A Digital Dial Recognition Method Based on Lightweight ResNet

Gongming Zheng (zgm831@yangtzeu.edu.cn)
Yangtze University
Zhiwei Wang
Yangtze University

Research Article

Keywords: meter detection, image recognition, ResNet, lightweight model, character recognition

Posted Date: January 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1224235/v1

A Digital Dial Recognition Method Based on Lightweight ResNet

Gongming Zheng * , Zhiwei Wang,

(National Demonstration Center for Experimental Electrical and Electronic Education, Yangtze University, Jingzhou, Hubei 434000, People's Republic of China)

Abstract: This paper proposes a lightweight residual network for highly accurate recognition of the numbers of wheel dials. The proposed residual network is based on an optimized ResNet-18 network with the depthwise separable convolution and Dropout function. In addition, a case study on automatic water meter readings based on captured images is presented. Compared to existing methods, the results demonstrate improved recognition accuracy as high as 99.8% and increased computation speed with reduced model parameter size.

Key words: meter detection; image recognition; ResNet; lightweight model; character recognition

1. Introduction

With the advancement of computer and communication technologies in modern society, automatic intelligent meter reading has gradually become a reality [1-2]. Embedded image acquisition and processing chips, characterized by small size, low cost, low power consumption, and high performance, can be widely used for automatically reading household and enterprise-level water meters [3-4]. Along with the development of artificial intelligence technology in recent years, computer vision and image processing technologies can detect digital regions and recognize readings in water meter images, making it possible to detect and recognize images in complex environments [5]. Meanwhile, the extensive application of deep learning, especially Convolutional Neural Network (CNN) [6-8] and Recurrent Neural Network (RNN) [9] in the field of optical character recognition and sequence recognition makes it possible to read water meter images under complex background and strong interference. In [10], the improved Lenet-5 network model was adopted to recognize digital characters, and combined features before and after asymmetric convolution were fused to improve the recognition ability of the network for double half-characters. In the full image recognition method [11], all images with the same text information are regarded as a category, and the image text sequence recognition is transformed into image classification, thus avoiding complex and error-prone single character location and image segmentation problems. In [12], the improved VGG16 network model was adopted to recognize digital characters, and the recognition rate of small sample sets was improved by the transfer learning method. The above methods can not meet the requirements of transplanting to the edge system in one or two aspects, such as processing speed, double-half character recognition rate, or model parameter size, etc.

The contribution of the presented study is: (1) Based on the traditional ResNet-18 network, the algorithm used depthwise separable convolution to optimize the residual module and the Dropout function to enhance the robustness of the network model. (2) By comparing and validating different network models, the algorithm proposed in this paper has significantly improved the accuracy and speed of double half-character recognition, and significantly reduced the number of model parameters. (3) The multiplication calculation times of this model are reduced by 83%, which lowers the computation requirement, and provides the possibility for deployment to embedded devices, and lays a foundation for AIoT engineering application.

2. Character image preprocessing

In the working environment of the print wheel digital meter, the images collected by the camera are subject to uneven illumination and noise interference [13]. In order to recognize the meter character more accurately, the image should be preprocessed before the character recognition to reduce the noise interference and segment the character image. The main preprocessing steps are shown in Figure 1. Firstly, Gaussian filtering was used to denoise the original image, then gray processing was delivered, and histogram equalization was used to enhance the contrast of the image. Finally, image character location and segmentation were carried out.

Fig.1. The steps of image pre-processing

2.1 Gaussian filtering

In the process of image acquisition of print wheel digital meter, the stains on the dial and the noise generated by the camera affect the image quality and interfere with subsequent image recognition [14]. Consequently, in the first place, Gaussian filtering was adopted to denoise the collected image, which has a noticeable denoising effect on Gaussian noise and salt and pepper noise while retaining the overall gray distribution characteristics of the image. δ is the standard deviation of normal distribution, which determines the smoothness degree. For any point (x, y) in the image, two-dimensional Gaussian filtering is shown in Equation (1):

