International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 2, Mar-Apr 2024
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

ROCK AND MINE PREDICTION USING MACHINE LEARNING ALGORITHM

[1.]Dr.K.Bala[2.]S.Poovarasan [3.]M.Prabhakaran [4.]S,Praveen [5.]S.Praveen Kumar

[1]Professor, School of Computing, Department, Of Computer Science and Engineering Bharath Institute Of Higher Education and Research, Chennai, India -600073.[2345]

Student, School of Computing, Department, Of Computer Science and Engineering ,Bharath Institute of Higher Education and Research, Chennai, India -600073

ABSTRACT~In the realm of marine robotics, underwater automatic target recognition (ATR) poses a significant challenge due to the intricate underwater environment. Current recognition techniques rely on manually crafted features and classifiers to identify targets, resulting in suboptimal recognition accuracy. Our study introduces an innovative approach to achieve precise multiclass underwater ATR using forward-looking sonar—Echoscope in conjunction with deep convolutional neural networks (dcnns). The entire recognition process, spanning from data preprocessing to network training and image recognition, was successfully executed. Initially, we curated a genuine Echoscope sonar image dataset. Leveraging the graph-based manifold ranking method in image preprocessing, we extracted the suspected target region, inspired by the human visual attention mechanism. Subsequently, we devised an end-to-end dcnnsmodel, dubbed echonet, for Echoscope sonar image feature extraction and recognition. Lastly, we devised a network training strategy based on transfer learning to address the issue of limited training data, employing mini-batch gradient descent for network optimization. Our experimental findings showcase the efficiency of our approach, with a recognition accuracy of 80.3% achieved in a nine-class underwater ATR task, surpassing conventional feature-based methods. This proposed methodology holds promise as a cutting-edge technology for enhancing the intelligent perception capabilities of autonomous underwater vehicles

I.INTRODUCTION

Accurate identification of targets is essential for underwater exploration and ocean development. Since the 1960s, naval departments have placed great importance on underwater target recognition. In recent years, the need for advanced underwater target recognition technology has grown significantly in civil and commercial sectors due to the global economic recovery. This includes tasks such as tracking and protecting endangered aquatic species, salvage operations, aquaculture, and underwater archaeology. However, the complex underwater environment and limitations in marine sensing have made accurate multiclass underwater automatic target recognition (ATR) a challenging issue.

The manuscript was reviewed and approved for publication by Changsheng Li, the associate editor overseeing the process. Various studies have focused on using sonar imaging systems, particularly side-scan sonar and synthetic aperture

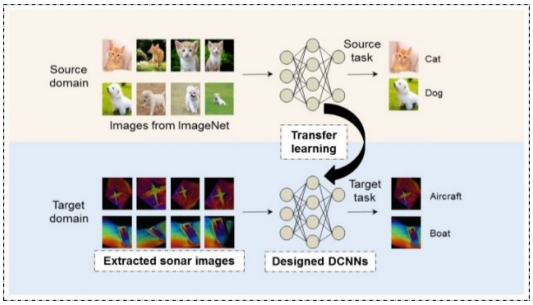
sonar (SAS), to recognize underwater targets through sonar images. Sonar images are preferred as they provide a clearer representation of underwater environments. The choice between side-scan sonar and SAS depends on the specific application. While side-scan sonar is suitable for imaging the seabed, it is not ideal for identifying floating objects or real-time recognition tasks. SAS faces challenges related to micro navigation and platform trajectory estimation, limiting its application due to high platform movement requirements. The real-time imaging sonar Echoscope has emerged as a significant innovation in underwater observation, enabling the generation of high-resolution images and the development of technologies for automatic underwater scene understanding. This paper addresses the complex task of underwater target recognition using the Echoscope sonar system.

Previously, manual features were utilized for visual object classification tasks, such as Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradient (HOG), and Fisher Vector. These features capture shape, texture, and color information and are combined with various classifiers as seen in previous works. While these features can perform well for specific data and tasks, they often lack generalization capability and require expertise and extensive trial and error for extraction. Underwater targets vary significantly in size, shape, texture, and background, even within the same class, posing challenges for conventional methods to accurately recognize multicass targets.

Deep learning has emerged as a prominent area of research in machine learning, aiming to automatically extract high-level features from large datasets through the learning process. The development of convolutional neural networks (CNNs) has been a major focus since the 1990s, with notable advancements in recent years. The introduction of deep CNNs by Krizhevsky et al. in 2012 demonstrated exceptional performance in image recognition tasks, leading to high expectations for their application in underwater target recognition. However, it is essential to consider the differences between sonar and optical images, as they present distinct characteristics that impact recognition processes.

Despite the progress in utilizing deep learning for underwater target recognition, there is still much to explore in this field. The success of deep CNNs in image processing has laid a foundation for further advancements, but challenges remain in adapting these techniques to underwater environments.

