images will not be replaced by the median due to the excessively large value taken by the filter. The median filter method not only removes the salt-and-pepper noise in the image, but also maintains the characteristics of the original image, which prevents the processed image from becoming blurred and the image quality from becoming poor.

2.3. Annotation of image datasets

Google's TensorFlow deep learning platform (Baylor et al., 2017) was used to annotate the image datasets of various machines and abnormal images. The training set folder and the validation set folder contained the 13 types of images listed in Table 1. In each folder, the original image data were stored and converted to the TFRecord format using the pre-compiled script provided by the TensorFlow platform. The first byte of the data format represented the type of the image, and the remaining bytes represented the basic information of the image. At this stage, the complete image annotation dataset of agricultural machinery had been constructed, which laid a foundation for subsequent research on the automatic recognition of agricultural machinery images.

3. Agricultural machinery image recognition algorithm

3.1. Inception_v3 network

Unlike other classic convolutional neural networks such as AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan and Zisserman, 2014), and ResNet (He et al., 2016), the Inception_v3 network (Szegedy et al., 2016) has been successfully applied in many fields due to its excellent classification performance and relatively low complexity. The inception module designed in the Inception_v3 network improved the utilization of the parameters. As shown in Fig. 6, designed based on the idea of *Network In Network* (NIN) (Lin et al., 2013), the inception module has four branches, each of which uses a 1 × 1 convolution, which not only crosses channels to organize information and improve the expressive ability of the network but also performs dimension-increase and dimension-reduction to the output channel. The inception module connects the highly correlated nodes through the 1 × 1, 3 × 3, and 5 × 5 convolution kernels and the 3 × 3 maximum pooling layers of different sizes in the four branches, thereby constructing an efficient sparse structure in accordance with the Hebbian principle (Song et al., 2000).

3.2. Convolutional neural network

Fig. 7 shows the structure of AMTNet, which was designed in this study to improve the Inception_v3 network, and Table 2 shows the detailed network parameters. The front part of AMTNet is a common structure consisting of five convolutional layers (serial No.: C1 to C5) and two maximum pooling layers (serial No.: S1 to S2). The convolutional layer uses the ReLu function with good convergence and low computational complexity as the nonlinear excitation function. The inception module was added to the middle part of the network. This module is composed of three different types of inception module substructures, with the number of each type of substructure being three, five, and three. In addition, with the idea of "factorization into small

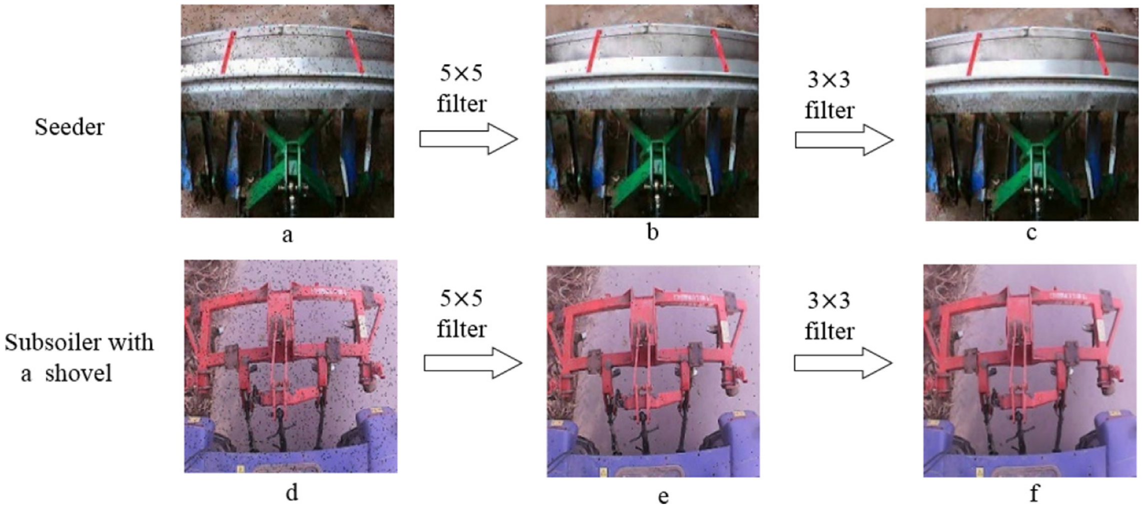


Fig. 5. Agricultural machinery images before and after image denoising.

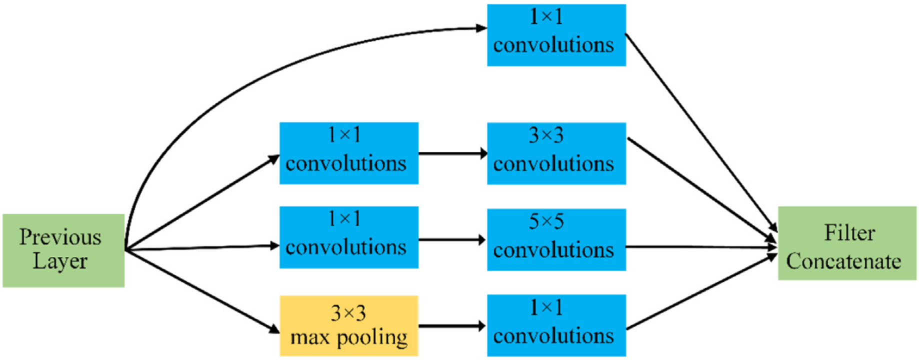


Fig. 6. The inception module.

convolutions" (Catani et al., 1991), more decoupling parameters were obtained by properly decomposing the convolution kernel, thereby speeding up the training. The purpose of each layer of convolution, pooling, or inception module in the front and middle of the network is to simplify the spatial structure and transform the spatial dimension into the channel dimension. The last part of the network consists of an average pooling layer, a Dropout layer to prevent overfitting, a fully connected layer (logits layer), and a softmax classifier. The probability values corresponding to the 13 types of machine images were calculated.