$$G(x,y)=\frac{1}{2\pi\delta^2}e^{-\frac{x^2+y^2}{2\delta^2}} \quad (1)$$

2.2 Histogram equalization

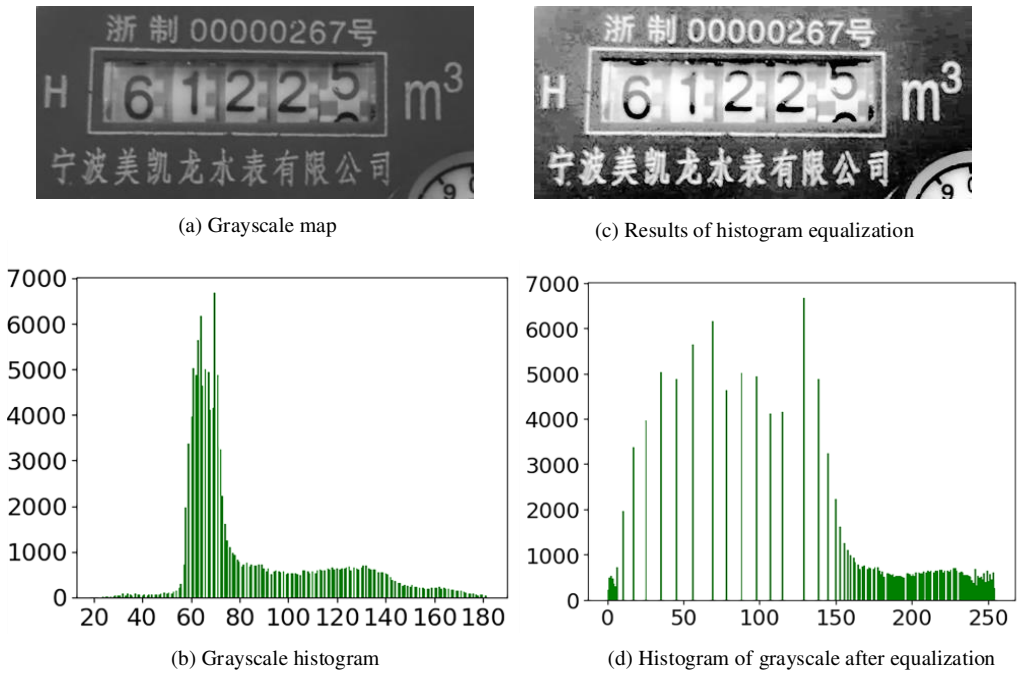
The print wheel meter images collected by the camera are RGB spatial color images, and the pixels of each image are composed of R, G, and B gray values. Color information has little effect on meter character recognition. First, gray processing was carried out to remove the color information in the image, so that each pixel has only one gray value to retain the gradient characteristics of the image, reduce the matrix dimension and improve the operation speed.

Gray processing cannot solve the problem that some areas of the collected image are too bright or too dark, so it is necessary to carry out histogram equalization of the image after gray processing, and solve the problem of too dark or too bright by adjusting the distribution of the gray value of the image, and improve the image contrast. The image histogram equalization algorithm mainly maps and converts the gray value of the image by inputting the histogram accumulative distribution function of the image. The mapping equation of the histogram equalization is shown in Equation(2). Where S_k represents the gray value after the current gray value k is mapped in the range of [0, L], L represents the maximum gray value of 255, n_i represents the number of pixel points with a gray value of k, and n represents the number of pixel points in the whole image.

$$S_k=L\cdot\sum_{i=0}^{k}\frac{n_i}{n} \quad (2)$$

This paper takes a water meter image as an example with gray processing and histogram equalization. Figure 2(a)is the image after gray processing, and Figure 2(b)is the gray histogram corresponding to Figure 2(a). The abscissa stands for the gray levels in the image, and the ordinate refers to the number of pixels of each gray level. Figure 2(c) is the image after histogram equalization.

Figure 2(d) is the gray histogram corresponding to Figure 2(c). It can be concluded from Figure 2 (b) that the gray pixel values in Figure 2(a) are distributed within [60, 180], indicating that the image in Figure 2(a) is dark. After histogram equalization, the gray pixel values in Figure 2(d) are distributed within [0, 255] . The visual inspection of Figure 2(c) indicates that the overall image brightness is improved, and contrast is enhanced.