In this research article, we present an innovative approach for underwater automatic target recognition (ATR) using deep learning techniques. Our aim is to enhance the precision of multiclass target recognition tasks in underwater environments. We have developed a comprehensive methodology that encompasses data preprocessing, network training, and image recognition. To achieve this, we have designed a state-of-the-art deep convolutional neural network (DCNN) model called EchoNet, along with an effective training strategy.



An outline of the proposed method for recognizing underwater targets is presented in Figure 1. The accurate identification of multiple target classes is achieved through the utilization of Echoscope sonar images and deep convolutional neural networks. Additionally, a training method based on transfer learning has been devised to address the issue of limited sonar image data.The purpose of this development is to address theissue of insufficient training data. Instead of relying on domain knowledge of sonar image feature extraction, the features are learned directly from the data itself. Additionally, we have created a dataset of Echoscope sonar images, which is accessible to the vision research community for testing sonar image recognition algorithms. To the best of our knowledge, this is the first work that utilizes Echoscope sonar images and DCNNs for underwater ATR. Through experimental results, we have demonstrated that our method significantly enhances the accuracy of underwater multiclass ATR in comparison to traditional feature-based classifiers.The remaining sections of this paper are organized as follows: Section II provides a detailed explanation of the EchoNet for sonar image recognition. In Section III and Section IV, we present and discuss the experimental results. Finally, in Section V, we offer concluding remarks and outline directions for future research.

II. ACCURATE TARGETRECOGNITION

The structure of the proposed precise underwater ATR technique is illustrated in Figure 1. The primary focus is on training our custom DCNNs for underwater target identification using Echoscope sonar images (depicted in the lower section of Figure 1). The following three crucial aspects will be elaborated upon: sonar image preparation, DCNNs model development, and network training through transfer learning. The upper section of Figure 1 involves training a DCNNs through conventional supervised learning using a vast image dataset (such as ImageNet [23]), with the trained model serving as the foundational network for transfer learning. The trained foundational network can be likened to the prior knowledge that a human acquires from past visual experiences, which aids in the learning process of the target network [24].

The Echoscope imaging sonar, created by Coda Octopus, stands as the most advanced commercial real-time sonar in the world. Boasting a horizontal and vertical resolution of 0.4◦, the Echoscope is capable of producing high definition images with an impressive maximum range.

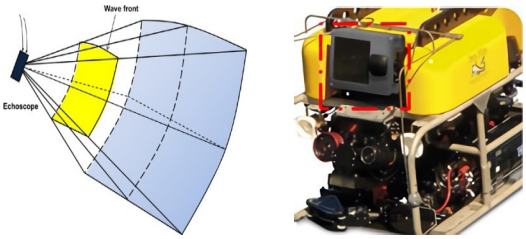


FIGURE2.OverviewoftheEchoscope:

(a)Beamenergydistributionof thephasedarrayimagingsonarsystem;

(b)AnunderwaterROVequipped withanEchoscope(markedwitharectangle)

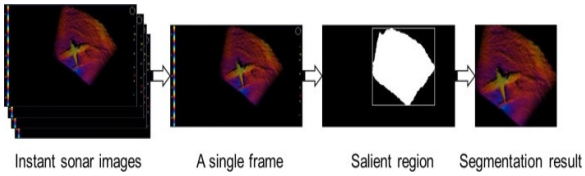


FIGURE3.Extractionoftheinterestedimageregio n.Underwaterscenesarefirstrecorded byE-UIS,saliencydetectionmethodviagraph-based manifold ranking is then applied to each single frame to get the interestedimageregion.Thesesegmentationresult sarefinallycollected fortrainingandtestingtheEchoNet.

The imaging sonar utilizes phased-array techniques to produce over 16,000 discrete beams simultaneously, resulting in range measurements that gather data points with known position and intensity (x, y, z, i) to form a comprehensive sonar image. Operating at a ping rate of up to 12 Hz, the Echoscope is capable of delivering successive image frames akin to video footage for monitoring both moving and stationary targets. A schematic diagram illustrating the Echoscope imaging principle is depicted in Fig. 2, along with an Echoscope mounted on a remotely operated vehicle (ROV).

By offering a high-resolution sonar image, the system presents a clear view of the underwater environment, facilitating the automatic identification of potential targets. The sonar data produced by the Echoscope allows for the creation of continuous underwater images.

The network architecture consists of several convolutional layers and 2 fully connected layers, inspired by the structure of AlexNet. The raw input is processed through these layers, with the final fully connected layer outputting to Softmax to create a probability distribution across n class labels, where n represents the category number of underwater targets. AVE pooling is chosen over MAX pooling after the initial convolutional layer to better handle the speckle-like characteristics of sonar images. In order to minimize connection parameters, only two fully connected layers are utilized, with dropout implemented solely in FC1.

1.CONVOLUTIONALLAYER

The output feature map yj can be obtained by combining convolutions with multiple input maps. It can be expressed as yj = f .Σxi∗kij + bjΣ (1), where ∗ represents the two-dimensional discrete convolution operator

and bj is an additive bias. The activationfunction f (x), known as Rectified Linear Units (ReLU), is applied to each convolutional layer and FC1 layer. It is important to note that both the convolution filter weights and biases are model parameters that require learning

2.POOLINGLAYER

The AVE pooling operation is employed to calculate the average value within a pixel's surrounding region, whereas the MAX pooling operation is utilized to determine the maximum value. By incorporating a pooling layer, the dimension of the feature maps can be reduced and a slight translation invariance can be introduced.

