Ship Detection Based on Compressive Sensing Measurements of Optical Remote Sensing Scenes

Shuming Xiao, Ye Zhang, and Xuling Chang

***Abstract*—The compressive sensing (CS)-based optical remote sensing (ORS) imaging system has been verified for the feasibility through numerical simulation experiments. The CS-based ORS imaging system can reduce the demand for sampling equipment, reduce sampling data, save storage space, and reduce transmission costs. However, it needs to reconstruct the original scene when facing the task of ship detection. The scene reconstruction process of CS is computationally expensive, high memory demanding, and time-consuming. In response to this problem, this article proposes an innovation pipeline to perform ship detection tasks, i.e., directly performing ship detection on CS measurements obtained by the imaging system, which avoids the process of scene reconstruction. To achieve the ship detection of CS measurements in the pipeline, we design a convolutional neural network-based algorithm, CS-CenterNet, which jointly optimizes the scene compression sampling phase and the measurements' ship detection phase. CS-CenterNet is divided into convolution measurement layer (CML), optimized hourglass network (OHgN), and optimized three-branch head network (OTBHN). First, CML without bias or activation function simulates the block compression sampling process in CS-based ORS imaging system, which performs convolutional coding on the scene to obtain the measurements. Second, OHgN extracts high-resolution feature information of measurements. Finally, OTBHN performs heat-map prediction, center-point offset prediction, and width–height prediction. We test the performance of CS-CenterNet using the HRSC2016 and LEVIR datasets. The experimental results show that the algorithm can achieve high-accuracy ship detection based on CS measurements of ORS scenes.**

***Index Terms*—Compressive sensing (CS), convolutional neural network (CNN), joint training optimization, ship detection based on compressive sensing measurements.**

I. Introduction

SHIP detection is the focus of research in the field of optical remote sensing (ORS). It has a wide range of civilian and military values, such as search and rescue, port management, marine environment monitoring, territorial security, and military reconnaissance [1], [2], [3]. With the significant improvement of the imaging resolution of the ORS imaging system, the amount of scene data acquired also increases dramatically. Therefore, to relieve the huge pressure of data storage and real-time transmission, the traditional ORS imaging system does not directly store and transmit the original scene information collected by the detector but compresses the data before transmitting to save time and space resources. However, the theoretical basis for data acquisition in this method is the Nyquist sampling theorem, which states that the underlying analog signal must be uniformly sampled at a sampling rate not less than twice the signal bandwidth to preserve signal information [4]. As a result, redundant information can only be discarded in the compression stage, which wastes the sampling resources acquired by the front-end using high-cost detectors.

Compressive sensing (CS) technology states that if the signal is sparse in a certain transform domain, the high-dimensional signal can be projected to a low-dimensional space through a measurement matrix irrelevant to the transform basis and can be accurately recovered with a sampling rate much lower than that required by the Nyquist sampling theorem [5]. Therefore, CS technology breaks through the bottleneck of the Nyquist sampling theorem and can collect scene data at a low sampling rate (much lower than the Nyquist sampling rate). And it can complete data compression at the same time as data collection. In addition, the CS reconstruction algorithm can ideally reconstruct the original data according to the collected sampling data under the premise that the original data is sparse [6], which relieves enormous pressure on data storage and real-time transmission.

The research works on imaging system [7], [8] have verified the feasibility of CS-based ORS imaging system through numerical simulation experiments. The imaging system simultaneously performs sampling and compression by hardware at the sensing stage via CS technology. Therefore, it can reduce the demand for sampling equipment, effectively reduce sampling data, save storage space, and reduce transmission costs. When CS-based ORS imaging system faces the task of ship detection, the information we are interested in is the location attribute of the ship. Fig. 1(a) shows the routine pipeline of CS-based ORS imaging system to perform ship detection tasks. First, the optical system compresses and samples the ORS scene to obtain CS measurements. Then, the original scene is reconstructed using an image reconstruction algorithm [9], [10]. Finally, the image-based ship detection algorithm [11], [12], [13], [14] is used on the reconstructed scene to get the ship detection result. However, the process of reconstructing the measurements to the original scene is computationally costly, memory demanding,

Manuscript received 4 May 2022; revised 3 July 2022 and 13 August 2022; accepted 20 September 2022. Date of publication 23 September 2022; date of current version 14 October 2022. This work was supported in part by CIOMP-Fudan University Joint Fund under Grant Y9S333T190 and in part by Jilin Provincial Institute Cooperation Program Fund under Grant 2020SYHZ0031. *(Corresponding author: Ye Zhang.)*

Shuming Xiao is with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China (e-mail: 15636041235@163.com).

Ye Zhang and Xuling Chang are with the State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China (e-mail: yolanda@spirits.ai; xuling.chang@ciomp.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3209024

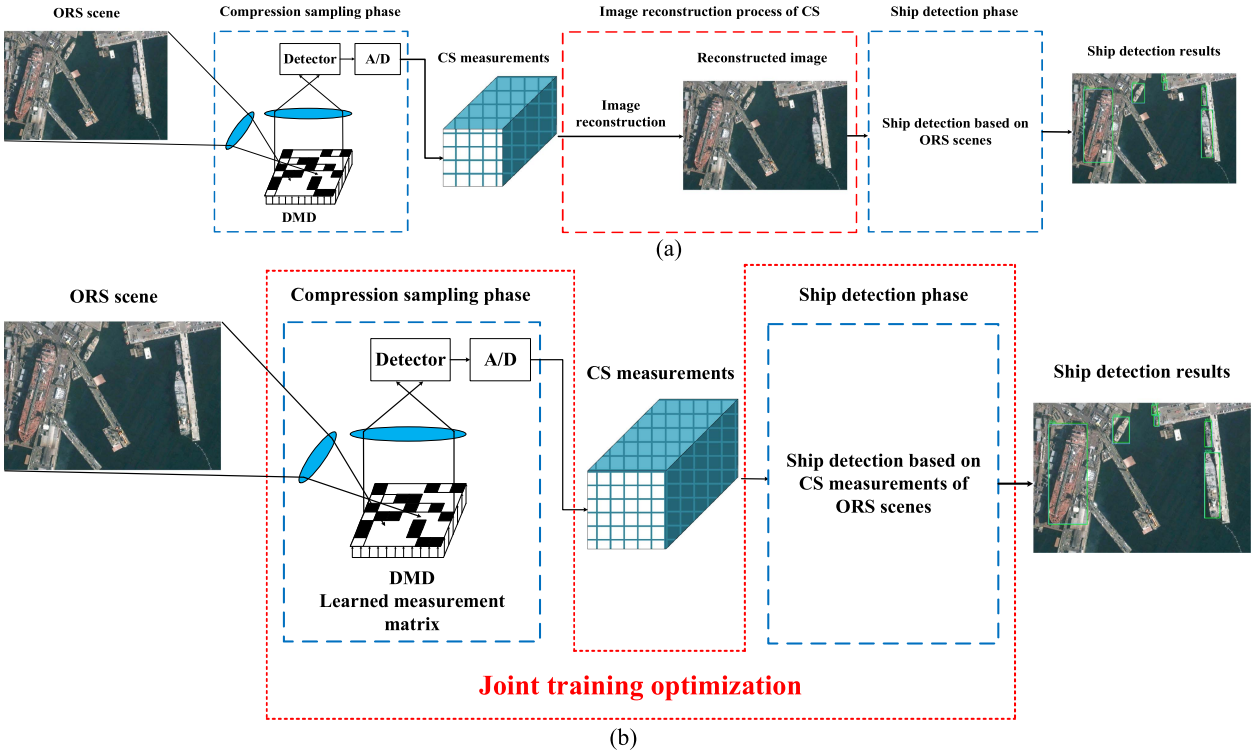


Fig. 1. Illustration of the pipeline of CS-based ORS imaging system to perform ship detection tasks, where digital mirror device (DMD) denotes a measurement matrix in the CS-based imaging system. (a) Routine pipeline method. (b) Our innovation pipeline method. (a) Routine pipeline of CS-based ORS imaging system to perform ship detection tasks. (b) Our innovation pipeline of CS-based ORS imaging system to perform ship detection tasks.

and time-consuming. Therefore, avoiding the process of scene CS reconstruction, i.e., directly performing ship detection on the measurements, can effectively solve the above problems.

In this article, when the CS-based ORS imaging system performs ship detection tasks, we innovatively propose a pipeline, as shown in Fig. 1(b). First, the same as step one in Fig. 1(a), the optical system compresses and samples the ORS scene to obtain the measurements. Then, the measurements-based ship detection algorithm is directly used for the measurements to obtain the ship detection result. This avoids the process of scene CS reconstruction.

Recently, there have been a lot of researches [15], [16], [17] on convolutional neural network (CNN)-based image CS. They use a convolutional measurement layer (CML) to obtain CS measurements of the scene. And the weight value of the CML convolution kernel after training is the learned measurement matrix (LMM). Thanks to the powerful self-learning ability of CNN, LMM can better retain the feature information of the image, thereby improving the quality of image reconstruction. Inspired by these researches, we use LMM in CS-based ORS imaging system to obtain measurements of the scene instead of the predefined measurement matrix (PMM). As shown in Fig. 1(b), LMM compresses and samples the scene, and then the measurements-based ship detection algorithm performs ship detection on the measurements. In this way, the joint training optimization of the scene compression sampling phase and the measurements' ship detection phase can be realized through the end-to-end training method. LMM can retain better scene features for subsequent measurements' ship detection, thereby realizing ship detection on CS measurements.

To directly carry out ship detection on CS measurements, we adopt the method shown in Fig. 1(b) and design a CNN-based algorithm, CS-CenterNet, which achieves high-precision ship detection on CS measurements by jointly training the scene compression sampling phase and the measurements' ship detection phase. The overall framework of CS-CenterNet is shown as in Fig. 2. To simulate the scene compression sampling phase, we use CML without bias or activation function to measure the scene. CML not only adaptively generates the LMM from training scenes but also can be jointly trained with the measurements' ship detection network. Besides, since the physical features of ships are extremely compressed in measurements, an optimized hourglass network (OHgN) is introduced to extract high-resolution feature information of the measurements. Compared with the previous hourglass network (HgN) [18], the squeeze-and-excitation network (SENet) is added to OHgN, which enables OHgN to aim to focus on the salient areas that contain ships. Moreover, since the prediction of ship features is extremely difficult in measurements, an optimized three-branch head network (OTBHN) is introduced to perform heat-map prediction, center-point offset prediction, and width-height prediction. Compared with the previous three-branch head network (TBHN) [19], the feature refinement network (FRNet) is added to OTBHN, which enables OTBHN to improve ship detection accuracy.

Our main contributions are as follows.

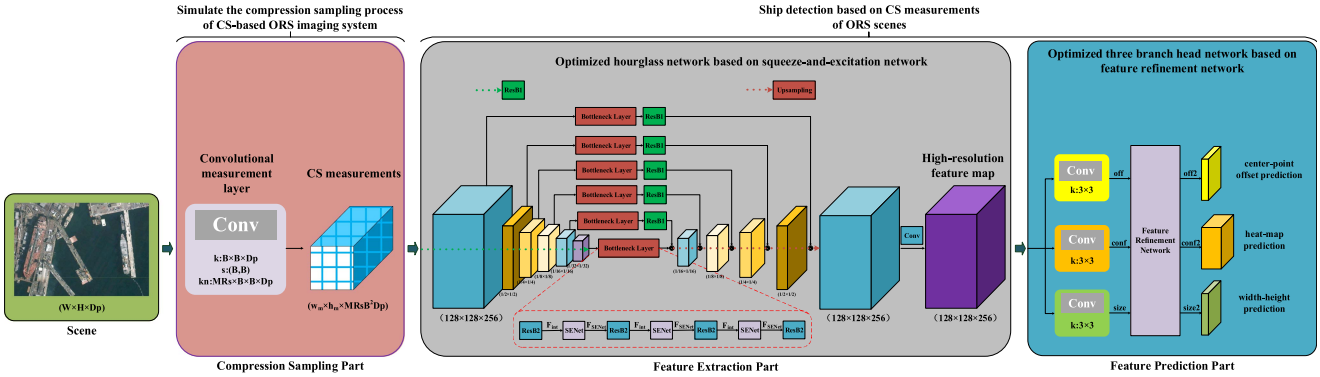


Fig. 2. Illustration of the overall framework of CS-CenterNet, including three key components: CML, OHgN, and OTBHN.

