IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 20, 2019, accepted February 3, 2020, date of publication February 13, 2020, date of current version February 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2973856

Visual Recognition Based on Deep Learning for Navigation Mark Classification

MINGYANG PAN[1], YISAI LIU[1], JIAYI CAO[1], YU LI[2], CHAO LI[1], AND CHI-HUA CHEN[3], (Senior Member, IEEE)

[1]Navigation College, Dalian Maritime University, Dalian 116026, China
[2]Changjiang Nanjing Waterway Bureau, Nanjing 210011, China
[3]College of Mathematics and Computer Sciences, Fuzhou University, Fuzhou 350108, China

Corresponding authors: Mingyang Pan (panmingyang@ dlmu.edu.cn) and Chi-Hua Chen (chihua0826@gmail.com)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 3132019400, and in part by the National Natural Science Foundation of China under Grant 61906043.

ABSTRACT Recognizing objects from camera images is an important field for researching smart ships and intelligent navigation. In sea transportation, navigation marks indicating the features of navigational environments (e.g. channels, special areas, wrecks, etc.) are focused in this paper. A fine-grained classification model named RMA (ResNet-Multiscale-Attention) based on deep learning is proposed to analyse the subtle and local differences among navigation mark types for the recognition of navigation marks. In the RMA model, an attention mechanism based on the fusion of feature maps with three scales is proposed to locate attention regions and capture discriminative characters that are important to distinguish the slight differences among similar navigation marks. Experimental results on a dataset with 10260 navigation mark images showed that the RMA has an accuracy about 96% to classify 42 types of navigation marks, and the RMA is better than ResNet-50 model with which the accuracy is about 94%. The visualization analyses showed that the RMA model can extract the attention regions and the characters of navigation marks.

INDEX TERMS Deep learning, image classification, multi-scale attention, navigation marks, ResNet.

I. INTRODUCTION

For vessels, Electronic Chart Display and Information System (ECDIS) is one of the most important systems to guarantee the safety of navigation. ECDIS means a navigation information system which with adequate back-up arrangements can be accepted as complying with the up-to-date chart required by regulations V/19 and V/27 of the 1974 SOLAS Convention, as amended, by displaying selected information from a system electronic navigational chart (SENC) with positional information from navigation sensors to assist the mariner in route planning and route monitoring, and if required display additional navigation-related information [1], [2]. ECDIS integrates information from SENC, GPS, Gyro, Radar, ARPA, AIS, and allows for monitoring of a ship's position in real-time in a single display. It significantly improve the situation awareness and spatial abilities at sea of mariner [3]. However, there are also risks involved with over-reliance on ECDIS. In a complex and rapidly changing marine environment, watchkeeping is still demanding for safe navigation. The watchkeeper should observe the environment, identify objects around and analyze the situation accordingly.

With the development of AI technology in recent years, various intelligent technologies have also been exploited to research smart ships [4], [5] and intelligent navigation [6], [7]. In order for a ship to autonomously navigate, basically, the first technologies are the ability to collect the information about the environment that occurs in and out of the ship [8]. For collecting outside environment information, ECDIS will still play an important role by providing hydrographic data and sensor data from Radar and AIS, however, one new challenge is to recognize objects from cameras, which is finally expected to replace the visual observation of watchkeeper.

In this field, deep learning techniques have been gradually applied especially in ship identification and classification. B. Liu et al. proposed a new ship tracking and recognition method based on Darknet network model and YOLOv3 algorithm which realize ship tracking and real-time detection and recognition of ship types, and solve the problem of ship

The associate editor coordinating the review of this manuscript and approving it for publication was Honggang Wang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

tracking and recognition in important monitored waters [9]. HX. Fu et al. established the structure of the marine target recognition system based on Faster RCNN, which improved the accuracy of ship identification and improved the efficiency of the algorithm [10]. Q. Oliveau et al. introduced a novel ship category recognition method based on semi-supervised learning; the strength of their method resides in its ability to leverage labeled and unlabeled observed data while being highly effective and efficient in order to handle unobserved ones [11]. QQ. Shi et al. proposed a new deep learning framework that combines low-level functions, which effectively utilizes useful information in optical images for ship classification [12]. QC. Fan proposed a method of ship target segmentation based on CP SAR image and a full convolution network (U-Net) for pixel level detection, which solves the problem that the traditional ship recognition method is difficult to select features [13]. MZ. Jiang presented a novel method for ship classification that uses synthetic-aperture-radar images to distinguish ships based on superstructure scattering features [14]. JW. Li proposed a new loss function and convolutional neural network model. The new loss function can maximize the intraclass compactness and interclass separation simultaneously [15]. SJ. Lee et al. proposed a state-of-the-art neural network based object detection algorithms which applied to detect ships from the images and videos taken on the sea [16].

Obviously, on addition to collision with other ships, there are many other types of navigation risks such as grounding, crossing track, entering a prohibited area and so on. So, detecting and recognizing the related features from the camera images is also very important. However, there is few research focused on this topic so far.

