A High-Precision Recognition Method of Circular Marks Based on CMNet Within Complex Scenes

Zhonghua Hong ⓘ, *Member, IEEE*, Zhupeng Li, Xiaohua Tong ⓘ, *Senior Member, IEEE*, Haiyan Pan, Ruyan Zhou, Yun Zhang, Yanling Han, Jing Wang, Shuhu Yang ⓘ, and Zhenling Ma ⓘ

Abstract—Accurate recognition of circular marks is crucial for calibration, object tracking, and three-dimensional reconstruction in videogrammetry. However, most existing studies were designed under single or relatively simple scenes. When the existing algorithms are applied to more complex scenarios, it will result in higher false detection and miss-detection rate. In this article, we present a high-precision recognition method based on a novel deep learning model, circular-MarkNet (CMNet) to solve this problem. The proposed network consists of three main steps: first, circular marks are detected using the improved YOLOv4 model to narrow the search region of the circular contour; the contour of the circular marks is then extracted based on the saliency object detection model BASNet; and finally, least square fitting is used to calculate the central pixel coordinate of the identified contour on the saliency map. The proposed method was tested under three complex scenarios with different characteristics and disturbances. The experimental results demonstrated that: the proposed CMNet can effectively recognize of circular marks within complex scenes, which reveals the superiority and generalization ability of the proposed method; the improved YOLOv4 can significantly enhance the detection accuracy of circular marks, which is crucial to the subsequent saliency courter detection and circle center identification; and CMNet achieved the best performance, with an RMSE of 0.0713 pixel, compared to the state-of-the-art methods.

Index Terms—Center positioning, circular mark, deep learning, high-speed videogrammetry, saliency detection.

I. INTRODUCTION

HIGH-SPEED videogrammetry is an efficient and low-cost engineering method that provides spatial information of objects by image acquisition and processing. Owing to its high-precision, noncontact, and nondamaging nature, it has been widely applied in civil engineering [1]–[3], environmental science [4], and industrial inspection [5]. For these

Zhonghua Hong, Zhupeng Li, Haiyan Pan, Ruyan Zhou, Yun Zhang, Yanling Han, Jing Wang, Shuhu Yang, and Zhenling Ma are with the College of Information Technology, Shanghai Ocean University, Shanghai 201306, China, and also with the Key Laboratory of Fisheries Information, Ministry of Agriculture, Shanghai Ocean University, Shanghai 201306, China (e-mail: zhhong@shou.edu.cn; m200711452@st.shou.edu.cn; hy-pan@shou.edu.cn; ryzhou@shou.edu.cn; y-zhang@shou.edu.cn; ylhan@shou.edu.cn; wangjing@shou.edu.cn; shyang@shou.edu.cn; zlma@shou.edu.cn).

Xiaohua Tong is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: xhtong@tongji.edu.cn).

applications, artificial marks posted on the measured object can help obtain spatial trajectory variation information and have yielded satisfactory precision. The most crucial factor is the automatic and high-precision recognition of artificial marks used to indicate points of interest. This is a significant pre-processing for subsequent high-precision camera calibration [6], displacement monitoring, and three-dimensional (3-D) reconstruction [7].

Several artificial marks have been widely used in videogrammetry. Owing to its exceptional properties of scale-, translation-, and rotation-invariance, the circular mark is more popular than others [8], [9]. Circular marks are divided into coded [10] and noncoded [11], [12]. Irrespective of type, traditional recognition methods consist of two steps: mark detection, which is used to narrow the search area of the circular contour; and center identification, which corresponds to the centroid positioning of the circle. Localization methods are used to identify the center of the circle in the detected area. In the past few decades, digital image processing has been used for localization, including Hough transform (HT)-based methods [13], centroid method [14]–[17], point-fitting methods [18]–[21].

The crucial issue focuses on the detection of the circular mark, which affects the precision of subsequent centroid positioning. Because coded information is readily identified, the coded mark is easily and automatically detected in the image. For non-coded marks, the detection relied on semiautomated and ellipse detection. Semi-automated may require the operator to box a search region for the locations of targets [2], [22], [23]. The ellipse detection methods [24], [25] used circle edge geometric features to complete the related detection tasks. In addition, some scholars have researched the extraction of marked regions. Guo et al. [26] used template matching to detect regions of interest containing artificial targets, while Ok [27] applied an region of interest (RoI) based method to focus on finding a specific circle in an area. However, traditional methods use shallow image information, including texture, edge, grayscale, etc., which are easily affected by the background and lighting, and indiscriminately misrecognize circular objects as circular marks in the image.

Deep learning methods, which have the advantage of detecting both shallow and deep features, have been widely used for image classification [28]–[30], target detection [31], [32] and image segmentation [33], [34]. For example, CSPDarknet53 combined cross stage partial with residual structure to extract different levels of features [35]. Path aggregation network (PANet) [36] added bottom-up path based on feature pyramid network (FPN)

[37] to better integrate shallow and deep features. Deep learning methods have been shown to be more effective than traditional methods. Therefore, they are the research focus for object detection tasks. In general, deep learning-based object detection can be divided into two categories: two-stage and one-stage methods. Region-based convolutional neural network (R-CNN) [38], fast R-CNN [39], and faster R-CNN [40] are the commonly used two-stage methods. They first generate the candidate object proposals and then the features from the proposals are extracted using a CNN. One-stage methods include SSD [41], YOLO series [42]–[45], and CenterNet [46]. They do not require the candidate object proposals, but directly regress the class scores and location coordinates of the object, so they have obvious advantages in efficiency. Zhou et al. [47] proposed a motion-blurred vision object recognition model based on a CNN. Shi and Zhang [48] used a faster R-CNN to locate and recognize the motion of a specially designed coded target in blurred images. Kinaz et al. [49] proposed a new deep CNN for the automatic detection and recognition of the coded target. However, these studies tested their detection algorithms only within laboratory conditions, utilizing a stable lighting environment. The performance of these methods in complex backgrounds remains to be explored.

Detected rectangular regions may accidentally contain interference information from the background, or noise, which will affect the center positioning of the circular mark. Salient object detection (SOD) methods can highlight the most prominent objects in an image and filter irrelevant interference information. Many studies have used SOD for auxiliary detection tasks. Han and Fu [50] proposed a saliency-based method to extract circular array objects from remote sensing images with high spatial resolution. Zhang et al. [51] used a saliency-guided sampling strategy to extract a representative set of patches from very high-resolution (VHR) images. Li et al. [52] fused a heat map with a saliency map to improve object detection performance.

From the above analysis, most existing methods are tested under synthetic and simple background environments. However, in actual high-speed videogrammetry, various complex environments exist, including indoor, outdoor, and circle-like interference. Precisely recognition of circular marks within complex scenes is still a challenging task due to the following reasons: The low light environment of indoor leads to blurred imaging and low contrast of the circular marks, which cannot be detected and located well. Likewise, overexposure leads to unclear outlines of some marks in outdoor scenes, making positioning more difficult. The background of the actual application is more complex, and there are many classes of circle-like interference, increasing the probability of false detection. These aforementioned factors increase the difficulty of mark detection, which result in bias and false detections by using existing algorithms. Therefore, it is necessary to develop a more robust algorithm. The objective of this article was to propose a high-precision circular mark recognition method based on circular-MarkNet (CMNet) for complex scenes. The proposed method adopted a coarse-to-fine strategy, and an object detection model is used to detect the circular mark in the image to narrow the search region of the circular contour in the coarse stage. During the implementation

of the fine strategy, we used a visual attention mechanism (VAM) to generate a saliency map of the detected rectangular region, focusing on the extraction of the circular contour. The central pixel coordinates of the identified contour are then calculated using the least squares fitting (LSF) method [53] on the saliency map. The main contributions of this article are as follows.

1) We propose a deep learning-based framework CMNet for recognition of circular marks within complex scenes. The network adopts a coarse-to-fine strategy and introduces a VAM based on a SOD network, BASNet. The false identification caused by background noise can be greatly reduced and sub-pixel level accuracy can be achieved.

2) We propose an improved YOLOv4 model for circular mark detection. The modified model used a large-scale feature map optimization structure and attention mechanism blocks (AMBs) between the neck and head to improve the accuracy of mark detection in the complex environment.

3) We generate a circular Mark recognition (CMR) dataset. The dataset contains three complex scenarios, such as indoor, outdoor, and circle-like interference. The experimental results reveal the superiority and generalization ability of the proposed method.

The rest of the article is organized as follows: Section II introduces the dataset. The details of the circular mark recognition methodology are described in Section III. Section IV presents and analyses a series of comparative experiments and Section V concludes the article.

