Dual-Attention-Driven Multiscale Fusion Object Searching Network for Remote Sensing Imagery

Haolong Fu, Qingpeng Li, *Member, IEEE*, Puhong Duan, *Member, IEEE*, Jiacheng Lin, Renwei Dian, *Member, IEEE*, Shutao Li, *Fellow, IEEE*, Xudong Kang, *Senior Member, IEEE*, and Zhiyong Li, *Member, IEEE*

***Abstract*—Object search is a challenging yet important task. Many efforts have been made to address this issue and achieve great progress in natural image, yet searching all the specified types of objects from remote sensing image is barely studied. In this article, we are interested in searching objects from remote sensing images. Compared to person search in natural scenes, this task is challenging in two factors: One is that remote image usually contains a large number of objects, which poses a great challenge to characterize the object features; another is that the objects in remote sensing images are dense, which easily yield erroneous localization. To address these issues, we propose a new end-to-end deep learning framework for object search in remote sensing images. First, we propose a multiscale feature aggregation module, which strengthens the representation of low-level features by fusing multilayer features. The fused features with richer details significantly improve the accuracy of object search. Second, we propose a dual-attention object enhancement module to enhance features from channel and spatial dimensions. The enhanced features significantly improve the localization accuracy for dense objects. Finally, we built two challenging datasets based on the remote sensing images, which contain complex changes in space and time. The experiments and comparisons demonstrate the state-of-the-art performance of our method on the challenging datasets.**

***Index Terms*—Attention-based module, deep learning, object searching, remote sensing image.**

I. Introduction

CURRENTLY, searching for designated objects in natural images has become increasingly popular in computer vision [1], [2], [3]. For example, searching a specific person through video surveillance is a very important computer vision application. There are similar needs in the remote sensing field. Searching designated objects from remote sensing images at different times and locations is of great significance to the national defense security fields and land resource management. However, there is also a lack of effective methods for remote sensing object searching.

In recent years, many person searching methods have been proposed [1], [3]. For example, Yan et al. [4] proposed a person searching network with contextual information to effectively improve the robustness of search results. Xiao et al. [5] proposed an individual aggregation network and accurately localized persons by learning to minimize variations in internal features. To share the research results of reidentification with person searching, Liu et al. [6] transferred the advanced person reidentification knowledge to the person searching model through a teacher-guided disentangling network, which significantly improved the person search performance. Most of the above methods improve the search performance by a shallow backbone or preserving shallow information.

As shown in Fig. 1, there is similarity between person searching and remote sensing object searching [7], [8]. They aim to search the given object from a larger number of images. In the meanwhile, there are some differences between person search in natural image and object search in remote sensing image. Specifically, the remote sensing object searching task often requires locating all the specified types of objects from a gallery containing many similar objects, rather than searching a single specific object. In addition, the objects in the remote sensing scene are numerous and dense. Person searching methods are prone to localization errors and detection confusion in remote sensing scenes.

In summary, there are still some problems in solving the remote sensing object searching problem as follows.

1) Current person searching methods mainly focus on extracting high-level semantic information to represent the person in natural images since an image in person search

Manuscript received 12 May 2022; revised 24 June 2022 and 29 August 2022; accepted 9 September 2022. Date of publication 16 September 2022; date of current version 27 September 2022. This work was supported in part by the National Key R&D Program of China under Grant 2021YFA0715203, in part by the National Key Research and Development Program of China under Grant 2018YFB1308604, in part by the National Natural Science Foundation of China under Grant U21A20518 and Grant 61976086, in part by the State Grid Science and Technology Project under Grant 5100-202123009A, in part by the Special Project of Foshan Science and Technology Innovation Team under Grant FS0AA-KJ919-4402-0069, in part by the National Natural Science Foundation of China under Grant 62201205, in part by the Changsha Natural Science Foundation of China under Grant kq2202170, and in part by the Changsha Natural Science Foundation under Grant kq2202171. *(Corresponding author: Zhiyong Li.)*

Haolong Fu, Jiacheng Lin, and Zhiyong Li are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: haolongfu@hnu.edu.cn; jcheng_lin@hnu.edu.cn; zhiyong.li@hnu.edu.cn).

Qingpeng Li and Xudong Kang are with the School of Robotics, Hunan University, Changsha 410082, China (e-mail: liqingpeng@hnu.edu.cn; xudong_kang@163.com).

Puhong Duan and Shutao Li are with the College of Electrical and Information Engineering and the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China (e-mail: puhong_duan@hnu.edu.cn; shutao_li@hnu.edu.cn).

Renwei Dian is with the School of Robotics and the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China (e-mail: drw@hnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3207302

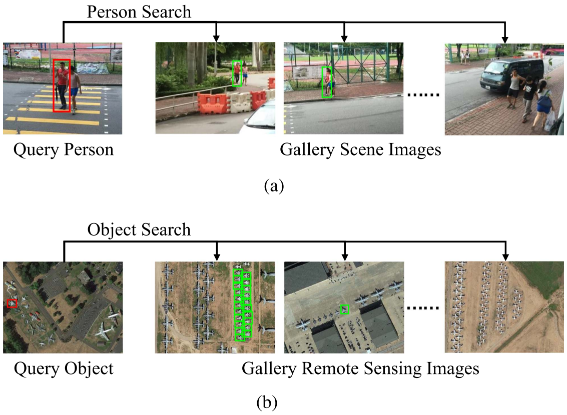


Fig. 1. Example illustrating the difference between remote sensing object searching and person searching. (a) Person searching in natural images. (b) Object searching in remote sensing images. (a) Example of Person Searching. (b) Example of Remote Sensing Object Searching.

datasets only contains a few persons. However, remote sensing images contain multiple objects. Therefore, the existing person searching methods cannot work well in characterizing the significant features of objects, which is prone to produce unsatisfactory search performance.