1) When CS-based ORS imaging system faces the task of ship detection, we innovatively propose the pipeline as shown in Fig. 1(b), which avoids the scene CS reconstruction process. And we design the CS-CenterNet to complete the pipeline, which implements ship detection on CS measurements of ORS scenes.
2) We convolutionally encode the scene using CML without bias or activation function to obtain CS measurements, which simulates the compression sampling process in CS-based ORS imaging systems.
3) Considering that the physical features of ships and backgrounds are extremely compressed in CS measurements, a novel OHgN is designed, which extracts the high-resolution feature information from CS measurements.
4) For feature prediction of high-resolution feature information, a novel OTBHN is designed, which refines ship features and improves detection accuracy.

The experimental results on the HRSC2016 dataset [20] and LEVIR dataset [21] demonstrate that CS-CenterNet implements excellent ship detection performance on CS measurements of ORS scenes, proving the feasibility of our model.

The rest of this article is summarized as follows. In Section II, we review the related works of CNN-based compression sampling in CS, CNN-based ORS image ship detection, and compressed learning: image processing in the measurement domain. Section III introduces the structure of CS-CenterNet in detail. Section IV shows the experimental results and results discussion. Finally, Section V concludes this article.

II. Related Works

A. CNN-Based Compression Sampling in CS

The compression sampling network can be connected to the reconstruction network for end-to-end training. And the kernel weight value of the trained CML is LMM. The CS measurements collected by LMM can obtain more image information, which is more beneficial for image reconstruction. Xiao et al. [17] proposed fused features and perceptual loss encoder-decoder residual network (FFPL-EDRNet), which connects CML and reconstruction network for end-to-end training. LMM in this model can improve the image reconstruction quality in CS-based ORS imaging system. Zhao et al. [22] proposed a region of interest (ROI)-aware compressive sensing network (ROI-CSNet), which achieves higher reconstruction quality in ROI while preserving scientific quality in the rest of the image. The measurement matrix in this model is also LMM. Shi et al. [15] proposed an image CS model using CNN (CSNet), which contains a sampling network and a reconstruction network. The measurement matrix in this network is also LMM. Shi et al. [16] proposed a multiscale model for image CS (SCSNet), which uses a sampling network to learn sampling operators and implement the compression sampling process. Shi et al. [23] proposed a novel video CS model based on CNN (VCSNet) for the CS of video to explore both intraframe and interframe correlations. The model still uses the sampling network to learn the sampling operator and implement the compression sampling process. All these CNN-based methods are used to implement the CS process of images or videos, i.e., signal compression sampling and measurements' reconstruction, rather than for ship detection on CS measurements.

B. CNN-Based Ship Detection on Images

The CNN-based ship detection on images can effectively learn complex features and achieve high-accuracy ship detection. Guo et al. [11] proposed a rotational Libra R-CNN, which adds the balanced feature pyramid module and the intersection over union-balanced sampling module to overcome the limitation of dense distribution and different scales. Wang et al. [12] proposed SDGH-Net, which avoids the overfitting problem through Gaussian heatmap regression. Wang et al. [13] proposed fused features and rebuilt (FFR) YOLOv3, which improves the speed and accuracy of ship detection in ORS images. Fu et al. [14] proposed a feature balancing and refinement network (FBR-Net), which achieves an excellent ship detection effect in the case of the wide diversity of scales and the strong interference of the nearshore background. Shan et al. [24] proposed the SiamFPN, which can realize visual object tracking in various maritime applications. All these CNN-based methods are used to improve the accuracy of image ship detection, rather than for ship detection on CS measurements.

C. Compressed Learning: Image Processing in the Measurement Domain

Compressed learning (CL) is a joint signal processing and machine learning framework, which can infer signals from a small number of CS measurements. Calderbank et al. [25] provide a theoretical basis for reasoning directly in the compressed domain. Moreover, the CL method proposed by them uses the support vector machine (SVM) classifier to realize image classification in the measurements, which has a high probability, and its real accuracy is close to the accuracy of the best linear threshold classifier in the data domain. Lohit et al. [26] proved that CNN can extract nonlinear features from CS measurements for image recognition based on the theory of Zisselman et al. [27] proposed an end-to-end method to solve CL, which is composed of fully connected layers and convolutional layers. In the training stage, the sensing matrix of the fully connected layers and the nonlinear inference of the convolutional layers are jointly optimized. Although these CL methods are actually image processing of CS measurements, their back-end processing only uses machine learning method for image classification, not complex target detection.

III. Methodology

CS-CenterNet is designed to construct a high-accuracy ship detection framework on CS measurements. The overall framework of CS-CenterNet is shown in Fig. 2, including three key components: 1) Scene compression sampling part: CML. 2) CS measurements feature extraction part: OHgN. 3) CS measurements feature prediction part: OTBHN. Section III-A–III-E starts with an overview, then the detailed implementations of the three key components in the model, and finally the detailed introduction of the joint training optimization of CS-CenterNet.

A. Overview of CS-CenterNet

To begin with, we describe the problem as follows. Given the ORS scene X to obtain the CS measurements Y, CML is used to compress sampling the scene X. The process simulates the compression sampling process of CS-based ORS imaging system. This process can be expressed as

$$Y = \mathrm{CML}(X) \tag{1}$$

where $\mathrm{CML}(\cdot)$ denotes the compression sampling process. Then, given the CS measurements Y, the backbone network is used to extract high-resolution convolutional features, and the feature prediction network is used to predict the category and location information of the ship. Therefore, we design a backbone network, OHgN, to extract high-resolution feature information F_{OHgN} of CS measurements. This process can be expressed as

$$F_{\mathrm{OHgN}} = \mathrm{OHgN}(Y) \tag{2}$$

where $\mathrm{OHgN}(\cdot)$ denotes the feature extraction process. We also design a feature prediction network, OTBHN, to refine the feature information F_{OHgN} and predict the ship information. This process can be expressed as

$$F_{\mathrm{OTBHN}} = \mathrm{OTBHN}(F_{\mathrm{OHgN}}) \tag{3}$$

Algorithm 1: CS-CenterNet.

Require: An initialized network $\mathrm{SN}_{\mathrm{CS-CenterNet}}$, a labeled ship detection dataset $\mathrm{Train}_{\mathrm{data}}$ and test data $\mathrm{Test}_{\mathrm{data}}$.
1: **Step 1: Train the network**
2: **repeat**
3: Randomly select a batch $\{X_i^{\mathrm{Train}_{\mathrm{data}}}\}_{i=1}^{N_{\mathrm{Train}}}$ from $\mathrm{Train}_{\mathrm{data}}$;
4: Optimize $\mathrm{SN}_{\mathrm{CS-CenterNet}}$ and update the network parameters Θ following (4).
5: **until** convergence
6: **Step 2: Process the detection results**
7: Obtain the CS measurements Y from the scene X utilizing CML with (1).
8: Obtain the high-resolution feature information F_{OHgN} from Y utilizing OHgN with (2).
9: Obtain the feature prediction results F_{OTBHN} from F_{OHgN} utilizing OTBHN with (3) and generate corresponding ship detection results.
Ensure: Optimized CML simulates the compression sampling process of CS-based ORS imaging system. What is more, OHgN and OTBHN constitute the ship detection model on CS measurements of ORS scenes.

where $\mathrm{OTBHN}(\cdot)$ denotes the feature prediction process. Finally, we jointly train the $\mathrm{CML}(\cdot)$, $\mathrm{OHgN}(\cdot)$, and $\mathrm{OTBHN}(\cdot)$ by learning all parameters in CS-CenterNet. Specifically, the overall network is trained using the loss function L_{det} and all parameters Θ are updated with (4).

$$\mathrm{SN}^{\Theta}_{\mathrm{CS-CenterNet}} = \underset{\mathrm{SN}_{\mathrm{CS-CenterNet}} \in \mathrm{SN}}{\arg\min} \max_{\Theta} L_{\mathrm{det}}(\mathrm{SN}_{\mathrm{CS-CenterNet}}, \Theta) \tag{4}$$

where $\mathrm{SN}^{\Theta}_{\mathrm{CS-CenterNet}}$ contains multiple model subsets with different structures.

Moreover, after the overall framework is jointly trained, the weight value in CML is the measurement matrix in CS-based ORS imaging system. The OHgN and OTBHN constitute the ship detection model on CS measurements of ORS scenes. In Algorithm 1, we present more details of CS-CenterNet.

B. Scene Compression Sampling Part

In the traditional compression sampling problem in CS theory, first, the scene needs to meet the sparse condition, and then the sampling matrix needs to meet the restricted isometry property (RIP). The existing sampling matrices are all signal-independent, and do not consider the characteristics of the sampled signal so that more information cannot be retained in measurements. The CNN-based method can solve the compression sampling problem in CS more effectively.

The prerequisite for ship detection on CS measurements is the acquisition of measurements. The CS-based ORS imaging system compresses and samples the scene with the measurement matrix to obtain CS measurements. Therefore, the key point is the design of the measurement matrix when simulating the compression sampling process of the imaging system. In the design of measurement matrix in this article, we refer to the compression sampling process in the related work [15] on CS

reconstruction, i.e., LMM is adopted. It is worth noting that, the weight value of CML convolution kernel after training is LMM. Therefore, we adopt a CML without bias or activation function to measure the scene to simulate the compression sampling process of the imaging system. MRs is the measurement rates in CS, i.e., the ratio of the compression measurement data obtained by the ORS imaging system based on CS to the original scene data.

The compression sampling process is shown in Fig. 3. As shown in Fig. 3(b), first, an ORS scene is divided into nonoverlapping $w_m \times h_m$ ($w_m \times h_m = \frac{W}{B} \times \frac{H}{B}$) blocks of size $B \times B \times Dp$, where W, H, and Dp are the width, height, and number of channels of the scene ($Dp = 3$) , respectively. B is the block size of the scene. Each image block can be denoted as $x_i^{B^2 Dp \times 1}$ in Fig. 3(a), where i is the label of the image block ($i = 1, 2, \ldots, w_m h_m$). Then, the CS measurements $y_i^{MRsB^2 Dp \times 1}$ of the image block $x_i^{B^2 Dp \times 1}$ are acquired using a measurement matrix Φ_{CML} of size $MRsB^2 Dp \times B^2 Dp$. This process can be expressed as

$$y_i^{MRsB^2 Dp \times 1} = \Phi_{\text{CML}}^{MRsB^2 Dp \times B^2 Dp} x_i^{B^2 Dp \times 1}. \tag{5}$$

Since the number of columns in the measurement matrix Φ_{CML} is $B \times B \times Dp$, the size of each convolution kernel in CML is also $B \times B \times Dp$, so that each convolution kernel outputs one measurement. Since the number of rows in the measurement matrix Φ_{CML} is $B \times B \times Dp$, we need $MR\text{s} \times B \times B \times D\text{p}$ convolution kernels in CML to obtain $MR\text{s} \times B \times B \times D\text{p}$ measurements. It should be noted that the stride of CML is $B \times B$ for nonoverlapping sampling. Furthermore, there is no bias or activation function in CML. As shown in Fig. 3(b), the output of each image block from CML is composed of $MR\text{s} \times B \times B \times D\text{p}$ feature maps.

$$Y = \text{CML}(X, W_{\text{CML}}) = W_{\text{CML}} * X \tag{6}$$

where $*$ denotes the elementwise convolution. X denotes the scene. W_{CML} denotes the weight value of CML, i.e., LMM in the CS. Y denotes the CS measurements of the scene.

Since the number of convolution kernels needs to satisfy the inequalities $MR\text{s} \times B \times B \times D\text{p} \geq 1$, the MRs can be any frequency larger than $\frac{1}{12}$ ($\frac{1}{12} \approx 8.33\%$). To avoid the contingency of scene compression sampling at a single MR, MRs will be directly taken as 25%, 10%, 4%, and 1% in the research works [10], [28] of CS. Therefore, the corresponding relationship between $B \times B$ strides and MRs is shown in Table I in this article.