3.FULLY CONNECTEDLAYERS

Serves as a categorizer within the entire network. The fully connected layers are calculated as Y6 = f (W6Y5 + B6) and Y7 = ψ (W7Y6 + B7), where Wl and Bl are matrices of the trainable parameters, ψ (X) [i] = eX[i].

C. NETWORK TRAINING WITH TRANSFER LEARNING

The training procedure of EchoNet aims to minimize the classification error on the training dataset. In other words, the model parameters θ are learned in order to minimize the cross-entropy cost function. The segmentation results can be obtained by normalizing the RGB channels separately using min-max normalization. The pixel values of the scene-level sonar images are also normalized to a range of [0, 1] to reduce any unwanted influence on target recognition. Additionally, the training process involves a weight decay term, λR(θ), which is a form of L2 regularization. Typically, the training process of DCNNs follows this pattern:

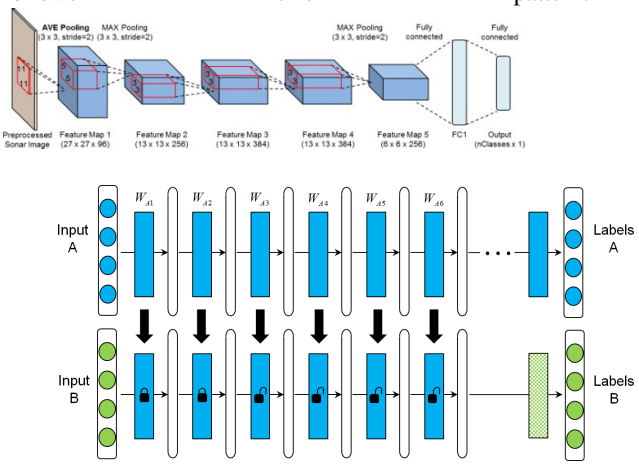


FIGURE5.ThetrainingmethodoftheEchoNet.Thelabeledrectangles(e.g.W_{A1})representweightslearnedforeachlayer,andcolorindicateswhichdatasettheweightswereoriginallytrainedon.Theellipsoidsbetweenrectanglesrepresentthefeaturemapsateachlayer

.

III.EXPERIMENTS ANDRESULTS

A. ECHOSCOPESONAR IMAGEDATASET

Our underwater target recognition method's effectiveness is assessed using an actual, measured sonar image dataset. Coda Octopus has carried out sea experiments for this purpose.

Resolution sonar image dataset. Table 1 provides a summary of each sea experiment and the corresponding number of sonar images obtained at the scene level. A total of 2,915 verified target images belonging to 9 classes were collected,resolution sonar image dataset. Table 1 provides a summary of each sea experiment and the corresponding number of sonar images obtained at the scene level. A total of 2,915 verified target images belonging to 9 classes were collected, preprocessed, and manually labeled. These images were then divided into 3 subsets: 900 images (100 images per class) for training, 450 images (50 images per class) for validation, and the remaining 1565 images for testing.

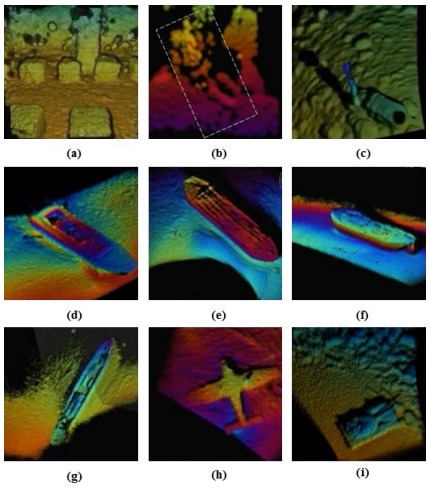


FIGURE 7. High-resolution sonar images of interested underwater targets.Fromtoplefttobottomright:(a)Cornerstone;(b)Diver(marked witharectangledottedline);(c)ROV;(d)Sunkenbarge1;(e)Sunken barge2;(f)Sunkenbarge3;(g)Shipwreck;(h)Sunkenplane;(i)Sunken militarytank

TABLE1. Details of the Echoscope sonar image dataset

Target	Sea experiments		Number of extracted images
	Year	Location	
Cornerstone	2009	USA	335
Diver	2010	UK	287
ROV	2011	UK	311
Sunken barge1	2012	UK	223
Sunken barge2	2012	UK	394
Sunken barge3	2012	UK	195
Shipwreck	2012	UK	370
Sunken plane	2012	UK	583
Sunken military tank	2012	UK	217

We conducted experiments using the high-resolution imaging sonar Echoscope between 2009 and 2012 in different geographical locations with varying environmentalconditions. Eachexperiment focused on a specific underwater target such as a cornerstone, diver, ROV, sunken barge, shipwreck, sunken plane, and sunken military tank, as illustrated in Fig. 7. The results of these experiments were used to create a high- The original size of the sonar images ranged from 150x150x3 to 240x240x3 pixels. To ensure consistency, all images were resized to 227x227x3 pixels before being inputted into EchoNet. The images presented in Fig. 7 were also resized to this standard size. The dataset contains sonar images that exhibit variations in target position, orientation, scale, colors, and textures within each class.

