SONAR IMAGE RECOGNITION BASED ON MACHINE LEARNING FRAMEWORK

MiaoDong[1] , HaiyangQiu[1, *], HuiWang[1], PengfeiZhi[1], ZihaoXu[1]

[1] Electronic Information School, Jiangsu University of Science and Technology, China - 1593829826@qq.com

Commission IV, WG IV/5

KEY WORDS: Sonar data; deep learning; Caffe framework; Inception-Resnet-v2 model; migration learning.

ABSTRACT:

In order to improve the robustness and generalization ability of model recognition, sonar images are enhanced by preprocessing such as conversion coordinates, interpolation, denoising and enhancement, and the transfer learning method under the Caffe framework of MATLAB as an interface is used respectively (mainly composed of 8 layers of network structure, including 5 convolutional layers and 3 full chain layers) And the transfer learning method under the Python deep learning framework Inception-Resnet-v2 model for sonar image training and recognition. First of all, part of the sonar image dataset (derived from the 2021 National Robot Underwater Competition online competition data), using MATLAB as the interface Caffe framework, the sonar image is trained to obtain a training model, and then through parameter adjustment, the convolutional neural network model of sonar image automatic recognition is obtained, and the transfer learning method can use less sonar image data to solve the problem of insufficient sonar image data, and then make the training achieve a higher recognition rate in a shorter time. When the training data is randomly sampled for testing, the sonar data recognition model based on the Caffe framework is quickly and fully recognized, and the recognition rate can reach 92% when the test sample does not participate in the training of sonar image data; The transfer learning method under the Inception-Resnet-v2 model of python deep learning framework is used to train recognition on sonar images, and the recognition rate reaches about 97%. Using the two models in this paper, it is feasible to identify sonar images with high recognition rate, which is much higher than traditional recognition methods such as SVM classifiers, and the two sonar image data recognition models based on deep learning have better recognition ability and generalization ability.

1. INTRODUCTION

Recognition sonar is the emission of high-frequency sound waves to the water, when the sound waves hit the underwater obstacles are returned, and then receive the echo, from the echo signal to form a sonar image. At present, the sonar image of the target can be obtained at a distance of several hundred meters, but with the maturity of sonar image acquisition technology, efficient and rapid processing of sonar images and identification of sonar images have become problems that need to be solved at present, and the automatic recognition of underwater targets has become a research hotspot. The traditional sonar image automatic recognition method is mainly to obtain the characteristics of the target image through experimental simulation, and then design a classifier to classify the sonar image. For this method, if you want to achieve a high recognition rate and accuracy rate, you must select a large number of features and a combination of various features, give different weights to adjust the parameters, low generalization ability, change the performance of this method of degrading, and train the manually designed features to the classifier robustness is not high.

With the development of artificial intelligence, deep learning methods have been widely used in many fields such as speech recognition, image recognition, image classification, natural language processing, and video analysis. At present, commonly used deep learning methods are deep belief network (DEEP belief network, DBN), convolutional neural network (CNN) and recurrent neural network (RNN), etc., of which the convolutional neural network model is most widely used in the field of vision and image recognition, compared with traditional images, sonar image recognition methods are also similar. It is mainly divided into feature extraction and recognition. In the traditional method of manual design of feature re-recognition method has been replaced by convolutional neural networks and other new algorithms, convolutional neural network methods in the field of optical image recognition is widely used, with the development of the application to radar target detection, but whether it is optical images or radar image training to have a large number of image data, and sonar images do not have a huge database, so the use of convolutional neural networks in sonar images is plagued by the problem of insufficient database. The overall architecture of a CNN consists of a series of stages of the convolutional and pooled layers. Includes local connections, sharing rights, pooling, and multi-tier usage. CNNs learn to extract features from images through their own regular learning, generate unique maps through repeated learning, and finally connect layers similar to existing hierarchical neural networks to produce the desired results.

This study uses two methods to train sonar images, first in MATLAB as the framework of Caffe interface, under the basis of the model, in the training process of CNN combined with the transfer learning method, through the adjustment of the network, weight adjustment training to obtain a training model. Secondly, the Python deep learning framework Inception-Resnet-v2 model is used to combine transfer learning to obtain a training model. The migratory network retains a large number of features of the original network, and only the last few layers of network structure need to be fine-tuned to reduce the difficulty of network training and the required image data.

In view of the application potential of CNN, this paper first pre-processes sonar data images to strengthen the contrast between the features and backgrounds of the images, so as to facilitate the subsequent extraction and recognition of features; then, on the basis of the Caffe model, the migration learning method and the Inception-Resnet-v2 model are combined with the transfer learning method to train samples; finally, the same image dataset and different image datasets are tested to obtain the

recognition accuracy, error rate and generalization ability of the two methods.

