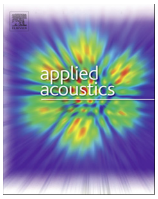
Contents lists available at ScienceDirect

Applied Acoustics

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



Underwater sonar image classification using adaptive weights convolutional neural network

Xingmei Wang [a], Jia Jiao [a], Jingwei Yin [b,c,d], Wensheng Zhao [e], Xiao Han [b,c,d,*], Boxuan Sun [a]

[a] College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
[b] Acoustic Science and Technology Laboratory, Harbin Engineering University, Harbin 150001, China
[c] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
[d] Key Laboratory of Marine Information Acquisition and Security, Harbin Engineering University, Harbin 150001, China
[e] College of Power and Energy Engineering, Harbin Engineering University, Harbin 150001, China

ARTICLE INFO

Article history:
Received 12 December 2017
Received in revised form 18 May 2018
Accepted 3 November 2018

Keywords:
Sonar image classification
Deep learning
Convolutional neural network
Deep belief network
Classification

ABSTRACT

As an important part of oceanographic surveying, underwater sonar image classification has attracted much attention. Most of the existing classification methods cannot be widely used in underwater sonar image classification. However, deep learning models can automatically extract underwater sonar image features to improve the classification accuracy through an internal network structure. In the present study, a novel deep learning model with adaptive weights convolutional neural network (AW-CNN) was proposed to classify underwater sonar images. To solve the random initialization of filter weights in a convolutional neural network (CNN), the generated weights of the deep belief network (DBN) were applied to adaptively replace the randomly trained filter weights of the CNN in the implementation process of the AW-CNN for underwater sonar image classification. Specifically, first, dimension conversion was accomplished by using the increment-dimension function to unify the inputs of the CNN and the DBN. Then, according to the dimension conversion, the internal fusion of the two models is realized, and replacement of the randomly trained filter weights of the CNN was completed. Finally, in order to further improve the classification accuracy, the local response normalization (LRN) function is proposed to normalize the adaptive weights in the network initialization. Compared with other models, the classification results demonstrate that the proposed AW-CNN approach has the capability to effectively and successfully divide sonar images into their relevant seabed classes, which is beneficial to finding mines and shoals and detecting the integrated degree of the dam bottom.

1. Introduction

Underwater object classification technology is an organic combination of sonar technology, signal detection theory, artificial intelligence, pattern recognition and computer technology [1]. With the development of sonar imaging technology, sonar images have been used extensively for underwater object classification, which has become a significant subject in the field of ocean development. A sonar image contains three kinds of regions: object-highlight regions, shadow regions and sea-bottom reverberation regions. An object-highlight region originates from the acoustic wave reflection from an object. The shadow region comes from a lack of acoustic backscatter behind the object. The remaining information consists of the so-called sea-bottom reverberation region [2]. The sonar image has the characteristics of low contrast, edge blur and high noise, which will seriously influence the underwater object classification. However, underwater object classification can help to find mines, submarines, and shoals and to detect the integrated degree of the dam bottom. Therefore, it has important practical significance both in military and civil fields.

Currently, many methods have been proposed to classify underwater objects in sonar images. Most methods have focused on machine learning. In this case, the main strategy has been to guide feature learning using feature extraction and classifier selection. Many methods that use features have been researched in sonar images, such as the Markov random field [3]and the Principle component analysis [4]. A method of seabed classification using the Self-organizing feature maps (SOFM) neural network was presented, and it is effective to some extent [5]. Martin et al. [6] introduced a multilayer perceptron to classify sonar images and solve

* Corresponding author at: Key Laboratory of Marine Information Acquisition and Security, Harbin Engineering University, Harbin 150001, China

E-mail address: hanxiao1322@hrbeu.edu.cn (X. Han).

https://doi.org/10.1016/j.apacoust.2018.11.003
0003-682X/© 2018 Elsevier Ltd. All rights reserved.

the uncertainty in the training stage. Khidkikar et al. [7]developed an approach using texture segmentation and classification for the real-time processing of side-scan sonar images. The sonar images are divided into four classes, including rocks, wreckage, sediment and the sea floor. Classifier selection is another crucial problem for sonar image classification. In 2009, a classifier for underwater sonar image fusion was proposed to complete underwater object classification [8]. Later, Lopera et al. [9]introduced an object classifier based on high resolution sonar images, and the classification performance is estimated by the ROC curve. Subsequently, the quality of sonar images was improved based on data production precision and processing efficiency to solve the classification problem, which combined the geometric correction method with image processing technology [10]. Meanwhile, the multiple-aspect problem was transformed into a multiple-instance learning problem, and mine-like object classification was completed according to the multi-instance classifiers [11].

In fact, as a typical classifier, the support vector machine (SVM) also has been applied to sonar images and gets better results. Among them, Li et al. [12] used the main characteristics of divers as the input feature vectors to the SVM, including average-scale, velocity, shape, direction and angle. It can obtain good classification results. In 2015, Karine et al. [13] proposed the wavelet coefficients to extract texture features, and then the sonar images were classified by the K-nearest neighbor algorithm and the SVM. Later, to improve the classification performance, Sion et al. [14] used the kernel function to extract features and the SVM to classify sonar images. Subsequently, a new classification method based on the kernel limit learning machine and the Principle component analysis (PCA) principle was proposed for sonar image classification [15]. This method has good stability and high classification accuracy. Although the above classification methods can effectively complete classification to some extent, they all adopt different feature extraction methods according to different sonar images. A specific feature extraction method cannot consider all the useful information of sonar images, causing these classification methods to be unable to be widely used. In addition, the sea environment is complex, and every underwater sonar imaging is different at the sea bottom. Even if one-class images also have various angles [16], further increasing the difficulty of sonar image classification.

In recent years, deep learning models have gradually attracted attention both domestically and abroad. It has become an important subject in the field of big data and artificial intelligence [17,18]. In 2006, Hiton et al. [19] proposed the unsupervised greedy algorithm based on the DBN, which solved the optimization problems related to deep structures. To further verify the advantages of the deep learning model, the deep structure of the multilayer auto-encoder was proposed [20]. The DBN and the CNN are applied not only to text and language but also to images, sounds and videos to convey human semantic information. The earliest image application was in the handwritten digit dataset [21]. The research of deep learning model was constantly studied and improved by scholars, and some achievements have made in the image classification. In 2010, Ciresan et al. [22] applied the simple CNN in the classification of handwritten digits and obtained high classification accuracy. Yalcin [23] applied the DBN to the classification of human activity by using RGB-D video sequences, which were divided into many images, and it obtained better results than other methods. Meanwhile, an HSI reconstruction model using the deep CNN was proposed, and the average accuracy was be increased by 30.04% [24]. An alternating deep belief network (ADBN) was also presented. It can obtain higher accuracy than the DBN and the SVM [25]. In 2016, Iftene et al. [26] used fine-tuning to preprocess the CNN and completed high-resolution images classification. Arsa et al. [27] developed a dimensionality reduction method by using the DBN for hyperspectral image classification. It is better than the PCA in hyperspectral image classification. Christian et al. [28] introduced an INCEPTION-V4 CNN in which the LRN normalization function was used to classify the ImageNet dataset, and it achieved a 3.08 percent top-5 error on the test set of the ImageNet classification (CLS) challenge. Later, Miki et al. [29] combined the ROI region with the CNN to augment CT images of teeth, which improved the results by 5% compared with the result without data augmentation. Li et al. [30] proposed a deep learning method for oil palm tree classification using high-resolution remote sensing images in Malaysia. And more than 96% of the oil palm trees can be classified correctly. In addition, Phillip et al. [31] presented transfer learning with the CNN for the classification of abdominal ultrasound images. It can construct effective classifiers for abdominal ultrasound images. Therefore, the DBN and the CNN of deep learning models have achieved good results in image classification through these literatures. However, using the deep learning models in underwater sonar image classification is still in the primary stage. The NATO STO Centre for Maritime Research and Experimentation first applied the CNN to underwater sonar image classification in 2016 [32,33], and the underwater sonar image dataset was collected by the laboratory.