3.3. Improvement of network training accuracy

3.3.1. Learning rate and optimization algorithm

As one of the most important hyper-parameters in neural network training, learning rate (LR) (Smith, 2017) is essential for the fast and efficient training of neural networks. In this study, the neural network was trained by exponentially decaying the learning rate, which not only made the network quickly approach a relatively ideal solution in the early stage of training, but also prevented the network from having too much fluctuation in the later stage, thus getting closer to the locally optimal solution. LR was calculated as follows:

$$decayed\_learning\_rate = learning\_rate * decay\_rate^{\frac{global\_step}{decay\_steps}} \qquad (1)$$

In Eq. (1), *learning_rate* is the initial learning rate, *global_step* is the global step of the attenuation calculation, *decay_rate* is the decay rate, and *decay_steps* is the decay speed (i.e., how many rounds are needed to iterate all the training data). The training data here is 100,000 images, and the *batch_size* each time is 32.

In this paper, the Adam optimization algorithm (Kingma and Ba, 2014) was used for network training. Featuring high computational efficiency and low memory requirements, the Adam optimization algorithm is easy to implement. Moreover, the algorithm is suitable for

Table 2
AMTNet parameters.

No.	Type	Size of the convolution kernel/Strides (annotated)	Input size
C1	Conv_1	5 × 5/1	64 × 64 × 3
C2	Conv_2	3 × 3/1	60 × 60 × 32
C3	Conv_3	3 × 3/1	58 × 58 × 32
S1	Pool_1	3 × 3/2	58 × 58 × 64
C4	Conv_4	3 × 3/1	28 × 28 × 64
C5	Conv_5	3 × 3/2	26 × 26 × 80
S2	Pool_2	3 × 3/1	12 × 12 × 192
A	Conv block_1	3 × Inception	12 × 12 × 288
B	Conv block_2	5 × Inception	5 × 5 × 768
C	Conv block_3	3 × Inception	5 × 5 × 1280
S3	Pool_3	2 × 2	2 × 2 × 2048
D	Linear	Logits	1 × 1 × 2048
E	Softmax	Value output by the classifier	1 × 1 × 13

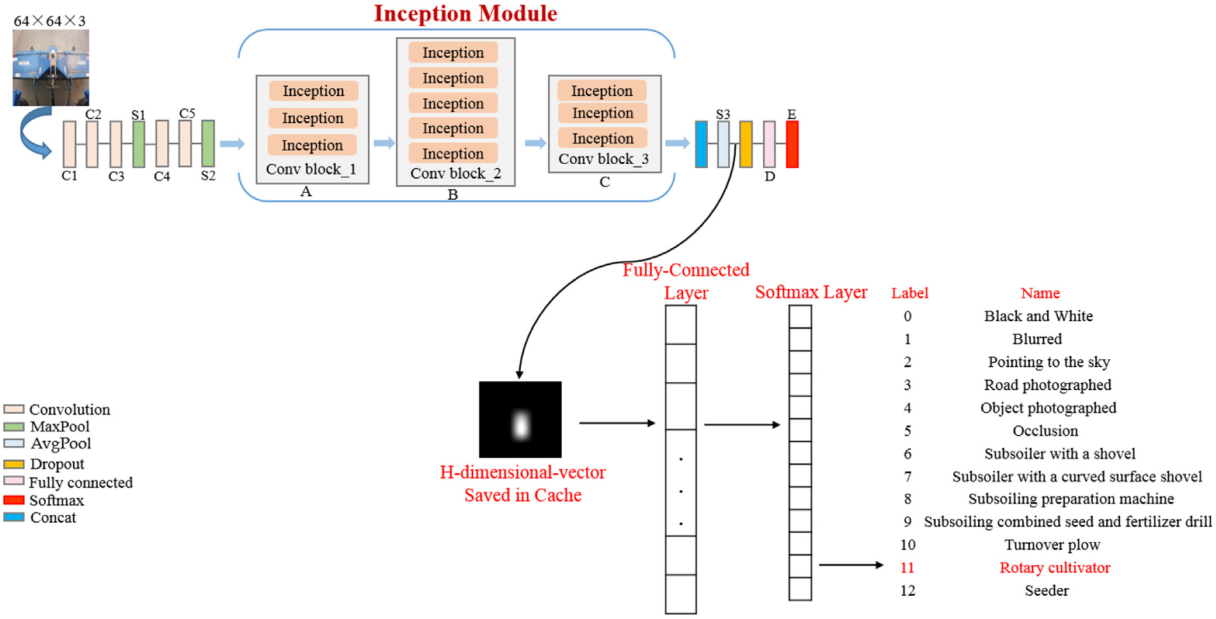


Fig. 7. AMTNet.