C.TRAINING AND TESTING THEECHONET

In this section, we initially outline the training specifics and then present the experimental outcomes of the proposed target recognition technique on the Echoscope sonar image dataset. Our DCNNs models are built on the efficient and practical open-source Caffe framework [30], with the

EchoNet architecture detailed in section II. B.

The training procedure of EchoNet is extensively discussed in section II. C. The AlexNet architecture serves as the base network, trained on a subset of 1000 optical images in each of 1000 categories from the ImageNet-2012 dataset, achieving a final top-1 error rate of 42.6% on the validation set. Subsequently, the parameters (WA1 WA6) of AlexNet are transferred to EchoNet. The first 2 layers of EchoNet are frozen, while the remaining layers are trained on the Echoscope sonar image dataset. Mini-batch gradient descent with a batch size of 45 is utilized for training EchoNet by back-propagating the classification error. The learning rate is set at 0.001, reduced by a factor of 2 every 200 iterations, along with weight decay of 0.005 and momentum of 0.9. The total number of training iterations is fixed at 1000, equivalent to 50 epochs.

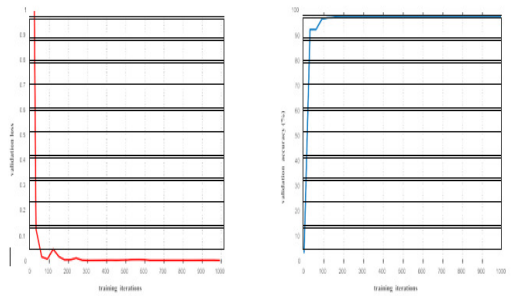
To provide a precise depiction of the results, we conducted the EchoNet training and testing experiment five times (each lasting approximately 1 hour), yielding an average iterations.Accuracyrisesquicklyatthefirst 100 iterations and tends to be stable after about 200 iterationstesting accuracy of 96.4% on the nine-class underwater target recognition task. The highest accuracy achieved was 97.3%, while the lowest was 94.4%. The validation loss versus training iterations curve and validation accuracy versus training iterations curve from one of the EchoNet training experiments are displayed in Fig. 8. Despite the substantial number of network parameters, overfitting is avoided due to the specialized training method developed, as well as the reduction of fully-connected layers. For instance, the confusion matrix for the best classification results is presented in Table 2,indicating minimal misclassification of highly similar samples.

We have observed some recent developments in this area.

C.METHODSCOMPARISON

We conduct comprehensive experiments to assess the effectiveness of the proposed EchoNet using four traditional classifiers in pattern recognition, along with two cutting-edge deep neural networks. The four traditional classifiers consist of the k-nearest neighbor (KNN) classifier, the multi-layer perceptron (MLP), and the nearest.result. In MLPClassifier, we set one hidden layer with 80 neurons and use stochastic gradient descent (SGD) to update the model weights. The learning rate is set to 0.1, and the maximum iteration is 1000. In SVC, the maximum iteration is also set to 1000, and the class weight is set to 'balanced'.

When using raw pixel values as input, the accuracy of the KNN classifier is 72.0%, while the NN classifier achieves an accuracy of 91.4%. By using the widely used baseline method of HOG SVM, we obtain an accuracy of 92.7%.



(a) (b)

FIGURE8.Training

curvesofEchoNet:(a)Validationlossvs.trainingiterations.Loss declinessharplyandtendsto be0afterabout300iterations;(b)Validationaccuracyvs.traini ng

TABLE 2**.** Confusion matrix of 9-class recognition results.

Category	Stone	Diver	ROV	Barge1	Barge2	Barge3	Ship	Plane	Tank	Accuracy
Stone	185	0	0	0	0	0	0	0	0	100.00%
Diver	0	137	0	0	0	0	0	0	0	100.00%
ROV	3	0	158	0	0	0	0	0	0	98.10%
Barge1	0	0	0	75	0	0	0	0	0	100.00%
Barge2	0	0	0	2	242	0	0	0	0	99.20%
Barge3	0	0	0	0	0	45	0	0	0	100.00%
Ship	0	0	0	0	0	0	220	0	0	100.00%
Plane	0	0	0	9	29	0	0	395	0	91.20%
Tank	0	0	0	0	0	0	0	0	57	100.00%

80.