Fig. 4 provides the frequency domain visualization results of PMM (Gaussian random matrix) and LMM on the HRSC2016 dataset at MRs $= 25\%$. Since the data dimension of the image block is 12 ($2 \times 2 \times 3$) and the MRs is 25%, the scale of the PMM is 3×12. The expression of the PMM is shown in (7) at the bottom of this page.

TABLE I
RELATIONSHIP BETWEEN $B \times B$ STRIDES AND MRS

$B \times B$	MRs
2×2	25% or 10%

Similarly, since the data dimension of the image block is 12 and the MRs is 25%, the scale of the LMM is $3 \times 3 \times 2 \times 2$ (the first 3 is obtained by $12 \times 25\%$, the second 3 is the depth of the convolution kernel, and 2×2 is the size of the convolution kernel). We select all three rows from each of LMM and PMM for visualization. To obtain a better visual effect, the frequency visualization is the result of the Fourier transform of each row of measurement matrix. It can be seen from Fig. 4 that the frequency of each PMM row (PMM[0], PMM[1], and PMM[2]) is randomly distributed, i.e., PMM will randomly sample scene information, while the frequency of each LMM row (Conv Kernel 0[nc], Conv Kernel 1[nc], and Conv Kernel 2[nc] with nc $= 0, 1, 2$) is a regular distribution, i.e., LMM will sample the specific frequency information of the scene. As we all know, the specific frequency sampling of the scene can better maintain the scene feature information than the random frequency sampling of the scene. Therefore, by training the compression sampling phase of the scene together with the ship detection phase of the CS measurements, LMM captures scene features information more efficiently than the PMM.

Fig. 5 shows an ORS scene and the CS measurements of the scene at MRs $= 25\%$. Fig. 5(b) is the measurements of PMM compression sampling of the scene at MRs $= 25\%$. Fig. 5(c) is the measurements of LMM compression sampling of the scene at MRs $= 25\%$. It can be seen from Fig. 5 that the CS measurements obtained by CML still retain the target position, target size, and shape information, while the CS measurements obtained by PMM destroy the target position, target size, and shape information.

C. CS Measurements Feature Extraction Part

Since the data volume of CS measurements is much lower than their corresponding original scenes, the feature extraction network of CS measurements needs to aggregate global information and multi-scale local information to obtain high-quality high-resolution feature information. We are inspired by the context refinement module (CRM) in [29] and adopt another existing unified framework structure HgN which captures and integrates information across all scales of measurements. Therefore, to extract the high-resolution feature information of measurements, we design the backbone network, OHgN, based on the HgN.

$$\begin{bmatrix} 0.0505 & 0.6554 & 0.4602 & -0.0601 & -0.0558 & -0.4984 & -0.8778 & 0.2321 & 0.4523 & 1.1642 & -0.3311 & -0.3638 \\ 0.0008 & -0.6213 & -0.2778 & -0.1211 & -0.2414 & -0.2510 & -0.1389 & -0.4900 & 0.2897 & -0.7125 & -0.1756 & -0.3116 \\ 0.8378 & 0.3088 & -0.6469 & -0.0353 & -0.9805 & -0.3985 & 1.5064 & -0.2152 & 0.4826 & 0.0819 & -0.3617 & -1.3199 \end{bmatrix}_{3\times 12} \tag{7}$$

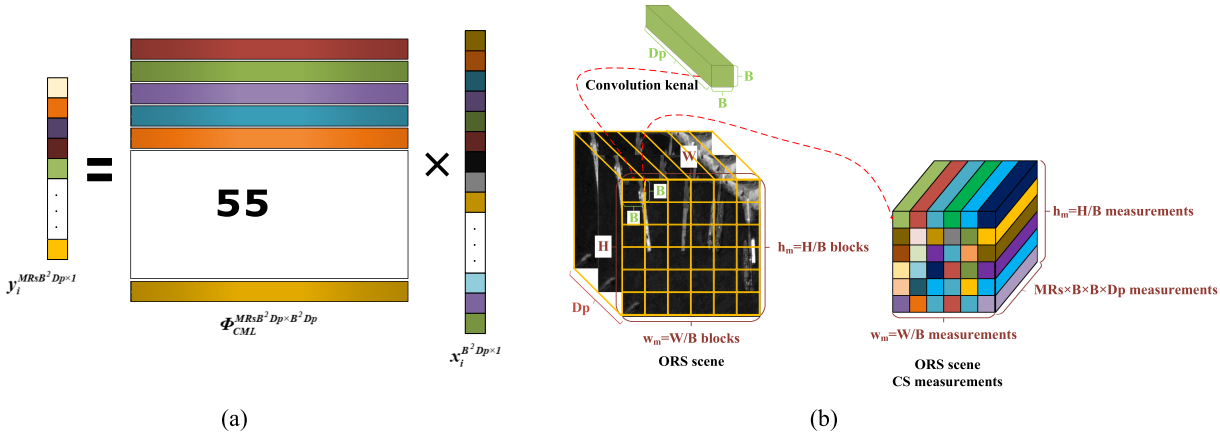


Fig. 3. Illustration of the compression sampling process. (a) The traditional compression sampling process in CS theory. (b) The compression sampling process in CS-CenterNet.

1) HgN: The structure of the HgN in this article is as shown in Fig. 6, where ResB1 denotes the residual block with $\frac{1}{2}$ downsampling, ResB2 denotes the residual block without downsampling, and Conv denotes the convolutional layer.

The CS measurement Y from CML is denoted as $Y \in R^{w_m \times h_m \times MRsB^2Dp}$, where $w_m \times h_m$ is the size of measurements. First, a ResB1 is adopted for feature extraction to obtain a $128 \times 128 \times 256$ feature map $C1$, and then ResB1 is adopted to perform five consecutive feature extractions. The results of five feature extractions are feature maps $C2, C3, C4, C5, C6$, respectively. The feature map sizes of $C2, C3, C4, C5, C6$ are reduced to $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$, and the feature map dimensions are increased to $256, 384, 384, 384, 512$. Afterward, features from $C1, C2, C3, C4, C5$ are extracted through BottleneckLayer (containing four ResB2 modules) and ResB1 to obtain features $C1a, C2a, C3a, C4a, C5a$. BottleneckLayer is used to extract feature $C6$ to obtain feature $C6a$.

To aggregate the feature information of two adjacent sizes, upsampling and cross-scale feature combination methods are used. $C6a$ and $C5a$ with the same size are added elementwise and then the nearest neighbor upsampling is performed to obtain $C5b$. $C5b$ and $C4a$ are added elementwise and the nearest-neighbor upsampling is also performed to obtain $C4b$, and so on to get the respective upsampling results $(C5b, C4b, C3b, C2b, C1b)$. The feature map sizes of $C5b, C4b, C3b, C2b, C1b$ are increased to $\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1$ and the feature map dimensions are reduced to $512, 384, 384, 384, 256$ in turn. After reaching the output resolution of 128×128, a 3×3 Conv is applied to generate the final high-resolution feature map.

In HgN, low-level, weak semantic features have rich location information, which is very useful for object positioning. High-level, strong semantic features have rich semantic information, which is very useful for object classification. Therefore, the two characteristics are fused in HgN. The advantage of HgN is that it can capture global and local features in a single unified structure. Therefore, the final feature information can contain almost all the critical points of the detected object.

Although HgN can extract high-resolution feature information from CS measurements, it cannot select the information that

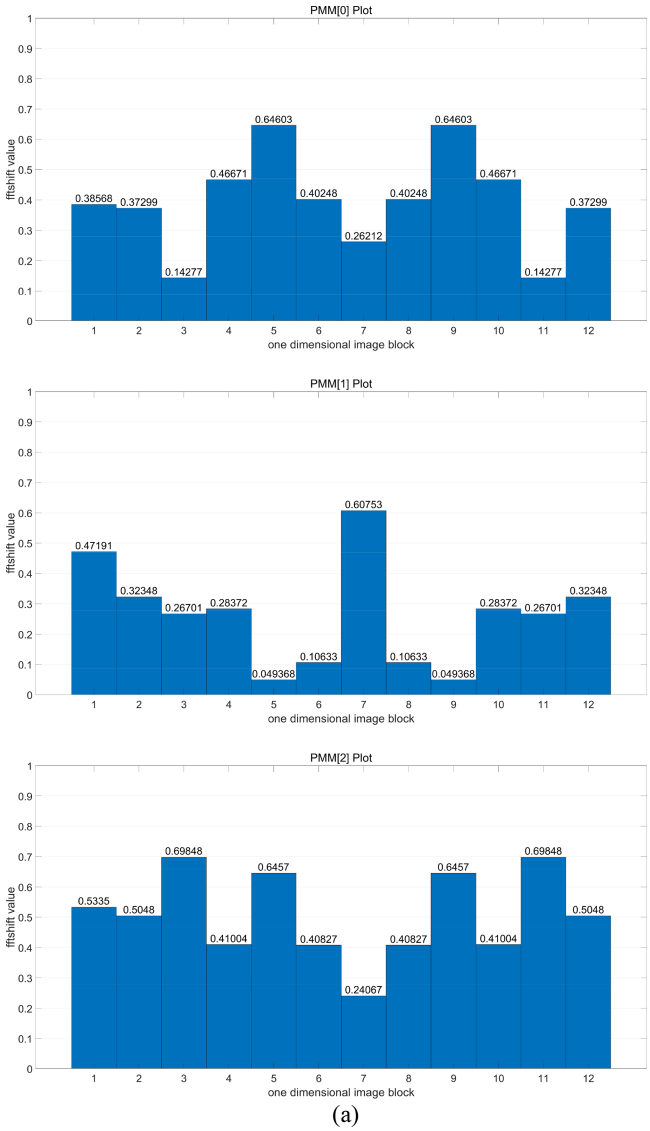


Fig. 4. Illustration of measurement matrix on the HRSC2016 dataset at MRs = 25%. (a) The frequency domain visualization results of PMM (Gaussian random matrix).