To solve the problem of the random initialization of filter weights in the CNN, which results in the model being unable to precisely classify underwater objects in sonar images, an AW-CNN based on deep learning was proposed in this paper. This method uses the increment-dimension function to accomplish the dimension conversion between CNN tensor and DBN vector. The internal fusion of the two models is realized according to dimension conversion, and the replacement of the randomly trained filter weights of the CNN was completed. To limit the maximum and minimum values of the input vectors, the AW-CNN was normalized by an LRN function. The proposed AW-CNN can achieve relatively high classification accuracy.

The proposed AW-CNN was applied on the original sonar image dataset, and the experimental results demonstrate that the AW-CNN has good effectiveness and convergence. The proposed AW-CNN has important theoretical and practical value.

2. The construction of the dataset

Underwater sonar images belong to a small sample and an unpublicized dataset. The dataset of this paper is from the laboratory collection acquired over the past years. This dataset was divided into six categories, including sand ripples, sunken ships, crashed planes, stones, tires and shoals. All sonar images are 100 × 100 pixels. Table 1 shows a detailed description of the dataset.

Considering the possible situation that the size of original dataset may cause over-fitting in deep learning models, the dataset was expanded using various transformations such as the HSV and HV channel transformation, the RGB channel transformation, sonar image rotation, sonar image with Gaussian noise and sonar image

Table 1
The detailed description of the dataset.

Categories	Sand ripple	Sunken ship	Crashed plane	Stone	Tire	Shoal
Numbers	432	1104	348	348	80	96

enhancement according to the characteristics of underwater sonar images and the complex seabed environment. After the construction of dataset, Figs. 1–3, show the various transformations of the underwater sonar image of a sand ripple, a sunken ship and a crashed plane, respectively.

The construction of dataset adequately considers the characteristics of underwater sonar images closer to the real situation of the sea bottom. It is beneficial to the subsequent classification research.

3. The proposed AW-CNN for underwater sonar image classification

The proposed AW-CNN can use the advantage of the DBN to quickly obtain a better feature matrix to adaptively adjust the distribution of filter weights in the CNN. The generated weights of the DBN were applied to replace the randomly trained filter weights of the CNN. The proposed AW-CNN can solve the problem of the random initialization of filter weights in the CNN, avoid falling into the local optimum, and improve the classification accuracy.

3.1. CNN

The CNN simulates the perceptual process of visual nerves in the cerebral cortex, and only a small number of neurons are active when it identifies an image. On the basis of this local receptive field, the CNN can use the convolution operation to effectively extract the features of images. Meanwhile, due to the characteristics of the feedforward neural network and local sensing, the CNN has good applicability in image classification, image recognition and even speech recognition. Therefore, it is considered as the most popular deep learning model for underwater sonar image classification.

The structure of the CNN includes the input layer, the convolutional layer, the pooling layer, the fully connected layer and the output layer. The convolutional layer is used to train the local receptive fields of the input underwater sonar images, which can further extract the abstract features of underwater sonar images. The underwater sonar image feature matrix is obtained by the convolution operation between the filter and underwater sonar image. The specific convolutional process is as follows:

$$z_j^m = f\left(\sum_{i\in T_j} k_{i,j}^m * z_i^{m-1} + b_j^m\right) \tag{1}$$

where$f(\cdot)$is the activation function, which is the rectified linear units (ReLU) in this paper, m is the number of layers, $k_{i,j}$is the filter, $b_j$is the bias, * is the convolution operation, and T_j represents a set that is used to store the input feature matrices.

The most important role of the pooling layer is to reduce the dimensions of underwater sonar images except to extract features. After reducing the dimension, the pooling layer can accelerate the running speed, solve the phenomenon of over-fitting and improve the stability and robustness of the network structure. In the pooling layer, max-pooling divides the receptive fields into a set of non-

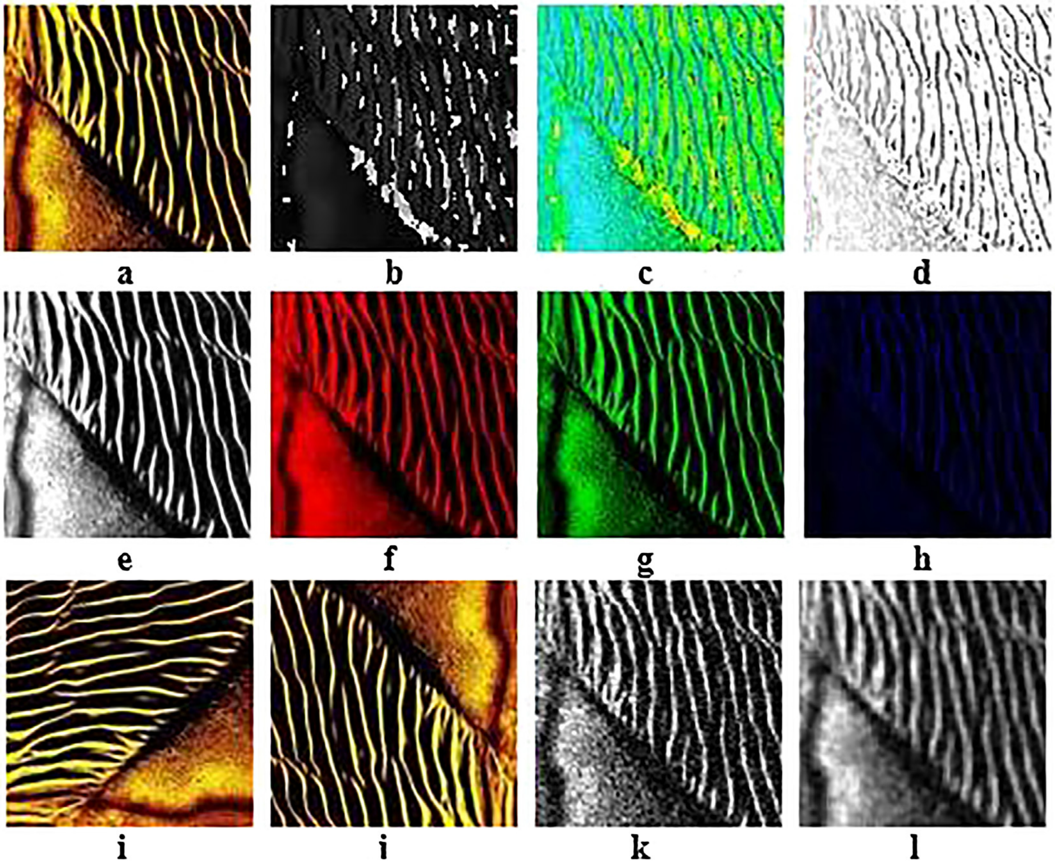


Fig. 1. Various transformations of the underwater sonar image of a sand ripple. (a) Original sonar image, (b) H channel transformation, (c) S channel transformation, (d) V channel transformation, (e) HV channel transformation, (f) R channel transformation, (g) G channel transformation, (h) B channel transformation, (i) rotating 90°, (j) rotating 180°, (k) sonar image with Gaussian noise, and (l) sonar image enhancement.