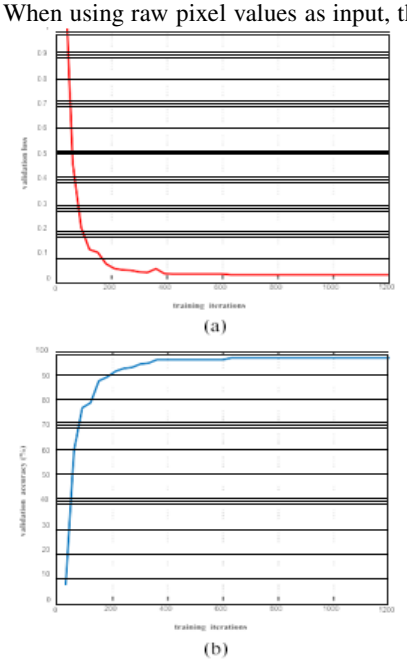
The Scikit-learn machine learning module is commonly used to implement the neighbor (NN) classifier and support vector machine (SVM). In the case of the sonar image dataset, it is divided into two subsets: 1350 images for training (150 images for each of the 9 classes) and the remaining 1565 images for testing. Two types of features are used as input for the classifiers: the original pixel value of the image and the HOG feature.

To begin with, we define two preprocessing functions. The first function flattens a 227x227x3 image into a row of pixels. The second function extracts the HOG feature from a resized 180x180x3 sonar image using the ft.hog function. The image is divided into 15x15 blocks, with each block containing 2x2 cells and each cell consisting of 6x6 pixels. We then extract the features from each image and store them in arrays.

Finally, we apply the KNeighborsClassifier, MLPClassifier, and SVC functions to evaluate the data.

For the KNN method, we vary the number of neighbors and keep track of the best result. In MLPClassifier, we set one hidden layer with 80 neurons and use stochastic gradient descent (SGD) to update the model weights. The learning rate is set to 0.1, and the maximum iteration is 1000. In SVC, the maximum iteration is also set to 1000, and the class weight is set to 'balanced'.

When using raw pixel values as input, the accuracy of the



(a)

(b)

KNN classifier is 72.0%, while the NN classifier achieves an accuracy of 91.4%. By using the widely used baseline method of HOG SVM, we obtain an accuracy of 92.7%.

In addition, EchoNet is compared with two well-known deep neural networks: AlexNet and GoogLeNet. We implement these networks using the Caffe framework and employ pre-trained models provided by Caffe. These models are then fine-tuned using our Echoscope sonar image training dataset. For the fine-tuning process, we use the same hyper-parameters as the EchoNet for AlexNet. As for GoogLeNet, each iteration of MBGD (Mini-Batch

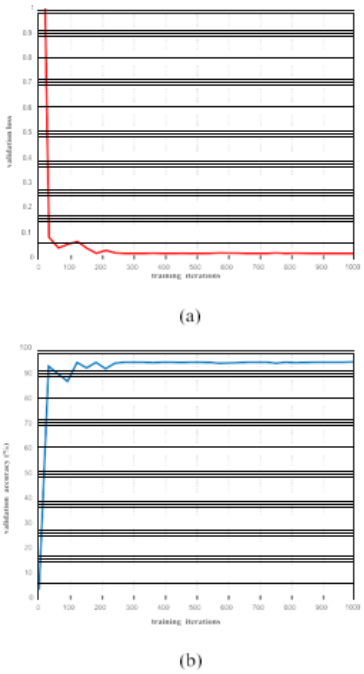
Method	Accuracy	Inference time per image (ms)
KNN_raw_pixel	72.0%	341.2
MLP_raw_pixel	89.3%	1.1
NN_HOG	91.4%	311.4
SVM_HOG	92.7%	133.7
AlexNet	94.1%	73.1
GoogLeNet	97.0%	186.9
EchoNet	97.3%	60.6

Gradient Descent) uses a batch size of 45, a momentum of 0.9, and a multiplicative weight decay of 0.005 per iteration. The learning rate is set to k=10. The accuracy of MLP is

FIGURE9.TrainingcurvesoftheAlexNet

(a)Validationlossvs.trainingiterations;

(b)Validationaccuracyvs.trainingiterations



(a)

(b)

FIGURE9.TrainingcurvesoftheAlexNet

(a)Validationlossvs.trainingiterations

(b)Validationaccuracyvs.trainingiterations

FIGURE10.TrainingcurvesoftheGoogLeNet:(a)Validationlossvs. trainingiterations;(b)Validationaccuracyvs.trainingiterations

GoogLeNet attains comparable recognition accuracy to EchoNet, although the training duration for GoogLeNet is ten times lengthier than that of EchoNet. In general, the methods based on Deep Convolutional Neural Networks (DCNNs) outperform the traditional hand-crafted features based methods.

TABLE3.TheresultsofcomparisonmethodsonthesonarimagedatasetIn addition to the accuracy of underwater multiclass target recognition, the real-time requirement is also a crucial aspect in ATR. Hence, we also evaluated the effectiveness of each method presented in Table 3.