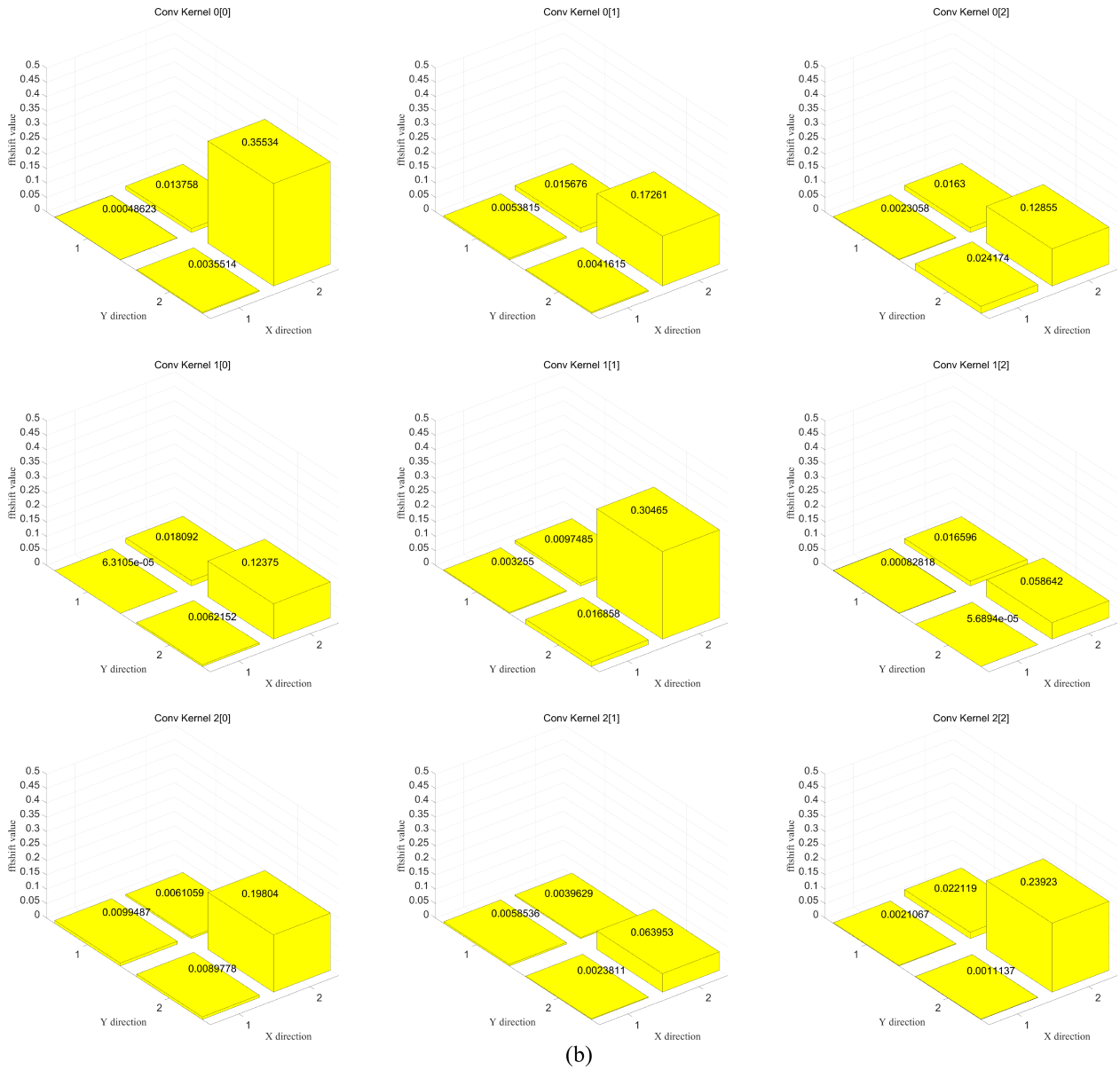


Fig. 4. *(Continued.)* Illustration of measurement matrix on the HRSC2016 dataset at MRs $= 25\%$. (b) The frequency domain visualization results of LMM.

is more critical to ships from measurements. However, SENet can devote more attention to the ships' area to obtain more detailed information about the ships, thereby suppressing other useless information. Therefore, we add SENet to HgN to refine the features and focus on the salient areas that contain ships.

2) SENet: The position where SENet is added to HgN is shown in Fig. 2. Specifically, SENet is added to the Bottleneck-Layer part of HgN. The feature processing process of SENet is shown in Fig. 7. The input feature of it is denoted as $F_{\text{int}} \in R^{(\frac{W}{s_{\text{se}}})\times(\frac{H}{s_{\text{se}}})\times C_{\text{se}}}$, where C_{se} is the channel dimension of feature F_{int}, and s_{se} is the corresponding downsampling ratio to the input scene ($s_{\text{se}} = 32$). Maxpool denotes the max-pool layer, Avgpool denotes the avg-pool layer, and Sig denotes the sigmoid function. First, it uses the Maxpool and Avgpool along the channel axis to generate features $F_{\max}, F_{\text{avg}} \in R^{(\frac{W}{s_{\text{se}}})\times(\frac{H}{s_{\text{se}}})\times 1}$. Then, it applies a 3×3 Conv and a Sig to get the feature $A_{\text{SENet}} \in R^{(\frac{W}{s_{\text{se}}})\times(\frac{H}{s_{\text{se}}})\times 1}$. The calculation of the above process is shown in (8). Finally, the feature A_{SENet} is multiplied by the initial feature F_{int} as shown in (9).

$$A_{\text{SENet}} = Sig\left(Conv_{3\times 3}\left(\left[MPool\left(F_{\text{int}}\right); APool\left(F_{\text{int}}\right)\right]\right)\right) \quad (8)$$

where $MPool(\cdot)$ and $APool(\cdot)$ denote the maximum pooling operation and average pooling operation, respectively. $Conv_{3\times 3}(\cdot)$ denotes the 3×3 convolution operation. $Sig(\cdot)$ denotes the sigmoid nonlinear operation.

$$F_{\text{SENet}} = A_{\text{SENet}} \times F_{\text{int}} \quad (9)$$

where $\times$ denotes the elementwise multiplication. Especially, the tensor dimension of $F_{\text{SENet}} \in R^{(\frac{W}{s_{\text{se}}})\times(\frac{H}{s_{\text{se}}})\times C_{\text{se}}}$ is the same as the tensor dimension of $F_{\text{int}} \in R^{(\frac{W}{s_{\text{se}}})\times(\frac{H}{s_{\text{se}}})\times C_{\text{se}}}$.

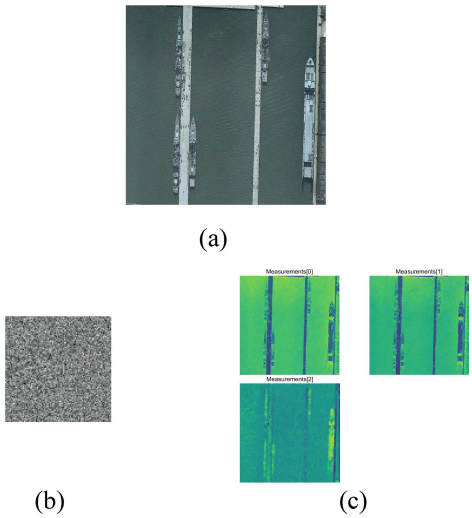


Fig. 5. Illustration of an ORS scene and the CS measurements of the scene at MRs = 25%. (a) An ORS scene. Its size is 512 × 512 × 3. (b) The measurements of PMM compression sampling of the scene. Its size is 256 × 256 × 3. (c) The measurements of LMM compression sampling of the scene, and the size of each is 256 × 256. (a) A scene. Its size is 512 × 512 × 3. (b) The measurements of PMM compression of the original image. Its size is 256 × 256 × 3. (c) The measurements of LMM compression of the original image It has three measurements, and the size of each is 256 × 256.

We select eight channels for visualization from the $128 \times 128 \times 256$ high-resolution feature map predicted by HgN and OHgN on two ORS scenes at MRs = 25%. The eight channels are the 1th channel, the 32nd channel, the 64th channel, the 96th channel, the 128th channel, the 160th channel, the 192th channel, the 224th channel, and the 256th channel, whose visualization results are shown in Fig. 8. It can be seen from Fig. 8 that the high-resolution feature map predicted by OHgN contains more ship target area information than HgN. This is because the SENet in OHgN can refine the features and focus more on the target area of the ship.

D. CS Measurements Feature Prediction Part

The traditional feature prediction network adopts the anchor box method to predict the category and location information of the target. However, using anchor boxes introduces many hyperparameters and design choices. These hyperparameters make network-tuning difficult and also increase network complexity and computational complexity. Recently, the research work on anchor-free [30] showed that the anchor-free method can eliminate the anchor problem and ensure detection accuracy. Therefore, we also predict the category and location information of ships based on an anchor-free method.

To adopt an anchor-free method for heat-map prediction, center-point offset prediction, and width–height prediction of high-resolution feature information, we design the feature prediction network, OTBHN, based on the TBHN.

1) TBHN: The structure of the TBHN in this article is as shown in Fig. 9. The feature F_{OHgN} from the OHgN is denoted as $F_{OHgN} \in R^{\left(\frac{W}{s_{tb}}\right)\times\left(\frac{H}{s_{tb}}\right)\times C_{tb}}$, where C_{tb} is the channel dimension of feature F_{OHgN} ($C_{tb} = 256$), and s_{tb} is the corresponding downsampling ratio to the input scene ($s_{\mathrm{tb}} = 4$) [19]. All branches in the TBHN have a $3 \times 3 \times 256$ Conv and a $1 \times 1 \times 256 \times T_a$ Conv with $a = 1, 2, 3$ ($T_1 = Cls$, $T_2 = 2$, $T_3 = 2$, where Cls is the number of categories). Especially, a 3×3 Maxpool is used to perform the equivalent nonmaximum suppression execution in the extraction branch of peak key points.

The illustration of the detection based on the center-point for TBHN is shown in Fig. 10. First, we independently extract the peak points on the heatmap of each category Cls. Then, we use $\hat{\mathrm{CP}}$ to denote the set of n detected center-points [$\hat{\mathrm{CP}} = (\hat{x}_i, \hat{y}_i)_{i=1}^{n}$, where $(\hat{x}_i, \hat{y}_i)$ denotes the predicted each key point position]. Finally, we get the coordinates of the upper left corner and the lower right corner of the prediction box and generate a horizontal box at this position. We denote the coordinates of the upper left corner and the lower right corner of the prediction box as follows:

$$\begin{aligned}&\left(\hat{x}_i + o\hat{x}_i - \frac{\hat{w}_i}{2}, \hat{y}_i + o\hat{y}_i - \frac{\hat{h}_i}{2}\right)\\&\left(\hat{x}_i + o\hat{x}_i + \frac{\hat{w}_i}{2}, \hat{y}_i + o\hat{y}_i + \frac{\hat{h}_i}{2}\right)\end{aligned} \tag{10}$$

where $(o\hat{x}_i, o\hat{y}_i)$ denotes the predicted position offset and $\left(\hat{w}_i, \hat{h}_i\right)$ denotes the predicted size.

Because Conv in TBHN is a fixed geometric structure, it limits its modeling of geometric deformation. To strengthen the ability of the feature prediction network to model deformation, we refer to the deformable convolution in [31]. By learning an additional offset, the deformable convolution makes the feature offset to focus on the target area of interest, which helps to solve the structural information between similar objects, thereby improving the accuracy of ship detection. Therefore, we design FRNet, which mainly uses deformable convolution to refine features. The FRNet is added to TBHN to refine the ship features and improve detection accuracy.

2) FRNet: The position where FRNet is added to TBHN is shown in Fig. 2. The structure of FRNet is shown in Fig. 11. The input features of it are denoted as conf $\in R^{\left(\frac{W}{s_{\mathrm{fr}}}\right)\times\left(\frac{H}{s_{\mathrm{fr}}}\right)\times Cls}$, off $\in R^{\left(\frac{W}{s_{\mathrm{fr}}}\right)\times\left(\frac{H}{s_{\mathrm{fr}}}\right)\times 2}$, and size $\in R^{\left(\frac{W}{s_{\mathrm{fr}}}\right)\times\left(\frac{H}{s_{\mathrm{fr}}}\right)\times 2}$, where s_{fr} is the corresponding downsampling ratio to the input scene ($s_{\mathrm{se}} = 4$). First, the *Sig* is applied to conf, and the result Sig(conf) is used to generate the adjusted feature conf1 $\in R^{\left(\frac{W}{s_{fr}}\right)\times\left(\frac{H}{s_{fr}}\right)\times Cls}$ with the feature conf as follows:

$$\mathrm{conf1} = \mathrm{conf} \times \mathrm{Sig(conf)}. \tag{11}$$

Then, we use the adjusted feature conf1 to generate the feature off1 and size1 with features conf and size. Afterwards, a 3×3 Conv is applied for feature off1, size1 , and conf1 to generate the features off11, size11 , and conf11. The calculation of the above process is as follows:

$$\mathrm{off11} = Conv_{3\times3}(\mathrm{conf1} + \mathrm{off}) \tag{12}$$

$$\mathrm{size11} = Conv_{3\times3}(\mathrm{conf1} + \mathrm{size}) \tag{13}$$

$$\mathrm{conf11} = Conv_{3\times3}(\mathrm{conf1}). \tag{14}$$

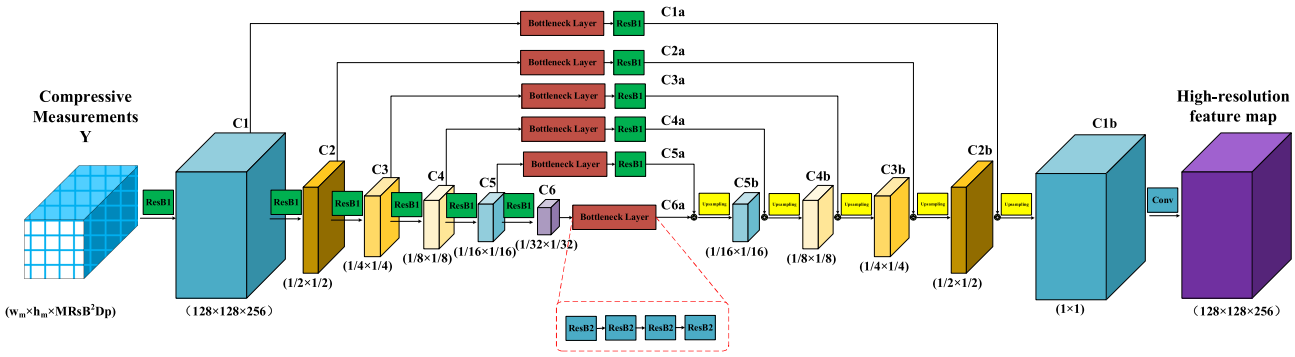


Fig. 6. Illustration of the structure of HgN.