Normally, such features are marked by aids to navigation. According to the dictionary of IALA (International Association of Marine Aids to Navigation and Lighthouse Authorities), aids to navigation is ''A device, system or service, external to vessels, designed and operated to enhance safe and efficient navigation of individual vessels and/or vessel traffic'' [17]. The term 'aids to navigation'' encompasses a wide range of floating and fixed objects, and consist primarily of buoys and beacons. Buoys are floating objects anchored to the bottom, their distinctive shapes and colors indicate their purpose and how to navigate around them. Beacons are structures permanently fixed to the sea-bed or land, they range from structures such as light houses, to single-pile poles. Both buoys and beacons may be called ''marks''.

Therefore, this study proposed a visual recognition model based deep learning to classify different types of navigation marks. And, the contributions of this study is highlighted as follows.

1) Exploiting deep learning technologies to recognize navigation marks.

2) Propose a fine-grained recognition model named RMA (ResNet-Multiscale-Attention) for navigation mark classification.

3) The RMA has an accuracy about 96% to classify 42 types of navigation marks, which is better than ResNet-50 model with 94% accuracy.

The remainder of the paper is organized as follows. Section II provides the overview of image recognition methods. Section III presents the proposed visual recognition model for navigation marks classification based on deep learning. The practical experimental results and discussion are illustrated in Section IV. Finally, conclusions and future work are given in Section V.

II. RELATED WORK

Image recognition is an important practical application field of applying deep learning algorithms. For classification, image recognition methods include two levels: general-level image classification and fine-grained image classification. The general-level image classification aims to classify images into different main categories such as ship, car, airplane and so on. And, fine-grained image classification normally aims to recognize sub-categories under some basic-level categories [18].

A. GENERAL-LEVEL IMAGE CLASSIFICATION

World image recognition competitions such as Kaggle, ImageNet, and web vision, attract schools and research institutions from all over the world every year. A. Krizhevsky et al. trained a large, deep convolutional neural network in 2012 to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes, and it achieved the best result [19]. C. Szegedy et al. proposed a deep convolutional neural network architecture called GoogLeNet, a 22 layers deep network, which achieved the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 [20]. KM. He et al. presented a residual learning framework to ease the training of networks that are substantially deeper than those used previously. On the ImageNet dataset they evaluated residual nets with a depth of up to 152 layers-8× deeper than VGG nets but still having lower complexity [21]. G. Huang et al. introduced the Dense Convolutional Network (DenseNet), which connects each layer to every other layer in a feed-forward fashion. DenseNet alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters [22]. J. Donahue et al. developed a novel recurrent convolutional architecture suitable for large-scale visual learning which is end-to-end trainable. And this model may have advantages when target concepts are complex and/or training data are limited [23]. TH. Chan et al. proposed a very simple deep learning network for image classification called the PCA network (PCANet) which can be extremely easily and efficiently designed and learned. And the PCANet has the potential to serve as a simple but highly competitive baseline for texture classification and object recognition [24].

J. Mansanet et al. proposed a new model called Local Deep Neural Network (Local-DNN), which is based on two

key concepts: local features and deep architectures. This model learns some difficult problems (such as the gender recognition problem) from small overlapping regions in the visual field using discriminative feed forward networks with several layers [25]. Y. Liu et al. proposed a novel semi-supervised classifier called discriminative deep belief networks (DDBN). For unsupervised learning, DDBN can preserve the information well from high-dimensional features space to low-dimensional embedding. For supervised learning, through a well-designed objective function, the back-propagation strategy of DDBN can directly optimize the classification results in training dataset by refining the parameter space [26]. HY. Shi et al. proposed a framework called the hypergraph-induced convolutional network to explore the high-order correlation in visual data during deep neural networks [27].

B. FINE-GRAINED IMAGE CLASSIFICATION

Different from the general-level object classification problem, fine-grained image classification is a quite challenging task due to the high degree of similarity among subcategories. Models of fine-grained image classification have made great progress in recent years.

N. Zhang et al. proposed Part-based R-CNN which uses the R-CNN algorithm to detect object-level and its local regions on fine-grained images. The proposed network structure contains both all features and local features with stronger discriminability, so it attained a higher accuracy [28]. S. Branson et al. proposed using the DPM algorithm to obtain the prediction points of the part annotation, and the Pose Normalized CNN performs a gesture alignment operation on the part-level image blocks, and extracts convolution features of different layers for different levels of image blocks of the fine-grained image [29]. X.-S. Wei proposed a novel Mask-CNN model without the fully connected layers, which consists of a fully convolutional network. Thanks to discarding the parameter redundant fully connected layers, the Mask-CNN has a small feature dimensionality and efficient inference speed by comparing with other fine-grained approaches [30]. T. Xiao et al. proposed he two-level attention models in deep convolutional neural network for fine-grained image classification, which focus on two different levels of features, object level and part level information. Also, the models do not require the data set to provide these annotation information, and relies entirely on its own algorithm to complete the detection of objects and local areas [31]. M. Simon et al. presented an approach that is able to learn part models in a completely unsupervised manner, without part annotations and even without given bounding boxes during learning [32]. T.-Y. Lin et al. proposed bilinear models, a recognition architecture that consists of two feature extractors whose outputs are multiplied using outer product at each location of the image and pooled to obtain an image descriptor. This architecture can model local pairwise feature interactions in a translationally invariant manner which is particularly useful for fine-grained categorization [33]. BL. Zhou et al. proposed a general technique called Class Activation Mapping (CAM) for CNNs with global average pooling. This enables classification-trained CNNs to learn to perform object localization, without using any bounding box annotations. Class activation maps allow us to visualize the predicted class scores on any given image, highlighting the discriminative object parts detected by the CNN [34].