2. DATA PROCESSING

2.1 Data sources

The sonar image data comes from the 2021 National Robot Underwater Competition online competition data, the training set is marked with categories, and the target types include cube, ball, cylinder, human body, tire, circle cage, square cage, and metal bucket. The training set is stored in 0-7 folders under the train folder, the name is the category, and the test set is under the test folder.

2.2 pretreatment

2.2.1 Build a sonar image: From the sonar strafing can be seen, the image composed of the echo signal is a flat sector, first of all, each pixel in the plane rectangular coordinate system is converted into a polar coordinate system, and the converted pixels form a fan sonar pixel map, the principle is Figure 1.1.1, assuming that P(x,y) is any pixel point under the cartesian coordinate system, for the pixel points under the corresponding polar coordinates, any point P in the following figure is an example: Figure 1 a is the original pixel data map, assuming that the total distance of the detection of the original data is high h=5, width w= 3, Figure b is the process of output data size conversion from a Cartesian coordinate system to polar coordinates, and the calculation formula is:

$$L_{\text{length}} = \text{h} = 5 \tag{1}$$

$$L = L_{\text{start}} + L_{\text{length}} \tag{2}$$

$$\text{h}_{\text{out}} = L \tag{3}$$

$$\text{W}_{\text{out}} = 2 * L\sin(\text{angle}) \tag{4}$$

where L_{start} is the starting length of the sample, L_{length} is the total distance of the probe, and are the length and width of the converted image, respectively.
Figure C calculates the coordinates corresponding to each pixel point after the pixel map is converted from a planar cartesian coordinate system to a polar coordinate system, and the calculation formula is:

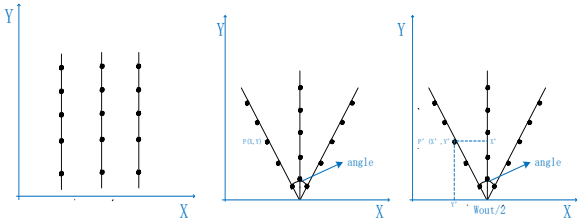
$$(L_{\text{start}} + 3) * \cos(\text{angle}) = \text{y}' \tag{5}$$

$$(L_{\text{start}} + 3) * \sin(\text{angle}) + \text{W}_{\text{out}}/2 = \text{x}' \tag{6}$$

From the above formula, $\text{p}'(\text{x}', \text{y}')$ is the coordinate corresponding to $\text{p}(\text{x}, \text{y})$ in the polar coordinates. Deriving the coordinates of point P under polar coordinates, equation (7) obtains the pixel value at $\text{p}'(\text{x}', \text{y}')$

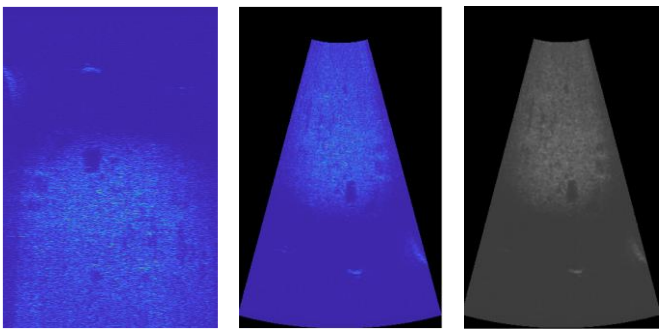
$$\text{im}_{out}(\text{y}', \text{x}' > 3) = \text{im}_{\text{in}}(\text{y}, \text{x} > 3) \tag{7}$$

where im_{out} and im_{in} are output and input pixel values, respectively.



(a) (b) (c)

Figure 1. Coordinate conversion schematic.



(a) Original (b) Polar plot (c) Grayscale plot.

Figure 2. Coordinate conversion rendering.

2.2.2 Image interpolation: After the image is converted by coordinates, the pixels are converted from the Cartesian coordinate system to the polar coordinate system, and due to the rounding operation during the conversion process, some areas are not filled and some pixels are lost. In order for the image to be closer to the original image, and the image is uniform and does not leave blanks, the blank space where the pixels are missing is interpolated using continuous uniform interpolation. First determine whether the interpolated point is within the sector, the calculation formula is:

$$\text{y} = \text{y}' \tag{8}$$

$$\Delta\text{x} = \text{y} * \tan(\text{angle}) \tag{9}$$

$$\rho = \sqrt{(\text{x} - \text{x}_0)^2 + \text{y}^2} \tag{10}$$

If $\text{x} - \Delta\text{x} < \text{x}' < \text{x}_0 + \Delta\text{x}$ and $\rho_{L_{\text{start}}} < \rho < \text{h}_{\text{out}}$ are determined to be within the sector, the interpolation point is determined to be in the sector, and then equation (11) calculates the signal strength of the interpolated point. For any , find two nodes adjacent to X , $X(i)$ 、 $X(i+1)$, make $X(i) \le X \le X(i+1)$, and use formula (11) to interpolate the formula to calculate.