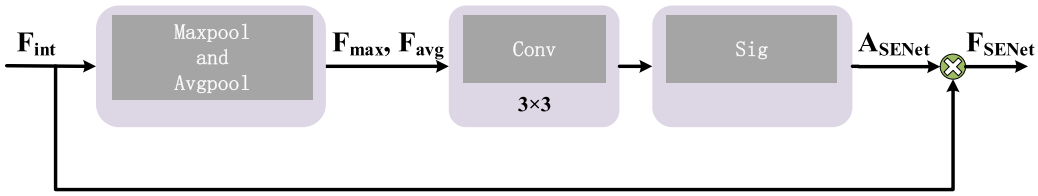


Fig. 7. Illustration of the structure of SENet.

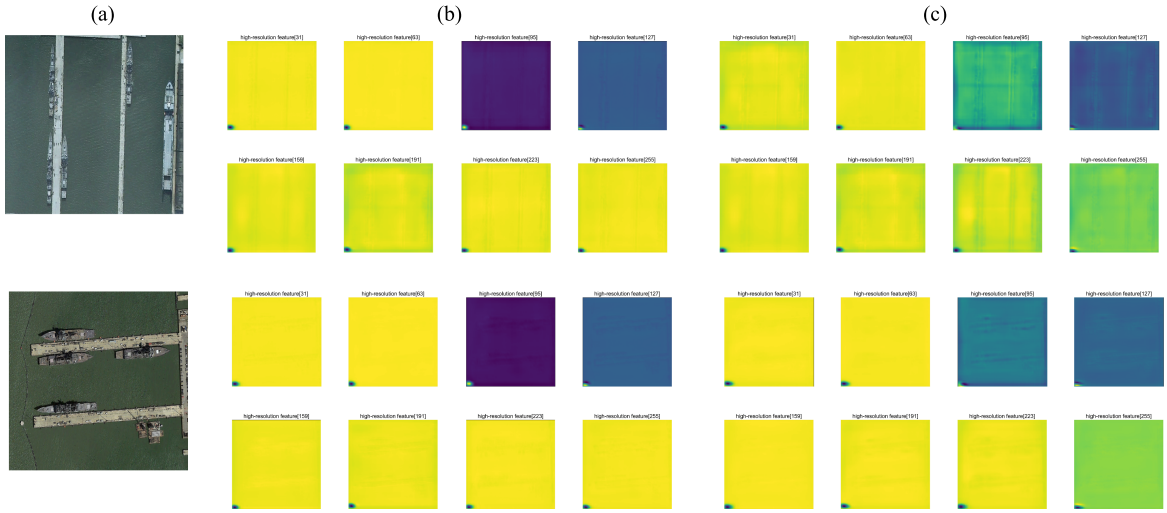


Fig. 8. Illustration of the eight channels for visualization from the $128 \times 128 \times 256$ high-resolution feature map predicted by HgN and OHgN at MRs $= 25\%$. The eight channels are the 1th channel, the 32nd channel, the 64th channel, the 96th channel, the 128th channel, the 160th channel, the 192th channel, the 224th channel, and the 256th channel. (a) Original scenes. (b) The eight channels for visualization from the high-resolution feature map predicted by HgN. (c) The eight channels for visualization from the high-resolution feature map predicted by OHgN.

Finally, we adopt deformable convolution [32] for FRNet. The kernel off set fields ($\text{offset}_{\text{off11}}$, $\text{offset}_{\text{conf11}}$, and $\text{offset}_{\text{size11}}$) of the three features ($off11$, conf11, and size11) were originally generated, respectively, by using a 1×1 Conv. Afterwards, a 3×3 deformable Conv is applied to offsets (offset_{off1}, $\text{offset}_{\text{conf1}}$, and $\text{offset}_{\text{size1}}$) to obtain the refined features $\text{conf2} \in R^{\left(\frac{W}{s_{fr}}\right)\times\left(\frac{H}{s_{fr}}\right)\times Cls}$, $off2 \in R^{\left(\frac{W}{s_{fr}}\right)\times\left(\frac{H}{s_{fr}}\right)\times 2}$, and $\text{size2} \in R^{\left(\frac{W}{s_{fr}}\right)\times\left(\frac{H}{s_{fr}}\right)\times 2}$ for further classification, center-point offset, and prediction box width–height.

The calculation of all process is as follows:

$$(\text{conf2}, off2, \text{size2}) = FRNet(\text{conf}, off, \text{size}) \tag{15}$$

where $FRNet(\cdot)$ denotes the operation of FRNet module, and the features $\text{conf2}, off2, \text{size2}$ denote the output of the features after refining treatment of FRNet, respectively.

We visualize the heat maps predicted by the TBHN and OTBHN on two ORS scenes at MRs $= 25\%$. The visualized result is shown in Fig. 12. It can be seen from Fig. 12 that OTBHN can locate the ship position more accurately than

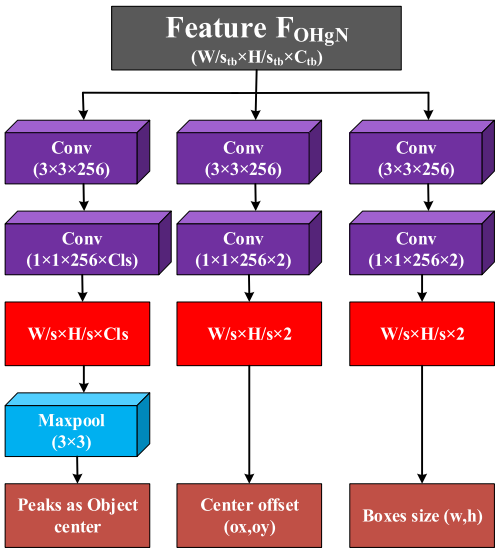


Fig. 9. Illustration of the structure of TBHN.

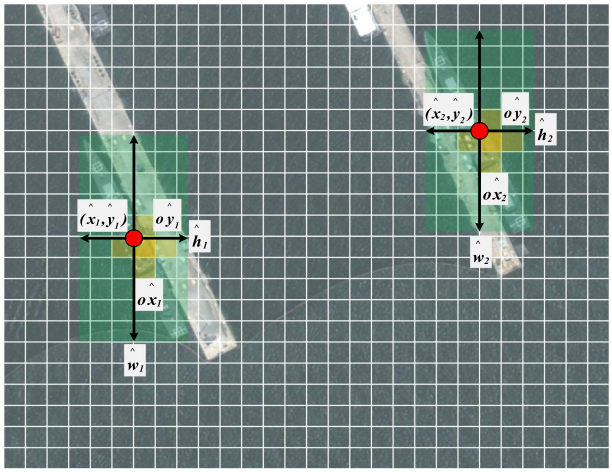


Fig. 10. Illustration of the detection based on the center-point.

TBHN. This is because FRNet is added to OTBHN, and the deformed convolution in FRNet refines the features.

E. CS-CenterNet Joint Training Optimization

1) Loss Function: Our training loss function consists of three parts.

$$L_{\text{det}} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}}. \tag{16}$$

where L_k, L_{size}, and L_{off} denote the heat-map loss, center-point offset loss, and width-height loss, respectively. λ_{size} and λ_{off} are hyperparameters. Inspired by Zhou et al. [19], we set their values to 0.1 and 1, respectively.

Considering the imbalance between negative and positive samples, the focal loss [33] is adopted for L_k. The calculation method for L_k is as follows:

$$L_k = -\frac{1}{N}$$

TABLE II
DATASET DIVISION

Dataset	Dataset division	Number of images
HRSC2016	Training set	1176
	Validation set	168
	Test set	336
LEVIR	Training set	1037
	Validation set	149
	Test set	296

$$\times \begin{cases} \sum_{xyz} \left(1-\hat{Y}_{xyc}\right)^{\alpha} \log\left(\hat{Y}_{xyc}\right) & if \\ \sum_{xyz} \left(1-Y_{xyc}\right)^{\beta} \left(\hat{Y}_{xyc}\right)^{\alpha} & Y_{xyc=1} \\ \log\left(1-\hat{Y}_{xyc}\right) & \text{otherwise} \end{cases} \tag{17}$$

where Y_{xyz} denotes the key point heatmap of the target; $\hat{Y}_{xyz}$ denotes the key point heatmap of the network output ($\hat{Y}_{xyz} \in [0,1]$); N denotes the number of key points in the scene; and α and β are hyperparameters. Inspired by Zhou et al. [19], we set their values to 2 and 4, respectively.

L_{size} and L_{off} adopt the $L1$ loss, and they can be formulated as

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^{N} |\hat{S}_{pk} - s_k| \tag{18}$$

where s_k denotes the true length and width of the target and $\hat{S}_{\text{pk}}$ denotes the predicted length and width of the target.

$$L_{\text{off}} = \frac{1}{N} \sum_{p} \left| \hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right| \tag{19}$$

where p denotes the center-point of the target box, $\tilde{p}$ denotes the center-point of the predicted box, $\hat{O}_{\tilde{p}}$ denotes the output offset of the network, and R denotes the downsampling multiple ($R = 4$ [19]).

2) Joint Training Optimization: Since joint training optimization plays a vital role in the detection performance of ships, we train the CML and the measurements' ship detection network by learning all parameters in the model. The set of all parameters in the model can be expressed as $\Theta = \{W_{\text{CML}}, W_{\text{OHgN}}, W_{\text{OTBHN}}\}$, where W_{OHgN} denotes the network parameters of OHgN and W_{OTBHN} denotes the network parameters of OTBHN. And the process of joint training is to obtain the optimal network parameters Θ. During training, the input and output of CS-CenterNet are the scene information and the ships' location information (box_xmin, box_xmax, box_ymin, box_ymax), respectively, i.e., the training samples are represented as {scene, ship location information}. After the training optimization, optimized CML simulates the compression sampling process of CS-based ORS imaging system. What's more, OHgN and OTBHN constitute the ship detection model on CS measurements of ORS scenes.

As shown in Fig. 13, the black arrow denotes the joint training process of the scene compression sampling part and the measurements' ship detection part. The red arrow denotes the test process of the measurements' ship detection. First,

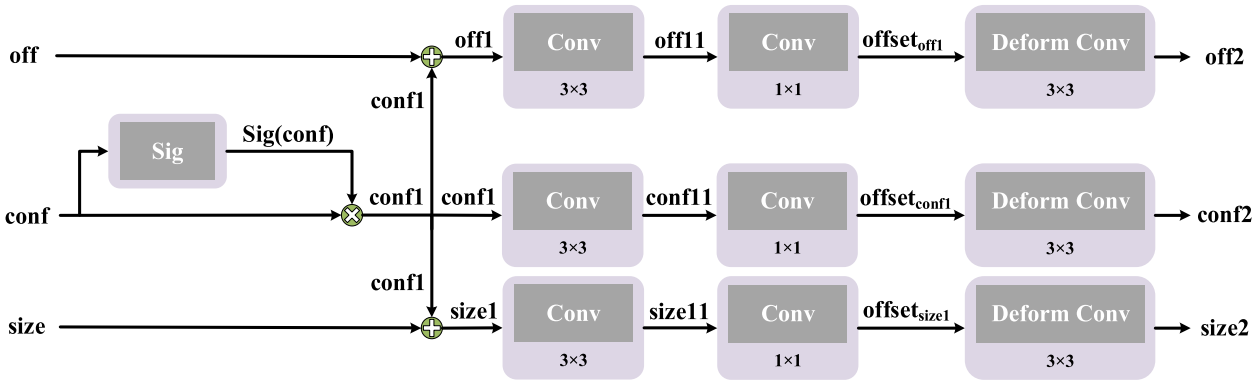


Fig. 11. Illustration of the structure of FRNet.