II. DATASET

The nonretroreflective targets used in this article were a simple white circular mark on a black backing, which were labeled at critical points on the measured object. The circular mark images were collected at Tongji University, including experiments such as the collapse of civil structures, butt joints, experimental models of frame shakers, etc. Images were acquired using a CamRecord CL600×2 high-speed camera (Optronis, German, and 1280 × 1024 pixels image resolution) and a Basler ACA 2040–180 KM (Basler, German, and 2048 × 2048 pixels image resolution). To verify the effectiveness of the proposed method under various conditions, images were captured in three different scenarios: indoor, outdoor, and circle-like inference scenes. The camera was located 1.0–8.5 m from the object, and images were acquired under different illumination conditions, including high and low light conditions. In this article, 1095 images of circular marks, including different angles of circular marks on various scales, were collected as the experimental dataset.

Fig. 1 shows three scenarios of the CMR dataset. Indoor scenes usually have low light intensities, and image quality is significantly affected, making the edges of some circular marks unclear [see Fig. 1(a)]. In outdoor scenes, owing to strong illumination, some marks are overexposed, resulting in fuzzy edges of the circular marks [see Fig. 1(b)]. The image in Fig. 1(c) was taken under dark indoor conditions, and the background was complex with many circle-like inferences, including holes, bolts, light bulbs, etc. Table I gives a detailed description of the scenes.

TABLE I
SCENE DESCRIPTIONS

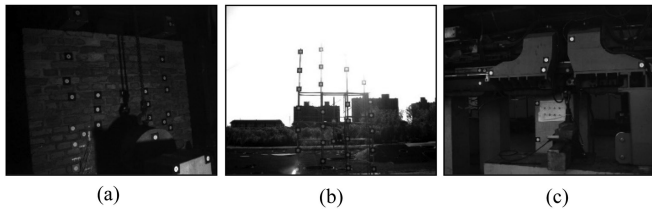| Scenes	Training set		Test set		Interference factor
	No. Images	No. Marks	No. Images	No. Marks	
Indoor	210	3382	70	1253	low light, unclear edges,
Outdoor	350	3418	117	1344	overexposed, small-sized
Circle-like	261	7003	87	2327	light bulbs, bolts, holes etc,
Total	821	13803	274	4924	different angles and multi-scale

Fig. 1. Three different scenes designed in the experiment. (a) Indoor. (b) Outdoor. (c) Circle-like background.

For the object detection model training, the CMR dataset was split into training and test sets in a ratio of 8:2. Subsequently, LabelImg software was used to label the circular marks in the image in the PASCAL VOC format. For saliency object detection model training, 2020 images of the circular mark regions were collected as the dataset. Labelme software was used to segment the circle contour.

III. METHODOLOGY

The flowchart of the proposed high-precision recognition method is shown in Fig. 2, which was constructed around the neural network CMNet. The method consists principally of three components, including circular mark detection, saliency contour extraction and circle center identification. The circular mark detection model was used to extract the region of the circular mark from the images based on the improved YOLOv4 (I-YOLOv4). The boundary-aware saliency detection network, BASNet [54], was used to generate the saliency map to focus on circular contour, while avoiding background noise interference, and the subpixel center coordinate was calculated on the saliency map of the circular contour by LSF.

A. Detection of Circular Mark Based on the Improved YOLOv4

The CMR dataset used in this article has more small-sized objects; thus, YOLOv4's [45] high-level detector head was unable to predict small-sized circular marks. In addition, interference from the complex background, including circle-like objects (light bulbs, bolts, holes, etc.), low light, and overexposed circle marks, caused errors using the original YOLOv4 model. Therefore, we made two adjustments to the original network structure. First, a larger-scale feature map optimization structure was employed on YOLOv4's neck and head to make the model robust with small marks. Second, AMBs were embedded to pay

more attention to the channel and spatial feature information to enhance the detection capability of circular marks in complex environments.

1) Feature Map Optimization: The original YOLOv4 has three output layers with down-sampling of 32, 16, and 8 times. Generally, the receptive field refers to the region that maps back to the input image. Thus, the larger receptive field is owned by the deeper network layer. Larger receptive field feature map is used to detect large marks. Low-level shallow feature maps retain more spatial information; therefore, it is more suitable for detecting minor circular marks. Consequently, it is necessary to design a new network structure with finer feature maps to detect small targets effectively. Fig. 4 shows the original structure and three newly designed structures with different redirected necks and heads. We tested different redirected necks and heads and found that the 4-in and 3-out structure worked best. Therefore, a larger-scale feature map optimization structure, with 4-in and 3-out, was adopted in this article.

The 16×16 feature map of the original YOLOv4 is responsible for detecting large objects, and feature maps with resolutions of 32×32 and 64×64 are responsible for detecting mid-sized and small targets, respectively. Because there are more small targets in the CMR dataset, we made corresponding improvements. In the neck network, we added four times down-sampling with the original three-scale feature map from the backbone for feature fusion. For the detector head, the 32 times down-sampling has too large a receptive field to regress the predicted circular-mark-sized targets. Therefore, this was deleted. Similar to the neck, we added 4 times down-sampling to detect minor circular marks. This redirected structure displays as 4-in and 3-out of the neck network. The sizes of the three-scale output layer of the detection network were changed from the original 16×16, 32×32, and 64×64 to 32×32, 64×64, and 128×128 to improve the detection accuracy of circular marks (see Fig. 3).

2) Attention Mechanism Block: The AMB consists of some particular convolution layers. It does not change the size of the feature map but can enhance target feature expression to increase detection ability. Therefore, an AMB can be easily inserted into the current object detection model. Fig. 3 shows the structure of the I-YOLOv4 model, where three AMBs are embedded after the three-scale output layer.

Given an intermediate feature map $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$ as input, AMB sequentially assigns a channel attention map $\mathbf{M_C} \in \mathbb{R}^{1 \times 1 \times C}$ and a spatial attention map $\mathbf{M_S} \in \mathbb{R}^{H \times W \times C}$. The channel and spatial attention mechanism process can be
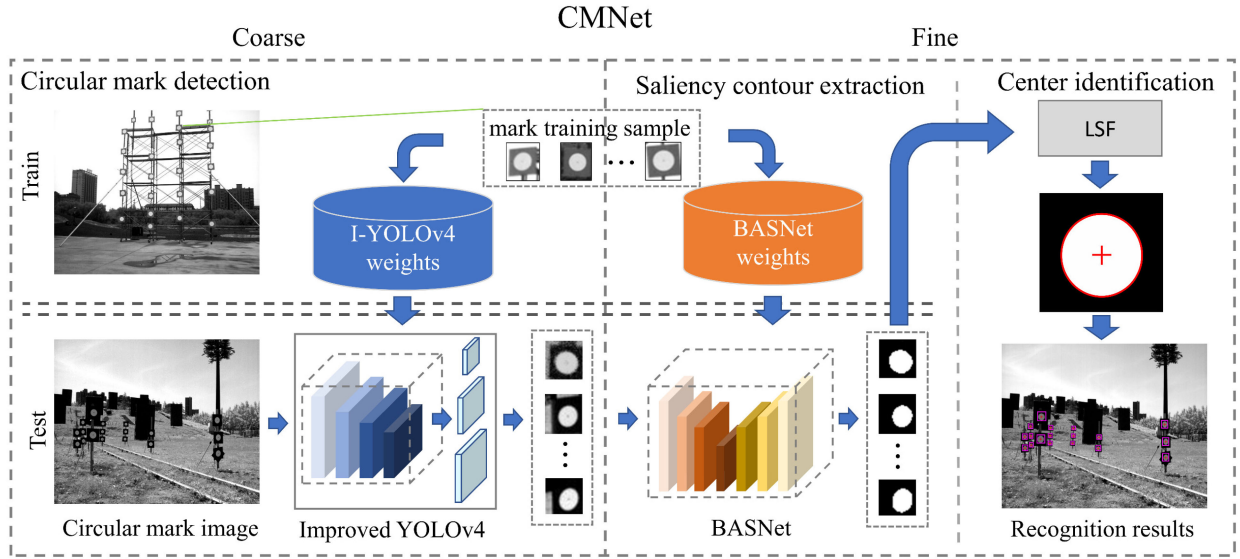
Fig. 2. Flowchart of CMNet, consisting of three components, including circular mark detection, saliency contour extraction and circle center identification. The center of each circular mark shown in the image has been marked with a red plus (+).
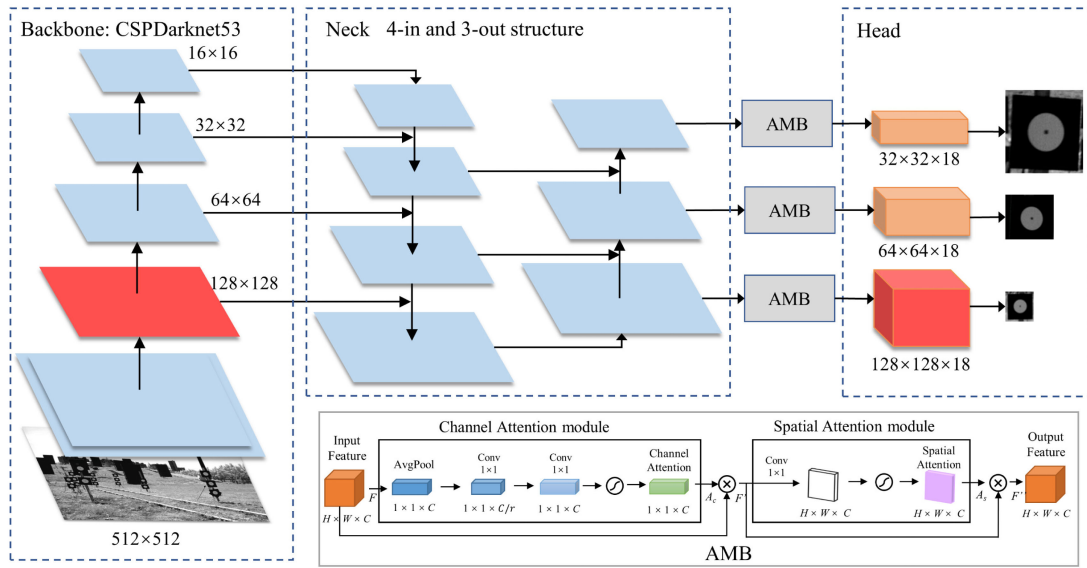