solving the instability of large noise and sparse gradients. The update rules of the algorithm are as follows:

$$t \leftarrow t + 1 \tag{2}$$

$$g_t = \nabla_\theta f_t(\theta_{t-1}) \tag{3}$$

where $f_t(\theta)$ represents the gradient of θ, that is, the partial derivative vector of f_t from θ under the time step t, and the initial parameters $t = 0$, $\theta_0 = 0$.

$$p_t \leftarrow \alpha_1 p_{t-1} + (1 - \alpha_1) g_t \tag{4}$$

$$q_t \leftarrow \alpha_2 q_{t-1} + (1 - \alpha_2) {g_t}^2 \tag{5}$$

The biased first-order and second-order moments were estimated in Eqs. (4) and (5), where α_1 and α_2 represent the exponential decay rate and initial parameter of the moment estimation $p_0 = 0$, $q_0 = 0$, respectively.

$$\hat{p}_t \leftarrow p_t / (1 - \alpha_1^t) \tag{6}$$

$$\hat{q}_t \leftarrow q_t / (1 - \alpha_2^t) \tag{7}$$

The first and second moment deviations were corrected in Eqs. (6) and (7).

$$\theta_t \leftarrow \theta_{t-1} - \lambda^* \hat{p}_t / (\sqrt{\hat{q}_t} + \varepsilon) \tag{8}$$

The value of θ was updated in Eq. (8), where λ is equivalent to the initial learning rate *learning_rate*, and ε is a small constant with a stable value.

Table 3 shows the values of the above hyper-parameters for the training of AMTNet in this study.

3.3.2. Dataset enhancement (C-DCGAN)

Sufficient data training samples are usually required to train deep learning network. In general, the larger the total amount of data, the more efficient the trained network will be. However, due to the actual production and promotion of agricultural machinery, there are relatively fewer image samples of some specific machines. To avoid the low recognition rate caused by incomplete data for certain types of agricultural machinery after network training, it is necessary to conduct dataset enhancement. The types of agricultural machinery lacking samples in this study mainly included the crawler self-propelled rotary cultivator, Tangshan corn seeder, Jinyuan subsoiler with a shovel, and Lugeng 1S-310B Subsoiler with a curved surface shovel. To solve the problem of insufficient samples of these specific machines, based on the advantages of Conditional Generative Adversarial Networks (CGAN) (Isola et al., 2017) and Deep Convolution Generative Adversarial Networks (DCGAN) (Goodfellow et al., 2014), the Conditional-Deep Convolution Generative Adversarial Networks (C-DCGAN) (Zhu et al., 2018), was established in this study to assist in the generation of sample data and conduct dataset enhancement. The principle of the algorithm is as follows:

$$\underset{G}{min}\underset{D}{max}V(D, G) = E_{x\ p_{data}(x)}[logD(x|y)] + E_{z\ p_z(z)}[log(1 - D(G(x|y)))] \tag{9}$$

Table 3
AMTNet hyper-parameters.

Hyper-parameter	Value
learning_rate	0.001
global_step	100,000
decay_steps	3125
decay_rate	0.94
batch_size	32
α_1	0.9
α_2	0.999
ε	10^{-8}

In Eq. (9), the essence of the algorithm lies in the optimization of the maximum and minimum. $E_{x\ p_{data}(x)}[logD(x|y)]$ indicates the probability of whether the sample obtained by inputting the real data x and the class label y into the discriminator D is real or not. $E_{z\ p_z(z)}[log(1 - D(G(x|y)))]$ denotes the probability of whether the sample obtained by inputting the random noise z and the class label y into the generator G is real or not.

Fig. 8 shows the C-DCGAN of the agricultural machinery. The 100-dimensional noise vector and the condition y (the class label) were input into the generator network, which then converted the noise vector into the vector of a similar feature map. The 100-dimensional noise vector was transformed into a 16,384-dimensional vector via the fully connected layer, forming a 4 × 4 × 1024 feature map through a reshaping process. Four sets of spatially sampled deconvolution were adopted to change the vector into a feature map whose width and height were twice as large as those of the original. The number of channels of the feature map was continuously reduced, and the size of the output feature map was 64 × 64 × 3. Eqs. (10) and (11) show the calculation process of deconvolution:

$$a = (i + 2p - k) \text{mod} s \tag{10}$$

$$o = s(i - 1) + k - 2p + a \tag{11}$$

Given the input condition y of the discriminator network and the size of the true and false feature maps (i.e., 64 × 64 × 3), the four sets of spatially sampled deconvolution were used to transform the feature map into a vector whose width and height were one-half of those of the original. The number of channels of the feature map continuously increased, and two-category data was obtained through the processing of the fully connected layer and the logits function. Eq. (12) shows the calculation process of convolution:

$$o = \left\lfloor \frac{i + 2p - k}{s} \right\rfloor + 1 \tag{12}$$

In Eqs. (10)–(12), where parameter a fills the number of zero and i (o) inputs (outputs) the size of the feature map. The size of the convolution kernel k is five, the step size s is two, and the value of p (padding), which is the same in each dimension, is two.