(a) Grayscale map

(c) Results of histogram equalization

(b) Grayscale histogram

(d) Histogram of grayscale after equalization

Fig.2. Comparison graph of histogram equalization

2.3 Character location and segmentation

In the print wheel digital water meter image, the area that needs to be processed and analyzed is the digital display area of water quantity. Therefore, it is necessary to locate the area from the image and, separate it from the background, and then perform single-character segmentation on the numeric characters in the area. The primary process of character location and segmentation is as follows:

Firstly, binarization is performed on the image. The image binarization algorithm mainly converts the gray image into a binarized image with a pixel value of 0 and 255 through an appropriate threshold, and it can still reflect the overall and local features of the image. This paper uses the OTSU algorithm (OTSU image binarization algorithm [15]) to binarize the collected water meter image. The binarized threshold value is determined by between-class variance, which separates the foreground and background in the image t. Therefore, it can overcome the influence of image brightness and contrast. Figure 3(a) shows the results.

Secondly, the digital display area of water quantity is connected through corrosion and expansion algorithm in morphological operation. And the contour of the area is identified to locate the digital display area of water quantity and separate it from the background, with the segmentation results shown in Figure 3(b).

Finally, horizontal and vertical projections were performed on the segmented digital display areas, respectively. According to the projected data in the horizontal and vertical directions, upper and lower boundaries and single character segmentation were performed, with the segmentation results shown in Figure 3(c).

(a) Graph of binarization result (b) Graph of character positioning results (c) Graph of character segmentation results

Fig.3. Graph of character segmentation and positioning results

3. Network model construction

3.1 ResNet

The ResNet network proposed by Kaiming He et al won the ImageNet Image Recognition Challenge in 2015 [16]. This network is mainly to reduce the occurrence of gradient dispersion or gradient explosion in the deep neural network to ensure that the accuracy of the network model does not decrease [17-19]. ResNet network model is mainly composed of multiple residual modules shown in Figure 4. In the network training, when the accuracy of network model converges, this module can change the following neural network layer into the identity mapping layer, so as to maintain the optimal state of the network [20], avoiding the problem that the training accuracy decreases with the increase of neural network layers.

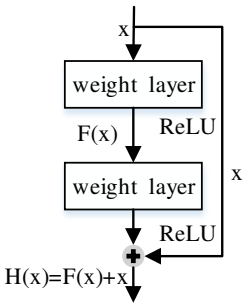


Fig.4. Sample graph of residual module

As can be seen from the figure above, the output result of the residual module is:

$$\mathrm{H}(x)=\mathrm{F}(x)+x \tag{3}$$

where x is the input value, $\mathrm{F}(x)$ is the residual mapping, and $\mathrm{H}(x)$ is the output value. An identity mapping is when $\mathrm{F}(x)=0$, $\mathrm{H}(x)=x$. Meanwhile, the residual mapping equation can be deduced from (1) :

$$\mathrm{F}(x)=\mathrm{H}(x)-x \tag{4}$$

According to (2), the network reaches the optimal state when F(x) infinitely approaches 0. Even with the increase of network layers, the training accuracy of the network model does not decrease. Existing classic ResNet network models can be classified into ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152 according to the different network layers [21].

3.2 Depthwise separable convolution

Depthwise separable convolution (DSC) is composed of DW (Depthwise) and PW (pointwise). The DW is to convolve the input image channel by channel, that is, one convolution kernel convolves one channel; the PW is to convolve point by point through the convolution kernel size of $1\times1\times M$, where M is the number of input image channels. Figure 5 is the block

diagram of the comparison between depthwise separable convolution process (a) and the traditional convolution process (b), where Kernel is the convolution kernel size and Filters are the number of convolution kernels. According to the convolution principle, in the case that both the convolution step size and the filling parameter are 1, the equations for calculating the number of parameters and the number of multiplication operations of the convolution layer are shown in (5) and (6) :

$$P_C = K_{wh} \times N \times C + N \tag{5}$$

$$N_m = K_{wh} \times N \times C \times O_{wh} \tag{6}$$

In (5) and (6), P_C is the number of parameters, N_m is the number of multiplication operations, K_{wh} is the product of the width and length of the convolution kernel, N is the number of convolution kernels, C is the number of input image channels, and O_{wh} is the product of the width and length of the output image.