Fig. 3. Structure of the I-YOLOv4.

summarized as follows:

$$\mathbf{F}' = F \otimes \mathbf{M_C}(F)$$
$$\mathbf{F}'' = \mathbf{F}' \otimes \mathbf{M_S}(F') \tag{1}$$

where $\otimes$ represents the element-wise multiplication and $\mathbf{F}''$ denotes the final adjusted output. Fig. 3 shows the specific calculation process of the AMB attention map.

The channel attention module (CAM) in the AMB is arranged before the spatial attention module (SAM). The squeeze and excitation (SE) [55] block is a CAM that applies attention to objects from the perspective of channel features. It can suppress

background information and highlights foreground characteristics by adaptively re-weighting channel-wise features. In this article, the SE block was used as a CAM to decrease the error detection of circular marks. Specifically, the feature map of each channel was transformed into a matrix of size $1 \times 1 \times C$ by average pooling, and the channel attention $\mathbf{M_C}$ was obtained after two 1×1 convolutional layers. The calculation process for $\mathbf{M_C}$ can be expressed as follows:

$$\mathbf{M_C} = \sigma\left(C^{1\times1}_{r=1/16}\left(CM^{1\times1}_{r=16}\left(\text{AvgPool}(\mathbf{F})\right)\right)\right) \tag{2}$$

where $CM^{1\times1}_r$ denotes the convolution operation before the Mish activation function, with 1×1 representing the size of the
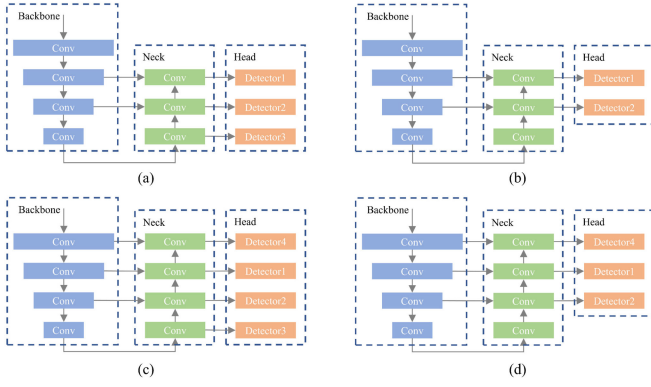
Fig. 4. Different detection layer structures. (a) *Structure 1*: Original YOLOv4 model. (b) *Structure 2*: Remove the 32 times down-sampling detector head. (c) *Structure 3*: A new detection scale is added to the original YOLOv4. (d) *Structure 4*: This structure is same as (c), except that the 32 times down-sampling detector head is removed. (a) Structure 1, 3-in, and 3-out. (b) Structure 2, 3-in, and 2-out. (c) Structure 3, 4-in, and 4-out. (d) Structure 4, 4-in, and 3-out.
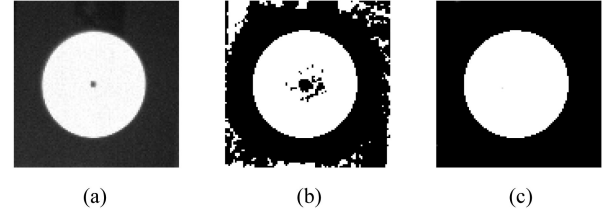

Fig. 5. Circular mark images. (a) Original circular mark region detected by the I-YOLOv4. (b) Binary image is obtained by adaptive local threshold method. (c) Saliency map of circular mark is generated by BASNet. (a) Circular mark region. (b) Binary image. (c) Saliency map.

convolution kernel, and r representing the reduction ratio. $C_r^{1\times1}$ is the convolution operation, which has the same superscript and subscript as $CM_r^{1\times1}$. σ denotes the sigmoid function.

SAM pays attention to objects at the spatial scale. Generally, the foreground occupies much fewer pixel than the background. Therefore, more attention should be paid to the foreground region. This article used a mask of the same size and depth as the input feature map to generate a spatial attention map $\mathbf{M_S}$. Specifically, the mask used to generate spatial attention is produced by 1×1 convolution layer. The calculation process for $\mathbf{M_S}$ can be expressed as follows:

$$\mathbf{M_S} = \sigma\left(CB_{r=1}^{1\times1}\left(\mathbf{F'}\right)\right) \quad (3)$$

where $\mathbf{F'}$ denotes the intermediate feature map of the CAM, $CB_r^{1\times1}$ represents the convolution operation before the batch normalization operation, and has the same superscript and subscript as $CM_r^{1\times1}$.

B. Extraction of Saliency Circular Mark Contour Based on BASNet

For identification of the circular mark center, the preprocessing consists of two steps, namely, image binarization, and edge detection from binary on the contour. Binarization is performed by converting grayscale image pixel to zero or one using the adaptive local threshold method [20], [56]. However, as shown in Fig. 5(b), the detected rectangular region usually contains background noise other than circular mark, which will subsequently affect the location of the center of the circular marks. In this article, we adopt a SOD model, BASNet, as a VAM to filter out the background noise while retaining the circular sub-pixel contour information.

The network structure of BASNet is shown in Fig. 6, which consists of a prediction module (PM) and a residual refinement module (RRM). The U-Net [57] structure is employed in the PM. The encoder extracts the feature map through the basic resblocks adopted from ResNet-34 [58]. Both encoder and decoder have

six levels. Each level feature map of the encoder is concatenated with the up-sampling output from the previous level and its corresponding level in the encoder. The output from the PM is a coarse map, in which the boundary of the mark is rough. The RRM then refines the saliency map of the PM by learning the residuals between the predicted saliency map and the ground truth. As same as the PM, it also has encoder and decoder phases. Unlike the PM, both the encoder and decoder have four levels. The final output is a refined saliency map that preserves the sub-pixel contour of the circular mark and removes background noise [see Fig. 5(c)].