C. APPLICATIONS OF IMAGE CLASSIFICATION

In recent years, various image recognition methods have been applied to different fields.

W. Hu et al. used a deep convolutional neural network to classify hyperspectral images directly in the spectral domain [35]. J. Ker et al. introduced the machine learning algorithms as applied to medical image analysis, focusing on convolutional neural networks, and emphasizing clinical aspects of the field [36]. GB. Liang et al. introduced the recurrent neural networks (RNNs). Specifically, they combined the CNN and RNN in order to propose the CNN-RNN framework that can deepen the understanding of image content and learn the structured features of images and to begin end-to-end training of big data in medical image analysis [37]. LQ. Yu et al. proposed a novel method for melanoma recognition by leveraging very deep convolutional neural networks (CNNs) to solve the high degree of visual similarity between melanoma and non-melanoma lesions, and the existence of many artifacts in the image [38]. PM. Cheng et al. evaluated transfer learning with deep convolutional neural networks for the classification of abdominal ultrasound images [39]. Q. Dou et al. proposed a novel automatic method to detect CMBs from magnetic resonance (MR) images by exploiting the 3D convolutional neural network (CNN). Their method can take full advantage of spatial contextual information in MR volumes to extract more representative high-level features for CMBs, and hence achieve a much better detection accuracy [40]. JM. Haut et al. introduced a new visual attention-driven technique for the HSI classification. Specifically, they incorporated attention mechanisms to a ResNet in order to better characterize the spectral-spatial information contained in the data [41]. Y. Gu et al. proposed a deep active learning framework that enables the active selection of both informative queries and reliable experts, which acquire high-quality, and large-annotated biomedical datasets [42].

As mentioned above, there are some researches about the application of image recognition methods on ship identification and classification [8]–[12]. There are also some methods to identify ships using fine-grained. H.C Shin et al. proposed a method to improve the accuracy of ship classification by using region of interest and artificial neural network [43]. B. Solmaz et al. used a multitasking learning framework and proposed a novel loss function to fine-grained classify and identify marine and land vehicles [44]. YB. Dong proposed a fine-grained ship classification framework for high-resolution SAR images based on the depth residual network, and built a deeper network and residual modules, and applyed batch normalization to maintain the output of activation

FIGURE 1. Some types of navigation marks in the Yangtze river.

function [45]. F. Bousetouane presented a CNN surveillance pipeline for vessel localization and classification in maritime video. The bright spot is to use the CNN function and the support vector machine classifier with linear kernel for fine-grained classification for object verification [46].

III. VISUAL RECOGNITION OF NAVIGATION MARKS

This section firstly describes the types of navigation marks, then applies ResNet and attention mechanism to construct a visual recognition model called RMA (ResNet-Multiscale-Attation) for classifying navigation marks.

A. TYPES OF NAVIGATION MARKS

Normally, navigation marks are prescribed across the world by the IALA. In 1977, the IALA endorsed two maritime buoyage systems, region A (IALA A) covers all of Europe and most of the rest of the world, whereas region B (IALA B) covers only the Americas, Japan, the Philippines and Korea. The differences between these two systems are few, the most striking difference is the direction of buoyage. All navigation marks within the IALA system are classified into 6 categories: lateral, cardinal, isolated danger, safe water, new wreck and special.

However, in this paper, a more complex buoyage system adopted in the Yangtze river of China is focused. There are 4 main categories of marks which be divided into 27 sub-categories, then be further classified into 97 types according their function, shape, colour and topmark.

Figure 1 shows the images of part of them. During daytime, the identification of marks can be accomplished by observing with shape, colour scheme, auxiliary features or markings (name, number, etc). As shown in Figure 1, some of them are generally same in global appearance, and distinguished only by the subtle and local differences. For example, the marks with code of ''DCZYTHFB'' and ''ZADCCMFB'' have similar boat-shaped body, a subtle difference is that ''DCZYTHBF'' has vertical stripe in the middle part. ''DCZYTHBF'' is a type of the sub-category of safe water marks which indicate there is navigable water all around the mark including the end of a channel or mid channel. However, ''ZADCCMFB'' is a type of the sub-category of lateral buoys with a boat-shaped body which mark the port (left) side of the channel. The marks with code of ''ZAZVXCMFB'', ''YAGXCMFB'', ''SDGXGXFB'', ''GXJXFB'', ''ZVXWXSYFB'' and ''ZVXZTHFB'' all have a blue boat-shaped body, the differences are mainly in the topmark. The marks with code of ''ZUXZYFB'', ''ZUXTSFB'', ''ZUXJZPMFB'' and ''ZUXWXSYFB'' all have yellow cylindrical body, their differences are mainly in the marking.