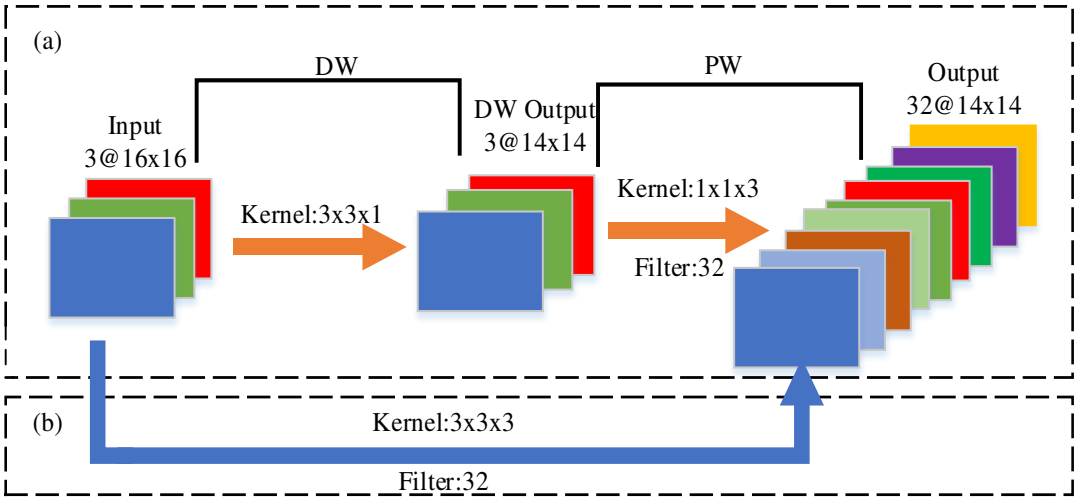


Fig.5. Comparison of depth separable convolution and traditional convolution methods

In Figure 5(a) about DSC, the DW is adopted with a convolution kernel with the size of 3×3×1.The input image size and output image features are 16×16×3 and, 14×14×3, respectively. In the process of single-channel convolution, K_{wh} = 9, N = 1, C = 1, O_{wh}= 196. In Figure 5(b), the traditional convolution uses 32 convolution kernels with a size of 3×3×3 to convolve with images with a size of 16×16×3, and the output image feature size is 14×14×32. In the traditional convolution process, K_{wh} = 9, N = 32, C = 3, O_{wh}= 196.Then, according to (5) and (6), The parameter numbers and multiplication times are calculated. The results are shown in table 1.

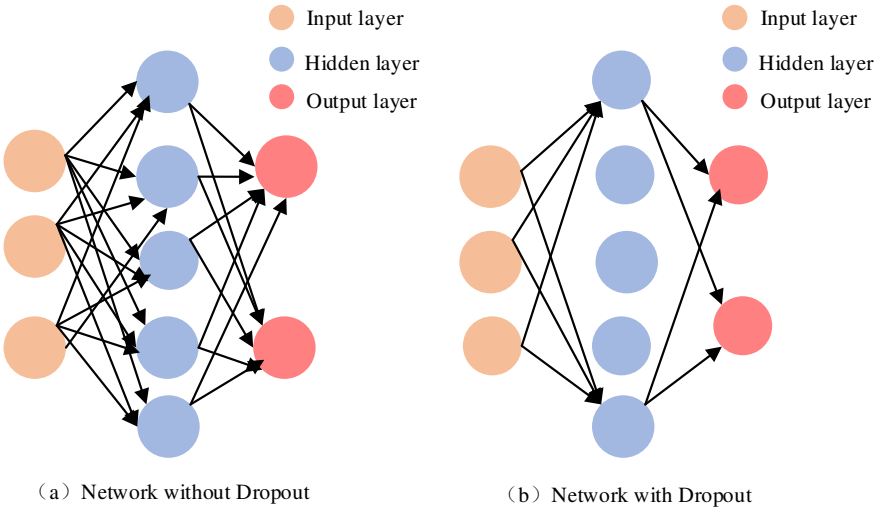
Tab.1. Comparison of parameter numbers and multiplication times