C. Identification of Circular Mark Center Based on Least Squares

The circle edge points can be recognized quickly and accurately on the saliency map using Canny operator. In this article, the LSF is used to fit the center of the ellipse to achieve sub-pixel positioning. The general expression for the ellipse equation is as follows:

$$f\left(\alpha, X\right) = \alpha X = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (4)$$

where A, B, C, D, E, and F are the five elliptical parameters and α represents (A, B, C, D, E, F), and $X_i = (x_i^2, x_iy_i, y_i^2, x_i, y_i, 1)$. According to the principle of least squares, the curve-fitting problem can be solved by minimizing the sum of the squared algebraic distances

$$f\left(A, B, C, D, E, F\right)$$
$$= \sum_{i=1} \left(Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F\right)^2. \quad (5)$$

The five elliptic parameters of A, B, C, D, E and F in (5) can be obtained by calculating the first-order partial derivative and setting it to 0. If the center coordinate of the ellipse is $P(x_0, y_0)$, the calculation formula can be expressed as

$$\begin{cases} x_0 = \left(2BC - AD\right) / \left(A^2 - 4B\right) \\ y_0 = \left(2D - AD\right) / \left(A^2 - 4B\right). \end{cases} \quad (6)$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Design and Environment

The training process was carried out on an NVIDIA GeForce GTX 1080ti GPU with 12 GB of memory, an AMD Ryzen 7
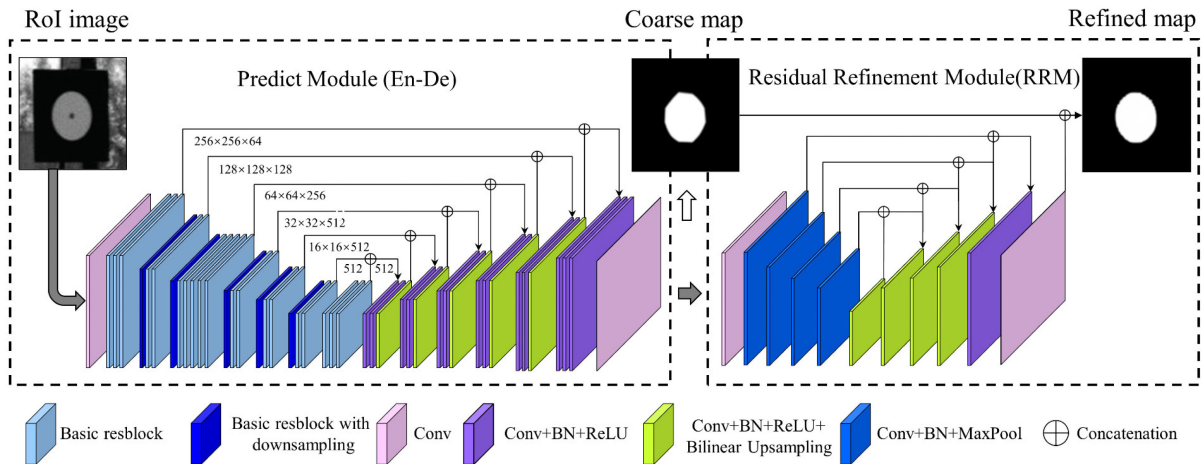
Fig. 6.   BASNet network structure.

2700 eight-core processor, and a memory size of 48 GB. The I-YOLOv4 algorithm was implemented in Darknet. During the training process, batch size was set to 64. The initial learning rate was set to 0.013, momentum was set to 0.949, and decay was set to 0.0005. BASNet was trained using the derived target binary images. During the training phase, images in the training set were resized to 256 × 256 pixels. The initial learning rate was set to 0.01, and batch size was set to 8. The training converges after 60 000 iterations.

B. Comparison of Circular Mark Detection Network

First, comparison experiments of the four structures based on the YOLOv4 network are introduced in this section. We then compared the performance of each strategy of the best structure with that of YOLOv4. Finally, the improved network model was compared with various detection models to verify the effectiveness of object detection in circular mark images. AP, precision, recall and detection efficiency (detection time / number of images) are evaluation metrics.

1) Structure: We compared the four structures of the neck and the head. Structure1 was the original YOLOv4 model, called 3-in and 3-out (see Fig. 4(a)]. In structure 2 [see Fig. 4(b)], we removed the 32 times down-sampling feature maps, and only used 16 and 8 times to detect circular marks. To use a high-resolution feature map to detect small targets, structure 3 [see Fig. 4(c)] extracts a 4 times down-sampling into the neck. Synchronously, we added detector 4 for receiving the same size feature map from the neck. Structure 4 [see Fig. 4(d)] was similar to structure3, which was 4-in, but only used three higher-resolution feature maps. Comparison experiments were conducted in the same environment. The detection accuracies of these four structures are given in Table II.

The four different structures obtained feature maps of different scales from the same backbone network (CSPDarknet53) for feature fusion. As given in Table II, structure4 yielded the best result, with an AP of 97.22%. The accuracy of structures 2, 3, and 4 were improved by 2.26%, 0.57%, and 2.67%,

TABLE II
DETECTION ACCURACY FOR FOUR NECK-HEAD STRUCTURES

Neck-head structure	AP (%)	Precision (%)	Recall (%)
Structure 1	94.96	82.90	**95.99**
Structure 2	96.65	96.71	92.89
Structure 3	94.55	88.83	94.51
Structure 4	**97.22**	**97.23**	94.62

TABLE III
ABLATION EXPERIMENT FOR MARK DETECTION

Model	Structure4	SAM	CAM	AP (%)
YOLOv4				94.96
	✓			97.22
	✓	✓		97.67
	✓		✓	98.25
	✓	✓	✓	**98.60**

respectively, compared to structure 1. The reason for this improvement was that the high-resolution feature map added by structure4 contained richer spatial information. After fusion with the feature maps of the other three scales, circular marks were detected under various conditions. Structure1 and structure 3 both had a low-resolution detection layer. Owing to the lack of sufficient spatial information, it does not perform well on detection of small circular marks, leading to precisions of only 94.96% and 94.55%, respectively. Because structure 2 had two detection layers and lacked the ability to detect small targets, recall was the lowest (92.89%) of the four structures (see Table II).

2) Ablation: Table III gives comparisons among each strategy and the best neck-head structure (YOLOv4 with structure4) before adding. Compared with YOLOv4 with structure 4, the accuracy is improved 0.45% and 1.03% by adding SAM and CAM, respectively. After inserting AMB (CAM + SAM), the detection accuracy was improved by 1.38%. This indicated
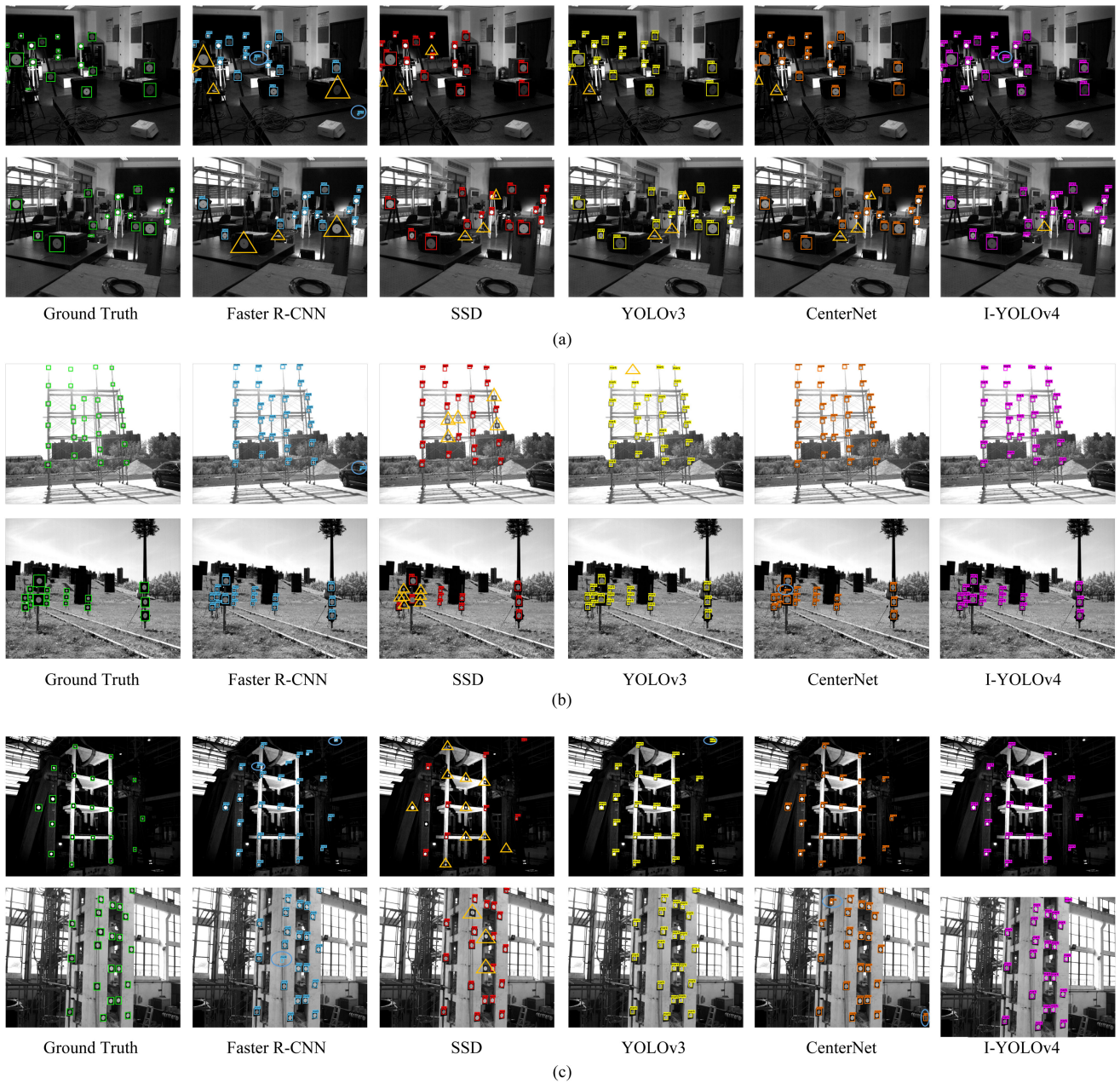
Fig. 7. Comparison of circular mark detection results for different deep learning object detection models within complex environments. The detection confidence threshold is uniformly set to 0.5. The green color represents the ground truth of the circular mark. The blue, red, yellow, orange, and purple color rectangles indicate corresponding detection results for Faster R-CNN, SSD, YOLOv3, CenterNet, and the I-YOLOv4. Blue circles and yellow triangles denote the false alarms and missing marks, respectively. (a) Indoor scenes. (b) Outdoor scenes. (c) Circle-like scenes.

that the channel attention and spatial attention mechanisms introduced more semantic and spatial information. In summary, the I-YOLOv4 improved accuracy by 3.64% compared to the original YOLOv4.