$$y \approx y(i) + [x - x(i)][y(i+1) - y(i)] / [x(i+1) - x(i)] \tag{11}$$

2.2.3 Image denoising:Sonar image in the process of receiving the signal, will be subject to a variety of different degrees of noise interference, there are many ways to remove noise, this paper uses Lee filtering to reduce the noise of the sonar subgraph, Lee filter directly weights the pixels of the window preprocessing area. Throughout the filtering process, the entire image is first transformed into the same frequency domain, then filtered and then inverted. Since the multiplication and mean squared error of the noise are minimal, the Equation 12Lee filtering algorithm is available:

$$I = \bar{n}R + \bar{R}(n - \bar{n}) \tag{12}$$

where R represents the original image without noise and n represents noise independent of the distribution of the original image. The letter I indicates the image that is noisy.
Suppose is a linear combination of the means of and , which can be represented by Equation 13

$$\hat{R} = a\bar{I} + b\bar{I} \tag{13}$$

Calculate the absolute value of the difference between the R and R estimates and the squared and expected values to get the value of the mean squared error of R . Let a and b return to zero, respectively, find the minimum value, calculate the values of a and b respectively, and substitute the values of a and b back to Equation 13, respectively, and the final result is shown in Equation 14:

$$R = \bar{I} + k(I - \bar{I}) \tag{14}$$

where the value of the coefficient k is calculated by Equation 15:

$$k = \frac{1 - \bar{I}^2 * \sigma_n^2 / \mathrm{var}\,(I)}{1 + \sigma_n^2} \tag{15}$$

Using the standard deviation factor of image I , C_I as a measure, two thresholds $C_{\min}$ and $C_{\max}$ are set to achieve segmented denoising. The value of 5 is calculated by Equation 16:

$$C_I = \sqrt{\mathrm{var}(I)}/\bar{I} \tag{16}$$

The standard deviation factor of C_I indicates how different the image is within the current window: when C_I is larger, the part may be at the edge of the image or isolated pixels. Therefore, when $C_I \le C_{\min}$, the image is in a flat area, and the mean filter should be used for filtering; When $C_I \ge C_{\max}$, the edge information of the image is preserved. When the C_I value between the two thresholds is small, the noise effect in this part of the image is the largest, and a weighted enhanced Lee filtering method is used to improve the signal-to-noise ratio of the peak. Calculates the weight of the current area of the image and weights the pixel value of the image. The mathematical expression of the entire algorithm is as follows:

$$\hat{R} = \begin{cases} \bar{I} & C_I \le C_{min} \\ I * W_2 + (1 - W_2) & C_{min} \le C_I \le C_{max} \\ I & C_I \ge C_{max} \end{cases} \tag{17}$$

$$W_2 = \frac{1 - (1 - W_1) * I_{\mathrm{med}}}{1 + \sigma} \tag{18}$$

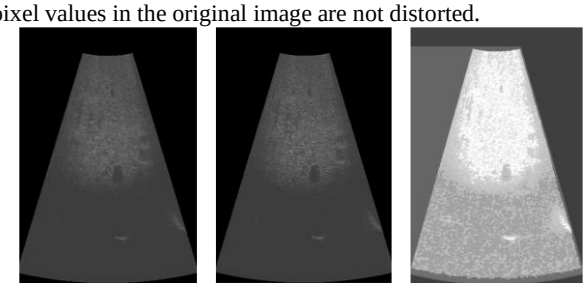
$$W_1 = \exp\left(-T\frac{C_I - C_{min}}{C_{max} - C_I}\right) \tag{19}$$

where I_{med} represents the value of the data in the current window, which is the difference between the entire image standard, T is adjustable to adjust the rate of change of , and the value of in this article is 1.

2.2.4 Image enhancemen: After the construction of the sonar image and the interpolation denoising, the sonar image is colorless compared to the optical image. The corresponding value of each pixel represents the intensity of the sonar echo, the stronger the signal, the more obvious the pixel value, the figure is replaced by white, but most of the sonar images have low gray value and poor contrast, which is not conducive to subsequent analysis and processing, so the image will be enhanced.
Purposefully emphasize the overall or local characteristics of the image, turn the original unclear image into a clear or emphasize the pixel features we need, expand the difference between the features of different objects in the image, suppress unwanted background features, improve image quality, strengthen image interpretation and recognition effects, and meet the needs of analysis. Image enhancement using histogram linear rollovers, with the formula 20:

$$F(x,y) = T[f(x,y)] = a^* f(x,y) + b \tag{20}$$

where $f(x, y)$ is the gray value of the original image, $T(\bullet)$ is the mapping function, and $F(x, y)$ is the gray value of the transformed image. A is the enhancement factor, b is the offset factor, b is often 0, a as far as possible to make the image pixel values evenly distributed and ensure that the high pixel values in the original image are not distorted.