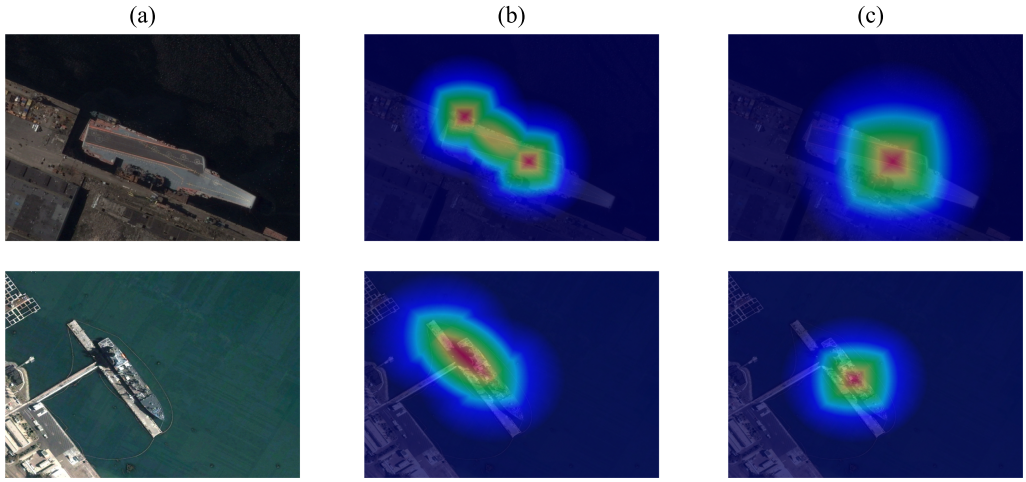


Fig. 12. Illustration of heat maps predicted by TBHN and OTBHN on two ORS scenes at MRs = 25%. (a) Original scenes. (b) The heatmap predicted by the TBHN. (c) The heatmap predicted by the OTBHN.

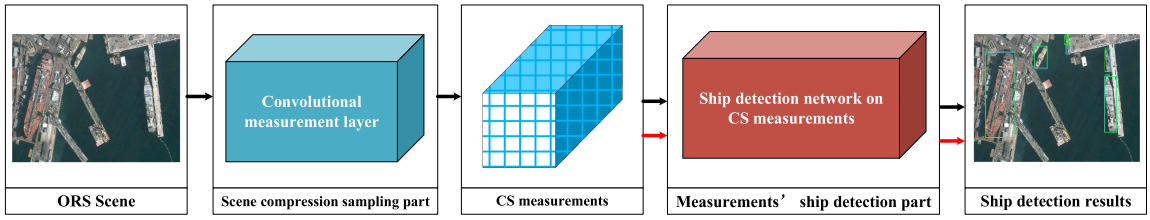


Fig. 13. Illustration of the joint optimization process of CS-CenterNet.

the trained CML compresses and samples the ORS scene, and then the measurements' ship detection network detects the ship information on CS measurements.

IV. Experiment

A. Dataset

We evaluate our model on two public ORS scene datasets: The HRSC2016 dataset [20] and LEVIR dataset [21]. There are 1680 images in the HRSC2016 dataset. In the experiment, the number of samples of the training set, validation set, and test set is divided as shown in Table II. The training set, validation set, and test set contain 1176, 168, and 336 images, respectively. Some ORS scenes in the dataset are shown in Fig. 14(a). There are 1482 images in the LEVIR dataset for ship object. In the experiment, the number of samples of the training set, validation set, and test set is divided as shown in Table II. The training set, validation set, and test set contain 1037, 149, and 296 images, respectively. Some ORS scenes in the dataset are shown in Fig. 14(b).

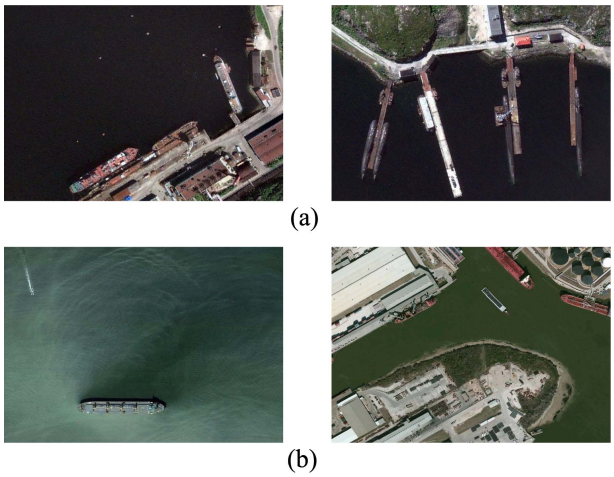


Fig. 14. Some ORS scenes of the HRSC2016 and LEVIR datasets. (a) Some ORS scenes of the HRSC2016 dataset. (b) Some ORS scenes of the LEVIR dataset.

TABLE III
EXPERIMENTAL ENVIRONMENT

System	Windows10
RAM	48.0 GB
CPU	4.10 GHz Intel processor
GPU	GeForce RTX 3070, memory 8 G
DL framework	Keras (backed as TensorFlow)

TABLE IV
TRAINING PARAMETERS

Convolution initialization	Gaussian distribution with a standard deviation of 0.001 [10]
Optimizer	Adam [34]
Learn rate	0.001
Batch size	2

B. Implementation Details and Parameters

The training and testing of CS-CenterNet require a relatively high hardware environment, so we use the experimental environment in Table III to train the model. Since the size of the image data in the experiment is different, it is uniformly adjusted to 512×512 before inputting the model.

Table IV shows the parameter settings during the CS-CenterNet training process. In particular, considering the 8G limitation of GPU memory, we set the batch size to 2. The initial learning rate of 0.001 is reduced by half every 10 epochs.

C. Evaluation Metrics

Different from the target detection models based on Intersection over Union (IoU) [35], [36], [37] to define positive samples and negative samples, CS-CenterNet defines positive and negative samples as follows. The position where the center of the ground truth box falls is a positive sample, and the remaining positions are negative samples. If the prediction result of a positive sample is also a positive sample, it is defined as a true positive (TP). If the prediction result of a negative sample is a positive sample, it is defined as a false positive (FP). If the prediction result of a positive sample is a negative sample, it is defined as false negative (FN).

Precision and recall rate are usually used as the evaluation criteria for target detection, and their calculation methods are shown in (20) and (21).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{20}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{21}$$

However, because the precision rate and recall rate are numerically contradictory, we add F1 and AP value as evaluation indicators. F1 is a comprehensive indicator of the imbalance between precision and recall. The AP reflects the overall quality of the network, which defines the average precision under a set of equidistant recall rates $S = \{0, 0.01, \ldots, 1\}$. In this article, we calculate AP when the IoU threshold is 0.5. The calculation methods of F1 and AP are as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{22}$$

$$\text{AP} = \frac{1}{101} \sum_{r \in S} \text{Precision}|_{\text{Recall}=r}. \tag{23}$$

TABLE V
SHIP DETECTION RESULTS IN THE HRSC2016 DATASET

Model	P(%)	R(%)	F1	AP(%)
CenterNet for ORS images	83.51	90.33	0.8679	89.90
CS-CenterNet for CS measurements of ORS scenes (MRs = 25%)	**84.35**	**92.19**	**0.8810**	**90.76**

The best results are highlighted in bold.

TABLE VI
SHIP DETECTION RESULTS IN THE LEVIR DATASET

Model	P(%)	R(%)	F1	AP(%)
CenterNet for ORS images	**70.93**	77.32	0.7399	74.72
CS-CenterNet for CS measurements of ORS scenes (MRs = 25%)	70.60	**78.20**	**0.7421**	**75.44**

The best results are highlighted in bold.

D. Comparison With Ship Detection Based on ORS Images

To test the effect of CS-CenterNet on the ship detection of CS measurements, we need to compare it with the ship detection model based on ORS images. Since CS-CenterNet refers to CenterNet [19], whose backbone network is ResNet50, we compare the ship detection performance of CS-CertenNet with CenterNet. In particular, CS-CenterNet is a ship detection model based on the CS measurements of ORS scenes while CenterNet is a ship detection model based on ORS images.

Tables V and VI show the ship detection results in the HRSC2016 and LEVIR datasets, respectively. Figs. 15 and 16 show the ship detection effect of some images of the models in the HRSC2016 and LEVIR test sets, respectively. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN.