Fig. 9 shows the 4 × 4 sample images generated by the agricultural machinery C-DCGAN designed in this study. As the number of Epochs increased, the definitions of the 16 sample images became higher and higher. The specific dataset enhanced by the C-DCGAN can make up for the inefficiency of the agricultural machinery identification network caused by the imbalance of sample categories.

4. Experiment and analysis

4.1. Network training experiment and result

AMTNet was trained on two NVIDIA GeForce GTX 1080 GPUs. The loss function converged to 0.01 after 100,000 iterations of network training. The recognition accuracies of Top_1 and Top_5 on the validation set were 97.83% and 100%, respectively. To better test the classification effect of AMTNet, AMTNet was compared with the ResNet_50 network and the Inception_v3 network. The same agricultural machinery image annotation dataset and hyper-parameters were used to train the three networks separately. The results of the validation set after network training are shown in Table 4.

According to the results in Table 4, (1) in terms of recognition accuracy, the accuracy of Top_1 and Top_5 in AMTNet network is higher than that in other networks; (2) in terms of recognition efficiency, it only takes AMTNet network 0.15 s to recognize an agricultural machinery image from the validation set. In that, its performance is superior to that of the other networks; (3) in terms of training time, it takes 12 h to train AMTNet network, which is faster than the other networks. In terms of comprehensive efficiency, AMTNet was more in

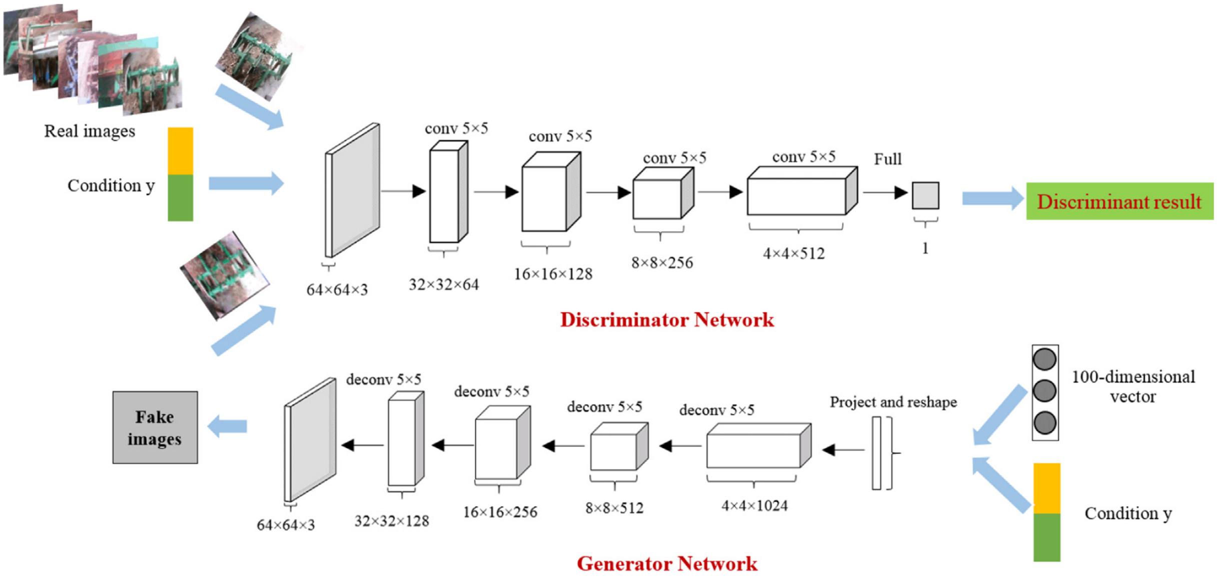


Fig. 8. The C-DCGAN of agricultural machinery.

line with the practical application requirements of high accuracy, low cost, and ideal effect proposed by the agricultural machinery operation supervision service system.

4.2. Network test and result

Since there are large differences between the numbers of various machines in the training set and the validation set, and the accuracy is not high enough to describe the actual application performance of AMTNet; In this study, a test set was used to assess the network generalization ability of AMTNet after the last hyper-parameter training. Abnormal images of agricultural machinery with different pitch angles, brands, and illuminations taken in Anhui Province in October 2018 were selected as the test set. Two hundred images of each of the 13 types of machines were selected, totaling 2600 images. Two methods were used for evaluation in the AMTNet network test experiment: receiver operating characteristic (ROC) curve (Brown and Davis, 2006) and area under the curve (AUC) (Pruessner et al., 2003), and confusion matrix and F1-score.

4.2.1. ROC curve and AUC

The abscissa of the ROC curve is the false positive rate (FPR), and the ordinate is the true positive rate (TPR). The AUC refers to the sum of the areas under the curve. Among the image recognition evaluation accuracy indexes, the ROC curve can minimize the interference caused by different test sets and more objectively measure the performance of the network itself.

The calculation methods of FPR and TPR are as follows:

$$\mathrm{FPR} = \frac{FP}{N} \tag{13}$$

$$\mathrm{TPR} = \frac{TP}{P} \tag{14}$$

where P is the number of true positive samples, N is the number of true negative samples, TP is the number of positive samples predicted by the classifier among the P positive samples, and FP is the number of positive samples predicted by the classifier among the N negative samples.