The classifier design typically comprises two components: training and test (inference). In practical terms, the focus is more on the inference efficiency of a classifier. Table 3 displays the average inference time taken by each recognition method on a single sonar image, calculated by averaging the total test time of 1565 test images. The test platform utilized is based on a dual-core Intel processor. It is important to note that since the first four methods are developed in Python, and the last three methods run in Caffe, these runtimes cannot be directly compared, but they generally indicate the efficiency of each method. The Caffe framework primarily relies on C, which is significantly more efficient than Python.

Regarding specifics, the time complexity of the KNN and NN algorithms increases with the growth of training sample size and feature dimension. Each test sample must be compared with all training samples, resulting in extensive distance calculations. The MLP method, with its shallow structure and straightforward computation, proves to be highly efficient. On the other hand, theSVM_HOG method requires independent feature extraction (50 ms) before utilizing SVM for image recognition, leading to reduced recognition efficiency. As for the last three DCNNs-based methods, their forward pass floating-point operations (FLOPs) are approximately 720M, 1550M, and 700M respectively. Although each model requires hours for training, they do not consume much time during testing due to the benefits of an end-to-end model structure and efficient numerical computation. As illustrated in Table 3, EchoNet only takes 60.6 ms to test one sonar image. Given that the maximum refresh rate of the Echoscope is 12 frames per second, EchoNet proves to be sufficiently fast to process each frame in real-time.

IV.DISCUSSION

In this section, we mainly discuss the impact of imagenoise, network architecture and training method on the recognition performance, and further analyze the reason for the success of transfer learning through parametervisualization

A. EFFECT OF IMAGE NOISE ONRECOGNITION

InFig.7,wehaveshownsomesonarimagesgenerated bythe Echoscopeduringseaexperiments,whichareofgood quality.

TABLE 4.Recognition results of The image datasets polluted by noise.

Method	Accuracy	
	Noise variance 0.01	Noise variance 0.1
MLP_raw_pixel	62.6%	30.9%
SVM_HOG	80.8%	56.0%
AlexNet	91.1%	88.4%

In this section, we aim to examine the impact of image noise on the accuracy of our method's recognition. We introduce two types of zero-mean white Gaussian noises with variances of 0.01 and 0.1 respectively. These noises are artificially added to the normalized real-measured sonar images, resulting in two simulated sonar image datasets that are contaminated by noise. Fig. 11 displays a selection of representative sonar images.

To evaluate the performance of our method, EchoNet, as well as other comparison methods, we conduct experiments on the polluted image datasets. The results of these experiments are presented in Table 4. The experimental settings remain consistent with those described in subsection III. C, and each data point represents the average value of five experimental results.

From Table 4, it is evident that both EchoNet and AlexNet are capable of achieving satisfactory recognition accuracy even in the presence of high levels of noise. Conversely, traditional feature-based methods are vulnerable to noise, leading to a significant reduction in their accuracy. This further highlights the advantage of using DCNNs. Additionally, as the noise level increases, the recognition accuracy of each method decreases. Therefore, it may be beneficial to consider employing image noise reduction techniques during sonar image preprocessing to enhance the effectiveness of our application.

B.AVE POOLING VS. MAXPOOLING

In Section II, we have provided an overview of the architecture of the designed DCNNs. In this architecture, the first pooling layer utilizes the average approach instead of the commonly used maximum approach. Now, we aim to evaluate a network that shares a similar structure with EchoNet, but with a modification in the first pooling layer to use MAX pooling. We conduct experiments using the same training method and hyper-parameters, and the results are presented in Fig. 12. With the MAX pooling approach, the average testing accuracy of the five experiments is 92.6%, with a maximum of 96.2% and a minimum of 89.3%. On the other hand, the AVE pooling approach achieves an average testing accuracy of 96.4%, with a maximum of 97.3% and a minimum of 94.4%. These results indicate that utilizing AVE pooling in the first pooling layer is more effective for Echoscope sonar image processing in our method.

C.TRAINING WITH DIFFERENT LOCKEDLAYERS

In our training approach, we enforce the first two layers of

the EchoNet to remain fixed, while the remaining layers are permitted to adapt and learn throughout the training procedure. Presently, we utilize the notation NL to represent the number of locked layers, and we delve into the influence of NL value on the performance of the network. The NL values are selected from the set {0, 1, ..., 6}, and we train new networks accordingly.

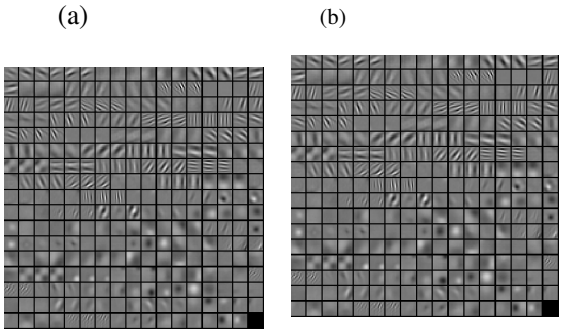


FIGURE11. Sonarimagespollutedbyzero-meanwhiteGaussiannoise:

Noise variance is 0.01; (b) Noise variance is 0.1.