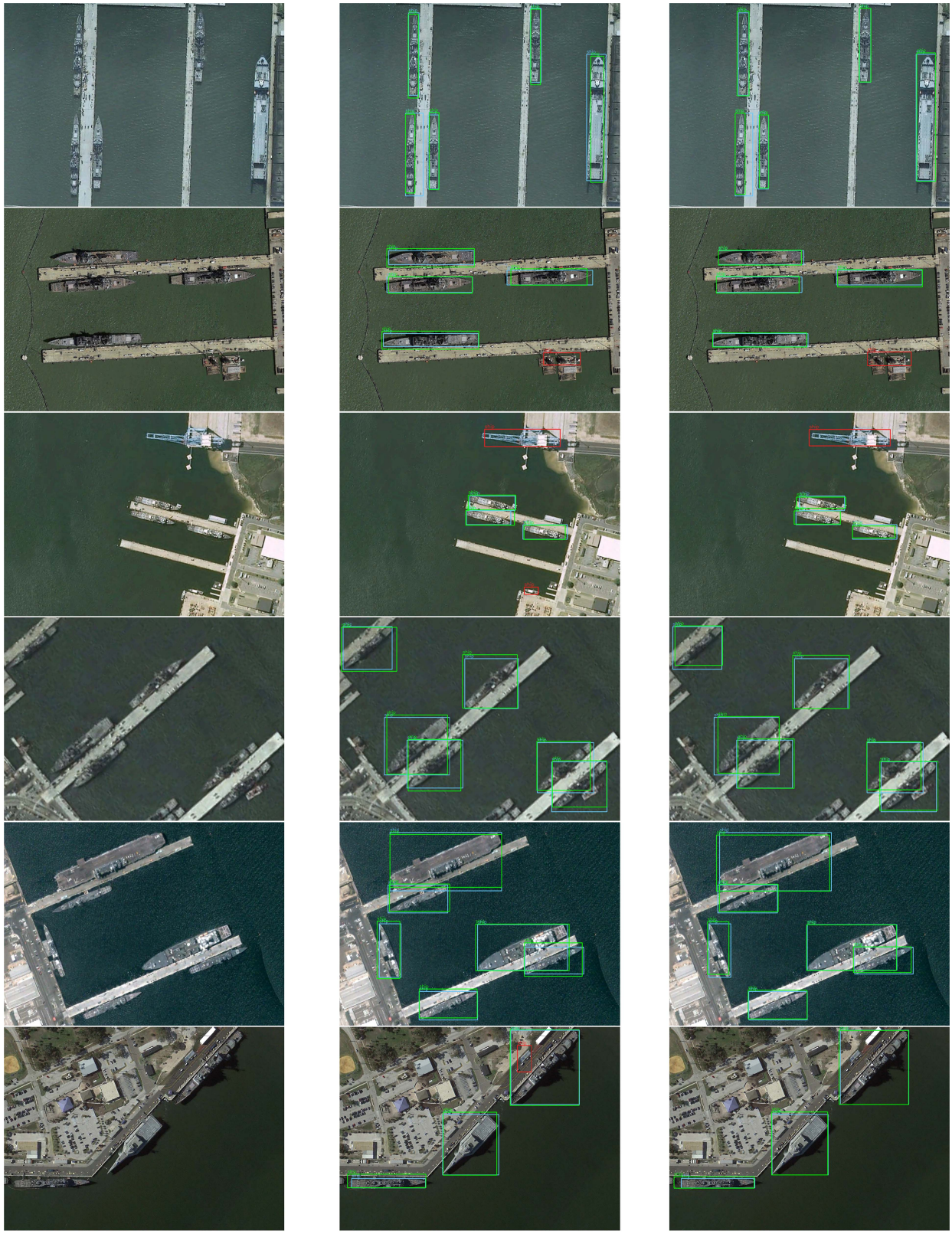
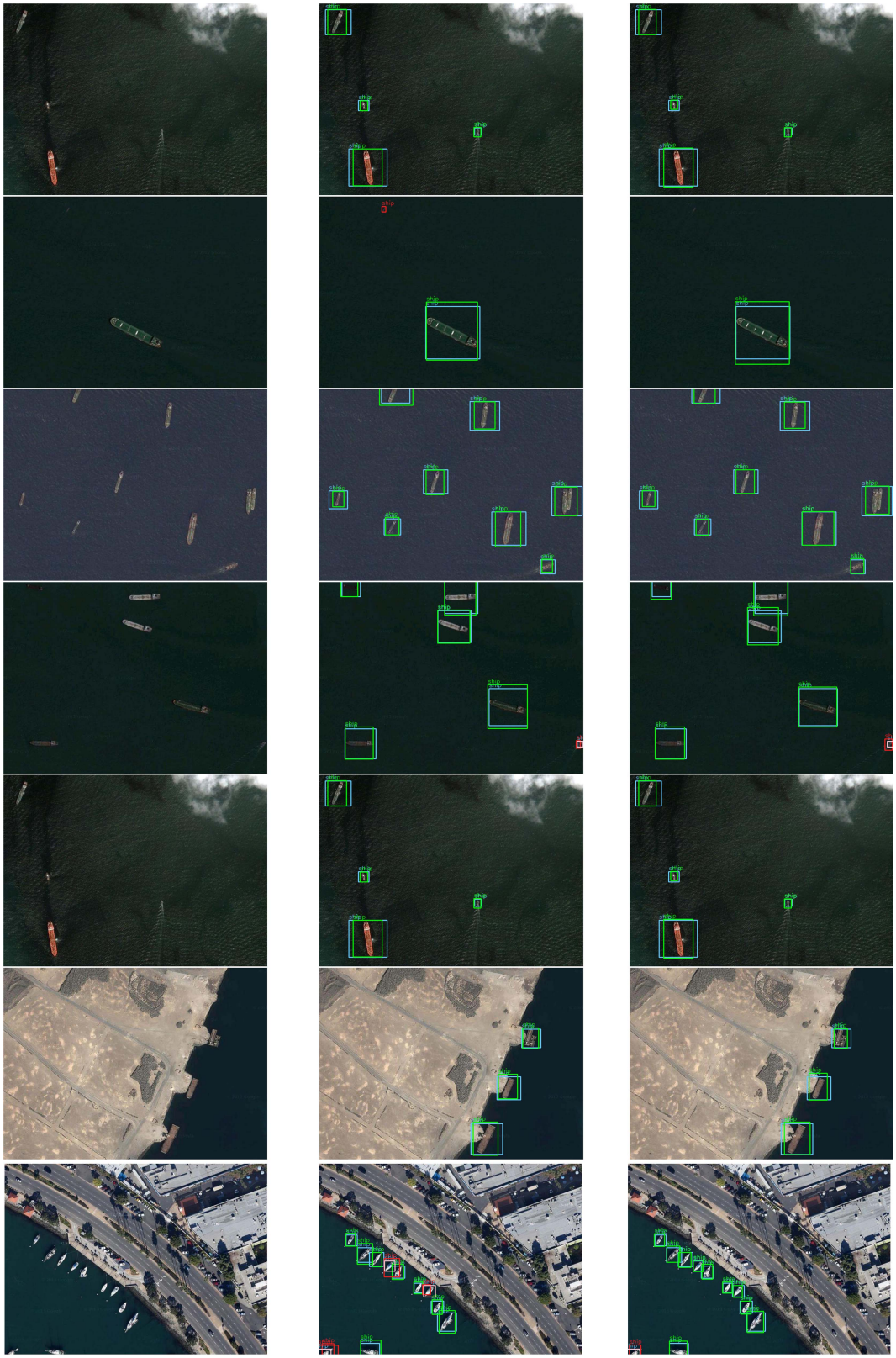


Fig. 15. Comparison of ship detection results under different methods in the HRSC2016 test sets. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN. (a) Original scenes. (b) Ship detection results of CenterNet based on images. (c) Ship detection results of CS-CenterNet at MRs $= 25\%$ based on CS measurements.



(a) (b) (c)

Fig. 16. Comparison of ship detection results under different methods in the LEVIR test sets. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN. (a) Original scenes. (b) Ship detection results of CenterNet based on images. (c) Ship detection results of CS-CenterNet at MRs $= 25\%$ based on CS measurements.

TABLE VII
PARAMETER SIZE OF THE CS-CENTERNET AND CENTERNET

Model	Total params
CenterNet for ORS images	95 503 020
CS-CenterNet for CS measurements of ORS scenes (MRs = 10%)	101 518 481
CS-CenterNet for CS measurements of ORS scenes (MRs = 25%)	101 523 625

According to Table V, the CS-CenterNet at MRs = 25% score is 84.35% in terms of detection precision, and the CS-CenterNet score is 92.19% in terms of recall. The precision and recall of the CS-CenterNet can get its F1, which is 0.8810. In terms of AP, the CS-CenterNet score is 90.76%. The P, R, F1, and AP of CS-CenterNet at MRs = 25% is higher than that of CenterNet. According to Fig. 15, we can find that CS-CenterNet has better ship detection performance in terms of visual quality.

According to Table VI, the CS-CenterNet at MRs = 25% score is 70.60% in terms of detection precision, and the CS-CenterNet score is 78.20% in terms of recall. The precision and recall of the CS-CenterNet can get its F1, which is 0.7421. In terms of AP, the CS-CenterNet score is 75.44%. The R, F1, and AP of CS-CenterNet at MRs = 25% is higher than that of CenterNet and the P of CS-CenterNet at MRs = 25% is basically the same as that of CenterNet. According to Fig. 16, we can find that CS-CenterNet has better ship detection performance in terms of visual quality.

It is worth noting that the quantitative indicators of the ship detection results in the LEVIR are lower than that of HRSC2016, because the ship targets in the LEVIR are smaller and denser.

Although the data volume of the CS measurements is only 25% of the original scenes, CS-CenterNet has detection P, R, F1, and AP that will not decrease but slightly increase compared with the CenterNet. This is because the backbone network HgN can fully extract the feature information in the CS measurements and the SENet added to it can improve the accuracy of ship detection. In addition, FRNet in THBN can refine ship features, which again improves the accuracy of ship detection. According to Figs. 15 and 16, we can find that CS-CenterNet has better ship detection performance in terms of quantitative indicators visual quality.

In short, for the innovation pipeline to complete the ship detection task in CS-based ORS imaging system, CS-CenterNet can directly detect ships on CS measurements while ensuring the quality of the detection.

In addition, we also tested the parameter size of the CS-CenterNet and CenterNet. Table VII shows the parameter size of them.

The parameter quantity of CS-CenterNet is much higher than that of CenterNet because CS-CenterNet detects CS measurements while CenterNet detects scenes. Therefore, the complexity of the former feature detection network OHgN and feature prediction network OTBHN is higher than that of the latter feature detection network ResNet50 and feature prediction network TBHN.

(a) (b)

Fig. 17. Ship detection results with HgN compared to ResNet50. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN. With HgN, the correct ship position can be detected effectively. (a) ResNet50+TBHN. (b) HgN+TBHN. (a) Ship detection results on CS measurements of OHgN+TBHN. (b) Ship detection results on CS measurements of HgN+TBHN.

In addition, the parameter quantity of CS-CenterNet at MRs = 25% is lower than that of CS-CenterNet at MRs = 10%, which is caused by different data dimensions of CS measurements. The input of CS-CenterNet at MRs = 25% is the measurements with dimension $256 \times 256 \times 3$, and the input of CS-CenterNet at MRs = 10% is the measurements with dimension $256 \times 256 \times 1$.

E. Ablation Studies

To verify the effect of the HgN, SENet, and FRNet modules in CS-Center, we conduct ablation studies on these three modules. The MRs of the experiment in this subsection are set to 25%.

1) Effect of HgN: To evaluate the performance of HgN, we conduct ablation experiments on HgN, and the corresponding experimental results are shown in the first and second rows of Table VIII and Fig. 17. In the first and second rows of Table VIII, "ResNet50 [38]+TBHN" and "HgN [18]+TBHN" can analyze the HgN performance. It can be seen that P increased by 4.47%, R increased by 9.66%, F1 increased by 0.0671, and AP increased by 11.16%. According to Fig. 17, we can find that using HgN as the backbone has better ship detection performance in terms of visual quality, especially the correct ship position can be detected effectively.

Therefore, using HgN as the backbone can achieve better detection accuracy. This is because HgN can capture global and local features from CS measurements.

2) Effect of SENet: To evaluate the performance of SENet, we conduct ablation experiments on SENet, and the corresponding experimental results are shown in the second and third rows of Table VIII and Fig. 18. In the second and third rows of Table VIII, "HgN+TBHN" and "OHgN+TBHN" denote that the backbone network is different to analyze the SENet performance. The detected comprehensive indicators F1 and AP are improved. According to Fig. 18, we can find that using OHgN as the backbone has better ship detection performance in terms of visual quality, especially the false detection of ships can be effectively reduced.

Therefore, using OHgN as the backbone can achieve better detection accuracy. This is because SENet in OHgN can focus on the salient areas that contain ships in the compressive measurements.

TABLE VIII
SHIP DETECTION RESULTS ON CS MEASUREMENTS OF ORS SCENES IN THE HRSC2016 DATASET AT MRS = 25%

Model	P(%)	R(%)	F1	AP(%)
ResNet50+TBHN	72.55%	82.53%	0.7722	79.80%
HgN+TBHN	77.02%	92.19%	0.8393	90.96%
OHgN+TBHN	77.64%	**92.94%**	0.8462	**91.63%**
OHgN+OTBHN	**84.35%**	92.19%	**0.8810**	90.76%

The best results are highlighted in bold.

Fig. 18. Ship detection results with OHgN compared to HgN. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN. With SENet, the false detection of ships can be effectively reduced. (a) HgN+TBHN. (b) OHgN+TBHN. (a) Ship detection results on CS measurements of HgN+TBHN. (b) Ship detection results on CS measurements of OHgN+TBHN.

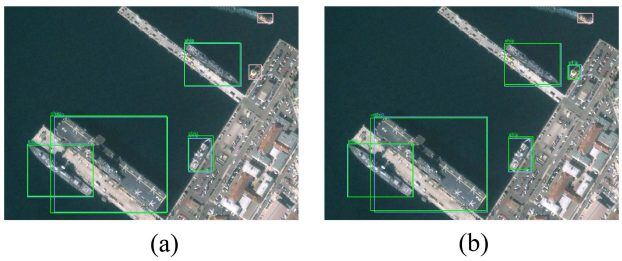


Fig. 19. Ship detection results with OTBHN compared to TBHN. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN. With FRNet, more small ships can be detected effectively. (a) OHgN+TBHN. (b) OHgN+OTBHN. (a) Ship detection results on CS measurements of OHgN+TBHN. (b) Ship detection results on CS measurements of OHgN+OTBHN.

3) Effect of FRNet: To evaluate the performance of FRNet, we conducted ablation experiments on FRNet, and the corresponding experimental results are shown in the third and fourth rows of Table VIII and Fig. 19. In the third and fourth rows of Table VIII, "OHgN+TBHN" and "OHgN+OTBHN" can analyze the FRNet performance. It can be seen that the P and F1 of the network are greatly improved. According to Fig. 19, we can find that using OTHBN as the prediction has better ship detection performance in terms of visual quality, especially more small ships can be detected effectively.

Therefore, using OTHBN as the prediction can achieve better detection accuracy. This is because FRNet can refine the ship features.