B. RESNET-50 MODEL

The ResNet-50 network is adopted as the basic network of the proposed RMA model.

ResNet is a residual learning framework with the advantages of easy optimization and small computational burden. The residuals are designed to solve the problem of degradation and gradients, so that the performance of the network can be increased while increasing the depth. The ResNet network can not only greatly deepen the number of layers of the convolutional neural network, but also effectively solve the problem of increased training error caused by layer stacking.

In the original CNN architecture, the neural network input is x F(x), the output is $F(x)$, the target to be fitted is $H(x)$, and the training goal is to let $F(x) = H(x)$. However, in the ResNet architecture, the input is x, the output becomes $F(x)+x$, the target to be fitted becomes $H(x)-x$, and the training target is $F(x)+x = H(x)-x$. And the $H(x)-x$ is the so-called residual [21].

To make constant changes to the shallow networks, it is needed to train $F(x) = x$, so the target becomes to $F(x)+x = x$, which is equivalent to $F(x) = 0$. This is much simpler than the original training goal, because the parameter initialization in each layer of the networks is generally biased towards 0. Compared to updating the parameters of the network layer to learn $H(x) = x$, the update parameter of the redundant layer to learn $F(x) = 0$ can converge more quickly.

As showed in Figure 2, ResNet-50 contains 49 convolutional layers and one fully connected layer. The CONV is a convolution operation, BatchNorm is a batch regularization process, Relu is an activation function, MAXPOOL and AvgPOOL are two pooling operations. The second to fifth stages represent residual blocks, in which CONV BLOCK represents a residual block with added dimensions, ID BLOCK represents a residual block which does not change the dimension, and each residual block contains three convolutional layers. So there are 1+3*(3+4+6+3)=49 convolutional layers in the whole network.

After the continuous convolution operation of the residual block, the channel number of the image pixel matrix is deeper and deeper. After the flatten, the image pixel matrix size is changed to batch_size$\times$4096, and is finally input into the fully connected layer FC, and the corresponding class probability is output through the softmax layer.

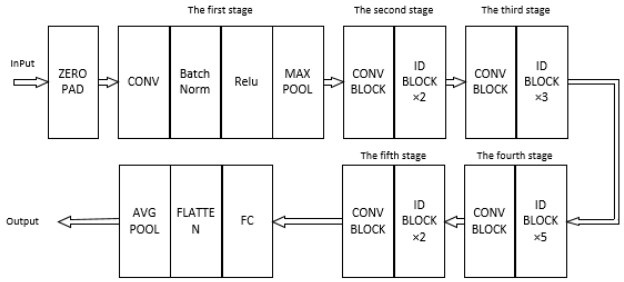


FIGURE 2. Structure diagram of ResNet-50.

C. RESNET-MULTISCALE-ATTENTION MODEL

Because some types of navigation marks are distinguished only by subtle differences, their visual recognition is a kind of fine-grained image classification in some extent. In addition, navigation marks usually appear as small objects in the images. Therefore, to achieved better classification result of navigation marks, this paper improves the ResNet-50 by adding a multiple scale attention mechanism.

It is well known that, in image classification, low-level features contain less semantic information, but more details and more accurate position of targets, conversely, the high-level features contain richer semantic information, but less details and relative rough position of targets. Normally, the general-level image classification networks including ResNet-50 use only the top-level features to do classification, so them performs not so good with the tasks of fine-gained level.

In this paper, to make up the details missing of the top-level features, a multiple scale attention mechanism is proposed. Features obtained from different stages of ResNet are integrated to form an attention matrix, which is used to notice the favorable region for classification by an element-wise multiplication with the input image. The enhanced images then are input to the second ResNet to finish final classification. The structure of the model called RMA (ResNet-Multiscale-Attention) is shown in Figure 3.

1) MULTISCALE ATTENTION MECHANISM

The first layer of the structure is designed to form an attention matrix to capture the notice regions. Its basic network is a ResNet-50 with parameters pre-trained on ImageNet. Feature maps with three channels, output from the second stage, the third stages and the last stages, noted as *f1*, *f2* and *f3* respectively, represent three different scale details from more to less. The features from same scale are convoluted to form a new feature, noted as *F1*, *F2* and *F3*. They are then upsampled to get $p1$, $p2$ and $p3$ with Equation (1).

$$O = s\,(i-1) + k - 2p \tag{1}$$

O is the size after upsampling, i is the input image size, s is the step size, k is the convolution kernel size, and p is the padding size.