(a) Grayscale plot (b) Denoising (c) intensifier

Figure 3. Image enhancement.

3. ALGORITHM DESCRIPTION

3.1 Image data and methods

Since the training of the model requires a large number of image extraction features, in order to avoid overfitting, the image is pre-processed such as data enhancement and denoising, and the preprocessed sonar images are randomly selected as the test set, and the rest are used as the training set, and the target types include cube, ball, cylinder, human body, tire, circlecage. Square cage, metal bucket 8 categories.

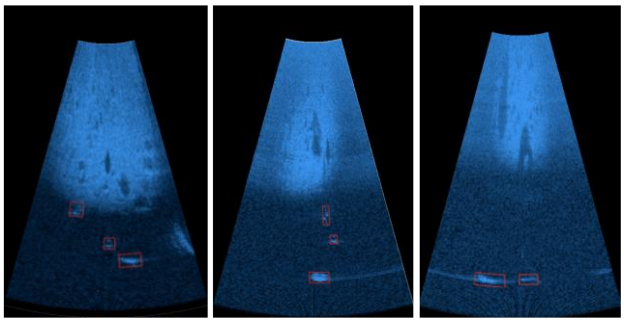


Figure 4. Target type.

a is a front-view sonar image, red is the target label box, from left to right is square cage, ball, tire.
b is the front-view sonar image, red is the target label box, from top to bottom is the human body, the ball, the circle cage .
c is the front-view sonar image, red is the target label box, from left to right is the tire , metal bucket.

3.2 The MATLAB interface is used to establish the Caffe sonar feature recognition model and the transfer learning model

caffe is a clear and efficient deep learning framework, is a C++/CUDA architecture, supports command line, python and MATLAB interfaces; can be seamlessly switched between CPU and GPU, Caffe's basic workflow is designed based on a simple assumption built on the neural network, all computation is represented in the form of layers, and what the network layer does is input data and then output the results of the calculations.
The caffe command-line interface can be used to learn the model, test the score of the run model, and represent the final result of the network output as a percentage, detect system performance and measure the relative execution time of the model, this command performs model detection by layer-by-layer timing and synchronization.

Based on the Caffe framework, this paper uses the migration learning method to fine-tune parameters and establish a sonar image recognition model. The structure of the Caffe Net sonar image feature model is shown in Figure 5, which is mainly composed of 8 layers of network structure, of which 5 convolutional layers and 3 full chain layers.

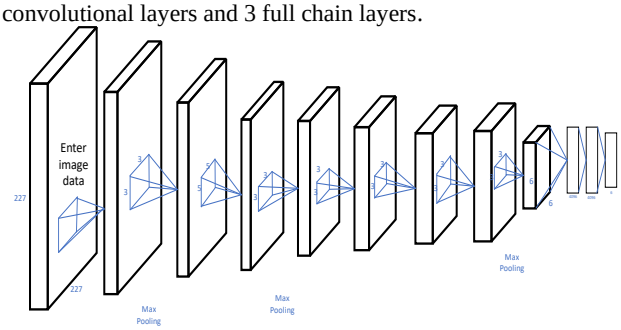


Figure 5. Caffe Net sonar image feature model structure

Enter the picture data (227 * 227 * 3), the first layer of convolution of the input picture such as the picture is convoluted, and then the convoluted picture is activeivation, normalization, and pooling, etc., etc. will not change the image size and therefore will not change. The first two convolutional layers integrate convolution, activation, pooling and normalization, the third and fourth convolutional layers only include convolution and activation, the full chain layer includes activation and dropout operations, the dropout layer weakens the fitting effect of the deep neural network, taking the default value of 0.5, and the output of the final fully connected layer is the SoftMax layer of class 8. Among them, the excitation function uses ReLU, convolution is mainly used to extract the features of each small part of the picture, and the output matrix after the pooling layer input is multiplied by the original data output of the convolutional layer and the corresponding convolution, reducing the number of training parameters, reducing overfitting only retains the useful part, and the output is 8 classes.