Label	DSC		Traditional convolution
	DW	PW	
K_{wh}	9	1	9
N	1	32	32
C	1	3	3
O_{wh}	196		
P_c	10*3	128	896
	158		
N_m	1764	18816	169344

In summary, compared with traditional convolution, the number of parameters and multiplication operations of depthwise separable convolution is significantly reduced. Specifically, the number of parameters decreased by 83%, from 896 to 158, and the times of multiplication operations decreased by 88%, from 169344 to 20580.It effectively reduces the number of operation parameters and improves the network operation speed.

3.3 Dropout function

The role of Dropout function [22] is to make the results of neural units be discarded with a certain probability during the training of deep learning, that is, the output is set to 0, and the weight is not updated. Since Dropout is to discard network neurons randomly, the network structure may differ, after each training. Therefore, the Dropout function can be regarded as model average, which means to average the prediction results of different models generated after training by corresponding weights. Figures 6-(a) and (b) show a network without and with Dropout function. The number of neurons in the hidden layer is five and two, respectively. In practical applications, the Dropout function is to randomly discard network neurons, thus weakening the joint adaptability between network neuron nodes, enhancing generalization ability and robustness, and effectively reducing the probability of overfitting.



（a）Network without Dropout　　（b）Network with Dropout

Fig.6. Comparison of networks with or without Dropout

3.4 Algorithm model construction

The network structure in this paper is shown in Figure 7. Based on the ResNET-18 network structure, the following improvements are made in this paper: 1) The number of residual modules is reduced to 4, to reduce the complexity of the network. 2) In the residual module, traditional convolution was replaced by depthwise separable convolution, and after convolution, the layer of batch normalization (BN) is added to improve the operation speed, convergence speed, and generalization ability of the network. 3) The Dropout function is added between fully connected layers to reduce the occurrence of overfitting. In this paper, the input image size of each channel is 32×32 ,respectively. The image size remains uncharged with convolutional layer and the number of channels becomes 16. After that, the image size becomes 2×2, and the number of channels is 256. Finally, with three fully connected layers, the Dropout function and Softmax function, the probability of 20 categories is

obtained. The number with the highest probability is the predicted value.

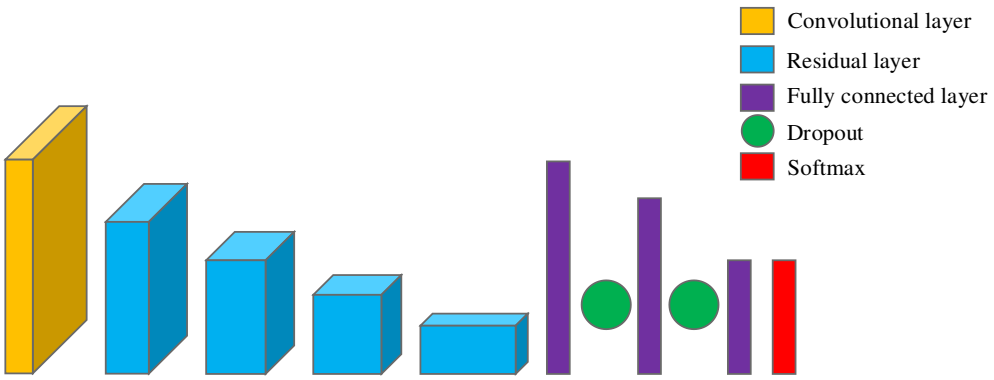


Fig.7. Structure of proposed network

The key parameters of the network model are shown in Table 2. Filters are the number of convolution kernels, W is the width of the input image, H is the height of the input image, K is the convolution kernel size, S is the step size, P is the Padding parameter, θ is the parameter of the Dropout function, and F is the fully connected layer or Softmax output depth.