2) Different from person search in natural image, remote sensing image object search is a challenging task since the distance between objects in remote sensing images is very small. In this situation, it is difficult to locate all the dense objects by using person search methods. The main reason is that these methods fail to emphasize informative features of dense objects.
3) The dataset for object search is scarce in remote sensing field. Although there are many datasets for object detection or segmentation from remote sensing images, such as DOTA [9], DIOR [10], and SZTAKI-INRIA [11], these datasets cannot directly applied in the object search problem, since most of objects in these datasets are not of the same type.

To address these problems, we design a new remote sensing object searching network and propose two modules for it. The contributions of this study can be summarized as follows.

1) A dual-attention-driven multiscale fusion object searching network is proposed for remote sensing images. To the best of our knowledge, this is the first time to study the object searching task in remote sensing community.
2) The proposed searching network is comprised of two modules: a) the multiscale feature aggregation (MSFA) module enhances the detail representation of features by fusing multilayer features, which improves the performance of object search in remote sensing images; and b) the dual-attention object enhancement (DAOE) module is proposed to focus on essential salient objects and suppress unnecessary ones, which contributes to making more accurate and precise landmark searching.
3) To verify the effectiveness of the proposed method, two object searching datasets are built by ourselves, which help to promote the development of object searching methods in remote sensing field.

The rest of this article is organized as follows. In Section II, we briefly review related work about several relevant topics. In Section III, we present the dual-attention-driven multiscale fusion object searching network and discuss its architecture and loss function. In Section IV, we present extensive experimental results. Finally, Section V concludes this article.

II. Related Work

The proposed object searching method organically integrates detection and reidentification. Thus, we briefly introduce object detection and reidentification.

A. *Object Detection*

Most detection algorithms can be divided into two-stage and single-stage methods according to the strategy of generating proposals [10]. The two-stage detection algorithm first generates a series of candidate boxes as samples and then classifies samples through the convolutional neural network. Common algorithms include R-CNN [12], Fast R-CNN [13], and Faster R-CNN [14]. The second category is represented by the YOLO series algorithm [15], [16], [17], [18] and SSD [19], which transform the positioning problem of the object box into a regression problem for processing. In addition, because candidate boxes are not needed, these methods have a marked advantage in inference time during testing. However, these object detectors are designed for images in natural scenes, the resulting object's direction of remote sensing images typically exhibits high uncertainty, and object scales varies widely. These problems lead to these methods not having good adaptability to remote sensing images.

For remote sensing image object detection, many scholars have proposed solutions to address these problems [20], [21], [22], [23]. The problem of uncertain object direction can be solved by rotating the frame scheme [24], [25], [26]. Cheng et al. [27] proposed a new optimization function by introducing rotation-invariance regularization and Fisher discriminant regularization to CNN features to solve the problem of low detection accuracy caused by object rotation in remote sensing images. Zhou et al. [28] proposed an encoder–encoder structure, where the rotation-sensitive feature maps are used for regression and the rotation-invariance feature maps are used for classification. Chen et al. [29] proposed a new pixel-IoU loss to effectively improve the detection performance. Xie et al. [30] proposed a remote sensing object detector. The accuracy can be comparable to that of the two-stage detector and its speed can be comparable to that of the one-stage detector. However, these rotating object detection methods only solve the problem of remote sensing object rotation, particularly improving the detection accuracy in dense scenes. However, they do not involve the negative impact of common factors such as illumination and weather in remote sensing images. Conversely, the feature pyramid networks (FPNs) proposed by Lin et al. [31] provide a good solution to the problem of scale disunity and environmental impact. In some studies, more complex pyramid structures are constructed to integrate multiscale feature layer information [32], [33], [34]. To improve detection accuracy, Yang et al. [35] propose a sampling fusion network by fusing a multilayer feature with effective anchor sampling, which effectively improves detection accuracy.

However, due to the doubling of the number of anchor points, the model's efficiency is low. Liu et al. [36] improve the feature representation ability of the backbone, adaptively combining multiscale features, and effectively reducing the interference of the background to the object, but this method has little effect on small objects.

Thus, most of these methods require a deep network structure to extract high-level semantic information, leading to a lack of low-level information for reidentification. However, the particularity of the searching task requires the unity of opposites between high-level and shallow information.

B. Reidentification

In recent years, due to the wide application of reidentification tasks in video surveillance and object tracking, many scholars have investigated reidentification in detail [37], [38], [39]. However, their research objects are more focused on pedestrians in natural image scenes. For example, by considering the spatial dependence in both interimages and intraimages, Si et al. [40] constructed a new spatially driven network that achieves good performance on multiple classic key indicators. Huang et al. [41] proposed a new full-scaled deep discriminant learning model, which considered the three concepts of depth, width, and cardinality concurrently. Under the condition of obtaining considerable accuracy, the structural complexity of the model and the difficulty of training were reduced. However, that study lacks background interference, leading to a large gap compared with the real scene, which limits the application scope. Therefore, to explore the real-world applications of pedestrian reidentification, researchers proposed a person searching task that aims to simultaneously locate and identify a person from the raw image [42], [43], [44]. For example, Han et al. [45] designed a trident network by dividing the person searching task into three parts: detection, reidentification, and part classification. Concurrently, the reidentification and part classification network weighted the gradient of backpropagation based on the quality of person detection. However, the network structure of this method is complex, which reduces computational speed. Li and Miao [46] account for the fact that the detection and reidentification in person searching is a gradual process through two subnetworks for sequential processing, and the contextual information is used to enhance reidentification. Although this method improves the searching speed, it fails to unify detection and identification tasks, and the two-step structure is still too complex.