F. Discussion

1) Ship Detection Performance of CS-CenterNet Under Different MRs: When CS-based ORS imaging system obtains scene compression sampling data, MRs denote the ratio of the amount of compressive measurement data obtained by the imaging system to the amount of original scene data. As explained in Section III-B, the size $w_m \times h_m \times MRsB^2Dp$ of the CS measurements is related to the MRs in the block compression sampling. Here, we test the effect of ship detection performance of CS-CenterNet under different MRs. The result is shown in Table IX. From Table IX, we can find that the ship detection performance of CS-CenterNet at MRs $= 10\%$ is worse than the ship detection performance of CS-CenterNet at MRs $= 25\%$.

TABLE IX
EXPERIMENT RESULTS OF CS-CENTERNET UNDER DIFFERENT MRS IN HRSC2016 DATASET

MRs	P(%)	R(%)	F1	AP(%)
25%	**84.35%**	**92.19%**	**0.8810**	**90.76%**
10%	81.85%	88.85%	0.8521	87.52%

The best results are highlighted in bold.

TABLE X
EXPERIMENT RESULTS OF CS-CENTERNET UNDER DIFFERENT $B \times B$ AT MRS = 25% IN HRSC2016 DATASET

$B \times B$	P(%)	R(%)	F1	AP(%)
2×2	**84.35%**	**92.19%**	**0.8810**	**90.76%**
4×4	83.74%	89.96	0.8674	88.86%

The best results are highlighted in bold.

This is because as the amount of acquired scene data decreases, the features of ships in the CS measurements decrease, which leads to a decrease in the detection performance of ships.

2) Ship Detection Performance of CS-CenterNet Under Different $B \times B$: In CS-CenterNet, we adopt CML to perform block compression sampling processing on the ORS scene. As explained in Section III-B, the resolution size $\frac{W}{B} \times \frac{H}{B}$ of the CS measurements is related to the block size $B \times B$ in the block compression sampling. Here, we test the effect of ship detection performance of CS-CenterNet under different $B \times B$. It can be seen from Table X that the block size of 2×2 will obtain the best ship detection results.

This is because the resolution of the CS measurements obtained by the block size of 2×2 is $\frac{W}{2} \times \frac{H}{2}$, which is higher than the resolution $\frac{W}{4} \times \frac{H}{4}$ of the CS measurements obtained by the block size of 4×4. The high-resolution CS measurements are more conducive to the backbone network to extract ship feature information, and improving the accuracy of ship detection.

3) Failure Cases of Our CS-CenterNet: Fig. 20(a)–(d) shows some failure cases of our CS-CenterNet at MRs $= 25\%$. CS-CenterNet is not sensitive to small ships. For example, the two failure cases (a) and (b) reflect that the prediction box of CS-CenterNet cannot detect some small ships. Moreover,

(a) (b) (c) (d)

Fig. 20. Some failure cases of CS-CenterNet at MRs = 25%. The blue box denotes the ground truth, the green box denotes TP, the red denotes FP, and the pink denotes FN. (a) Failure case reflects that the prediction box cannot detect some small ships. (b) Failure case reflects that the prediction box cannot detect some small ships. (c) Failure case reflects that the prediction box incorrectly detects long objects as ship objects. (d) Failure case reflects that the prediction box incorrectly detects small bases on the shore as ship objects. (a) Prediction box cannot detect some small ships. (b) Prediction box cannot detect some small ships. (c) Prediction box incorrectly detects long objects as ship objects. (d) Prediction box incorrectly detects small bases on the shore as ship objects.

CS-CenterNet is sensitive to ship-like objects. For example, the two failure cases (c) and (d) reflect that the prediction box of CS-CenterNet incorrectly detects long objects and small bases on the shore as ship objects.

In future work, it is of great significance to set appropriate hyperparameters according to the ship object and improve the distinguishing ability of the model.

4) Limitations: To simulate the compression sampling process of CS-based ORS imaging system, we use CML to compress the scene to obtain CS measurements. However, this method of acquisition is ideal. In future work, we will obtain the CS measurements on the physical platform of a CS-based ORS imaging system.

V. Conclusion

This article proposes an efficient model, CS-CenterNet, for ship detection on CS measurements of ORS scenes. Specifically, our model uses CML to perform convolutional coding on the scene to obtain the CS measurements, which simulates the block compression sampling process in CS-based ORS imaging system. A OHgN is designed, which can effectively extract the high-resolution feature information of measurements. A OTBHN is designed, which can refines the ship features and perform feature prediction with high accuracy. Experiments based on the HRSC2016 dataset show that the detection precision of our model for the detection of ships with measurements in ORS scenes is 84.35%, the recall is 92.19%, the F1 value is 0.8810, and the AP value is 90.76%. Therefore, it can achieve high-accuracy ship detection on CS measurements of ORS scenes. In the future, we will try to perform experiments of measurements' ship detection on the physical platform of CS-based ORS imaging system (such as the CS-based ORS camera). What is more, we will continue to study the basic theories of deep learning (DL) to better design the network structure and the detection accuracy of small ships.

References

[1] Y. D. Yu, X. B. Yang, S. J. Xiao, and J. L. Lin, "Automated ship detection from optical remote sensing images," in *Proc. Int. Conf. Adv. Mater. Microw. Opt.*, Bangkok, Thailand, 2011, pp. 785–791.

[2] Z. K. Liu, H. Z. Wang, L. B. Weng, and Y. P. Yang, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, Aug. 2016.

[3] G. Melillos et al., "The use of remote sensing for maritime surveillance for security and safety in Cyprus," in *Proc. Conf. Detection Sens. Mines, Explosive Objects, Obscured Targets XXV*, vol. 11418, 2020, pp. 141–152.

[4] S. K. Jayaweera and S. K. Jayaweera, *NYQUIST Sampling Theorem (Signal Processing for Cognitive Radios)*. Hoboken, NJ, USA: Wiley, 2015, pp. 704–710.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] R. G. Baraniuk, "A lecture on compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[7] J. Y. Liu and J. B. Zhu, "Design of remote sensing imaging system based on compressive sensing," *J. Syst. Eng. Electron.*, vol. 32, no. 8, pp. 1619–1624, 2010.

[8] L. Jiying, Z. Jubo, Y. Fengxia, and Z. Zenghui, "Theoretical frameworks of remote sensing systems based on compressive sensing," in *Proc. ISPRS Tech. Commission VII Symp. – 100 Years ISPRS – Advancing Remote Sens. Sci.*, Vienna, Austria, 2010, pp. 77–81.

[9] S. Lohit, K. Kulkarni, R. Kerviche, P. Turaga, and A. Ashok, "Convolutional neural networks for noniterative reconstruction of compressively sensed images," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 326–340, Sep. 2018.

[10] H. T. Yao, F. Dai, S. L. Zhang, Y. D. Zhang, Q. Tian, and C. S. Xu, "DR2-Net: Deep residual reconstruction network for image compressive sensing," *Neurocomputing*, vol. 359, pp. 483–493, Sep. 2019.

[11] H. Y. Guo, X. Yang, N. N. Wang, B. Song, and X. B. Gao, "A rotational Libra R-CNN method for ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5772–5781, Aug. 2020.

[12] Z. Q. Wang, Y. Zhou, F. T. Wang, S. X. Wang, and Z. Y. Xu, "SDGH-Net: Ship detection in optical remote sensing images based on Gaussian heatmap regression," *Remote Sens.*, vol. 13, no. 3, Feb. 2021, Art. no. 499.

[13] Q. Wang, F. Shen, L. Cheng, J. Jiang, and Z. Mao, "Ship detection based on fused features and rebuilt YOLOv3 networks in optical remote-sensing images," *Int. J. Remote Sens.*, vol. 42, no. 2, pp. 520–536, 2021.

[14] J. M. Fu, X. Sun, Z. R. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.

[15] W. Shi, F. Jiang, S. Liu, and D. Zhao, "Image compressed sensing using convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, pp. 375–388, 2019.

[16] W. Z. Shi, F. Jiang, S. H. Liu, and D. B. Zhao, "Multi-scale deep networks for image compressed sensing," in *Proc. 25th IEEE Int. Conf. Image Process.*, Athens, Greece, 2018, pp. 46–50.

[17] S. M. Xiao, S. J. Wang, and L. Chang, "Image reconstruction based on fused features and perceptual loss encoder-decoder residual network for space optical remote sensing images compressive sensing," *IEEE Access*, vol. 9, pp. 50413–50425, 2021.

[18] A. Newell, K. U. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Comput. Vis.*, vol. 9912, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Lecture Notes in Computer Science, Cham, Switzerland: Springer, 2016, pp. 483–499.

[19] X. Zhou, D. Wang, and P. Krhenbühl, "Objects as points," 2019, *arXiv:1904.07850*.

[20] Z. Liu, H. Wang, L. Weng, and Y. Yang, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, Aug. 2016.

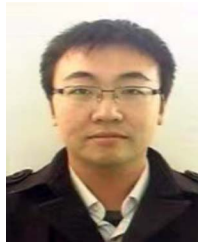
[21] Z. X. Zou and Z. W. Shi, "Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1100–1111, Mar. 2018.
[22] Z. F. Zhao, X. M. Xie, C. Y. Wang, S. Y. Mao, W. Liu, and G. M. Shi, "ROI-CSNet: Compressive sensing network for ROI-aware image recovery," *Signal Process. Image Commun.*, vol. 78, pp. 113–124, Oct. 2019.
[23] W. Z. Shi, S. H. Liu, F. Jiang, and D. B. Zhao, "Video compressed sensing using a convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 425–438, Feb. 2021.
[24] Y. X. Shan, X. M. Zhou, S. H. Liu, Y. F. Zhang, and K. Huang, "SiamFPN: A deep learning method for accurate and real-time maritime ship tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 315–325, Jan. 2021.
[25] R. Calderbank, S. Jafarpour, and R. Schapire, "Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain," 2009.
[26] S. Lohit, K. Kulkarni, and P. Turaga, "Direct inference on compressive measurements using convolutional neural networks," in *Proc. 23rd IEEE Int. Conf. Image Process.*, Phoenix, AZ, USA, 2016, pp. 1913–1917.
[27] E. Zisselman, A. Adler, and M. Elad, "Compressed learning for image classification: A deep neural network approach," in *Processing, Analyzing and Learning of Images, Shapes, and Forms: Pt 1*, Handbook of Numerical Analysis, vol. 19, R. Kimmel and X. C. Tai, Eds., 2018, pp. 3–17.
[28] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "ReconNet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 449–458.
[29] X. Sun, P. J. Wang, C. Wang, Y. F. Liu, and K. Fu, "PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 50–65, Mar. 2021.
[30] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "Foveabox: Beyound anchor-based object detection," *IEEE Trans. Imag. Process.*, vol. 29, pp. 7389–7398, 2020.
[31] Q. B. He, X. Sun, Z. Y. Yan, and K. Fu, "DABNet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5601216.
[32] J. F. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, New York, NY, USA, 2017, pp. 764–773.
[33] T. Y. Lin, P. Goyal, R. Girshick, K. M. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[35] W. Liu et al., "Ssd: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Springer, Cham, 2016, pp. 21–37.
[36] 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, 2015.
[37] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[38] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, 2016, pp. 770–778.

Shuming Xiao was born in Weifang, China, in 1997. He received the B.Eng. degree in mechanical engineering and automation from Northeast Agricultural University, Harbin, China, in 2019. He is currently working toward the Ph.D. degree in mechatronics engineering with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China.

His research interests include machine vision, compressive sensing, deep learning, and image processing.

Ye Zhang was born in Changchun, China, in 1981. She received the B.Eng. degree in information science from Jilin University, Jilin, China, in 2003, and the Ph.D. degree in mechatronics engineering from the University of Chinese Academy of Sciences, Changchun, China, in 2008.

She is currently a Professor with the State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. Her research interests include computer vision, pattern recognition, and machine learning.



Xuling Chang was born in Changchun, in 1990. He received the B.Eng. degree in software engineering from Jilin University, Jilin, China, in 2012, and the M.D. degree in computer science from Shanghai Jiaotong University, Shanghai, China, in 2015.

He is currently an Assistant Professor with the State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His research interests include computer vision and image processing.