3) Performance Comparison With Other Detection Models: A qualitative and quantitative representation was provided for the Faster R-CNN [40], SSD [41], YOLOv3 [44], CenterNet [46] and the I-YOLOv4.

Fig. 7 shows the validation of the mark detection results for different deep learning detection models under complex scenes, including indoor and outdoor scenes and circle-like background scenes. Faster R-CNN and SSD did not reliably detect multiscale marks in the images, missing large targets in the vicinity. Faster R-CNN, SSD, CenterNet, and YOLOv3 missed detections of occluded and oblique marks. Faster R-CNN and the I-YOLOv4 incorrectly detected the background as a mark. For the outdoor scenes in Fig. 7, the detection results of most models were good, except for SSD. Some problems were apparent with YOLOv3 and CenterNet. YOLOv3 missed the overexposed and occluded marks, and CenterNet incorrectly detected circle-like cables as marks. Faster R-CNN, SSD, CenterNet, and YOLOv3 detected complex backgrounds as marks, including

TABLE IV
CIRCULAR MARK DETECTION FOR DIFFERENT OBJECT DETECTION MODELS

Model	AP (%)	Precision (%)	Recall (%)	Detection efficiency (ms)
Faster R-CNN [40]	90.68	92.23	87.95	90.9
SSD [41]	92.46	94.56	46.23	**18.5**
YOLOv3 [44]	93.99	94.92	92.81	69.7
CenterNet [46]	96.64	96.41	93.29	54.9
I-YOLOv4	**98.60**	**96.91**	**97.25**	23.1

light bulbs and pipes [see Fig. 7(c)]. In contrast, the I-YOLOv4 exhibited good detection performance under our experimental conditions.

Table IV gives the comparison of the AP, precision, recall, and detection efficiency for the five models employed in this article. Our improved method achieved the highest AP of 98.60% and the I-YOLOv4, which used a new neck-head structure (4-in and 3-out), exhibited excellent feature abstraction and feature fusion capabilities to deal with the detection of small circular marks (see Table IV). The modified YOLOv4 with AMB accurately detected circular marks in complex scenarios. CenterNet exhibited good detection accuracy (AP of 96.64%), but detection required 54.9ms per image, and recall was 3.96% lower than that of the I-YOLOv4. CenterNet employs keypoint estimation to find the center point and regresses it to other attributes. The centers of dense circular marks overlap after feature map down-sampling, resulting in CenterNet failing to detect dense and occluded circular marks in images. YOLOv3 uses FPN for feature fusion and the backbone network for feature extraction, whose feature abstraction ability is weaker than that of PANet [36] and CSPDarknet53. In addition, the structure of YOLOv3 is similar to that of the original YOLOv4. Thus, the AP was 4.16% lower than that of the I-YOLOv4. Faster R-CNN and SSD use only a single feature layer for object prediction, which cannot cope with the detection of multiscale scenes in the CMR dataset. Furthermore, because there is no feature fusion of multiple hierarchical feature maps, low-level features lack sufficient semantic information, and high-level features lack sufficient spatial information, making it difficult to accurately detect and locate circular marks in some complex backgrounds.

In addition, we also compared the receiver operation characteristics (ROC) of different models. ROC is an important metric that can be used to evaluate the detection effect under the same false positive [59], [60]. ROC The higher the true positive rate (TPR), the better the detection effect. ROC curves are acquired using Monte-Carlo simulations [61], [62], which is done on the CMR dataset. As given in Table IV and Fig. 8, the recall of SSD is only 46.23%, but the precision is 92.46%. SSD has both low FP and TP, causing the ROC curve to be close to the x-axis. CenterNet and YOLOv3 exhibit similar performance on ROC curves. The I-YOLOv4 works best by having lower number of FP and higher TPR.
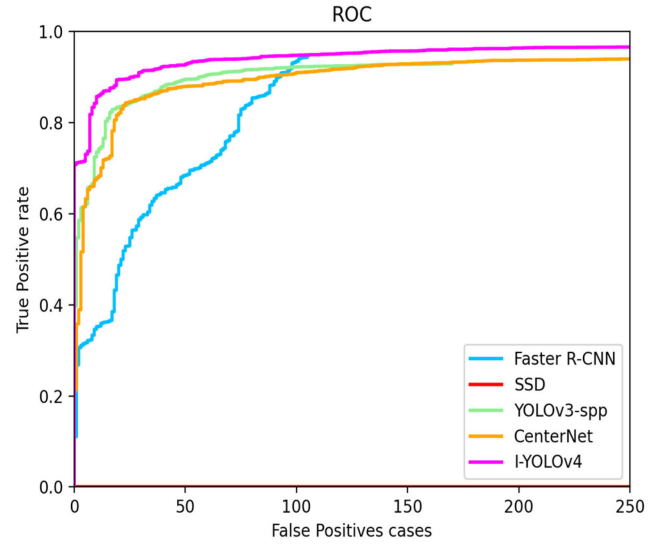

Fig. 8. ROC curves of different detection models in the CMR dataset. The blue solid, the red solid, the yellow solid, the orange solid, the purple solid are the ROC curves of faster R-CNN, SSD, YOLOv3, CenterNet, and I-YOLOv4 with a threshold of 0.5, respectively.

C. Comparison of Center Identification

1) Evaluation Metrics: We used high-precision total station (SOKKIA NET05AX) to obtain the 3-D coordinates of the center. According to the camera calibration parameters and exterior orientation elements, the collinear equation is used to obtain the image pixel coordinates of the circular mark centers as the ground truth values. On the one hand, we use mean absolute error (MAE) and root-mean-square error (RMSE) to measure the accuracy of mark center localization. On the other hand, precision, recall and F-measure are used to verify algorithms' ability to recognize circular marks in real images, and they are defined as

$$\text{Precision} = \frac{\text{TPs}}{\text{TPs} + \text{FPs}} \qquad (7)$$

$$\text{Recall} = \frac{\text{TPs}}{\text{TPs} + \text{FNs}} \qquad (8)$$

$$\text{F} - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (9)$$

TABLE V
RECOGNITION RESULTS

Methods	Precision (%)	Recall (%)	F-measure (%)
Arc-Support [24]	83.30	93.21	87.98
AAMED [25]	86.28	95.86	90.81
Proposed without VAM	79.69	95.07	86.70
Proposed	**98.77**	**99.37**	**99.07**

where TPs, FPs, and FNs are numbers of true positives, false positives, and false negatives, respectively. We define points with RMSE less than 0.5 pixel as TP.

2) Compared Methods: We compared the performance of the proposed method with three state-of-the-art methods, which are the centroid search algorithm [17], Arc-Support [24] and arc adjacency matrix-based ellipse detector (AAMED) [25]. The centroid search algorithm integrated into the PhotoModeler Scanner software can achieve high-precision positioning of marks, which is currently widely used in videogrammetry [2], [22], [23]. Arc-Support utilized rich geometric features and arc-support line segments to complete the ellipse detection tasks. AAMED detected ellipse robustly by constructed a digraph-based arc adjacency matrix (AAM) for arc pairing. In addition, the role of the VAM is also explored by comparing with our method without the VAM.

3) Accuracy Comparison of Center Identification: First, we compare mark recognition rate of aforementioned methods. Since centroid search algorithm requires manual selection of marks, we did not include it in the recognition comparison experiment. Then, we compare the accuracy of methods for locating the center of the circle.