AUC can be obtained by summing the areas of the various parts under the ROC curve formed by connecting the points whose coordinates are $\{(x_1, y_1), (x_2, y_2), \cdots(x_m, y_m)\}$ in a specific order. Thus, AUC was calculated as follows:

$$AUC = \frac{1}{2}\sum_{i=1}^{m-1}(x_{i+1} - x_i)\cdot(y_i + y_{i+1}) \tag{15}$$

where the coordinate values x_i and y_i are the values of the above-mentioned FPR and TPR, respectively.

The ROC curve is often used as one of the most important indexes for two-category data (Jerez-Aragonés et al., 2003). As for multi-classifications in this study, there were m test samples (m = 2600) and n categories (n = 13). After training, the probability of each test sample under each category was calculated to obtain a matrix $\boldsymbol{P}$ with the shape

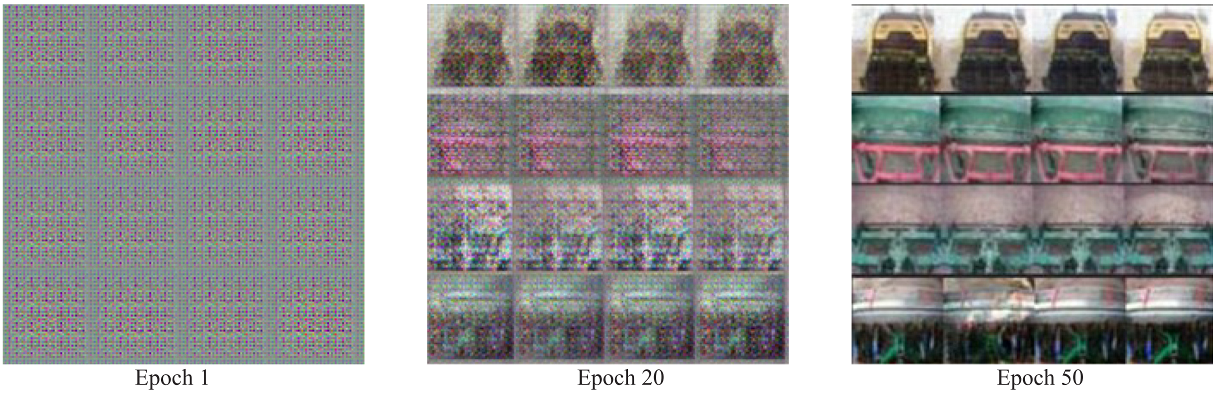


Fig. 9. Sample images generated by the C-DCGAN.

Table 4
Performance of the three convolutional neural networks.

No.	Network	Size of the input image	Top_1 accuracy rate (%)	Top_5 accuracy rate (%)	Recognition efficiency (s/image)	Training time (h)
1	Resnet_50	224 × 224	97.58	99.82	0.23	13
2	Inception_v3	299 × 299	97.62	99.80	0.36	16
3	AMTNet	64 × 64	97.83	100	0.15	12

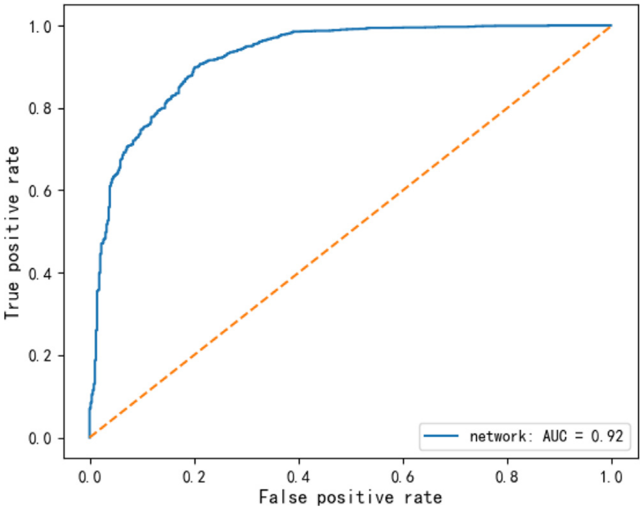


Fig. 10. ROC curve and AUC of agricultural machinery images.

being [m, n], and each row was sorted by class label, indicating the probability of a test sample under each category. Accordingly, the category of each test sample was converted into a form similar to the binary system, and each position was sorted by the class label to mark whether it belonged to the corresponding category; thus, obtaining a label matrix $\boldsymbol{L}$ of [m, n]. The probability (columns in matrix $\boldsymbol{P}$) that m test samples belong to each category could be obtained. Therefore, the FPR and TPR under each threshold were calculated according to each column in the probability matrix $\boldsymbol{P}$ and the label matrix $\boldsymbol{L}$, thereby plotting a ROC curve. In this way, a total of n ROC curves were plotted. Finally, the n ROC curves were averaged to obtain the final ROC curve and AUC. Fig. 10 shows the ROC curve and AUC of the classification of agricultural machinery images in this network test.