In fact, the upsampling is a procedure of deconvolution. Furthermore, all upsampled features are pooled to a same size, and connected together as following formula to form

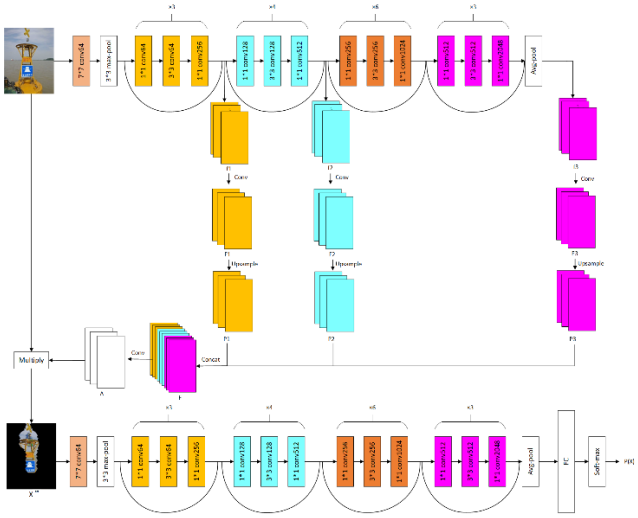


FIGURE 3. Structure of ResNet-Multiscale-Attention model.

fusion features F, which are hoped to be a multi-scale characteristic expression of the navigation mark.

$$F = concat[AVG\,(p1)\,, AVG(p2), AVG(p3)] \tag{2}$$

where, $AVG\,(\cdot)$ represents the average pooling operation and $concat\,(\cdot)$ represents the stitching operation.

In order to eliminate the aliasing effect of the upsampling, a convolution is then performed on F to form an attention matrix A with the same shape of input images.

$$A = conv\,(F) \tag{3}$$

Furthermore, a multi-scale attention enhanced input image X^{att} can be obtained by element-wise multiplication of the attention matrix A and the inputting image as following formula:

$$X^{att} = x \odot A \tag{4}$$

where $\odot$ represents the element-wise multiplication.

2) CLASSIFICATION

The second layer of the structure is the classification module which also based on ResNet-50 network. The input of the network is the X^{att}. The final output is the probability distribution $p\,(x)$ of different types of navigation marks, which is obtained from the following formula:

$$p\,(x) = f(W_c * X^{att}) \tag{5}$$

W_c is the parameters used in ResNet-50, $f\,(\cdot)$ represents the operations in the fully connected layer, where the convolution features are mapped onto feature vectors, which are further converted by softmax function into the final classification probability of the navigation mark type items.

3) TRAINING DETAILS

In order to improve the accuracy of the model and accelerate training process, the loss function and the optimizer are improved.

Methods such as image flipping, panning, and noise addition are used to enhance the dataset. However, the image number of different navigation types are still imbalanced. To deal with the imbalanced dataset, the loss function is improved to enable a type with fewer samples to contribute more to the loss function. The improved loss function is as follows:

$$loss = -\sum_i w \times y_i \times \log(logits_i) + (1 - y_i) \times \log(1 - logits_i) \quad (6)$$

where, the w, multiplied to the discrimination of the positive sample, is a coefficient calculated according to the dataset in advance. The less the number of samples, the larger the weight of that type which will therefore enlarge its contribution to the loss function.

In addiction, in order to make the model converge faster, the moment is added to the SGD optimizer. The improved optimizer function looks like following:

$$\begin{aligned} v &= M_u^* v - LR^* dx \\ x+ &= v \end{aligned} \quad (7)$$

For training the dataset of navigation marks, the variable v is set to 0, and the momentum parameter M_u is set to 0.9. With momentum, the change of parameters will be increased in any direction with a continuous gradient, therefore speed up the training process.

IV. EXPERIMENTS AND RESULTS

A. DATASET OF NAVIGATION MARKS

10260 images of 42 types of marks which installed right now in the downstream segment of the Yangtze river have been collected while regular maintenance. Figure 4 shows some images of the dataset of navigation marks. The labels of ''DCZYTHFB'', ''ZADCCMFB'', etc. are the codes of different types of navigation marks, which are used as the supervision information.

The images of each type are then divided into training dataset and test dataset with a ratio of 8:2, that is about 8208 images was used for model training, and about 2052 images was used to verify the top-1 accuracy.

B. TEST RESULTS

Experiments were carried out in the above data set, and the GoogleNet, VGG-16, VGG-19, AlexNet, SqueezeNet, and ResNet-50 were used to do comparison with the proposed RMA model. To check the effect of multiscale attention, the RMA model with two scales (output from the third stages and the last stages), named RMA-2, and the RMA model with three scales (output from the second stage, the third stages and the last stages), named RMA-3, were tested receptively.

The test results obtained from the validation set are shown in Table 1.

From the comparison, the RMA model proposed in this paper is proved can improve the accuracy of navigation marks classification. And the RMA model with three scales is a little better than that with two scales. That means the more details was noticed, the higher the classification accuracy.

FIGURE 4. Some images of the dataset of navigation marks.

TABLE 1. Test results.