Sonar images commonly used in the total number of pictures is not large, the use of a small number of training samples to train may cause network overfitting, first of all, the weights of the model are initialized, using the method of migration learning, the convolutional neural network is regarded as a feature extractor, trained on the existing data set, the model with good classification is adjusted, the recognition model of the sonar image feature is obtained, and then the target data set is tested.

Using migration learning for sonar image feature recognition, first select a dataset with a large number of labels to train the pre-trained network, move the parameters of some of its layers to the target network, remove the last fully connected layer in the original network, and add a fully connected layer to the migrated network, and train the new network with labeled data to complete the migration. According to the initialization parameters, the image is backpropagated to calculate the error and adjust the weight parameters, and the loss function is minimized by continuous calculation, and the resulting weight parameters are used as the weight parameters of the final model.

3.3 The Inception-ResNet-v2 model is combined with a transfer learning model

3.3.1 Inception module: Sonar data image pixel position differences, sizes are not equal, it is difficult to choose the appropriate convolutional kernel size, deep networks are easy to overfit, simple stacking of large convolutional layers consume computing resources, so run a filter with multiple sizes at the same level, so the network essence is wider, not editing, so the design of the Inception module, with three filters of different sizes.

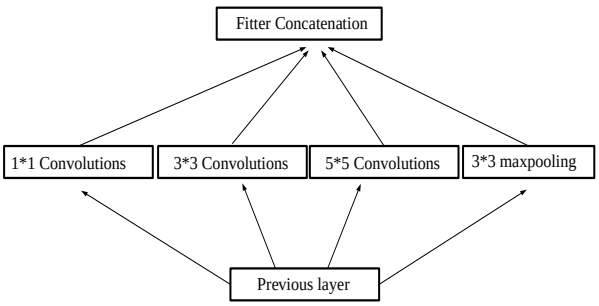


Figure 6. Original Inception module.

3.3.2 ResNet algorithms and transfer learning: Deep residual networks (deep residual networks, ResNet) is a convolutional, by deepening the network level to improve the accuracy of the network model structure, network deepening training accuracy declines, as the network layer is deeper, its performance tends to be more saturated, and even begin to decline rapidly, so with the deep residual network to introduce an identity fast connection directly skip one or more layers, improve computing efficiency, mainly by the data input layer, convolutional computing layer, ReLu excitation layer, pooling layer 3. It is composed of a fully connected layer and an output layer. The working principle of ResNet is shown in Figure 7, the image features are extracted by the convolutional layer, output to the excitation layer, and then the pooling layer is linearly mapped and then the pooled layer is processed to reduce dimensionality, and finally the previous discrete three-dimensional features are transformed by the fully connected layer into global features that can reflect the image.

The actual residuals are not equal to zero, so the residual function can learn new features in the stacking layer, making the convolutional structure more effective use of multi-layer network information, according to the pooled image through the full connection layer can get the probability matrix of 1000 image features, but this article only needs to divide it into 8 categories, so the transfer learning model is used, all network weights in ResNet except the fully connected layer are fixed, and only the last layer of the fully connected layer is replaced with "8 fully connected" with random weights "Fully connected layer, get a probability matrix of 8 image features, and train this layer, as shown in Figure 9. The transfer learning model can make the deep convolutional neural network model have higher accuracy and training speed in a short period of time, and it is more suitable for the requirements.

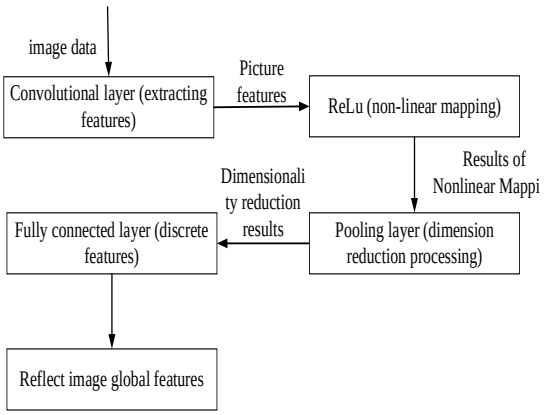


Figure 7. ResNet working principle diagram.