As given in the Table V, the recall of Arc-Support and AAMED are 93.21% and 95.86%, respectively. Both methods misrecognize many circle-like objects, such as light bulbs, pipes, auto wheels, etc., and the precision is only 83.30% and 86.28%, respectively. The ellipse detection methods cannot distinguish between circular marks and general circle-like objects. Due to the background noise interference after binarization, the precision of our method without VAM is only 79.69%. From point10 of Fig. 10(a), point12 of Fig. 10(b), and point14 of Fig. 10(c), it can be seen that the proposed method without VAM has deviations and false alarms in the positioning of the circular mark. After introducing the VAM, our proposed method can effectively recognize most of the circular marks in the image, and the precision and recall reach 98.77% and 99.37%, respectively. This proves that VAM has improved the recognition rate of the circular marks.

Both centroid search and our proposed method can identify the center of the circular mark very well (see Fig. 10). In scene 1, arc-support, AAMED and the proposed method without VAM cannot identify the overexposed circular marks from the location of the yellow triangle. In scene 2, from the position of the blue circle, the white background was recognized as the center of the circle without VAM. In scene 3, the proposed method was not
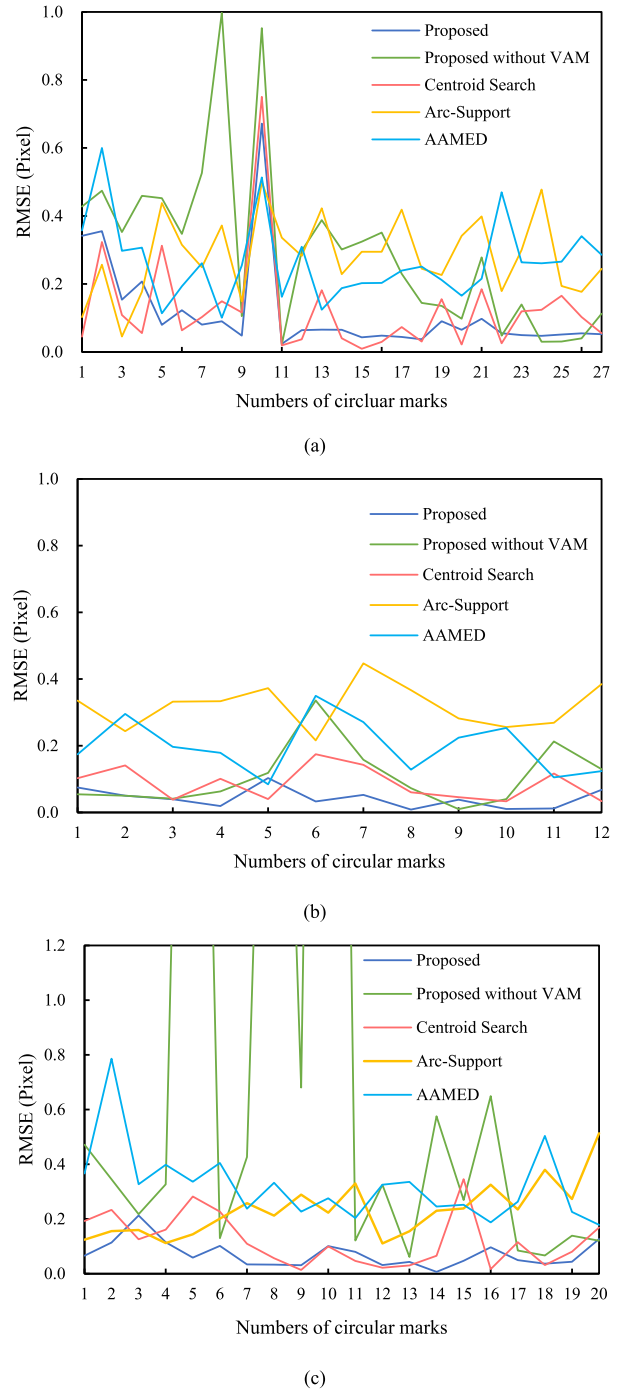

Fig. 9. Comparison of the RMSE results. (a) Scene 1. (b) Scene 2. (c) Scene 3.

influenced by the light bulb. However, owing to the absence of VAM and the small size of the circular mark, there was a large deviation in the identification of the method without VAM, as revealed by the location of the blue circle (see Fig. 10). Arc-support and AAMED missed some points due to the smaller size of the circular mark and noise in Fig. 10 scene 3.

Fig. 9 and Table VI show that the proposed method achieved the best result, with a mean RMSE of 0.0796 pixel. Without
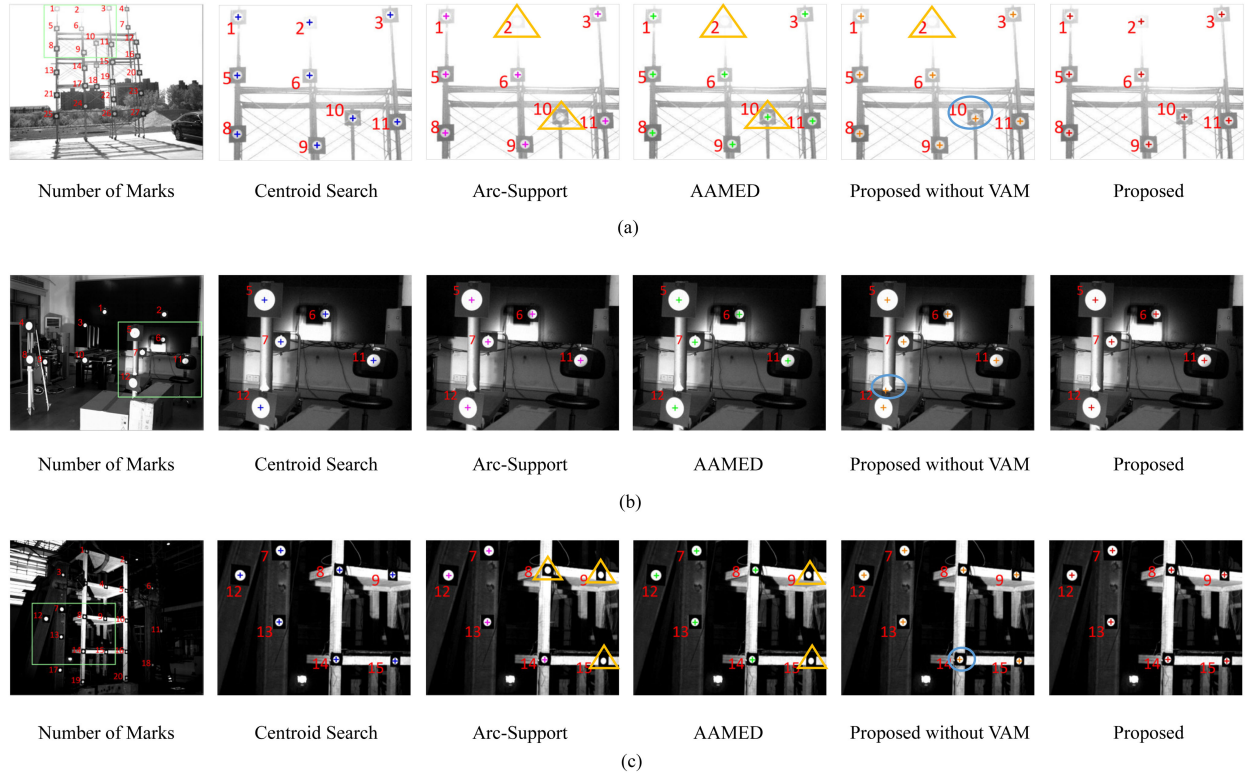
Fig. 10. Comparison of the identification results of the circle mark center for three recognition methods under different scenarios. The three columns of images on the right show the identification results in the green box area in the images in the first column. The blue, purple, green, orange and red "+" indicate corresponding identification results for Centroid Search, Arc-Support, AAMED, Proposed without VAM and our proposed method. Blue circles and yellow triangles represent the false alarms and missing marks, respectively. (a) Scene 1, outdoor scenes. (b) Scene 2, indoor scenes. (c) Scene 3, complex background scene.