Existing searching methods often perform one-to-one positioning and reidentification of pedestrian objects; only a single object with the same id appears in the image to be searched. However, objects with the same id often appear repeatedly in remote sensing scenes, and remote sensing images also pose new challenges, such as scale changes and weather effects.

III. Methodology

A. Method Overview

In this section, the proposed object searching method is introduced in detail. As shown in Fig. 2, the proposed object searching model consists of the following components: an MSFA module and a DAOE module. Specifically, we use the MSFA module to strengthen the feature representations by fusing more low-level features, and then, the extracted features are fed into the reidentification task and the detection task. In addition, the DAOE module is used to select task-related features and capture more spatial details, which contributes to making more accurate and precise landmark search. The following subsections elaborate on the details.

B. MSFA Module

As far as we know, the FPN structure is widely used to extract multiscale features of images because it can fuse feature maps with strong high-level semantic information and feature maps with weak low-level semantic information but rich spatial information. Although the FPN structure can fuse different levels of features, the simple merger is suboptimal due to there being a conflict between low-level and high-level information in the object searching. Therefore, the AFA feature aggregation module [1] inspires us to find a more suitable feature extraction module for remote sensing images.

Thus, we use the MSFA module, as shown in Fig. 2. The primary idea of this module is to fuse high-level semantic information to low-level semantic information, thus obtaining the low-level semantic information that can adapt to both reidentification and detection. Specifically, we use the $\{S_2, S_3, S_4\}$ feature maps from the Res-50 backbone, and MSFA outputs $\{C_2\}$. We only use $\{C_2\}$ to reidentification and detect, instead of using the characteristics of each layer as in the original FPN. Although this design will affect detection performance, it unifies the reidentification and detection tasks. We will show in Section IV that the proposed method achieves a good tradeoff between reidentification and detection subtasks.

Due to the broad imaging range of remote sensing imaging scenes, there will likely encompass a large number of dense objects. The reidentification subtask requires more detailed information to identify the objects. In the proposed method, we designed an MSFA module to improve feature representation ability. Specifically, we use 3×3 deformable Conv to extract features. The primary function of 3×3 deformable Conv is to reduce the channels of feature maps and adaptively adjust the receptive field on the obtained features that can pay more attention to the object itself, thereby reducing background interference. And then, a concatenation operation is used to fuse the top-down feature maps, which is an important step to connect high-level semantic information with low-level semantic information. Finally, we use 3×3 Conv to fuse the connected feature maps, thus generating feature mappings containing more detailed information for the reidentification and detection tasks. With the above three steps, we obtain the fusion features with more attention on object details.

C. DAOE Module

After processing the MSFA module, we obtain a multiscale feature map of the input remote sensing scene, which is used for both the reidentification and detection subtasks. In the object