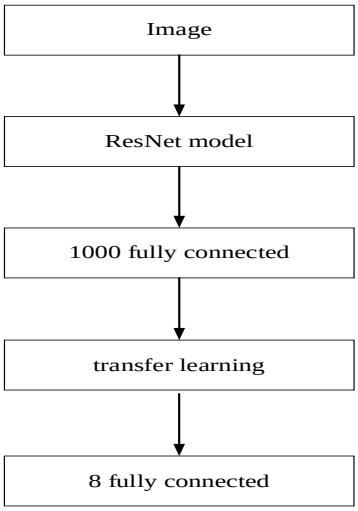


Figure 8. Transfer learning model.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Training results of Mat Caffe convolutional neural network and transfer learning model based on image features

The training model and test are all run by MATLAB as caffe interface, and this experiment mainly trains the migration model on the three parameters of the number of trainings, the number of network data submitted each time, and the learning rate. Adjusting the number of submission networks per commit is good for solving the problem of excessive amount of data, and the optimal number of commit networks can achieve the best balance between memory efficiency and memory capacity, and the model performance and speed are optimal.The random picture set assigns each type of picture according to 80% training set and 20% test set, and sets the step to 1 in each iteration to take out loss and accelerator, and the test results are as shown in Figure (9):

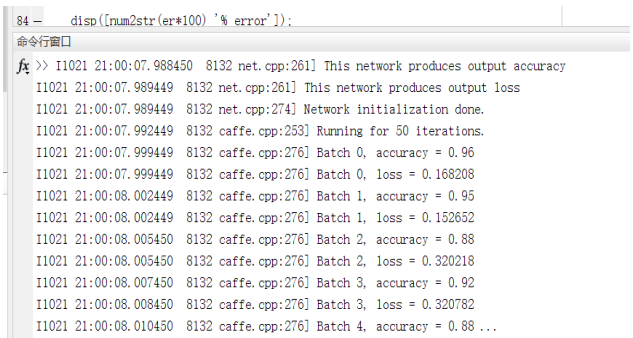


Figure9. Training model results.

caffe provides the basic input function, but this article imported the picture with the input itself, in order to test the model, a part of the training data as test data, test the trained network recognition rate, get the results as shown in Figure 10, in the test 33 times has reached 100% accuracy.

```
Training on single GPU.
Initializing image normalization.
|=========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |    (hh:mm:ss)  |   Accuracy   |     Loss     |      Rate       |
|=========================================================================================|
|       1 |           1 |       00:00:00 |       5.47% |       3.2156 |      5.0000e-05 |
|       1 |          50 |       00:00:01 |      56.25% |       1.3366 |      5.0000e-05 |
|       2 |         100 |       00:00:02 |      60.94% |       1.1112 |      5.0000e-05 |
|       2 |         150 |       00:00:03 |      75.78% |       0.7464 |      5.0000e-05 |
|       3 |         200 |       00:00:04 |      74.22% |       0.7222 |      5.0000e-05 |
|       4 |         250 |       00:00:06 |      88.28% |       0.4253 |      5.0000e-05 |
|       4 |         300 |       00:00:07 |      86.72% |       0.4561 |      5.0000e-05 |
|       5 |         350 |       00:00:08 |      89.06% |       0.3768 |      5.0000e-05 |
|       6 |         400 |       00:00:09 |      93.75% |       0.3156 |      5.0000e-05 |
|       6 |         450 |       00:00:10 |      91.41% |       0.3044 |      5.0000e-05 |
|       7 |         500 |       00:00:12 |      91.41% |       0.2672 |      5.0000e-05 |
|       8 |         550 |       00:00:13 |      96.09% |       0.1679 |      5.0000e-05 |
|       8 |         600 |       00:00:14 |      96.09% |       0.2026 |      5.0000e-05 |
|       9 |         650 |       00:00:16 |      96.09% |       0.1934 |      5.0000e-05 |
|       9 |         700 |       00:00:17 |      96.88% |       0.1153 |      5.0000e-05 |
|      10 |         750 |       00:00:19 |      96.09% |       0.1510 |      5.0000e-05 |
|      11 |         800 |       00:00:20 |      99.22% |       0.1020 |      5.0000e-05 |
|      11 |         850 |       00:00:22 |      99.22% |       0.0918 |      5.0000e-05 |
|      33 |        2500 |       00:01:01 |      99.22% |       0.0182 |      5.0000e-05 |
|      33 |        2550 |       00:01:02 |     100.00% |       0.0146 |      5.0000e-05 |
|      34 |        2600 |       00:01:03 |     100.00% |       0.0188 |      5.0000e-05 |
|      34 |        2650 |       00:01:04 |     100.00% |       0.0122 |      5.0000e-05 |
|      35 |        2700 |       00:01:06 |     100.00% |       0.0142 |      5.0000e-05 |
|      36 |        2750 |       00:01:07 |     100.00% |       0.0131 |      5.0000e-05 |
|      36 |        2800 |       00:01:08 |     100.00% |       0.0114 |      5.0000e-05 |
|      37 |        2850 |       00:01:09 |     100.00% |       0.0079 |      5.0000e-05 |
|      38 |        2900 |       00:01:10 |     100.00% |       0.0103 |      5.0000e-05 |
|      38 |        2950 |       00:01:11 |     100.00% |       0.0100 |      5.0000e-05 |
|      39 |        3000 |       00:01:13 |     100.00% |       0.0065 |      5.0000e-05 |
|      40 |        3050 |       00:01:14 |     100.00% |       0.0114 |      5.0000e-05 |
|      40 |        3100 |       00:01:15 |     100.00% |       0.0092 |      5.0000e-05 |
|      41 |        3150 |       00:01:16 |     100.00% |       0.0120 |      5.0000e-05 |
|      42 |        3200 |       00:01:17 |     100.00% |       0.0051 |      5.0000e-05 |
|      42 |        3250 |       00:01:18 |     100.00% |       0.0081 |      5.0000e-05 |
|      43 |        3300 |       00:01:19 |     100.00% |       0.0100 |      5.0000e-05 |
|      43 |        3350 |       00:01:21 |     100.00% |       0.0104 |      5.0000e-05 |
|      44 |        3400 |       00:01:22 |     100.00% |       0.0078 |      5.0000e-05 |
|      45 |        3450 |       00:01:23 |     100.00% |       0.0054 |      5.0000e-05 |
|      45 |        3500 |       00:01:25 |     100.00% |       0.0056 |      5.0000e-05 |
|      46 |        3550 |       00:01:26 |     100.00% |       0.0060 |      5.0000e-05 |
|      47 |        3600 |       00:01:28 |     100.00% |       0.0077 |      5.0000e-05 |
```