Tab.2. Key parameters of the network model

Type	Parameter
Input	*Filters: 1， W: 32， H: 32*
Convolution	*Filters: 16， K: 3， S: 1， P: 1×1*
BN	--
Residual—1	*Filters: 32， K: 3， S: 2， P: 1×1*
Residual—2	*Filters: 64， K: 3， S: 2， P: 1×1*
Residual—3	*Filters: 128， K: 3， S: 2， P: 1×1*
Residual—4	*Filters: 256， K: 3， S: 2， P: 1×1*
FC1	*F:1024*
Dropout-1	$\theta = 0.5$
FC2	*F:512*
Dropout-2	$\theta = 0.5$
FC3	*F:20*
Softmax	*F:20*

4. Experimental results and analysis

4.1 Experimental platform and data set

The experiment in this paper is carried out under the PaddlePaddle deep learning framework, and the hardware environment is Baidu AI Studio cloud platform, including dual-core CPU, 8G memory, Tesla V100 GPU card, 16G video memory, 100G hard disk and PaddlePadd2.0 architecture.

This paper's data set is obtained using an OV5640 camera to collect the image of the digital dial in the the print wheel digital water meters, including a total of 25,000 dial images. After the image screening and preprocessing, 15,000 water meter character images are selected as the data set, with an image size of 32×32. The complete character image and double half-character image account for 50% respectively. The proportion of training set data and test set data is 8:2. The full character image and

double half-character image are marked respectively, the full character is marked with its value [0-9], and the double half-character label is marked with [10-19], as shown in Figure 8.

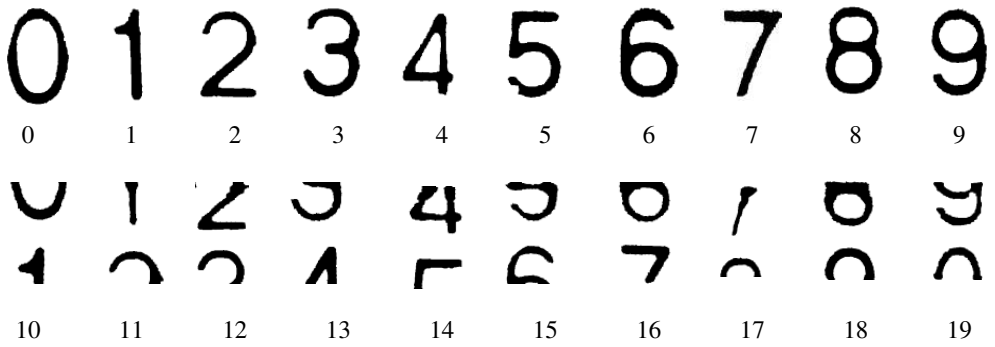


Fig.8. Example of data tagging

4.2 Network training parameters

ReLU function is used as the activation function in the network model, which is

$$y = \begin{cases} x & x > 0 \\ 0 & x \le 0 \end{cases} \tag{7}$$

Adaptive Moment Estimation (Adam [23]) algorithm is used in network training, which dynamically adjusts the learning rate of each parameter in the network by using the first-order moment estimation and second-order moment estimation of the gradient. It is assumed that the first-order moment estimation and second-order moment estimation of the gradient are δ_t and λ_t respectively, γ represents the learning rate, ε is the smoothing term set to prevent the denominator from being zero, and the parameter update equation at t+1 is shown in (8). The hyperparameters of Adam algorithm selected in this paper are: learning rate is γ = 0.0012, the exponential decay rate of the first-order moment estimation of the gradient is β_1 = 0.9, the exponential decay rate of the second-order moment estimation of the gradient is β_2 = 0.999 and parameter is ε = 1×10^{-8}. After 8,000 rounds of iterative training for the network model under the above parameters, the error value is less than 1×10^{-4}, and then the network model is judged to be in the fitting state; that is, the network model training can be stopped.

$$\theta_{t+1} = \theta_t - \gamma \times \frac{\hat{\delta}_t}{\sqrt{\hat{\lambda}_t} + \varepsilon} \tag{8}$$

4.3 Experimental analysis

4.3.1 Analysis of network parameters

In this paper, the convolution kernel size and Dropout parameter values of deep neural networks are tested under different conditions, as shown in Table 2. When the convolution kernel size is 3×3, and 5×5, and the Dropout parameter is 0.2, 0.5, and 0.7, the character recognition accuracy of the dial is determined. Among them, ACC represents the character recognition accuracy of the dial. K-Size represents the convolution kernel size, and θ represents the Dropout parameter value. It can be seen from the data in Table 3 that the convolution kernel size of 3×3, and the Dropout parameter of 0.5

selected for the network in this paper are most effective.