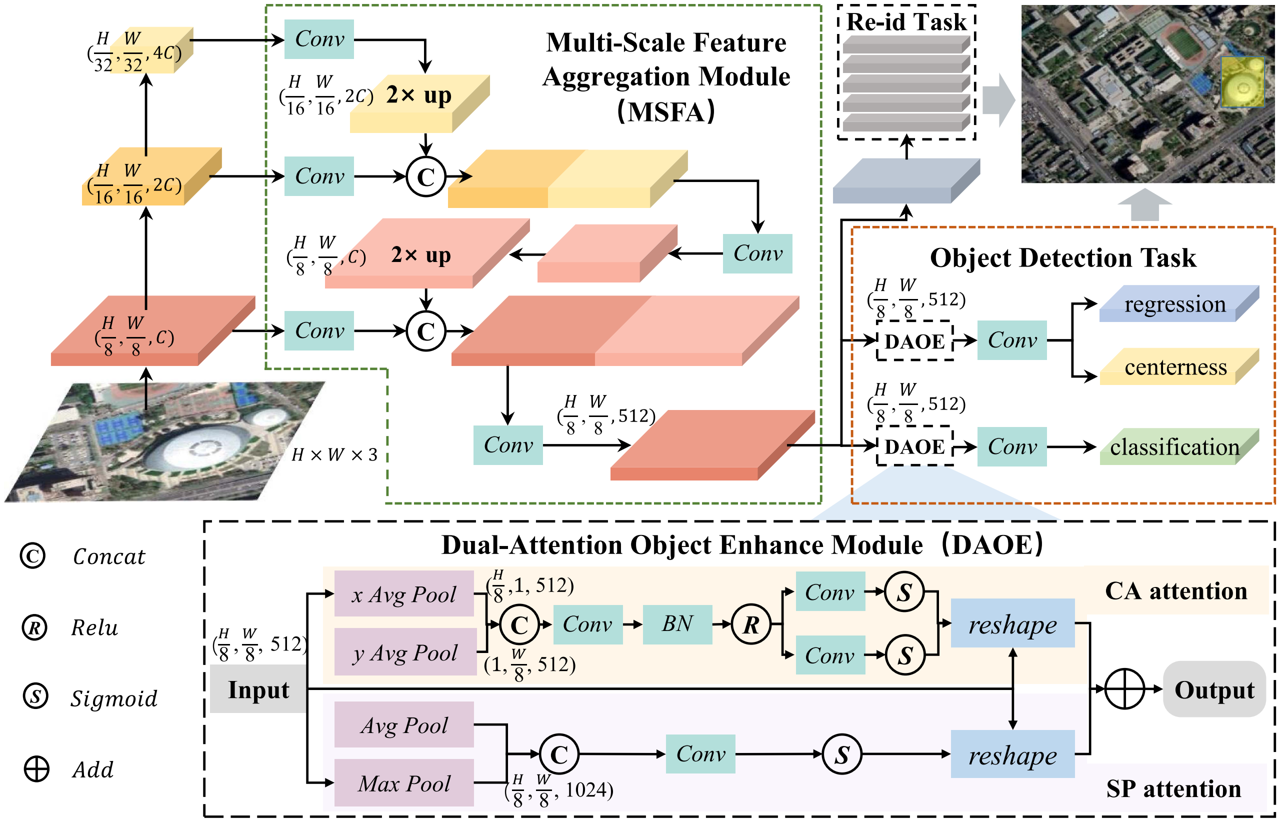


Fig. 2. Architecture of the proposed method, which includes three primary steps. First, we use the MSFA module to extract features for reidentification and detection. Second, the feature is flattened directly for the reidentification module. Third, the features are used for detection after the DAOE module.

searching task, accurate detection results will markedly improve the searching speed and accuracy. However, the uniqueness of the searching task leads to fewer feature layers for detection subtasks, which affects the accuracy of the detection subtask. Therefore, to obtain accurate search results in complex remotely sensed images, we propose new strategies to enhance the accuracy of object feature representation, thus obtaining more precise landmark. Inspired by Yang et al. [47], extracting rich global context information from multiscale maps is conducive to improving the ability to distinguish different elements in the scene. Therefore, we use the DAOE module to optimize feature representations from both the channel and spatial perspectives. As shown in Fig. 2, the DAOE module is composed of two parallel branches: the channel domain (CA attention) and the spatial domain (SP attention).

CA attention: After extensive testing, we found that CA attention [48] has better feature optimization capabilities. Thus, CA attention is used as the channel attention module in the proposed model. We first revisit CA attention. Specifically, the global pooling is broken down and converted to a one-to-one feature code. Given input $\mathbf{X}$, the pooling kernel of size $(H, 1)$ or $(1, W)$ is used to encode each channel along with horizontal and vertical coordinates, respectively. Therefore, the output of the cth channel at height H can be formulated as

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leqslant j < W} x_c(h, i). \tag{1}$$

Similarly, the output of the cth channel at width w can be written as

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leqslant j < H} x_c(j, w). \tag{2}$$

Then, this transformation is concatenated, and the 1×1 convolutional transformation function F_1 is used, yielding

$$\mathbf{f} = \delta\left(F_1\left(\left[\mathbf{z}^h, \mathbf{z}^w\right]\right)\right) \tag{3}$$

where $[\cdot, \cdot]$ is the concatenation operation along the spatial dimension and δ is a sigmoid function. Another two 1×1 convolutional transformations F_h and F_W are used to separately transform f^h and f^w to tensors with the same channel number as the input $\mathbf{X}$, yielding

$$\mathbf{g}^h = \delta\left(F_h\left(\mathbf{f}^h(\mathbf{X})\right)\right) \tag{4}$$

$$\mathbf{g}^w = \delta\left(F_w\left(\mathbf{f}^w(\mathbf{X})\right)\right) \tag{5}$$

where δ is the sigmoid function. The outputs $\mathbf{g}^h$ and $\mathbf{g}^w$ are then expanded and used as attention weights. Finally, the output of the CA attention module y_{ca} can be written as

$$y_{\text{ca}}(i, j) = x(i, j) \times g_c^h(i) \times g_c^w(j). \tag{6}$$

After obtaining sufficient global context information, the CA attention module also contains certain spatial information, but the spatial information is too weak. Thus, we must add spatial information to improve the optimization effect of the feature map.

SP attention: We use spatial attention to enhance the optimization effect of the CA attention module on the feature map. To learn the spatial weight relationships effectively, we first generate two feature descriptors of size $(H \times W \times 1)$ for each spatial position through global average pooling and global max-pooling operations. Next, the above two feature descriptors are concatenated, and then, the 7×7 convolutional transformation function F_2 is used, yielding

$$g_s(i,j) = \delta\left(F_2\left(\text{Avg}\left(x(i,j)\right), \text{Max}\left(x(i,j)\right)\right)\right)\right) \tag{7}$$

where $[\cdot,\cdot]$ is the concatenation operation along the spatial dimension and δ is the sigmoid function. The output of the SP attention module y_{sp} can be written as

$$y_{\text{sp}}(i,j) = x(i,j) \times g_s(i,j). \tag{8}$$

Finally, we combine the maps of the two branches to obtain the output of the module as

$$y(i,j) = y_{\text{ca}}(i,j) + y_{\text{sp}}(i,j). \tag{9}$$

From these processes, the final feature map is used for the detection subtask, which fuses the strong low-resolution semantic information and features with weak high-resolution semantic information but rich spatial information.