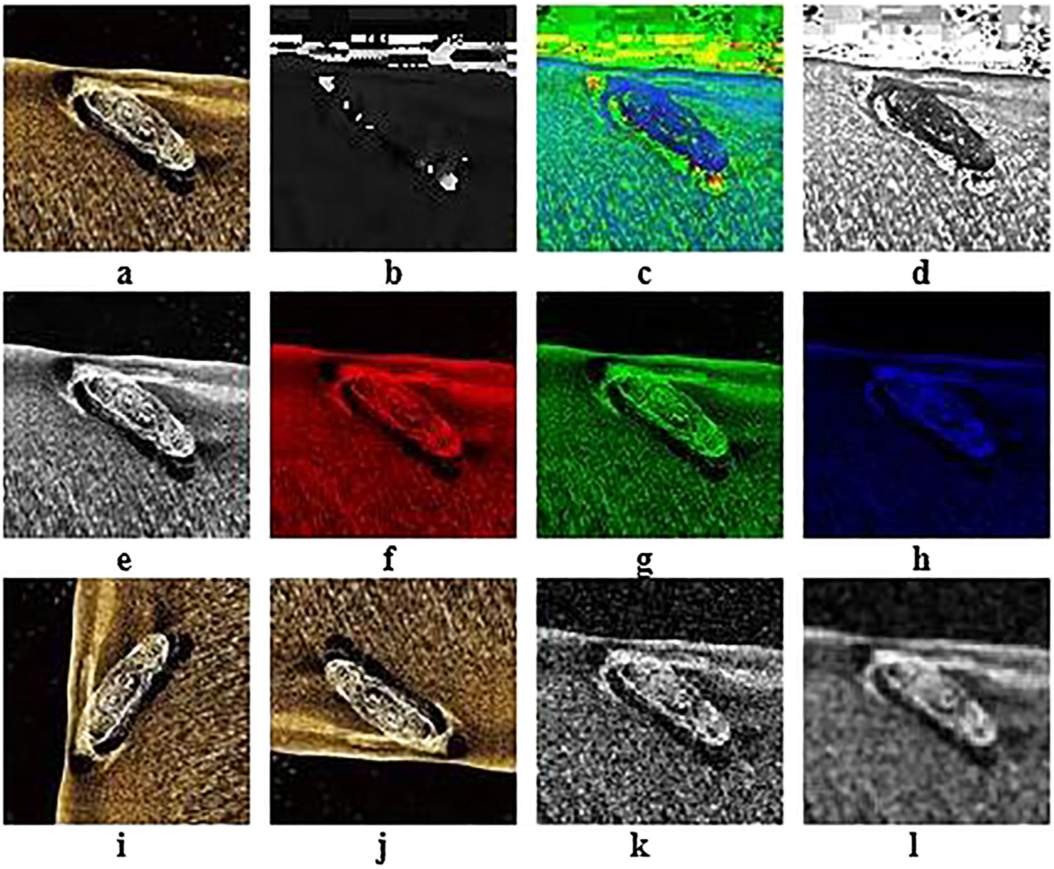


Fig. 2. Various transformations of an underwater sonar image of a sunken ship. (a) Original sonar image, (b) H channel transformation, (c) S channel transformation, (d) V channel transformation, (e) HV channel transformation, (f) R channel transformation, (g) G channel transformation, (h) B channel transformation, (i) rotating 90°, (j) rotating 180°, (k) sonar image with Gaussian noise, and (l) sonar image enhancement.

overlapping matrices and outputs the maximum value. The max-pooling process is as follows:

$$q_j^m = f\left(down\left(q_i^{m-1}\right)\cdot\omega_j^m + b_j^m\right) \tag{2}$$

where ω are the weights, and $down(\cdot)$ is the pooling function.

The fully connected layer is a more abstract layer. It integrates the global information of the whole image. This layer connects each neuron in the pooling layer to each neuron in the output layer. Therefore, the number of parameters is the most in the fully connected layer. The fully connected layer and softmax function are generally composed of the classification structure, which is used as the output layer.

For the vectors of the fully connected layer, the output layer can output classes with a certain probability. The underwater sonar images were divided into six classes by the softmax function in the output layer, and the specific softmax function is as follows:

$$f(x_g) = \frac{e^{x_g}}{\sum_{n=1}^{N} e^{x_n}} \tag{3}$$

where N is the number of classes in the sonar image dataset. If the value of x_g is the largest in the images, the component of the map approximates 1, and the other approximates 0.

In Eq. (3), the softmax function is expressed as the loss function. It is given as follows:

$$J(\omega) = -\frac{1}{c}\left[\sum_{s=1}^{c}\sum_{n=1}^{N} 1\{y^{(s)} = N\}\log\frac{e^{\omega_n^T x^{(s)}}}{\sum_{n=1}^{N} e^{\omega_n^T x^{(n)}}}\right] \tag{4}$$

where c is the mini-batch size $1\{\cdot\}$ represents the marking function $(1\{a\ true\ statement\} = 1,\ and\ 1\{a\ false\ statement\} = 0)$. $\left\{\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(c)}, y^{(c)}\right)\right\}$ is a training set of c labeled samples.

According to the characteristics of underwater sonar images, the CNN architecture was obtained by constantly adjusting the parameters in this paper. The CNN architecture consists of one input layer, four convolutional layers, two fully-connected layers and one output layer. Each convolutional layer is followed by a pooling layer, a ReLU activation function and an LRN function. The optimization proceeds using the Adam optimizer with a learning rate of 0.0004, a β_1 of 0.9, a β_2 of 0.999 and a batch size of 64 images with shuffling. All training images are 100 × 100 pixels. The specific CNN architecture and its parameters are shown in Fig. 4.