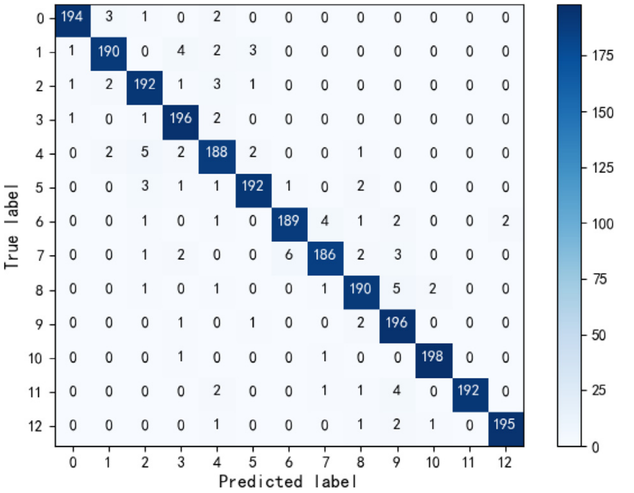
Fig. 10 shows that the mean AUC value under the ROC curve is 92% and the curve is close to the upper left corner, indicating the high true positive rate and less errors of classification using AMTNet network. In this case, the ability of the AMTNet network to classify the 7 types of machinery images and 6 types of abnormal images met the practical application requirements of agricultural machinery operation supervision service systems. The ROC curve is smooth, which indicates there is not too much over-fitting after AMTNet network training.

4.2.2. Confusion matrix and F1-score

Confusion matrices are mainly used to compare a target result and measured value in image recognition accuracy evaluation. If $\boldsymbol{C}[s,t]$ was a confusion matrix, the t^{th} column represented the prediction category, and the total number for each column is the number of data predicted to be a part of that category; the s^{th} row is the actual category the data are in, and the total number for reach row is the actual number of that category. The value in the matrix is the number of samples actually in category s but incorrectly determined to be in category t.

F1-score is used to evaluate the performance of classified networks. It considers both the precision A and the recall B of the test to compute the score. Eq. (16) shows the process used to calculate F1-score:

$$\text{F1 - score} = \frac{2 \times A \times B}{A + B} \tag{16}$$



(0-Black and White 1-Blurred 2-Pointing to the sky 3-Roaded photographed 4-Object photographed
5-Occlusion 6-Subsoiler with a shovel 7-Subsoiler with a curved surface shovel
8-Subsoiling preparation machine 9-Subsoiling combined seed and fertilizer drill
10-Turnover plow 11-Rotary cultivator 12-Seeder)

Fig. 11. Visual confusion matrix.

In Eq. (16), A is the number of correct positive results divided by the number of all positive results returned by the classifier, and B is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

On the test set, 13 types of agricultural machinery images were tested using AMTNet. Fig. 11 is the visual confusion matrix of the experimental results.

As shown in Fig. 11, the Precision, Recall and F1-score of 0–12 classes of agricultural machinery were calculated, the results of which are shown in Table 5. Precision denotes the value of the diagonal line of the confusion matrix divided by the sum of corresponding columns of the class; Recall denotes the value of the diagonal line of the confusion matrix divided by the sum of corresponding rows of the class.

As shown in Table 5, the mean precision of AMTNet on the test set was 0.96, which indicates that the network has a strong ability to recognize negative samples; the mean recall was 0.96, which indicates that the network has a strong ability to recognize positive samples; the mean F1-score was 0.96, which indicates that the network has high accuracy for recognizing agricultural machinery images. Being able to accurately recognize most machine types, operation scenes, non-machinery images, and light and shadow interference indicates that this network has good robustness and is highly practical.

4.2.3. Error analysis

The images of agricultural machinery incorrectly recognized were collated and analyzed (Fig. 12). The main reasons for incorrect recognition are as follows:

(1) In some cases, due to the installation position of the vehicle-mounted terminal camera and the pitch angle of the camera, an image of only a part of a machine was taken, which resulted in recognition error. For instance, the subsoiling preparation machine was recognized as a subsoiler with a curved surface shovel (Fig. 12a), and the seeder was recognized as a subsoiling

Table 5
Precision, recall, and F1-score.

Machinery category	0	1	2	3	4	5	6	7	8	9	10	11	12	Average value
Precision	0.98	0.96	0.94	0.94	0.93	0.96	0.96	0.96	0.95	0.92	0.99	1	0.99	**0.96**
Recall	0.97	0.95	0.96	0.98	0.93	0.96	0.95	0.93	0.95	0.98	0.99	0.96	0.98	**0.96**
F1-score	0.98	0.95	0.95	0.96	0.93	0.96	0.95	0.94	0.95	0.95	0.99	0.98	0.98	**0.96**

preparation machine (Fig. 12b).

(2) Since some of the machines were covered by soil, straw, clothes, and billboards, the images were incorrectly recognized. As shown in Fig. 12c and d, the rotary cultivator was recognized as a subsoiler with a shovel.

(3) Some machines with similar appearances were also wrongly recognized. As shown in Fig. 12e and f, the subsoiler with a shovel was recognized as a subsoiler with a curved surface shovel. The only difference between the two was the subsoiling shovel at the bottom.

(4) Information such as color and texture in some datasets failed to be comprehensively collected. For example, the abnormal images where objects were photographed were incorrectly recognized as the rotary cultivator (Fig. 12g, h, and i).

As indicated by the above error analysis, the network has some shortcomings. The network has an insufficient ability when it comes to the local recognition of a small number of agricultural machines. In addition, the images cannot be correctly and effectively recognized when large areas of the machine surface are covered by objects.