D. Loss Function

To train the proposed module, two parts of the loss function are used for the reidentification and detection subtasks. For reidentification loss, the TOIM loss proposed by Yan et al. [1] shows good performance in the reidentification task. They proposed a specifically designed triplet loss to improve the OIM loss. Specifically, the OIM loss stores the feature centers of all the labeled identities in a lookup table, $\mathbf{V} \in \mathbb{R}^{D\times L} = \{v_1, \ldots, v_L\}$. A circular queue $\mathbf{U} \in \mathbb{R}^{D\times Q} = \{u_1, \ldots, u_Q\}$ containing the features of $\mathbf{Q}$ unlabeled identities is maintained. At each iteration, given an input feature $\mathbf{x}$ with label i, the OIM loss computes the probability of $\mathbf{x}$ belonging to the identity i and is calculated as

$$p_i = \frac{\exp\left(v_i^T\right)/\tau}{\sum_{j=1}^{L} \exp(v_j^T x)/\tau + \sum_{k=1}^{Q} \exp(v_k^Q x)/\tau}. \tag{10}$$

The objective of the OIM loss is to minimize the expected negative log-likelihood

$$L_{\text{OIM}} = -\mathbf{E}_x[\log p_t], \quad t = 1, 2, \ldots, L. \tag{11}$$

For the specifically designed triplet loss, S vectors are sampled from one object, and then, $\mathbf{X_m} = \{x_{m,1}, \ldots, x_{m,S}, v_m\}$ and $\mathbf{X_n} = \{x_{n,1}, \ldots, x_{m,S}, v_n\}$ are described by the candidate feature sets for the object with identity labels m and n. Given $\mathbf{X_m}$ and $\mathbf{X_n}$, positive pairs can be sampled within each set, while negative pairs are sampled between the two sets. The triplet loss can be calculated as

$$L_{\text{tri}} = \sum_{\text{pos,neg}} [M + D_{\text{pos}} - D_{\text{neg}}]. \tag{12}$$

Finally, the TOIM loss is the summation of these two terms

$$L_{\text{TOIM}} = L_{\text{tri}} + L_{\text{OIM}}. \tag{13}$$

For the detection loss, we used the FCOS loss (L_{det}) to train the proposed detection head. The details are as follows:

$$L_{\text{det}} = \frac{1}{N_{\text{pos}}} \sum_{x,y} L_{\text{cls}}\left(p_{x,y}, c^*_{x,y}\right) + \frac{\lambda}{N_{\text{pos}}} \sum_{x,y} \mathbb{I}_{c^*_{x,y}>0} L_{\text{reg}}\left(t_{x,y}, t^*_{x,y}\right) \tag{14}$$

where L_{cls} is the focal loss, L_{reg} is the IOU loss, N_{pos} is the number of positive samples, and λ, where 1 is the balance weight for L_{reg}. The summation is calculated over all the locations on the feature maps F_i. $\mathbb{I}_{c^*_i>0}$ is the indicator function, with 1 if $c^*_i > 0$ and 0 otherwise.

Finally, the total loss is the summation of these three terms

$$L_{\text{total}} = L_{\text{tri}} + L_{\text{OIM}} + L_{\text{det}}. \tag{15}$$

Using this loss function, we optimize the parameter settings of the multiscale extraction network based on training data, thus obtaining the multiscale representation of the image scene with a smaller semantic gap.

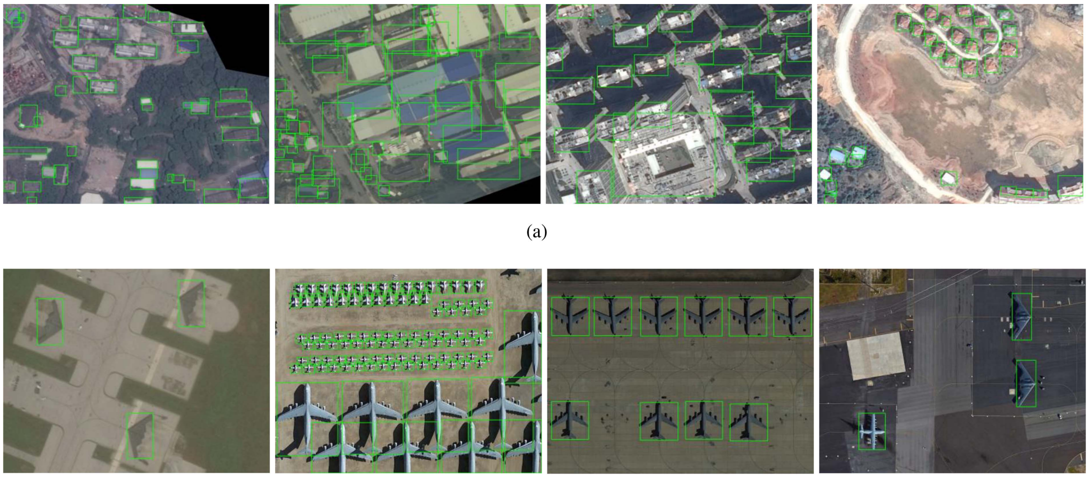
IV. Experiments

In this section, we first introduce the experimental details, including two self-built datasets and some implementation details. Then, ablation experiments are performed for the two modules proposed in this article, and the influence of each module on the final results is analyzed. Finally, the proposed method is compared with several other advanced methods to verify the performance of the proposed algorithm.