FIGURE12. RecognitionresultsofAVEpoolingapproachandMAX poolingapproach

Please take note that during NL 0, all layers are able to participate in training, while during NL 6, only the last fully connected layer is permitted to learn. We conduct four experiments for each NL value. Apart from varying training strategies, the structure of each network remains identical to what was mentioned in Section II. All experiments utilize training data and test data from the same dataset, as specified in Table 1. The experimental results are depicted in Fig. 13, where we observe that the average recognition accuracy varies with NL.

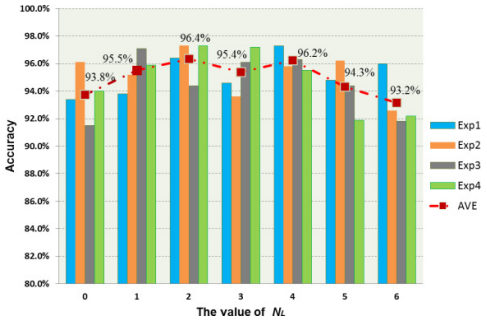


FIGURE13.Experimentalresults

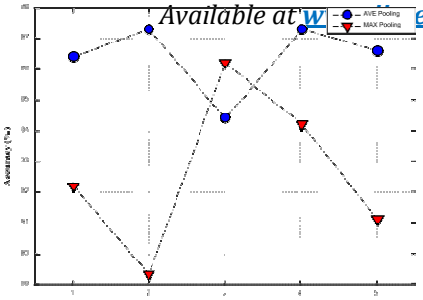
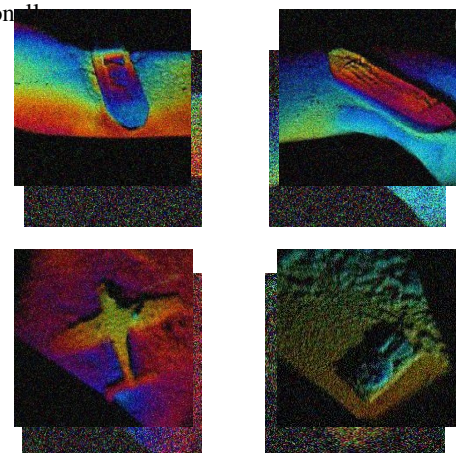


FIGURE14.Filterweightsvisualizationofthefirstconvolution



Filtersonlytrainedonopticalimages(N_L=2);(b)Filterstrainedonopticalimagesandfine-tunedbysonarimages(N_L=0tsoftheEchoNetwithdifferenttraining strategies

Upon examining the designed DCNNs, it becomes evident that NL 2 yields the most favorable outcome. This implies that solely fine-tuning the target network (NL 0) may not yield optimal performance. The decision to lock or unlock the initial l layers of the target network may rely on the scale of the target dataset and the number of parameters in the first m layers [32].

D.VISULIZATION OF LEARNED FILTERS

1. Based on the visualization technique proposed in [33], we showcase the filter weights of the initial convolutional layer from the EchoNet with NL values of 0 and 2 in Figure 14. The comparison between the two sets of filters visually reveals minimal discrepancies. To quantitatively assess the similarity between the two sets of weights, we introduce the correlation coefficient of matrices, which is computed using the formula provided.

Matrix respectively. The correlation coefficient of these two matrixes is 0.99, which is consistent with the results shown in Fig. 14.

Network parameters trained on optical images are similar to those used for sonar image recognition, which

illustrates the generalization ability of DCNNs for image processing. The front layers of the DCNNs can be treated as a versatile featureextractor,soitisreasonabletotransfertheknowledge of optical image recognition to sonar imagerecognition

V.CONCULSION

This paper introduces an ATR method that combines forward-looking sonar images with deep convolutional neural networks to enhance the accuracy of underwater multiclass target recognition tasks. The proposed end-to-end DCNNs model, EchoNet, is specifically designed for this purpose, along with a corresponding training strategy that automatically extracts high-level features from sonar images during the learning process to facilitate target recognition. Additionally, a sonar image dataset comprising 2,915 images is created for testing sonar image recognition algorithms.

A series of experiments are conducted to explore the impact of network architectureand training methods on recognition performance, as well as to analyze thesuccess of transfer learning. The results indicate that the proposed method outperforms traditional classifiers in terms of accuracy, real-time performance, and noise resistance. Notably, the method achieves an accuracy of 80.3% in a nine-class underwater ATR task, surpassing four traditional classifiers and two deep neural networks.

The research highlights the potential of utilizing imaging sonar and DCNNs for underwater ATR, which is crucial for enabling underwater vehicles to autonomously navigate and perceive the ocean environment. It is anticipated that deploying DCNNs on unmanned platforms will become more feasible in the future as network architecture continues to be optimized and hardware computing acceleration technology advances.

REFERENCES

[1]T.Fei,D.Kraus, and A. Zoubir, ‘‘Contributions to automatic tar- get recognition systems for underwater mine classification,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 505–518, Jan. 2015. doi:10.1109/TGRS.2014.2324971.