In practical applications, the convolution operation of the CNN mainly depends on the filter weights, and the result of the convolution operation is the precondition of CNN classification accuracy. Hence, the filter weights are the key of the CNN. However, the initialization of filter weights is random, which makes it easy to fall into the local optimum and influence the final classification accuracy. Additionally, the DBN applies the softmax classifier in the last layer, which can receive the output feature vectors of the RBM as its input feature vectors and uses the BP algorithm fine-tune the RBM parameters of each layer to complete the classification. The process of fine-tuning equates to the initialization of the filter weights in the whole deep network. Therefore, the advantage of the DBN that can quickly obtain a better feature matrix was used to adaptively adjust the distribution of the filter weights in the

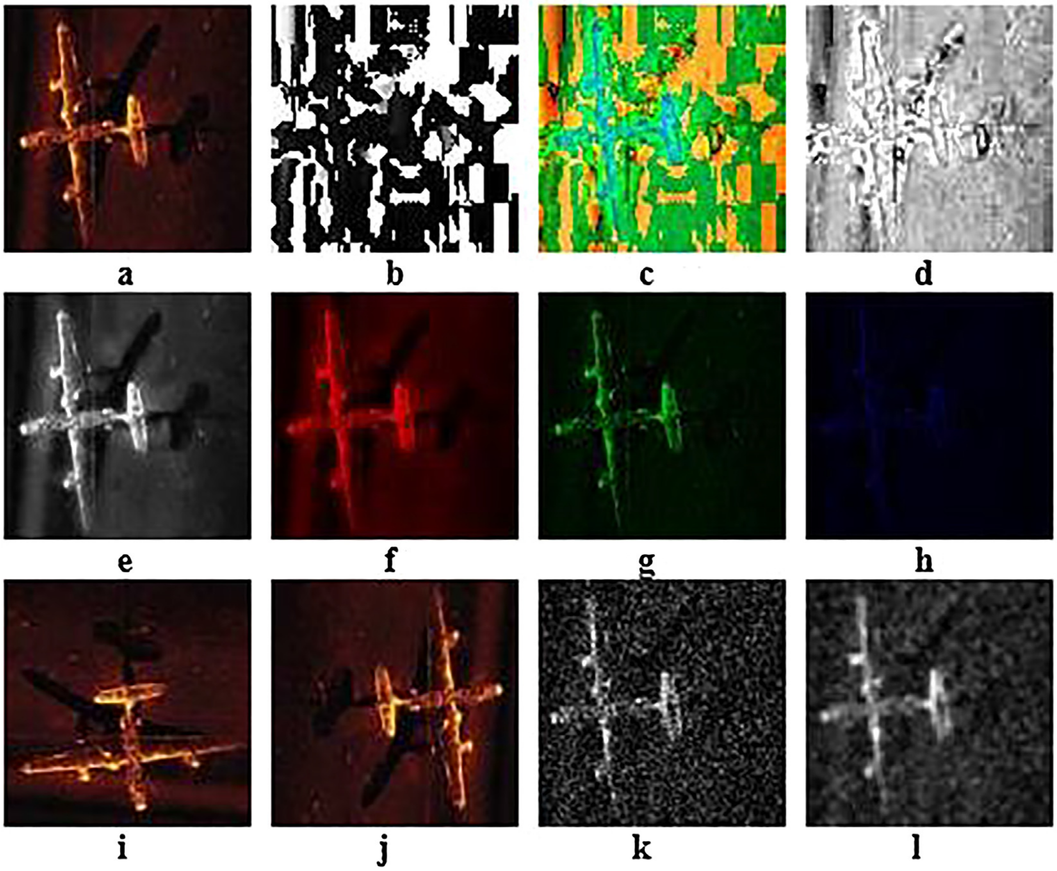


Fig. 3. Various transformations of an underwater sonar image of a crashed plane. (a) Original sonar image, (b) H channel transformation, (c) S channel transformation, (d) V channel transformation, (e) HV channel transformation, (f) R channel transformation, (g) G channel transformation, (h) B channel transformation, (i) rotating 90°, (j) rotating 180°, (k) sonar image with Gaussian noise, and (l) sonar image enhancement.

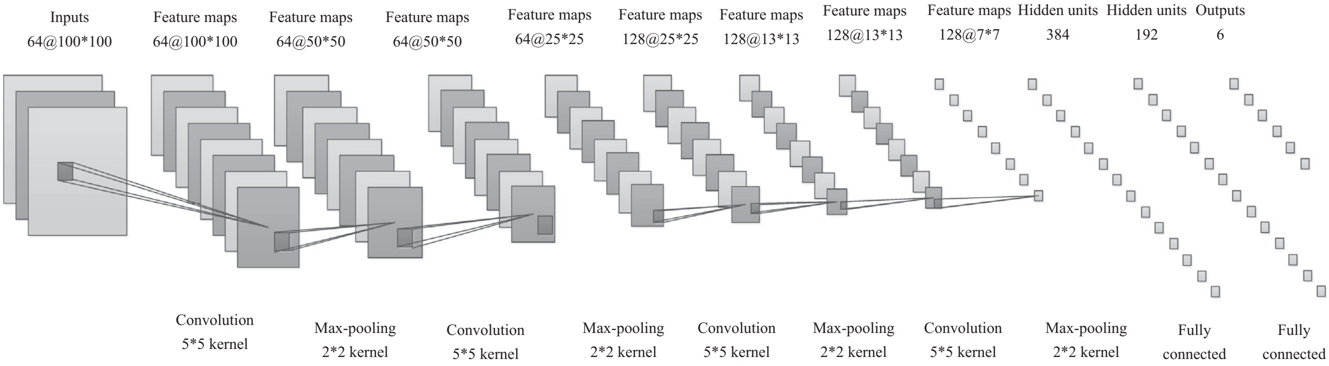


Fig. 4. CNN architecture diagram.

CNN. The generated weights in the DBN were applied to replace the randomly trained filter weights of the CNN.

3.2. Initialization of adaptive weights

The advantage of the generated weights in the DBN depends on its architecture and the whole classification process. There is a two-layer structure in the DBN in which one layer is the visible layer, and the other is the hidden layer. Each RBM consists of m visible units $v = (v_1, v_2 \ldots v_m)$ and n hidden units $h = (h_1, h_2 \ldots h_n)$, and each unit carries a random binary value $(v, h) \in \{0, 1\}^{m+n}$. The energy function of the RBM model is given as follows:

$$E(v,h) = -\sum_{i=1}^{m} b_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n} \omega_{ij} v_i h_j \tag{5}$$

where ω_{ij} represents the weights associated between v_i and h_j, and b_i and b_j are the biases of the visible units v_i and the hidden units h_j, respectively.

Through the energy function, the joint distribution probability function of (v, h) can be obtained as follows:

$$p(v,h) = \frac{e^{-E(v,h)}}{Z} \tag{6}$$

where Z is the normalization constant, and $Z = \sum_{v,h} e^{-E(v,h)}$.

The marginal distribution of the visible vector$v$ is assigned by the RBM, and it is expressed as follows:

$$p(v) = \frac{1}{Z}\sum_{h} e^{-E(v,h)} \tag{7}$$

The conditional probability of the visible units v and the hidden units h are respectively expressed as follows:

$$p(v_i = 1|h) = sig\left(b_i + \sum_{j=1}^{n} \omega_{ij} h_j\right) \tag{8}$$

$$p(h_j = 1|v) = sig\left(b_j + \sum_{i=1}^{m} \omega_{ij} v_i\right) \tag{9}$$

where $sig(x) = \frac{1}{1+e^{-x}}$ is the logistic sigmoid function.