A. Experimental Settings

1) Datasets: Remote sensing object searching must find specific subclass objects from a gallery that contains a large number of images. However, this task is novel, and there is no public dataset; thus, to verify the effectiveness of the proposed method, we design many experiments with two self-labeled datasets. The gallery size is set as 50 images (i.e., finding the specified subclass object from 50 remote sensing images). The first annotated dataset is a building dataset from rural to urban areas, in which the scale span of the object is large and the difference between classes is large. The other dataset primarily includes aircraft in various remote sensing scenes, which means that there is interference from the angle, time, and artificial facilities in the second dataset.

a) Building dataset: This dataset was collected from Google Earth and includes different places in Fujian Province, China, from typical cities and suburbs to rural areas. It should be noted that we refer to the labeling method of [49], which identifies irregular buildings and connected concrete floors as a whole. This dataset includes 2180 images and more than 100 000 labeled buildings. Some examples of buildings are shown in Fig. 3(a). In this study, we divided the dataset according to 7:1:2. A total of 1526 images were used for training, 218 images were used for verification, and the remaining images were used for testing.



(a)

(b)

Fig. 3. Examples of annotated images. The first row shows some examples from the building dataset, including buildings from rural to urban areas. The second row shows the corresponding examples in the aircraft dataset, including weather changes as well as scale changes. (a) Samples of the building dataset. (b) Samples of the plane dataset.

TABLE I
DATASET OVERVIEW

Item	The building dataset	The plane dataset
Category	103 marked buildings	19 types of plane
Data Type	RGB	RGB
Image Size	1280 × 960	1280 × 980
Target Size (Avg.)	57.75 × 48.48	98.07 × 98.23
Total Target	148177	5359
Total Image	2180	500

b) Plane dataset: The second dataset collected 500 images of aircraft from Google Earth. The contents of the dataset are shown in Fig. 3(b). Fig. 3 shows that the proposed data are challenging, including object scale changes and unfavorable conditions for object searching (e.g., poor lighting and poor weather). For this small dataset, we used 400 images for training, 62 images for testing, and the remaining images for validation.

In particular, the dataset created in this article contains various scale changes and has a wide range of environmental impacts, which is challenging for remote sensing object searching. More details are shown in Table I.

2) Evaluation Metrics: There is a marked difference between the proposed searching method and object detection. Traditional evaluation metrics cannot be fully applied to this task. Thus, we propose a new evaluation metric for searching tasks. Different from the average precision (AP), the searching method must count the number of false detections and missed detections in the entire gallery corresponding to the query. The AP_s index is defined as

$$\mathrm{precision}_s = \frac{\mathrm{TP}_g}{\mathrm{TP}_g + \mathrm{FP}_g} \tag{16}$$

$$\mathrm{recall}_s = \frac{\mathrm{TP}_g}{\mathrm{TP}_g + \mathrm{FN}_g} \tag{17}$$

$$\mathrm{AP}_s = \int_0^1 \mathrm{precision}_s\left(\mathrm{recall}_s\right) d\left(\mathrm{recall}_s\right) \tag{18}$$

where TP_g, FP_g, and FN_g denote true positive, false positive, and false negative counts from the entire gallery, respectively. The higher the AP_s value is, the better the searching result.

3) Implementation Details: The proposed model is implemented using PyTorch and MMDetection on an Nvidia RTX 5000 GPU. We set the batch size to 4 and use an SGD optimizer with a weight decay of 0.0005. The initial learning rate is set to 0.001 and is reduced by a factor of 10 at epochs 20 and 22, with a total of 24 epochs.

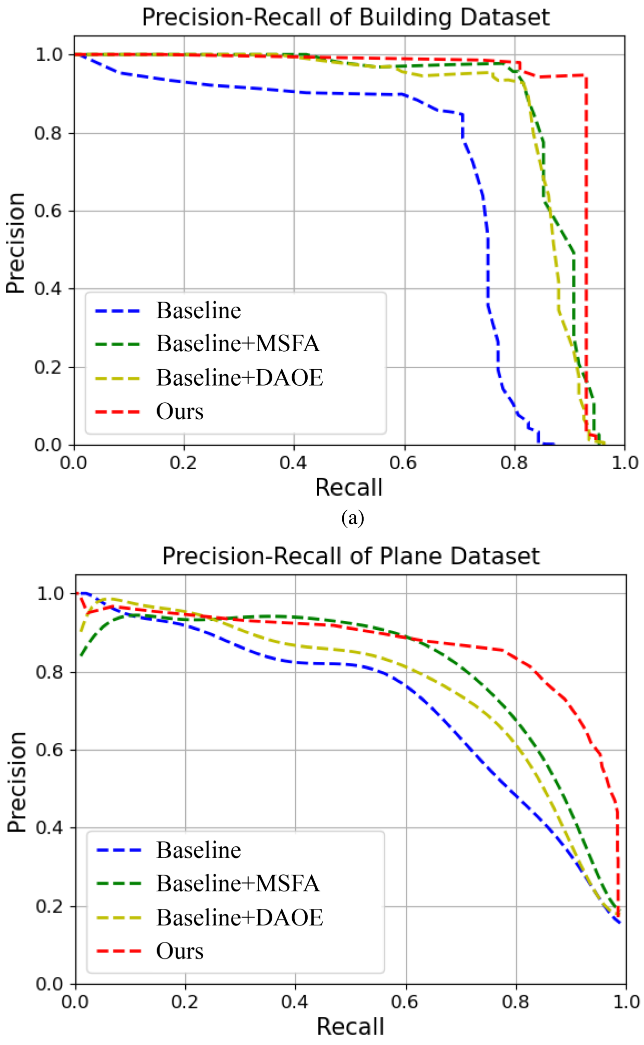
B. Results and Discussion

1) Ablation Experiments: To evaluate the performance of the proposed method and measure the contribution of each proposed module. The high-quality AlignPS structure proposed in the literature [1] was used as a baseline, and componentwise experiments were performed on the two datasets. The proposed modules were configured on different branches of the baseline, and their respective contributions to the final results were analyzed by comparing the evaluation of these modules before and after being used in Table II.