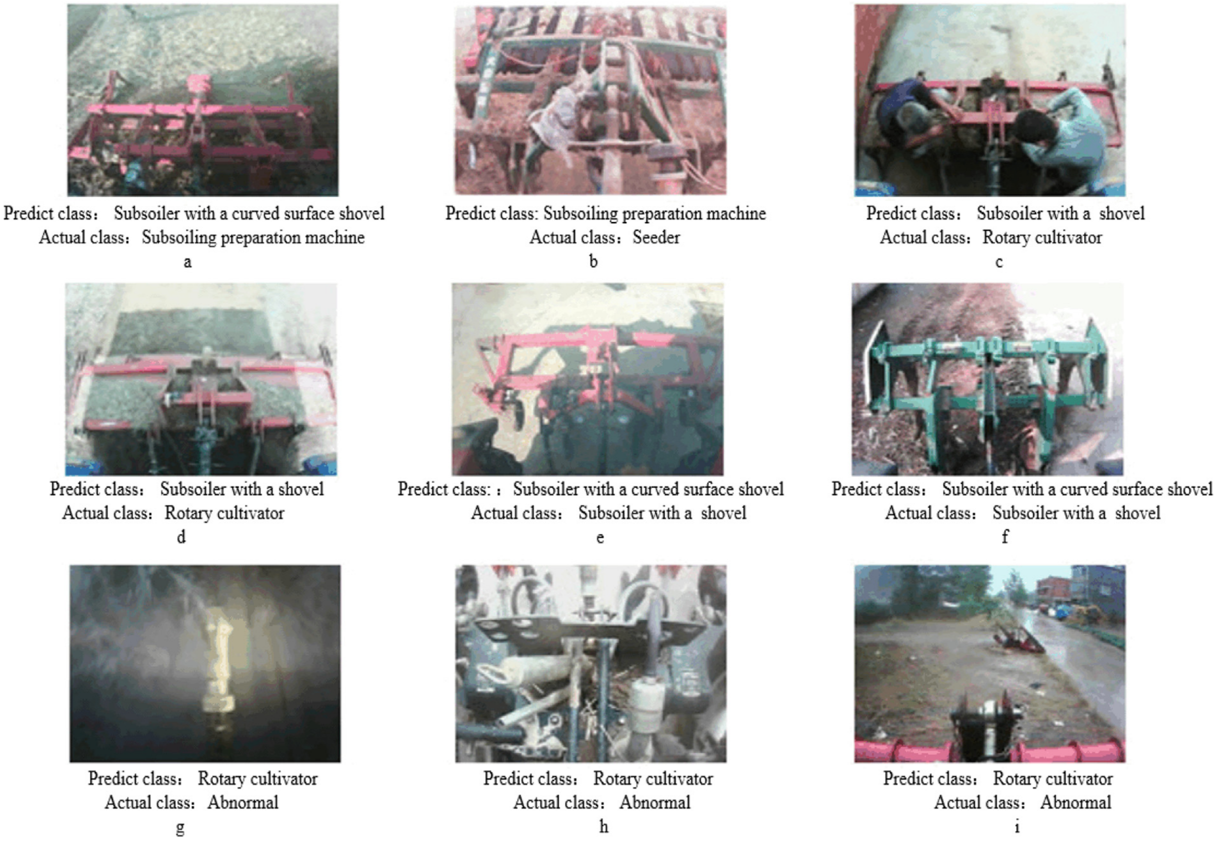
5. Conclusion

(1) Based on tens of millions of images in the agricultural machinery operation supervision service system, an agricultural machinery image annotation dataset containing the images of seven types of machines and six types of abnormal images was constructed. This dataset contained a total of 125,000 images, 100,000 of which were in the training set and 25,000 of which were in the validation set. To improve the quality of agricultural machinery images, image preprocessing functions, including image cropping, image color adjustment, elimination of motion blur, and image denoising, were carried out.

(2) According to the actual application requirements of the agricultural machinery operation supervision service system and the characteristics of the images, AMTNet was designed to improve the Inception_v3 network. In this study, two methods were applied to improve the accuracy after network training: decaying the learning rate and using the Adam optimization algorithm, and enhancing the C-DCGAN dataset.

(3) Compared with the classic networks, ResNet_50 and Inception_v3, AMTNet is superior in terms of recognition accuracy, recognition efficiency, and training time.

(4) To further test the performance of the AMTNet, 200 images of each of the 13 types of agricultural machinery images were selected as a test set. The mean values of both area under the curve and F1-score of AMTNet for recognizing images of various types of machines were 92% and 96%, respectively. The test results show that the


Predict class：Subsoiler with a curved surface shovel
Actual class：Subsoiling preparation machine
a

Predict class: Subsoiling preparation machine
Actual class：Seeder
b

Predict class：Subsoiler with a shovel
Actual class：Rotary cultivator
c

Predict class：Subsoiler with a shovel
Actual class：Rotary cultivator
d

Predict class：：Subsoiler with a curved surface shovel
Actual class：Subsoiler with a shovel
e

Predict class：Subsoiler with a curved surface shovel
Actual class：Subsoiler with a shovel
f

Predict class：Rotary cultivator
Actual class：Abnormal
g

Predict class：Rotary cultivator
Actual class：Abnormal
h

Predict class：Rotary cultivator
Actual class：Abnormal
i

Fig. 12. Incorrectly recognized agricultural machinery images.