For the biases $b_i$and b_j and the weight ω_{ij} in the RBM, the derivative of the logarithm likelihood probability are respectively expressed as follows:

$$\begin{aligned}\frac{\partial \ln p(v)}{\partial w_{ij}} &= p(h_i = 1|v)v_i - \sum_{v} p(v)p(h_j = 1|v)v_i \\ &= \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{mod\ el}\end{aligned} \tag{10}$$

$$\frac{\partial \ln p(v)}{\partial b_i} = v_i - \sum_{v} p(v)v_i = \langle v_i \rangle_{data} - \langle v_i \rangle_{mod\ el} \tag{11}$$

$$\begin{aligned}\frac{\partial \ln p(v)}{\partial b_j} &= p(h_j = 1|v) - \sum_{v} p(v)p(h_j = 1|v) = \langle h_j \rangle_{data} \\ &- \langle h_j \rangle_{mod\ el}\end{aligned} \tag{12}$$

where $\langle\cdot\rangle_{data}$ and $\langle\cdot\rangle_{mod\ el}$ are respectively the data distribution and the model expectation.

The specific classification process is shown in Fig. 5.

As seen from Fig. 5, in the DBN, each RBM can adjust the feature extraction matrix, and it is trained in an unsupervised way. Meanwhile, its features are mapped into different feature spaces and retain the feature information as much as possible. The architecture of the DBN has five layers in this paper: one input layer with 30,000 visible units, three hidden layers with 2000 units, 400 units and 80 units, respectively, and a 6-units softmax layer. Each layer uses a Sigmoid activation function. The optimization proceeds using the Gradient Descent optimizer with a learning rate of 0.5, and the same batch size of 64 images with shuffling. The DBN applies the softmax classifier in the last layer, which can receive the output feature vectors of the RBM as its input feature vectors, and uses the BP algorithm to fine-tune the RBM parameters of each layer to complete the classification. The process of fine-tuning equates to the initialization of the filter weights in the whole deep network.

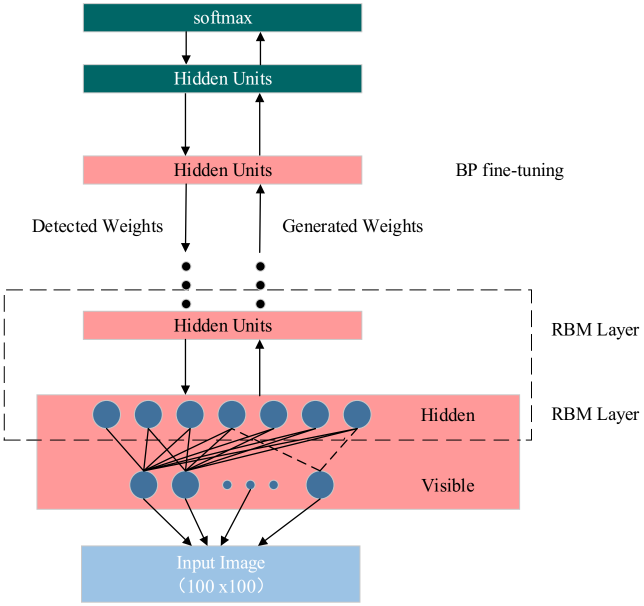


Fig. 5. The classification process of the DBN.

Because the generated weights of the DBN can compensate for the shortage of randomly trained filter weights in the CNN, the AW-CNN was proposed in this paper. However, the data type of the CNN is a tensor, and the data type of the DBN is a vector. Therefore, the AW-CNN needs to unify the dimension of the input data to realize the fusion of the model. In addition, the data was normalized by the LRN function after model fusion. The proposed AW-CNN can complete the replacement of randomly trained filter weights in the CNN, avoid falling into the local optimum, and improve the classification accuracy. The specific implementation process of the AW-CNN is as follows:

(1) *Dimension conversion*. To unify the input of two different deep learning models, the dimension conversion between the CNN and the DBN was accomplished using the increment-dimensional function. In the CNN, the most significant role of the pooling layer is to reduce the dimension except to extract the features. However, the feature extraction matrix is smaller than the original image. Meanwhile, the data type is a tensor, and its filter is a four-dimensional matrix. While the DBN does not change the size of the image when each RBM extracts the features, the data type is a vector, and the feature matrix is a two-dimensional matrix. Therefore, the unification of the dimensions is necessary in the proposed AW-CNN. The feature matrix was the output of the first convolutional layer, which is input into the DBN model. The size of the image was randomly cut as the size of the filter. The number of hidden units in the last layer of the DBN was the product of the length and the width in the filter, and the number of output units was 3. After the training of the DBN, the two-dimensional parameter matrix Wbetween the last two layers was obtained. The parameter matrix $W$was changed into a three-dimensional matrix by the increment-dimensional function. The former two dimensions represent the size of the filter, and the last dimension represents the number of color channels for the input image. The increment-dimensional function was used to increase the dimensions of Wuntil it equaled the input dimension of the CNN.

(2) *Model fusion.* On the basis of realizing the input of two different deep learning models, the internal fusion of the two models was completed according to the dimension conversion in the proposed AW-CNN. The generated weights of the DBN were applied to replace the randomly trained filter weights of the CNN. The data type of the weights was changed into a floating-point type.

(3) *Normalization processing*. After the model fusion, the weights were not normalized. This can influence the fitting situation, which led to no convergence and low classification accuracy. The value range of the weights is the main factor through the experiments in this paper. To limit the maximum and minimum value of the weights, an LRN function is proposed to normalize the adaptive weights in the network initialization to improve the classification accuracy. Combining the characteristics of underwater sonar images with the trained filter weights of the CNN, the normalized function is as follows:

$$t^i_{x,y} = a^i_{x,y} \Bigg/ \left(10 + \sum_{j=\max(0,i-2)}^{\min(M-1,i+2)} \left(a^j_{x,y}\right)^2\right) \tag{13}$$

where i represents the serial number of filters, which is the serial number of feature maps, x and y represent the positions of the filters, and M is the total number of filters.

The normalized weights preserved the advantages of the random normal truncated matrix. Meanwhile, it also preserved the trained matrix and the relations among the matrices in the DBN.

Therefore, the trained filter weights in the AW-CNN can make full use of the features that the filters extracted, which can solve the problem of the random initialization of filter weights in the CNN, avoid falling into the local optimum, and improve the classification accuracy.

4. Experimental results and analysis

This section shows the numerical experimental results to validate the effectiveness of the proposed AW-CNN for underwater object classification. The train set is three times as large as the test set in the experiments of this paper. To increase the reliability and stability of the experiments, Fig. 6 shows classification results of 30 repeated experiments after 600 iterations, which includes the AW-CNN, the CNN [28] and the DBN [19].