Figure 10. Training model test results.

Save the trained network, randomly read the pictures in the test set under the trained model, the picture size is represented by [width, height], take out the loss and arccuracy of each iteration to draw the image, get the results of Figure 11, the accuracy of the model can reach more than 92%, which is about 5% higher than the accuracy of the common deep learning.

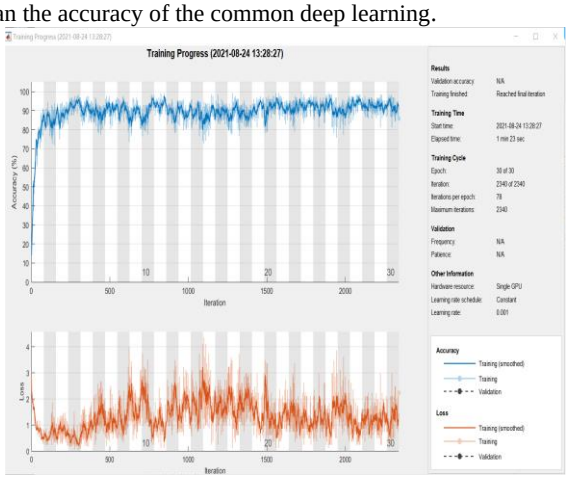


Figure 11. Model test result graph.

4.2 The Inception-ResNet-v2 model combines the experimental results of the transfer learning model

The model training and testing are run under Python's deep learning framework, and the experiment is also based on two datasets of sonar images, the two datasets are: training set train and test set test, which are also divided into 8 categories. The migration model is also mainly trained on the three parameters of the number of trainings, the number of network data submitted each time, and the learning rate. Using the pre-processed sonar image with inception-ResNet-v2 network training, and then using transfer learning to obtain the final model, the experimental results are as shown in Figure 12, the results show that the correct rate is 97.9% when iterating 1500 times, and the experimental results show that the overall recognition effect of the model is good.

```
After 0 training step(s), validation accuracy using average model is 0.0872
After 1000 training step(s), validation accuracy using average model is 0.977
After 1500 training step(s), test accuracy using average model is 0.9793
```

Figure 12. Inception-ResNet-v2 combined with transfer learning experimental results.

5. DISCUSS

In this paper, MATLAB is used as a caffe interface combined with transfer learning and Inception-ResNet-v2 model combined with transfer learning training to obtain two models to identify sonar images, and the accuracy rate is about 92% and 97%, respectively, which is higher than that of existing deep learning and classifiers. MATLAB as a caffe interface combined with the transfer learning model is relatively simple training test time is shorter, more suitable for simple picture recognition, in order to test the generalization ability of different models, the untrained sonar image is tested, the results show that the training model under the Caffe framework has a high recognition rate, and has better generalization ability for sonar data from different sources; Inception-ResNet-v2 model combined with the transfer learning model is more complex and takes longer but has high accuracy. Better solve the problem that the deep network is easy to overfit, the convergence speed is fast, and it has a good recognition effect and robustness.

REFERENCES

CAI Tao. Research on Sonar Image Target Active Recognition Technology[D].Southeast University, 2019.

CAI Weiming,PANG Haitong,ZHANG Yitao,ZHAO Jian,YE Zhangying. Identification Model of Farmed Fish Species Based on Convolutional Neural Network[J/OL].Journal of Fisheries Science:1-8[2021-08-28].