[2]S.Ntalampiras, ‘‘Hybrid framework for categorising sounds of Mysticete whales,’’ *IET Signal Process.*, vol. 11, no. 4, pp. 349–355, Jun. 2017. doi:10.1049/iet-spr.2015.0065.

[3]V.Myers and J. Fawcett, ‘‘A template matching procedure for automatic target recognition in synthetic aperture sonar imagery,’’ *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 683–686, Jul. 2010. doi:10.1109/LSP.2010.2051574.

[4]R.Fandos, A. M. Zoubir, and K. Siantidis, ‘‘Unified design of a feature- based ADAC system for mine hunting using synthetic aperture sonar,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no.5, pp. 2413–2426, May 2014.doi:10.1109/TGRS.2013.2260863.

[5]D.P. Williams, ‘‘Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks,’’ in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Cancun, Mexico, Dec. 2016, pp. 2497–2502.

[6]V.Murino and A. Trucco, ‘‘Three-dimensional image generation and processing in underwater acoustic vision,’’ *Proc. IEEE*, vol. 88, no. 12, pp. 1903–1948, Dec. 2000.

[7]N.Dalal and B. Triggs, ‘‘Histograms of oriented gradients for human detection,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886–893.

[8] J. McKay, V. Monga, and R. G. Raj, ‘‘Robust sonar ATRthrough Bayesianpose-corrected sparse classification,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5563–5576, Oct. 2017. doi:10.1109/TGRS.2017.2710040.

[9]M. D. Santos, P. O. Ribeiro, P. Núñez, P. Drews-Jr, and S. Botelho, ‘‘Object classification in semi structured enviroment using forward- looking sonar,’’ *Sensors*, vol. 17, no. 10, pp. 2235–2250, Sep. 2017. doi:10.3390/s17102235.

[10]G. E. Hinton and R. R. Salakhutdinov, ‘‘Reducing the dimensionality of data with neural networks,’’ *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006. doi:10.1126/science.1127647.

[11] Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ *Nature*, vol. 521, pp. 436–444, May 2015.doi:10.1038/nature14539.

[12] M. Z. Alom, T. M. Taha, and C. Yakopcic, ‘‘A state-of-the-art survey on deeplearningtheoryandarchitectures,’’*Electron.*,vol. 8,no.3,Mar.2019. doi:10.3390/electronics8030292.

[13]Y.LeCun, L. Bottou, Y. Bengio, and P. Haffner, ‘‘Gradient-based learn- ing applied to document recognition,’’ *Proc. IEEE*, vol.86, no.11, pp. 2278–2324, Nov. 1998.doi:10.1109/5.726791.

[14]A.Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘Imagenet classification with deep convolutional neural networks,’’ in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[15]Y. Sun, X. Wang, and X. Tang, ‘‘Deep learning face representation frompredicting 10,000 classes,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 1891–1898.

[16]D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang,

A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap,

F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, ‘‘Mastering the game of go without human knowledge,’’ *Nature*, vol. 550, no. 7676, pp. 354–359, Oct. 2017.

[17]H. Cho and S. C. Yu, ‘‘Real-time sonar imageenhancement for AUV-based acoustic vision,’’ *Ocean Eng.*, vol. 104, pp. 568–579, Aug. 2015.doi:10.1016/j.oceaneng.2015.05.037.

[18]X.Wang, J. Jiao, J. Yin, W.Zhao, X.Han, and B. Sun ‘‘Under- water sonar image classification using adaptive weights convolutional neural network,’’ *Appl. Acoust.*, vol. 146, pp. 145–154, Mar. 2019. doi:10.1016/j.apacoust.2018.11.003.

[19]G. E. Hinton, S. Osindero, and Y.-W. Teh, ‘‘A fast learning algorithm for deepbeliefnets,’’*NeuralComput.*,vol.18,no.7,pp .1527–1554,Jul.2006. doi:10.1162/neco.2006.18.7.1527.

[20]D. Jia, D. Wei, S. Richard, L.-J. Li, K. Li, and L. Fei-Fei, ‘‘Imagenet: A large-scale hierarchical image database,’’ *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[21]M.Oquab, L. Bottou, I. Laptev, and J. Sivic, ‘‘Learning and transferring mid-level image representations using convolutional neural networks,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 1717–1724.

[22]A. Davis and A. Lugsdin, ‘‘High speed underwater inspection for port and harbour security using coda echoscope 3D sonar,’’ in *Proc. OCEANS*, Washington, DC, USA, Sep. 2005, pp.2006–2011.

[23]C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, ‘‘Saliency detection via graph-based manifold ranking,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, Jun. 2013, pp. 3166–3173.

[24]H. Qiao, Y. Li, F. Li, X. Xi, and W. Wu, ‘‘Biologically inspired model for visual cognition achieving unsupervised episodic and semantic feature learning,’’*IEEETrans.Cybern.*,vol.46,no.10,pp. 2335–2347,Oct.2015. doi:10.1109/TCYB.2015.2476706.