As seen from Fig. 6, the classification accuracy of the AW-CNN is higher than that of the CNN and the DBN in 30 repeated experiments after 600 iterations. The CNN can obtain the feature matrix using the convolution operation between the filter and the underwater sonar image, which is more suitable for underwater sonar image classification than the DBN. However, the initialization of the filter weights is random in the CNN, which makes it easy to fall into the local optimum and influence the final classification accuracy. Additionally, the generated weights of the DBN can compensate for the shortage of randomly trained filter weights in the CNN. Therefore, the trained filter weights in the AW-CNN can make full use of the features that the filters extracted, which can solve the problem of the random initialization of filter weights in the CNN, avoid falling into the local optimum, and improve the classification accuracy.

Meanwhile, in order to further verify the effectiveness of the classification, the best experiment of classification accuracy in 30 repeated experiments for each model has been chosen. Tables 2 and 3 show the classification accuracy under the condition of the maximum number of iterations and different fixed number of iterations respectively. Table 2 shows the classification accuracy of the AW-CNN, the CNN and the DBN when the maximum number of iterations is 1500. Table 3 shows the classification accuracy of the AW- the CNN and the CNN when the fixed number of iterations are 600, 700, 800 or 1000, respectively.

Table 2
The classification accuracy under the maximum iterative times.

Models	AW-CNN	CNN	DBN
Classification accuracy (%)	85.50	81.10	60.20

Table 3
The classification accuracy under different fixed iterative times.

Iterative times	AW-CNN(%)	CNN(%)
600	85.50	80.20
700	85.50	80.20
800	85.50	80.50
1000	85.50	80.60

As depicted in Table 2, the classification accuracies of the AW-CNN, the CNN and the DBN are 85.5%, 81.1% and 56%, respectively. Therefore, the classification accuracy of the proposed AW-CNN is higher than that of the CNN and the DBN, while the classification accuracy of the DBN is far lower than that of the AW-CNN and the CNN. Therefore, the DBN is not suitable for sonar image classification.

From Table 3, the classification accuracy of the proposed AW-CNN is higher than that of the CNN under the fixed number of iterations, and it has a certain stability. Through the analysis of Tables 2 and 3, it can demonstrate that the proposed AW-CNN is obviously better than the CNN and the DBN in classification efficiency.

On the basis of Table 3, to demonstrate the convergence of the proposed AW-CNN in this paper, Fig. 7 shows the changes in the loss accuracy of the AW-CNN and the CNN under 600 iterations using Eq. (4).

It can be seen from Fig. 7 that the loss values of the AW-CNN and the CNN gradually converge after 300 iterations, and the convergence rate is relatively close. Therefore, the proposed AW-CNN has higher classification accuracy under a relatively close convergence rate from Figs. 6 and 7.

Similarly, in order to verify better feature extraction of the proposed AW-CNN, the visualization of the first convolutional layer in the AW-CNN and the CNN are respectively given in Figs. 8 and 9.

For comparison, the images of the texture structure in Fig. 8 are clearer than those of in Fig. 9. The trained filter weights in the AW-

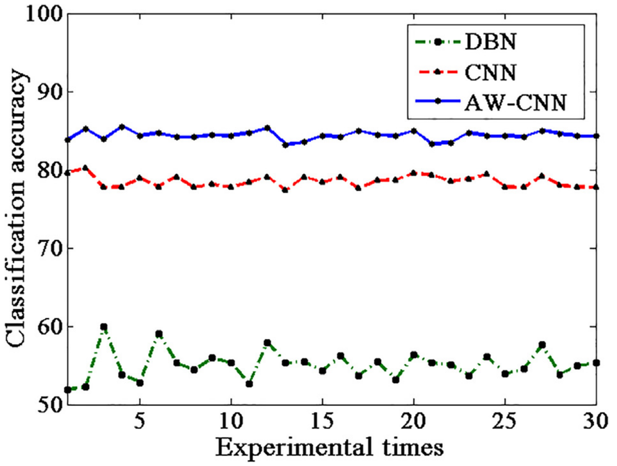


Fig. 6. The contrast diagram of the classification accuracy.

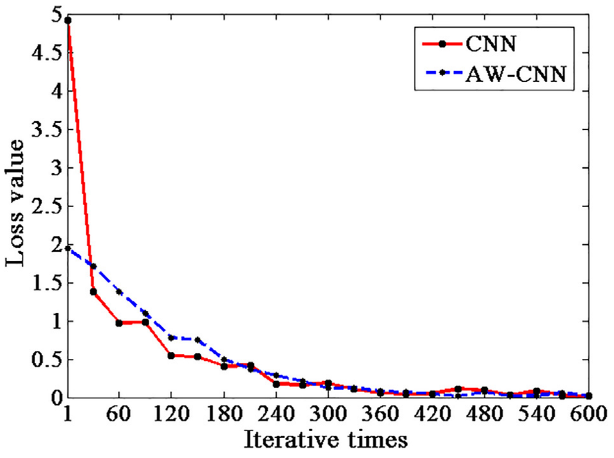


Fig. 7. The changes in the loss accuracy.

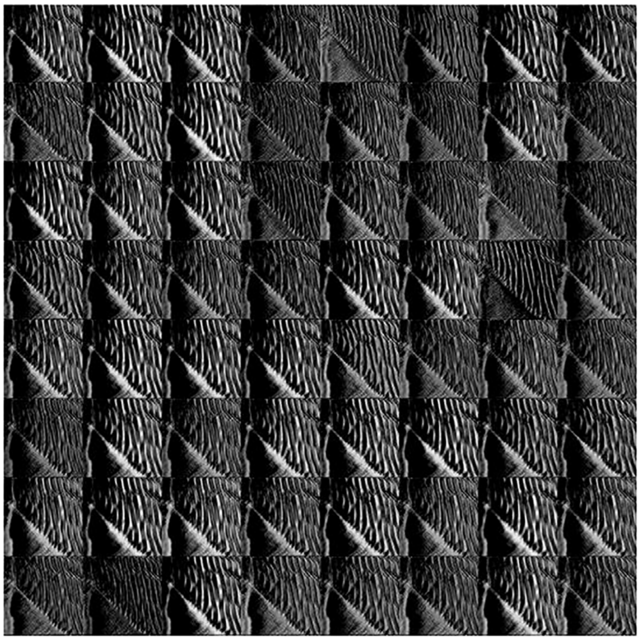


Fig. 8. The visualization of the first convolutional layer in the AW-CNN.

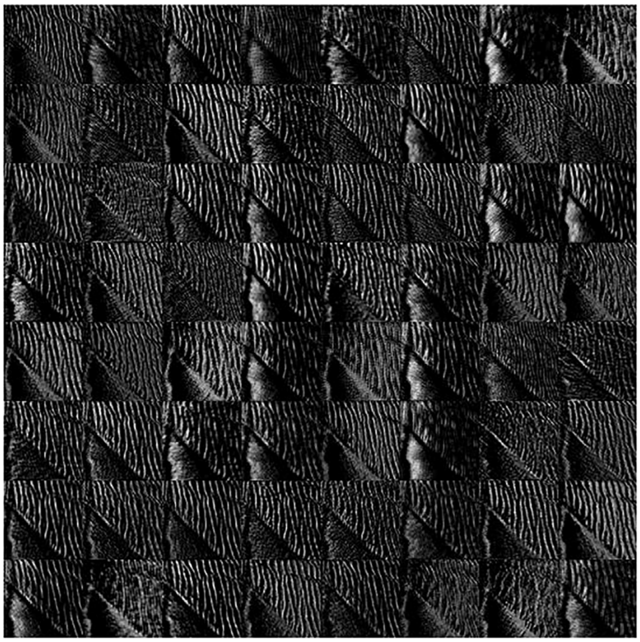


Fig. 9. The visualization of the first convolutional layer in the CNN.