AMTNet has good robustness to illumination, environmental changes, and small area occlusion, which meets the practical application requirements of intelligent supervision over agricultural machinery operation.

In conclusion, we constructed an agricultural machinery image annotation dataset according to agricultural machinery operation supervision service system. This was followed by designing an AMTNet and network training system, upon which automatic machinery recognition that satisfied practical demands was realized. Follow-up studies should focus on local recognition to further improve the network's capability to recognize machinery types.

Acknowledgements

This study is supported by the National Natural Science Foundation of China (Grant Nos. 31571563 and 31571564) and the National Key Research and Development Program of China (Grant Nos. 2017YFD0700600 and 2017YFD0700605).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2019.104978.

References

Baylor, D., Breck, E., Cheng, H.-T., Fiedel, N., Foo, C.Y., Haque, Z., Haykal, S., Ispir, M., Jain, V., Koc, L., 2017. Tfx: a tensorflow-based production-scale machine learning platform. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 1387–1395.

Brown, C.D., Davis, H.T., 2006. Receiver operating characteristics curves and related decision measures: a tutorial. Chemom. Intell. Lab. Syst. 80, 24–38.

Buades, A., Coll, B., Morel, J.-M., 2005. A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, pp. 60–65.

Catani, S., Ciafaloni, M., Hautmann, F., 1991. High energy factorization and small-x heavy flavour production. Nucl. Phys. B 366 (1), 135–188.

Dandan, Wang, Dongjian, He, 2019. Recognition of apple targets before fruits thinning by robot based on R-FCN deep convolution neural network. Trans. Chin. Soc. Agric. Eng. 35 (03), 156–163.

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., Roth, A., 2015. The reusable holdout: Preserving validity in adaptive data analysis. Science 349, 636–638.

Esakkirajan, S., Veerakumar, T., Subramanyam, A.N., PremChand, C., 2011. Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter. IEEE Sign. Process Lett. 18, 287–290.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Bing, X., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. International Conference on Neural Information Processing Systems.

Gribbon, K., Johnston, C., Bailey, D.G., 2003. A real-time FPGA implementation of a barrel distortion correction algorithm with bilinear interpolation. Image Vis. Comput. New Zealand 408–413.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134.

Jerez-Aragonés, J.M., Gómez-Ruiz, J.A., Ramos-Jiménez, G., Muñoz-Pérez, J., Alba-Conejo, E., 2003. A combined neural network and decision trees model for prognosis of breast cancer relapse. Artif. Intell. Med. 27, 45–63.

Kang, Kang, Zhongguo, Chen, Linfeng, Wang, Ziqiang, Tang, Meng, Jiang, 2018. Mobile agricultural equipment monitoring system based on internet of things. Jiangsu Agric. Sci. 46, 169–173.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neur. Inform. Process. Syst. 1097–1105.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.

Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 452–456.

Lin, M., Chen, Q., Yan, S., 2013. Network in Network. arXiv preprint arXiv:1312.4400.

Liu, S., Chen, L., Fan, X., Qu, Z., Yang, X., 2010. Combining pseudo-median filter and median filter to improve performance. In: 2010 3rd International Conference on Computer Science and Information Technology. IEEE, pp. 513–517.

Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2017. Notice of publishing "The 13th Five-Year Plan for the Development of Agricultural Mechanization in China" by the Ministry of Agriculture. [EB/OL].

Pruessner, J.C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D.H., 2003. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 28, 916–931.

Qiao, Y., Xin, X., Bin, Y., Ge, S., 2002. Anomaly intrusion detection method based on HMM. Electron. Lett. 38, 663–664.

Roffman, D., Hart, G., Girardi, M., Ko, C.J., Deng, J., 2018. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. Sci. Rep. 8, 1701.

Shaonong, Wang, Weidong, Zhuang, Xi, Wang, 2015. Research on agricultural machinery remote control management system. J. Agric. Mechaniz. Res. 37, 264–268.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

Smith, L.N., 2017. Cyclical learning rates for training neural networks. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 464–472.

Song, S., Miller, K.D., Abbott, L.F., 2000. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat. Neurosci. 3, 919–926.

Szegedy, Christian, et al., 2016. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Thanh Noi, P., Kappas, M., 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors 18, 18.

Xia, X., Xu, C., Nan, B., 2017. Inception-v3 for flower classification. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC). IEEE, pp. 783–787.

Yang, K., Liu, H., Wang, P., Meng, Z., Chen, J., 2018. Convolutional neural network-based automatic image recognition for agricultural machinery. Int. J. Agric. Biol. Eng. 11, 200–206.

Yang, Minli, 2011. Analysis on Situation of Agricultural Mechanization Development in the Next Five Years.

Yin, Y., Guo, S., Meng, Z., Qin, W., Li, B., Luo, C., 2018. Method and system of plowing depth online sensing for reversible plough. IFAC-PapersOnLine 51, 326–331.

Zhu, Y., Aoun, M., Krijn, M., Vanschoren, J., Campus, H.T., 2018. Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants. Computer Vision Problems in Plant Phenotyping (CVPPP2018) 1.