TABLE VI
MEAN OF MAE AND RMSE RESULTS

Method	Scene1		Scene2		Scene3		Mean of RMSE
	MAE	RMSE	MAE	RMSE	MAE	RMSE	
Centroid Search [17]	**0.1144**	**0.1150**	0.0858	0.0858	0.1167	0.1211	0.1073
Arc-Support [24]	0.2836	0.2839	0.3198	0.3196	0.2328	0.2334	0.2790
AAMED [25]	0.2650	0.2691	0.1987	0.1992	0.3205	0.3346	0.2676
Without VAM	0.2867	0.2954	0.1062	0.1072	1.0011	1.0188	0.4738
Proposed	0.1208	0. 1250	**0.0453**	**0.0425**	**0.0675**	**0.0713**	**0.0796**

VAM, there were many false and error identifications at some points [such as point7 and point10 in Fig. 10(a) and point 5, point14 in Fig. 10(c)]. After adding the VAM, the RMSE of the proposed method was reduced by 0.3942 pixel. In particular, in a complex background scene (scene 3), the RMSE was reduced from 1.0188 pixel to 0.0713 pixel. The RMSE of Arc-Support and AAMED is 0.2790 pixel and 0.2676 pixel, respectively. However, these methods cannot directly identify circular marks from the entire image, and need to use a sliding window to traverse the image, which reduces the efficiency of detection. Our proposed method was comparable to centroid search method, which manual selection is required. In summary, the proposed method achieved the best performance under all three complex scenarios.

4) Efficiency Comparison of Different Methods: As given in Table VII, the centroid search algorithm takes an average of 125.20 s per image to identify all circular marks. This is because the centroid search algorithm requires manual indication of search area, leading to low recognition efficiency. The recognition time of arc-support and AAMED is 2.29 and 3.91 s, respectively. I-YOLOv4 is a one-stage detector that can quickly extract and fuse deep and shallow features, making the recognition time of CMNet without VAM only 0.89 s. VAM performs boundary-aware detection on each circular mark region in the image, making the salient contour extraction stage more time-consuming. So, adding the VAM resulted in an increase in detection time per image of 3.36 s, taking 4.25 s to recognize.

TABLE VII
TIME COST COMPARISON ON THE CMD DATASET

Methods	Recognition times (s)
Centroid Search	125.20
Arc-Support	2.29
AMMED	3.91
Proposed without VAM	**0.89**
Proposed	4.25

V. CONCLUSION

In this article, a high-precision recognition method based on a novel deep learning model, CMNet, is presented. First, circular marks were detected based on the I-YOLOv4 model to narrow the search region of the circular contour. The contours of the circular marks were then extracted based on the saliency object detection model BASNet, and LSF is used to calculate the central pixel coordinate of the identified contour on the saliency map. Two major improvements were made based on the original YOLOv4 model. The first improvement was a large-scale feature map optimization structure, which was displayed as a 4-in and 3-out structure. The second modification was to insert AMBs between the neck and head to improve the accuracy of mark detection in a complex environment.

Three complex scenes with different characteristics and disturbances were used to evaluate the effectiveness and robustness of the method. These three scenes were indoor with low light, outdoor with extremely strong light and over-exposure, and a densely scene with multiple circle-like objects. The experimental results demonstrated the following.

1) The improved YOLOv4 significantly enhanced the detection accuracy of circular marks, which is crucial to the subsequent saliency courter detection and circle center positioning.
2) CMNet achieved the best performance with an RMSE of 0.0713 pixel, comparable to the precision of the commercial software PhotoModeler.
3) The precision, recall, and F-measure of center identification of our proposed method is 98.77%, 99.37%, and 99.07%, respectively.

The results indicate that our method exhibited a good ability for circular mark recognition in complex environments. In the future, we will try to propose an end-to-end circular mark recognition network with shared weights to realize the real-time circle center identification.

REFERENCES

[1] X. Liu et al., "High-speed videogrammetric measurement of the deformation of shaking table multi-layer structures," *Measurement,* vol. 154, 2020, Art. no. 107486.

[2] X. Liu, X. Tong, X. Yin, X. Gu, and Z. Ye, "Videogrammetric technique for three-dimensional structural progressive collapse measurement," *Measurement,* vol. 63, pp. 87–99, 2015.

[3] X. Tong et al., "Monitoring a progressive collapse test of a spherical lattice shell using high-speed videogrammetry," *Photogramm. Rec.,* vol. 32, no. 159, pp. 230–254, 2017.

[4] J. Vallet, B. Turnbull, S. Joly, and F. Dufour, "Observations on powder snow avalanches using videogrammetry," *Cold Regions Sci. Technol.,* vol. 39, no. 2/3, pp. 153–159, 2004.

[5] D. Karayel, M. Wieschoff, A. Özmerzi, and J. Müller, "Laboratory measurement of seed drill seed spacing and velocity of fall of seeds using high-speed camera system," *Comput. Electron. Agriculture,* vol. 50, no. 2, pp. 89–96, 2006.

[6] D. D. Lichti, D. Jarron, W. Tredoux, M. Shahbazi, and R. Radovanovic, "Geometric modelling and calibration of a spherical camera imaging system," *Photogramm. Rec.,* vol. 35, no. 170, pp. 123–142, 2020.

[7] S. Y. Lin, J. P. Mills, and P. D. Gosling, "Videogrammetric monitoring of as-built membrane roof structures," *Photogramm. Rec.,* vol. 23, no. 122, pp. 128–147, 2008.

[8] P. Rachakonda, B. Muralikrishnan, and D. Sawyer, "Metrological evaluation of contrast target center algorithm for terrestrial laser scanners," *Measurement,* vol. 134, pp. 15–24, 2019.

[9] Q. Yu et al., "A displacement-relay videometric method for surface subsidence surveillance in unstable areas," *Sci. China Technol. Sci.,* vol. 58, no. 6, pp. 1105–1111, 2015.

[10] M. R. Shortis and J. W. Seager, "A practical target recognition system for close range photogrammetry," *Photogramm. Rec.,* vol. 29, no. 147, pp. 337–355, 2014.

[11] R. Maalek and D. D. Lichti, "Automated calibration of smartphone cameras for 3D reconstruction of mechanical pipes," *Photogramm. Rec.,* vol. 36, no. 174, pp. 124–146, 2021.

[12] R. Maalek, D. D. Lichti, and S. Maalek, "Towards automatic digital documentation and progress reporting of mechanical construction pipes using smartphones," *Automat. Construction,* vol. 127, 2021, Art. no. 103735.

[13] M. R. Shortis, J. W. Seager, S. Robson, and E. S. Harvey, "Automatic recognition of coded targets based on a hough transform and segment matching," in *Videometrics VII*, vol. 5013, Bellingham, WA, USA: Int. Soc. Opt. Photon., 2003, pp. 202–208.

[14] Z.-J. Yu, C. Gang, R.-S. Che, C.-Y. Liu, and T. Wei, "Bilinear interpolation centroid algorithm used for circular optical target location," in *Proc. 2nd Int. Conf. Image Graph.*, 2002, vol. 4875, pp. 333–339.

[15] S. Dong, J. Ma, Z. Su, and C. Li, "Robust circular marker localization under non-uniform illuminations based on homomorphic filtering," *Measurement,* vol. 170, 2021, Art. no. 108700.

[16] Y. Li, H. Li, X. Gan, J. Qu, and X. Ma, "Location of circular retro-reflective target based on micro-vision," in *Proc. Int. Conf. Opt. Instrum. Technol., Optoelectron. Meas. Technol. Syst.*, 2020, vol. 11439, Paper 114390S.

[17] L. M. Galantucci, F. Lavecchia, G. Percoco, and S. Raspatelli, "New method to calibrate and validate a high-resolution 3D scanner, based on photogrammetry," *Precis. Eng.,* vol. 38, no. 2, pp. 279–291, 2014.

[18] S. Zheng, P. Chen, S. Liu, X. Ma, S. Gaol, and X. Tong, "A high-precision elliptical target identification method for image sequences," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 3374–3377.

[19] Y. Liu, X. Su, X. Guo, T. Suo, and Q. Yu, "A novel concentric circular coded target, and its positioning and identifying method for vision measurement under challenging conditions," *Sensors,* vol. 21, no. 3, 2021, Art. no. 855.

[20] R. Maalek and D. D. Lichti, "Robust detection of non-overlapping ellipses from points with applications to circular target extraction in images and cylinder detection in point clouds," *ISPRS J. Photogramm. Remote Sens.,* vol. 176, pp. 83–108, 2021.

[21] H. Li, "Multiple ellipse fitting of densely connected contours," *Inf. Sci.,* vol. 502, pp. 330–345, 2019.

[22] J. Leifer, B. Weems, S. C. Kienle, and A. M. Sims, "Three-dimensional acceleration measurement using videogrammetry tracking data," *Exp. Mechanics,* vol. 51, no. 2, pp. 199–217, 2011.

[23] X. Tong et al., "Acceleration of object tracking in high-speed videogrammetry using a parallel OpenMP and SIMD strategy," *Photogramm. Rec.,* vol. 34, no. 166, pp. 174–197, 2019.