No.	Model	Accuracy
1	GoogleNet	0.8881
2	VGG-16	0.8973
3	VGG-19	0.9035
4	AlexNet	0.9091
5	SqueezeNet	0.9208
6	ResNet-50	0.9414
7	RMA-2	0.9532
8	**RMA-3**	**0.9598**

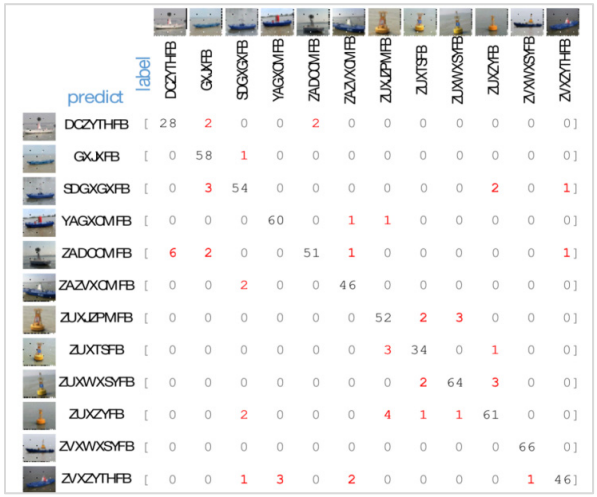
Figure 5 shows the confusion matrices of the misclassified images, the columns are the true types and the rows are the predicted types. The matrix includes 674 images belonging to twelve types from test dataset. The results show that the misclassification are mainly between types with slight difference such as ''DCZYTHFB'' and ''ZADCCMFB'', ''ZUXJZPMFB'' and ''ZUXTSFB'', ''GXJXFB'' and ''DCZYTHFB''. And the comparison between the Figure 5(a) and Figure 5(b) shows that RMA-3 decreases the total misclassified images from 54 to 41 with reduction in most types.

C. ANALYSIS WITH REGIONAL VISUALIZATION

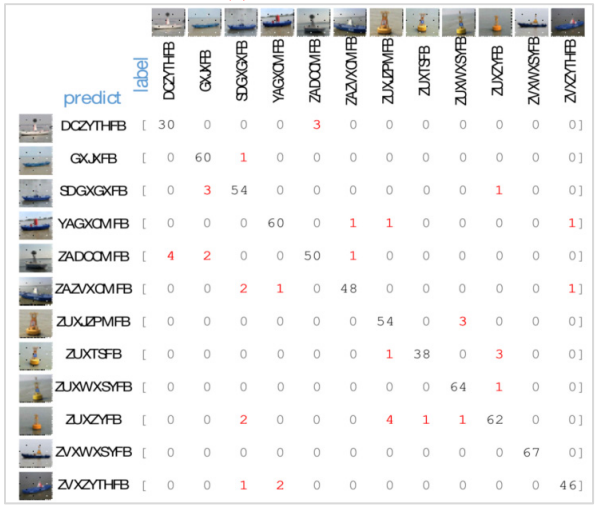
The purpose of the RMA model is to locate discriminative regions of the input image, and put more attention to these regions to identify the slight difference among some types of navigation marks, therefore improve the classification accuracy.

In order to verify the effect of the multiscale attention mechanism, the attention enhanced images X^{att} mentioned above are visualized. As shown in Figure 6, there are some samples of four types of navigation marks with similar appearance, labeled with ''ZUXJZPMFB'', ''ZUXTSFB'', ''ZUXWXSYFB'', and ''ZUXZYFB''. The image at right side of each sample is the related attention enhanced image, in which the pixels with low attention than a threshold are set to zero for increasing contrast.

From the visualization results, it can be found that the most important regions of marking plank, topmark, and middle structure that distinguish the four types were noticed exactly.



(a) ResNet-50



(b) RMA-3

FIGURE 5. Confusion matrix of the misclassified images.

FIGURE 6. Visualization of the attention enhanced images.

The visualization results verified that the RMA model is able to locate the attention regions and enhance the distinguishing characters, even in the case that navigation mark is a small object in the scene. And they also explain well why the RMA model can discriminate slight difference among similar navigation mark types, and achieve a better classification accuracy.

V. CONCLUSION AND FUTURE WORK

This paper exploits deep learning to research the recognition of navigation marks, and proposes a fine-grained classification model named RMA (ResNet-Multiscale-Attention) for navigation marks. Its attention mechanism is based on the fusion of feature maps with different scales which come from different stages of the ResNet. The RMA model does not need any additional supervision information except labels, and can to be trained end-to-end.

Experimental results on a dataset with 10260 navigation marks images show that the RMA has a accuracy about 96% to classify 42 types of navigation marks, it is better than ResNet-50 model with which the accuracy is about 94%.

Visualization results of attention enhanced images verified the ability of the RMA model of capturing the attention regions which is important to distinguish the slight difference among some similar types of navigation marks.