Tab.3. Analysis of network parameters

ACC/%	θ		
K-Size	0.2	0.5	0.7
3×3	99.60	99.80	99.60
5×5	98.80	99.00	98.20

4.3.2 Analysis of depthwise separable convolution

A comparative test is conducted on the network parameters size, operation, and reasoning speed using traditional convolution and depthwise separable convolution. The results are shown in Table 4: CNN-1 represents the network using traditional convolution; Size represents the parameter size of the corresponding network; TSpeed represents the average operation time of a single image of the corresponding network; PSpeed represents the average reasoning time of a single image of the corresponding network. The proposed netwaork model reduces the total parameter number by 56% compared to CNN-1, as shown in Table 4.Moreover, it increases the operation and reasoning speed by 21% and 25%, recpectively.

Tab.4. Analysis of deep separable convolution

Model	Size/MB	TSpeed/ms	PSpeed/ms
CNN-1	17.7	32	16
Proposed Network	7.71	25	12

4.3.3 Analysis of network training parameters

This experiment investigates the impact of different learning rates on the network model accuracy. The test results are shown in Figure 9. If the learning rate of the convolutional neural network is too low, the network parameters may be updated slowly, and network convergence may be slow. If the learning rate of the network is too high, the gradient may fluctuate violently during training. Therefore, it is difficult to find the direction of the fastest gradient descent. Based on the experimental test results, the learning rate of the network is set to 0.0012, making the network model stably achieve the shorter time, the greater smoothness and the higher accuracy more than others.

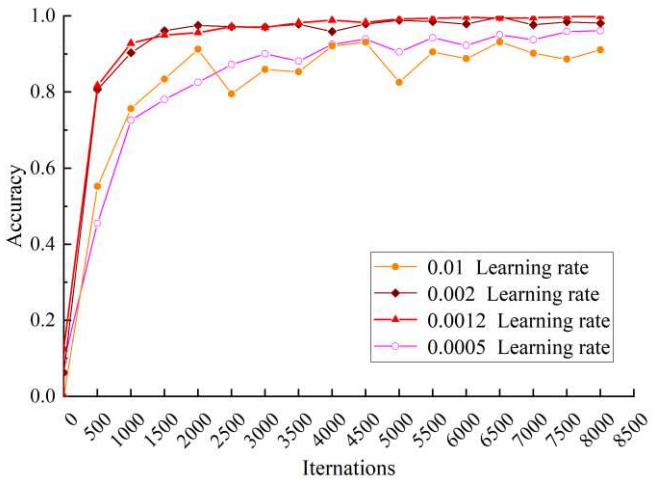


Fig.9. Accuracy of network models with different learning rates

4.3.4 Analysis of model performance

The proposed network model is compared to ResNET-18 network model and Lenet-5 network model in [10] to benchmark the performance. The comparison results are shown in Table 4. ACC represents the complete character recognition rate of the model; Double-ACC represents the double-half character recognition rate of the model; Size represents the parameter size of the model, and TSpeed represents the average operation time of a single image of the model; PSpeed represents the average reasoning time of a single image of the model.

Tab.5. performance benchmarking of different network models

Model	ACC/%	Double-ACC/%	T-Speed/ms	P-Speed/ms	Size/ MB
ResNet-18	98.40	97.72	42	19	20.18
Lenet-5	99.14	98.20	590	--	--
Proposed network	99.80	99.68	25	12	7.71

The benchmarking shown in Table 5 reveals an increased of 1.4% and 0.66% complete character recognition rate compared to ResNet-18 and Lenet-5, respectively. The increase becomes 1.96% and 1.48% in the double half-character recognition rate, respectively, with better anti-interference capability. Compared to the ResNet-18 network method, single imag's operation and reasoning time are reduced by 17ms and 7ms, respectively, implying significantly increased speed. By comparing the number of parameters of the three network models, the network model in this paper has the lowest number of parameters, less than 8MB, which is feasible for deployment to lightweight embedded devices.