[24] C. Lu, S. Xia, M. Shao, and Y. Fu, "Arc-support line segments revisited: An efficient high-quality ellipse detection," *IEEE Trans. Image Process.,* vol. 29, pp. 768–781, Aug. 2019.

[25] C. Meng, Z. Li, X. Bai, and F. Zhou, "Arc adjacency matrix-based fast ellipse detection," *IEEE Trans. Image Process.,* vol. 29, pp. 4406–4420, Jan. 2020.

[26] X. Guo, Y. Chen, C. Wang, M. Cheng, C. Wen, and J. Yu, "Automatic shape-based target extraction for close-range photogrammetry," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 583–587, 2016.

[27] A. O. Ok, "A new approach for the extraction of aboveground circular structures from near-nadir VHR satellite imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 6, pp. 3125–3140, Jun. 2014.

[28] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 5, pp. 4340–4354, May 2021.

[29] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[30] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–15, Nov. 2021, Art. no. 5518615.

[31] Z. Hong et al., "Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 14, pp. 6083–6101, Jun. 2021.

[32] G. Cheng, Y. Si, H. Hong, X. Yao, and L. Guo, "Cross-scale feature fusion for object detection in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.,* vol. 18, no. 3, pp. 431–435, Mar. 2021.

[33] F. Mu, J. Li, N. Shen, S. Huang, Y. Pan, and T. Xu, "Pixel-Adaptive Field-of-View for remote sensing image segmentation," *IEEE Geosci. Remote Sens. Lett.,* vol. 19, pp. 1–5, Jun. 2022, Art. no. 6512705.

[34] X. Zhang, J. Li, and Z. Hua, "MRSE-Net: Multi-scale residuals and SE-Attention network for water body segmentation from satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 15, pp. 5049–5064, Jun. 2022.

[35] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops,* 2020, pp. 390–391.

[36] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2018, pp. 8759–8768.

[37] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2017, pp. 2117–2125.

[38] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2014, pp. 580–587.

[39] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.,* 2015, pp. 1440–1448.

[40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.,* vol. 28, pp. 91–99, 2015.

[41] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.,* 2016, pp. 21–37.

[42] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2016, pp. 779–788.

[43] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2017, pp. 7263–7271.

[44] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018. [Online]. Available: https://arxiv.org/abs/1804.02767v1

[45] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020. [Online]. Available: https://arxiv.org/abs/2004.10934

[46] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019. [Online]. Available: http://arxiv.org/abs/1904.07850

[47] H. Zhou, M. Chen, L. Zhang, N. Ye, and C. Tao, "Measuring shape and motion of a high-speed object with designed features from motion blurred images," *Measurement,* vol. 145, pp. 559–567, 2019.

[48] Y. Shi and L. Zhang, "Design of Chinese character coded targets for feature point recognition under motion-blur effect," *IEEE Access,* vol. 8, pp. 124467–124475, 2020.

[49] V. Kniaz, L. Grodzitskiy, and V. Knyaz, "deep learning for coded target detection," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.,* vol. XLIV-2/W1, pp. 125–130, 2021.

[50] X. Han and Y. Fu, "Circular array targets detection from remote sensing images based on saliency detection," *Opt. Eng.,* vol. 51, no. 2, 2012, Art. no. 026201.

[51] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 4, pp. 2175–2184, Apr. 2015.

[52] M. Li, X. Zhao, J. Li, and L. Nan, "ComNet: Combinational neural network for object detection in UAV-Borne thermal images," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 8, pp. 6662–6673, Aug. 2021.

[53] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 21, no. 5, pp. 476–480, May 1999.

[54] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.,* 2019, pp. 7479–7489.

[55] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2018, pp. 7132–7141.

[56] D. Jarron, M. Shahbazi, D. Lichti, and R. Radovanovic, "Automatic detection and labelling of photogrammetric control points in a calibration test field," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.,* vol. XLII-2/W13, pp. 1673–1680, 2019.

[57] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention,* 2015, pp. 234–241.

[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2016, pp. 770–778.

[59] J. Ai, Y. Mao, Q. Luo, L. Jia, and M. Xing, "SAR target classification using the multikernel-size feature fusion-based convolutional neural network," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–13, Sep. 2021, Art. no. 5214313.

[60] J. Ai et al., "A fine PolSAR terrain classification algorithm using the texture feature fusion-based improved convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–14, Nov. 2021, Art. no. 5218714.

[61] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 57, no. 12, pp. 10070–10087, Dec. 2019.

[62] J. Ai et al., "Robust CFAR ship detector based on bilateral-trimmed-statistics of complex ocean scenes in SAR imagery: A closed-form solution," *IEEE Trans. Aerosp. Electron. Syst.,* vol. 57, no. 3, pp. 1872–1890, Jun. 2021.
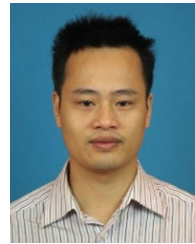
Zhonghua Hong (Member, IEEE) received the Ph.D. degree in GIS from Tongji University, Shanghai, China, in 2014.

Since 2019, he has been an Associate Professor with the College of Information Technology, Shanghai Ocean University, Shanghai, China. His research interests include 3D damage detection, photogrammetry, GNSS-R, coastal mapping, and deep learning.

Zhupeng Li received the B.S. degree in road bridge and river-crossing engineering from Southwest Jiaotong University, Sichuan, China, in 2019. He is currently working toward the M.S. degree in electronic information with Shanghai Ocean University, Shanghai, China.

His research interests include videogrammetry, deep learning, and object detection.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree from Tongji University, Shanghai, China, in 1999.

During 2001 to 2003, he was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006. From 2008 to 2009, he was a Visiting Scholar with the University of California, Santa Barbara, CA, USA. His research interests include trust in spatial data, photogrammetry and remote sensing, and image processing for high-resolution satellite images.

Haiyan Pan received the Ph.D. degree in surveying and mapping from Tongji University, Shanghai, China, in 2020.

She has been a Lecture with the College of Information Technology, Shanghai Ocean University. Her research interests include multispectral/hyperspectral image classification, multitemporal remote sensing data analysis, and change detection.
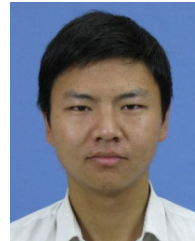
Ruyan Zhou received the Ph.D. degree in agricultural bioenvironment and energy engineering from Henan Agricultural University, Zhengzhou, China, in 2007.

From 2007 to 2008, she was with the Zhongyuan University of Technology. She has been an Associate Professor with the College of Information Technology, Shanghai Ocean University, Shanghai, China.

Yun Zhang received the Ph.D. degree in applied marine environmental studies from Tokyo University of Maritime Science and Technology, Tokyo, Japan, in 2008.

Since 2011, he has been a Professor with the College of Information and Technology, Shanghai Ocean University, Shanghai, China. His research interests include the study of navigation system reflection signal technique and its maritime application.

Yanling Han received the B.E. degree in mechanical design and manufacturing and the M.E. degree in mechanical automation from Sichuan University, Sichuan, China, and the Ph.D. degree in engineering and control theory from Shanghai University, Shanghai, China.

She is currently a Professor with the Shanghai Ocean University, Shanghai, China. Her research interests include the study of ocean remote sensing, flexible system modeling, and deep learning.

Jing Wang received the Ph.D. degree in biomedical engineering with the Department of Biomedical Engineering, Shanghai Jiaotong University, Shanghai, China, in 2014.

Since 2015, she has been a Lecturer with the College of Information Technology, Shanghai Ocean University. Her research interests include computer vision and medical image processing.

Shuhu Yang received the Ph.D. degree in physics of physics from the School of Physics, Nanjing University, Nanjing, China.

Since 2012, he has been a Lecturer with the College of Information Technology, Shanghai Ocean University, Shanghai, China. His research interests include hyperspectral remote sensing, evolution of the Antarctic ice sheet, and the use of navigational satellite reflections.
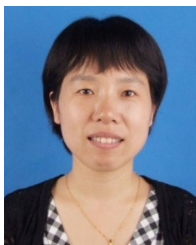
Zhenling Ma received the B.E. and the M.S.E. degrees in photogrammetry and remote sensing from Central South University, Changsha, China, in 2008 and 2011, respectively, and the Ph.D. degree in geomatics from Wuhan University, Wuhan, China, in 2015.

From 2016 to 2018, she was a Post-doctoral Research with the College of Marine Sciences, Shanghai Ocean University, Shanghai, China. Since 2018, she has been a Lecturer with the College of Information Technology.

Her research interests include georeferencing for satellite images and underwater photogrammetric engineering. She is currently a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE ACCESS, *Remote Sensing*, and *ISPRS International Journal of Geo-Information*.