CNN can make full use of the features that the filters extracted, which can solve the problem of the random initialization of filter weights in the CNN. It is demonstrated that the proposed AW-CNN can better extract features to improve the classification accuracy. Through the above experimental results, the proposed AW-CNN has better accuracy and effectiveness.

To further improve the classification accuracy, combining the AW-CNN with the characteristics of underwater sonar images, the underwater sonar images were preprocessed by a narrowband Chan-Vese model with an adaptive ladder initialization approach for detection [2] and a gray level co-occurrence matrix for feature extraction [34]. The preprocessed underwater sonar images replaced the original images, which formed the novel dataset. Figs. 10–12, show the image preprocessing of the underwater sonar image of a sand ripple, a sunken ship and a crashed plane, respectively.

Fig. 13 shows the classification results of the proposed AW-CNN with the original dataset and the preprocessed dataset in 30 repeated experiments after 600 iterations.

From Fig. 13, the classification accuracy of the AW-CNN with the preprocessed dataset is relatively higher than that of the AW-CNN. Because the sonar image has the characteristics of low contrast, edge blur and high noise, which will seriously influence the underwater object classification. The preprocessed dataset can make the object contour more prominent and more beneficial to classification.

Meanwhile, in order to compare the various deep learning models with the SVM of a typical non-deep learning model [12], the best experiment of classification accuracy in 30 repeated experiments for each model has been chosen. Table 4 shows the classification accuracy after 600 iterations.

As depicted in Table 4, the classification accuracy of all deep learning models is higher than that of the SVM. In addition, the classification accuracy of the AW-CNN is higher than that of the CNN and the DBN. The classification accuracy of the AW-CNN with the preprocessed dataset is the highest. Therefore, sonar image preprocessing is conducive to sonar image classification.

From the above comparative experiments, the following results can be drawn. The proposed AW-CNN can improve classification accuracy under a relatively close convergence rate in Figs. 6 and 7, Tables 2 and 3. Moreover, the images of the texture structure in Fig. 8 are clearer than those of in Fig. 9. This further verifies that the proposed AW-CNN has better accuracy and effectiveness. Meanwhile, the proposed AW-CNN with the preprocessed dataset can make the objects' contours more prominent and more beneficial to classification in Fig. 13 and Table 4. Therefore, the proposed AW-CNN can solve the problem of the random initialization of the filter weights in the CNN, avoid falling into the local optimum, and improve the classification accuracy. It has a certain effectiveness in underwater sonar image classification.

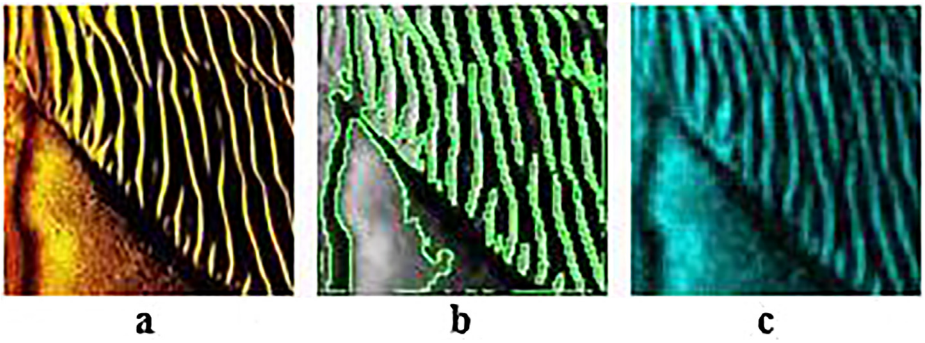


Fig. 10. Image preprocessing of a sand ripple. (a) Original image, (b) detection image, and (c) feature extraction image.

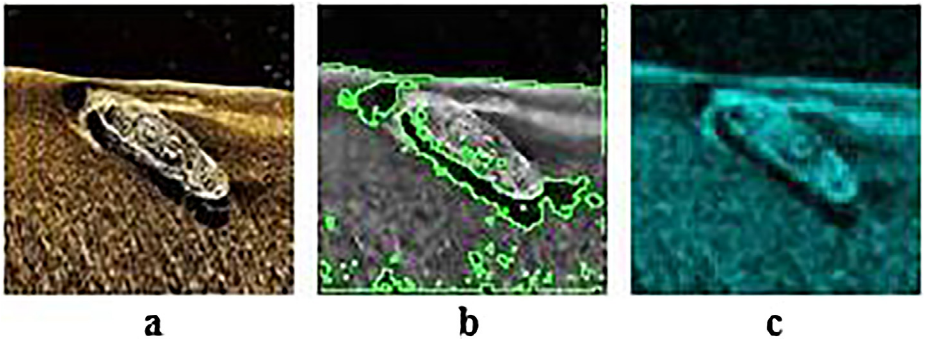


Fig. 11. Image preprocessing of a sunken ship. (a) Original image, (b) detection image, and (c) feature extraction image.

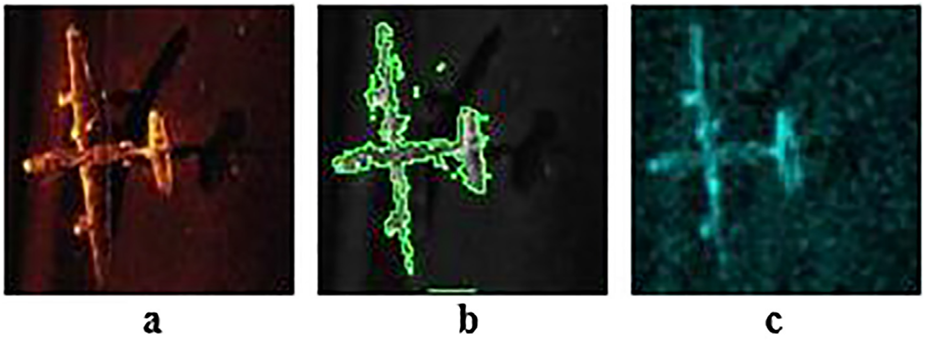


Fig. 12. Image preprocessing of a crashed plane. (a) Original image, (b) detection image, and (c) feature extraction image.