In the future work, in order to achieve higher classification accuracy, other good attention mechanisms for fine-grained image classification such as APN (Attention Proposal Network) will be investigated and be referred or combined to improve the RMA model. Furthermore, technologies of super resolution GAN (Generative adversarial network) will be studied and applied to identify navigation marks from a more far distance.

In addition, during the night, it is difficult to identify navigation marks by observing with their shapes from images. In this case, the features of the navigation mark's light are usually utilized to both identify it and ascertain its purpose. Therefore, capturing the light features including the colour and sign sequence from video, then recognizing the type of navigation marks will be another important research direction.

ACKNOWLEDGMENT

Thanks the Changjiang Nanjing Waterway Bureau of the People's Republic of China for providing the image dataset of navigation marks.

REFERENCES

[1] *Performance Standards for Electronic Chart and Display and Information System (ECDIS)*, document IMO Resolution A.817(19), 1995.
[2] *Revised Performance Standards for Electronic Chart and Display and Information System (ECDIS)*, document IMO Resolution MSC.232(82), 2006.
[3] G. Müller-Plath, "How does digital navigation on sailboats affect spatial abilities at sea?" *Int. J. Mar. Navigat. Saf. Sea Transp.*, vol. 13, no. 2, pp. 296–299, 2019.
[4] A. Garcia-Dominguez, "Mobile applications, cloud and bigdata on ships and shore stations for increased safety on marine traffic; a smart ship project," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Seville, Spain, Mar. 2015, pp. 1532–1537.
[5] Y.-L. Tang and N.-N. Shao, "Design and research of integrated information platform for smart ship," in *Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Banff, AB, Canada, Aug. 2017, pp. 37–41.
[6] J. Pandey and K. Hasegawa, "Autonomous navigation of catamaran surface vessel," in *Proc. IEEE Underwater Technol.(UT)*, Busan, South Korea, Feb. 2017, pp. 1–6.
[7] J.-Y. Zhuang, L. Zhang, S.-Q. Zhao, J. Cao, B. Wang, and H.-B. Sun, "Radar-based collision avoidance for unmanned surface vehicles," *China Ocean Eng.*, vol. 30, no. 6, pp. 867–883, Dec. 2016.