CAO Yudong,LIU Haiyan,JIA Xu,LI Xiaohui. A Review of Image Quality Evaluation Methods Based on Deep Learning[J/OL].Computer Engineering and Applications:1-11[2021-08-28].

CHEN Hao. Research on object recognition method of bathymetric side sweep sonar image[D].Harbin Engineering University, 2020.

Du Yang, Yang Rui, Chen Zhiyuan, Wang Lei, Weng Xiaodong, Liu Xiuheng. A deep learning network-assisted Bladder Tumor Recognition under Cystoscopy Based on Caffe Deep Learning Framework and EasyDL Platform.[J]. The international journal of medical robotics + computer assisted surgery : MRCAS, 2020.

Du Yang, Yang Rui, Chen Zhiyuan, Wang Lei, Weng Xiaodong, Liu Xiuheng. A deep learning network - assisted bladder tumour recognition under cystoscopy based on Caffe deep learning framework and EasyDL platform[J]. The International Journal of Medical Robotics and Computer Assisted Surgery, 2020, 17(1).

E O Kovalenko, A Sushchenko. Cloud Service for Sonar Signal Processing[J]. IOP Conference Series: Earth and Environmental Science, 2019, 272(2).

FU Suining,LU Zezhong,WANG Shunyao. An Improved Lee Filtering SAR Image Denoising Algorithm[J].Computer and Digital Engineering,2019,47(08):2018-2021.

GAO Qiang. Research on sonar image feature extraction method based on wavelet moment[D].Dalian University of Technology, 2020.

Gui Shui Yu，Ke Li. Watershed Image Segmentation Based on PSO and FCM[J]. Advanced Materials Research，2015，3701:
HAN Pengju. Research on sonar image registration based on convolutional neural network[D].Hangzhou Dianzi University, 2020.

JIANG Dayang. Design of Commodity Image Recognition System Based on Convolutional Neural Network[J].Journal of Beijing Institute of Industry and Technical College,2021,20(03):28-31.

Lin Chaojun，Shi Ying，Zhang Jian，Xie Changjun，Chen Wei，Chen Yue. An anchor-free detector and R-CNN integrated neural network architecture for environmental perception of urban roads[J]. Proceedings of the Institution of Mechanical Engineers，Part D: Journal of Automobile Engineering，2021，235(12).

Lin Chaojun，Shi Ying，Zhang Jian，Xie Changjun，Chen Wei，Chen Yue. An anchor-free detector and R-CNN integrated neural network architecture for environmental perception of urban roads[J]. Proceedings of the Institution of Mechanical Engineers，2021，235(12).

LU Jianhua. SAR Image Target Recognition Method For Fusion CNN and SRC Decision Making[J/OL].Infrared and Laser Engineering:1-8[2021-08-28].

Nalini K Ratha，Vishal M. Patel，Rama Chellappa. Deep Learning-Based Face Analytics[M].:2021-08-23.

QI Qianhui. Sonar image segmentation based on Markov airfield[D].Beijing Institute of Printing, 2021.

Qin Shanshan. Three-dimensional reconstruction of side-scan sonar image based on SFS method[D].Xi'an University of Technology, 2021.

QIU Zhibin,LIU Zhou,LIAO Caibo,YU Xiaobin. Water-repellent identification of composite insulators based on deep transfer learning[J/OL].High Voltage Technology:1-10[2021-08-28].

Seo Seunghwan，Lee JeJun，Lee RyongGyu，Kim Tae Hyung，Park Sangyong，Jung Sooyoung，Lee HyunKyu，Andreev Maksim，Lee KyeongBae，Jung KilSu，Oh Seyong，Lee HoJun，Kim Ki Seok，Yeom Geun Young，Kim YongHoon，Park JinHong. An Optogenetics-Inspired Flexible van der Waals Optoelectronic Synapse and its Application to a Convolutional Neural Network.[J]. Advanced materials (Deerfield Beach，Fla.)，2021.

Shi Hong，Shan Fang-jian，Cong Bo，Qiu Wei. An underwater ship fault detection method based on Sonar image processing[J]. Journal of Physics: Conference Series，2016，679(1).

WANG Qilin,WANG Hongjian,LI Qing,XIAO Yao,BAN Xicheng. Improved method of image feature extraction of side-scan sonar[J].Journal of Underwater Unmanned Systems,2019,27(03):297-304.

WANG Tao,PAN Guofu,ZHANG Jibo. Automatic Extraction Method of Target Contour of Side-Scan Sonar Image Based on K-means Clustering and Mathematical Morphology[J].Marine Sciences,2019,43(08):80-85.