5 Conclusions

A lightweight residual network is proposed in this paper to recognize the number of wheel dials. It is based on an algorithm using depthwise separable convolution to optimize the residual module and the Dropout function to enhance the robustness of the network model. The case study on water meter readings shows accuracy levels as high as 99.8% can be achieved, which is improved compared to traditional ResNet-18 and Lenet-5 methods. Moreover, the proposed method significantly reduces the

required number of model parameters and reduces the times of multiplication calculations by 83%, contributing to much faster computation speeds. It lays a foundation to develop a solution for deployment to embedded devices in practical AIoT engineering applications.

References

[1] Peng Kunfu. Research on pointer meter reading recognition method based on deep learning. China University of Science and Technology, 2020.

[2] Zhang Ran et al. Design of electric power remote meter reading system. Journal of Physics: Conference Series, 2021, 1802(3) : 032030-032039.

[3] Huang Yu. Design and realization of intelligent remote meter reading system software. University of Electronic Science and Technology, 2020.

[4] Qian Zhang et al. Smart city information acquisition system based on internet of things. Cluster Computing: The Journal of Networks, Software Tools and Applications, 2019, 22(5) : 9013-9025.

[5] Zhan Wei. Research on intelligent meter reading based on the internet of things. Beijing University of Posts and Telecommunications, 2018.

[6] Simonyan K, Zisserman A. Very deep Convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[7] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[8] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep Convolutional Neural Networks. Advances in Neural Information Processing Systems[C]. 2012: 1097-1105.

[9] Williams R J, Zipser D. A learning algorithm for continually running fully Recurrent Neural Networks. Neural computation, 1989, 1(2): 270-280.

[10] Shao Liang, Tu Junxiang, Yu Jie. Digital recognition of water meter based on combined features fusion of Lenet-5 network. Machinery manufacturing and automation, 2020,49(06): 189-192.

[11] Almazán J, Gordo A, Fornés A, et al. Word spotting and recognition with embedded attributes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(12): 2552-2566.

[12] Fan Rui, Jiang Pinqun, once upstream, Xia Haiying, Liao Zhixian, Li Peng. Design of lightweight convolution neural Network based on multi-scale parallel fusion.Journal of Guangxi Normal University (NATURAL SCIENCE EDITION), 2019,37 (03): 50-59.

[13] Li Tiantian. Wheel water meter reading recognition based on improved convolutional neural network. Chongqing Normal University, 2019.

[14] Hou Yuanyuan, He Ruhan, Liu Junping. Fusion color ferature and depth feature clothing image retrieval algorithm. Computer Applications and Software,2020,37(10):194-199.

[15] Otsu N . A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems Man & Cybernetics, 1079, 9(1):62-66.

[16] He K., Zhang X., Ren S., et al. Deep Residual Learning for Image Recognition In computer vision and pattern recognition. Las Vegas, NV, USA: IEEE press, 2016:770-778.

[17] Yan Le et al. Flow interval prediction based on deep residual network and lower and upper boundary estimation method. Applied Soft Computing Journal, 2021, 104.

[18] Luo Junli et al. Automatic identification of cashmere and wool fibers based on microscopic visual features and residual network model. Micron, 2021, 143.

[19] He Jia , Meng Weiwei , You Tingting. Intelligent Recognition of Vehicle Information in Surveillance Video. Journal of Physics: Conference Series, 2021, 1802(3) : 032049-032061.

[20] Li Ailian et al. Study on the status identification of slag steel under improved ResNet101 network. CHINA MEASUREMENT & TEST,2020,46(11):116-119+125.

[21] Tian Qiang, Jia Xiaoning. Vehicle logo recognition based on deep residual network. Journal of

Jilin University (Science Edition), 2021,59 (02): 319-324.
[22] SRIVASTAVAN, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
[23] KINGMA D P, BA J. Adam: a method for stochastic optimization.International Conference on Learning Representations[C]. 2014.