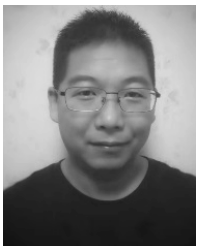
[8] I. Im, D. Shin, and J. Jeong, "Components for smart autonomous ship architecture based on intelligent information technology," in *Proc. Process. Int. Conf. Mobile Syst. Pervasive Comput.*, Gran Canaria, Spain, Jul. 2018, pp. 91–98.
[9] B. Liu, S. Z. Wang, J. S. Zhao, and M. F. Li, "Ship tracking and recognition based on Darknet network and YOLOv3 algorithm," *J. Comput. Appl.*, vol. 39, no. 6, pp. 1663–1668, 2019.
[10] H. Fu, Y. Li, Y. Wang, and P. Li, "Maritime ship targets recognition with deep learning," in *Proc. 37th Chin. Control Conf. (CCC)*, Wuhan, China, Jul. 2018, pp. 9297–9302.
[11] Q. Oliveau and H. Sahbi, "From transductive to inductive semi-supervised attributes for ship category recognition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Valencia, Spain, Jul. 2018, pp. 4827–4830.
[12] Q. Shi, W. Li, F. Zhang, W. Hu, X. Sun, and L. Gao, "Deep CNN with multi-scale rotation invariance features for ship classification," *IEEE Access*, vol. 6, pp. 38656–38668, 2018.
[13] Q. Fan, F. Chen, M. Cheng, S. Lou, R. Xiao, B. Zhang, C. Wang, and J. Li, "Ship detection using a fully convolutional network with compact polarimetric SAR images," *Remote Sens.*, vol. 11, no. 18, p. 2171, Sep. 2019, doi: 10.3390/rs11182171.
[14] M. Jiang, X. Yang, Z. Dong, S. Fang, and J. Meng, "Ship classification based on superstructure scattering features in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 616–620, May 2016.
[15] J. Li, C. Qu, and S. Peng, "Ship classification for unbalanced SAR dataset based on convolutional neural network," *J. Appl. Rem. Sens.*, vol. 12, no. 3, p. 1, Aug. 2018.
[16] S.-J. Lee, M.-I. Roh, H. Lee, J.-S. Ha, and I.-G. Woo, "Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks," in *Proc. Process. 28th Int. Ocean Polar Eng. Conf.*, Sapporo, Japan, Jun. 2018, pp. 726–730.
[17] *International Dictionary of Marine Aids to Navigation?* Accessed: Nov. 14, 2019. [Online]. Available: https://www.ialaaism.org/wiki/dictionary/index.php/Aid_to_Navigation
[18] X. He and Y. Peng, "Fine-grained image classification via combining vision and language," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 7332–7340.
[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 770–778.
[22] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.
[23] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, Apr. 2017.
[24] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.
[25] J. Mansanet, A. Albiol, and R. Paredes, "Local deep neural networks for gender recognition," *Pattern Recognit. Lett.*, vol. 70, pp. 80–86, Jan. 2016.
[26] Y. Liu, S. Zhou, and Q. Chen, "Discriminative deep belief networks for visual data classification," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2287–2296, Oct. 2011.
[27] H. Y. Shi, Y. B. Zhang, Z. Z. Zhan, N. Ma, X. B. Zhao, Y. Gao, and J. G. Sun, "Hypergraph-induced convolutional networks for visual classification," *IEEE Trans. Neural Netw. Learn.Systems*, vol. 30, no. 10, pp. 2963–2972, 2019.
[28] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based R-CNNs for fine-grained category detection," in *Proc. Eur. Conf. Comput. Vis.* (Lecture Notes in Computer Science), Zürich, Switzerland, vol. 8689, Sep. 2014, pp. 834–849.
[29] S. Branson, G. Van Horn, P. Perona, and S. Belongie, "Improved bird species recognition using pose normalized deep convolutional nets," in *Proc. Brit. Mach. Vis. Conf.*, Nottingham, U.K., Sep. 2014, pp. 1–14.
[30] X.-S. Wei, C.-W. Xie, and J. Wu, "Mask-CNN: Localizing parts and selecting descriptors for fine-grained image recognition," 2016, *arXiv:1605.06878*. [Online]. Available: http://arxiv.org/abs/1605.06878
[31] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 842–850.
[32] M. Simon and E. Rodner, "Neural activation constellations: Unsupervised part model discovery with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1143–1151.
[33] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1449–1457.
[34] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 2921–2929.
[35] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Jul. 2015, doi: 10.1155/2015/258619.
[36] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
[37] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," *IEEE Access*, vol. 6, pp. 36188–36197, 2018.
[38] L. Q. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Dec. 2017.
[39] P. M. Cheng and H. S. Malhi, "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images," *J. Digit. Imag.*, vol. 30, no. 2, pp. 234–243, Apr. 2017.
[40] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.
[41] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Visual attention-driven hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8065–8080, Oct. 2019.
[42] Y. Gu, M. Shen, J. Yang, and G.-Z. Yang, "Reliable label-efficient learning for biomedical image recognition," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2423–2432, Sep. 2019.
[43] H. C. Shin and K.-I. Lee, "Classification maritime vessel image utilizing a region of interest extracted and convolution neural network," *J. Korean Inst. Intell. Syst.*, vol. 29, no. 4, pp. 321–326, Aug. 2019.
[44] B. Solmaz, E. Gundogdu, V. Yucesoy, A. Koç, and A. A. Alatan, "Fine-grained recognition of maritime vessels and land vehicles by deep feature embedding," *IET Comput. Vis.*, vol. 12, no. 8, pp. 1121–1132, Dec. 2018.
[45] Y. Dong, H. Zhang, C. Wang, and Y. Wang, "Fine-grained ship classification based on deep residual learning for high-resolution SAR images," *Remote Sens. Lett.*, vol. 10, no. 11, pp. 1095–1104, Nov. 2019.
[46] F. Bousetouane and B. Morris, "Fast CNN surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios," in *Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Colorado Springs, CO, USA, Aug. 2016, pp. 242–248.

MINGYANG PAN received the Ph.D. degree in traffic information engineering and control from Dalian Maritime University, Dalian, China, in 2004. He is currently an Associate Professor with the Navigation College, Dalian Maritime University. He is also the Director of the Technical Institute of Navigation, Dalian Maritime University. His research activities are in the fields of electronic chart display and information systems (ECDIS), digital waterway systems, and intelligent waterway transportation systems.

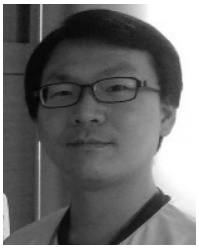
YISAI LIU received the bachelor's degree from Dalian Maritime University, in 2017. He is currently pursuing the master's degree with Dalian Maritime University. His research interests include data analysis and computer vision.

JIAYI CAO received the bachelor's degree from the Shenyang Institute of Engineering University, in 2018. He is currently pursuing the master's degree with Dalian Maritime University. His research interests include data augmentation and computer vision.

YU LI received the bachelor's degree from Southeast University, in 2005 and the master's degree in engineering, in 2014. Since 2005, he has been with the Changjiang Nanjing Waterway Bureau. His research mainly includes intelligent transportation and survey.



CHAO LI received the master's degree in traffic information engineering and control from Dalian Maritime University, Dalian, China, in 2007. He is currently a Senior Laboratory Technician with the Navigation College, Dalian Maritime University. His research interests focus on electronic chart display and information systems (ECDIS) and intelligent waterway transportation systems.



CHI-HUA CHEN (Senior Member, IEEE) is currently a Distinguished Professor with Fuzhou University and a Chair Professor with Dalian Maritime University, China. He has published over 270 academic articles and patents. His recent research interests include the Internet of Things, machine learning and deep learning, big data, cellular networks, and intelligent transportation systems.