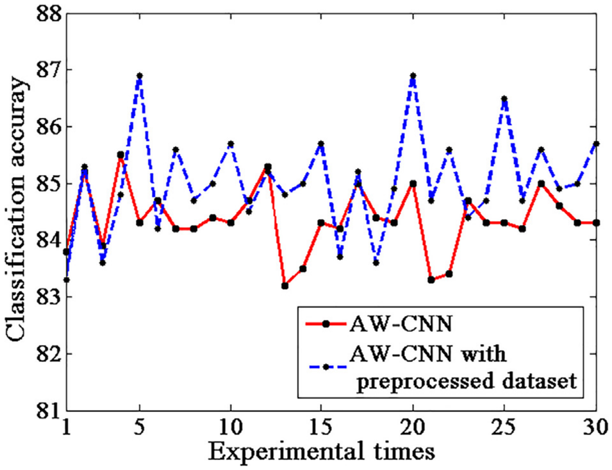


Fig. 13. The contrast classification diagram of preprocessed image.

Table 4
The classification accuracy of various models.

Models	AW-CNN with preprocessed dataset	AW-CNN	CNN	DBN	SVM
Classification accuracy(%)	86.90	85.50	80.20	60.00	52.00

5. Conclusion

In this paper, a novel AW-CNN was proposed to classify underwater sonar images. Through the proposed AW-CNN, the image features can be automatically extracted by the internal network structure, and the random initialization of the filter weights in the CNN can be effectively solved. In the process of implementing the AW-CNN, the generated weights of the DBN were utilized to adaptively replace the randomly trained filter weights of the CNN, which can achieve higher classification accuracy under a relatively close convergence rate in sonar image classification. To validate the classification accuracy of the AW-CNN, the simulation method and experimental research were used to analyze the classification accuracy of the AW-CNN under the maximum number of iterations and fixed numbers of iterations. The classification accuracies of the AW-CNN, the CNN and the DBN are 85.50%, 81.10% and 60.20%, respectively, under the maximum number of iterations. In addition, when the number of iterations is fixed 600, 700, 800 or 1000, the AW-CNN presents higher classification accuracy than the CNN. Moreover, the AW-CNN with the preprocessed dataset has the highest classification accuracy. Based on the above observations, the newly proposed classification method can effectively improve the classification accuracy of underwater sonar images, and thus it yields better results.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant, Nos. 41306086, 11404077, 61631008, 61471137, 50509059 and 51779061). This work was also supported by the Heilongjiang Province Outstanding Youth Science Fund (Grant No. JC2017017), The Fok Ying-Tong Education Foundation, China (Grant No. 151007), and the Marine Nonprofit Industry Research Subject (Grant No. 2013M531015). The authors are grateful to the guest editors and anonymous reviewers for their constructive comments based on which the presentation of this paper has been greatly improved.

References

[1] Yang L, Chen K. Performance comparison of two types of auditory perceptual features in robust underwater target classification. Acta Acust United Acust 2017;103(1):56–66.

[2] Wang X, Guo L, Yin J, Liu Z. Narrowband Chan-Vese model of sonar image segmentation: A adaptive ladder initialization approach. Appl Acoust 2016;113:238–54.

[3] Hodgetts MA, Greig AR, Fairweather A. Underwater imaging using Markov Random Fields with feed forward prediction. Underwater Technol 1999;23(4):157–67.

[4] Erkmen B, Yıldırım T. Improving classification performance of sonar targets by applying general regression neural network with PCA. Expert Syst Appl 2008;35(1):472–5.

[5] Zhao J, Zhang H. Seabed classification based on SOFM neural network. IEEE Comput Soc 2008:902–5.

[6] Martin A, Osswald C. Experts Fusion and Multilayer Perceptron Based on Belief Learning for Sonar Image Classification. IEEE 2008:718–23.

[7] Khidkikar M, Balasubramanian R. Segmentation and classification of side-scan sonar data. Lect Notes Comput Sci 2012;7506:367–76.

[8] Rominger C, Martin A, Khenchaf A, et al. Sonar image registration based on conflict from the theory of belief functions. IEEE 2009:1317–24.

[9] Lopera O, Dupont Y. Target classification from HR sonar images. Oceans 2013:1–6.

[10] Peng L. Efficient High-Precision Feature Extraction Method in Side Scan Sonar Image. Appl Mechan Mater 2014;678:197–200.

[11] Wang X, Liu X, Japkowicz N, et al. Automatic Target Recognition using multiple-aspect sonar images. IEEE 2014:2330–7.

[12] Li K, Li C, Zhang W. Research of Diver Sonar Image Recognition Based on Support Vector Machine. Advanced Mater Res 2013;785:1437–40.

[13] Karine A, Lasmar N, Baussard A, et al. Sonar image segmentation based on statistical modeling of wavelet subbands. Int Conf Comput Syst Appl 2015:1–5.

[14] Sion M, Teubler T, Hellbruck H. Embedded multibeam sonar feature extraction for online AUV control. OCEANS-IEEE 2016:1–4.

[15] Zhu M, Song Y, Guo J, et al. PCA and Kernel-based Extreme Learning Machine for Side-Scan Sonar Image Classification. IEEE 2017:1–4.

[16] Williams DP. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks. Int Conf Patt Recogn 2016:2497–502.

[17] Yu K, Jia L, Chen Y, Xu W. Deep learning: yesterday, today, and tomorrow. Comput Res Dev 2013;50(9):1799–804.

[18] Lecun Y, Bengio Y, Hinton GE. Deep learning. Nature 2015;521(7553):436–44.

[19] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neur Comput 2006;18(7):1527.

[20] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504.

[21] Cecotti H, Belaid A. Rejection strategy for convolutional neural network by adaptive topology applied to handwritten digit recognition. IEEE Comput Soc 2005:765–9.

[22] Cireşan DC, Meier U, Gambardella LM, et al. Deep, big, simple neural nets for handwritten digit recognition. Neural Comput 2010;22(12):3207–20.

[23] Yalçın H. Human activity recognition using deep belief networks. IEEE 2016:1649–52.

[24] Li Y, Xie W, Li H. Hyperspectral image reconstruction by deep convolutional neural network for classification. Pattern Recogn 2016;63:371–83.

[25] Shi T, Zhang C, Li F, et al. Application of alternating deep belief network in image classification. CCDC 2016:1853–6.

[26] M. Iftene Q. Liu Y. Wang Very high resolution images classification by fine tuning deep convolutional neural networks Proceed SPIE 2016

[27] Arsa DMS, Jati G, Mantau AJ, Wasito I. Dimensionality reduction using deep belief network in big data case study: Hyperspectral image classification. IEEE 2016:71–6.

[28] Christian S, Sergey I, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. Artif Intell 2016:1–7.

[29] Miki Y, Muramatsu C, Hayashi T, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. Comput Biol Med 2017;80:24–9.

[30] Li W, Fu H, Yu L. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sensing 2017;9(1):1–13.

[31] Phillip MC, Harshawn SM. Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images. J Digit Imaging 2017;30(2):234–43.

[32] Williams DP, Dugelay S. Multi-view SAS image classification using deep learning. IEEE Monterey 2016:1–6.

[33] Williams DP. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks. Pattern Recogn 2017:2497–502.

[34] Shin YG, Yoo J, Kwon HJ. Histogram and gray level co-occurrence matrix on gray-scale ultrasound images for diagnosing lymphocytic thyroiditis. Comput Biol Med 2016;75:257–66.