For the building dataset, Table II shows the evaluation results of the ablation experiments on the building dataset. The $\mathrm{AP}_{s\,50:95}$ evaluation index of the basic AlignPS network is 63.71%. $\mathrm{AP}_{s\,50:95}$ increases by 9.6% with the MSFA module compared to the AFA module in AlignPS [1]. When adding the DAOE module, $\mathrm{AP}_{s\,50:95}$ increased by 6.94%, which indicates that each module has a positive impact on the baseline architecture, and the addition of the attention module plays a more effective role in the final result. When the two modules are configured on the baseline network simultaneously, the final $\mathrm{AP}_{s\,50:95}$ increases by 10.87%. Fig. 6(a) shows the application effects of these ablation experiments on the building dataset. In Fig. 4(a), we show the precision–recall curve of different ablation experiments. As shown in these figures, the proposed method outperforms the other compared methods for searching accuracy, which contains each proposed module.

TABLE II
COMPARISON OF AP$_s$ (%) BETWEEN THE PROPOSED METHOD, BASELINE METHODS, AND ABLATION EXPERIMENT ON THE BUILDING DATASET AND THE PLANE DATASET

Settings	Module combination	$AP_{s\,building50}$	$AP_{s\,building75}$	$AP_{s\,building50:95}$	$AP_{s\,plane50}$	$AP_{s\,plane75}$	$AP_{s\,plane50:95}$
baseline	None	81.86%	70.11%	63.71%	60.63%	57.54%	45.31%
baseline	MSFA	91.33%	82.00%	73.31%	67.13%	60.00%	49.92%
baseline	DAOE	92.27%	81.33%	70.65%	69.93%	64.04%	50.41%
Our Method	MSFA + DAOE	**94.67%**	**84.33%**	**74.58%**	**73.71%**	**65.21%**	**51.65%**



(a)

(b)

Fig. 4. Comparison of PR curves on two datasets. (a) PR curves on the building dataset. (b) PR curves on the plane dataset.

Experimental results on the plane dataset are shown in Table II. Compared with the baseline architecture, the AP$_{s\,50:95}$ of the proposed method with the MSFA and DAOE modules increases 4.61% and 5.1%, respectively. The combination of these two modules yields a marked increase in AP$_{s\,50:95}$ of 6.34%. Fig. 6(b) and Table II show that the proposed final method performs better than the comparison method and ablation experiment. These results demonstrate the effectiveness and necessity of each proposed module.

In addition to the above analysis, we visualized the object feature maps of the baseline and the proposed method in the testing process. As shown in Fig. 5, compared with the feature maps used for reidentification in the baseline method, the proposed method not only contains the object region, but also contains more details information. For this figure, the proposed method focuses on not only the features of the fuselage part but also the wing and edge parts. The flattened features extracted from them are more suitable for reidentification tasks. In addition, for the detection task, the proposed method focuses on the object, and the region of interest can be clearly obtained. However, the baseline method is ambiguous for the feature map of dense objects and prone to localization errors.

2) Comparative Experiments: Because the proposed method is novel, comparison with existing methods is difficult. Therefore, to verify the advancement provided by the proposed method, the proposed method is compared with several existing pedestrian searching algorithms, including AlignPS [1], AlignPS+ [1], Roi-AlignPS [50], and SeqNet [46]. The quantitative results from the two datasets are shown in Table III. The proposed method achieves the best results on both the datasets compared to other methods; although the proposed method is not the fastest, it is also real time. Both the proposed method and the comparison method perform better on the building dataset than on the plane dataset. Because the objects of different models in the plane dataset are more similar, it is difficult to obtain more descriptions when the object scale is small characteristics of their information. Thus, more erroneous searching results appear on the aircraft dataset. Roi-AlignPS and SeqNet thus achieve better performances than AlignPS and AlignPS+ because the ROI module in the two-stage algorithm provides explicit feature expression, which reduces the probability of false object detection for subsequent detection and reidentification tasks. Deformable convolution provides a larger receptive field for the AlignPS+ algorithm, but it is unimportant for detection and reidentification tasks. Therefore, the AlignPS+ and AlignPS algorithms produce similar results on both datasets. However, they are all designed for natural scenes, which lack good search results for large numbers and dense objects in remote sensing scenes, leading to the inability to obtain comparable results with the proposed method. In contrast, this article aggregates the object features from top to bottom through the MSFA module to merge deep semantic information and shallow appearance features and improves the performance of the reidentification subtask through the powerful description ability of the fusion features. As shown in Fig. 6, the proposed method can accurately locate the object and reduce misjudgment compared

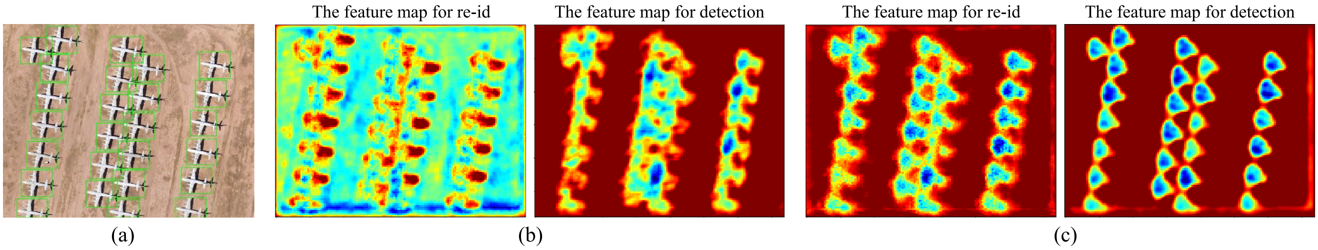

The feature map for re-id
The feature map for detection
The feature map for re-id
The feature map for detection
(a)
(b)
(c)

Fig. 5. Visualization of the feature map. (a) Original image of the input. (b) Feature visualization of the baseline method. (c) Feature visualization of the proposed method. (a) Original image. (b) Baseline. (c) The proposed method.

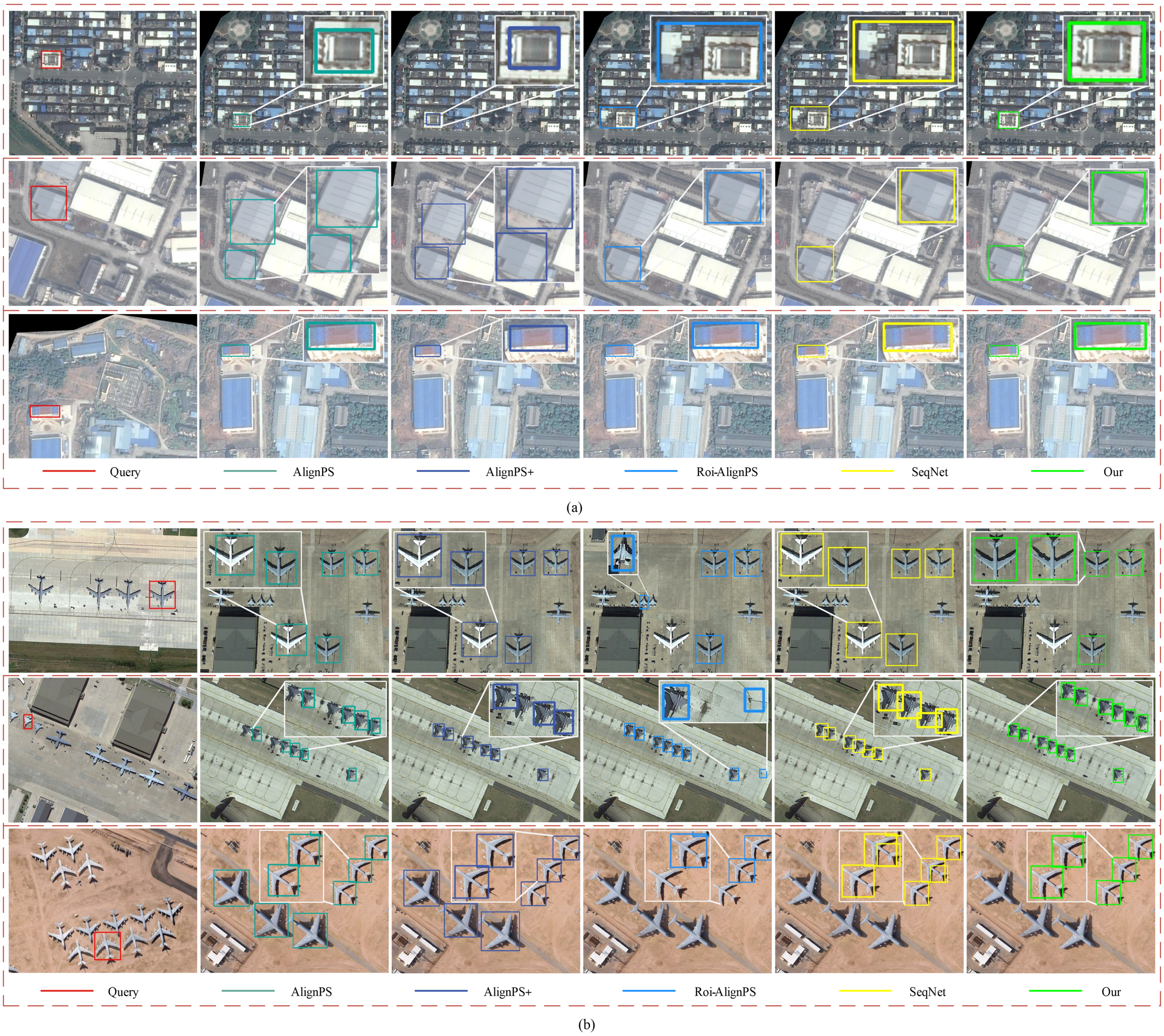


Fig. 6. Results of remote sensing object searching, only partial results are shown. (a) Searching result on the building dataset. (b) Searching result on the plane dataset.

to other methods. In addition, to address the negative impact of dense location on the remote sensing object searching task, we enhance the accuracy of feature expression from both channel and spatial dimensions, thus improving the performance of the detection subtask with dense object locations. To be specific, the proposed method can obtain more accurate bounding boxes in the searching results. Therefore, the proposed method achieves the best performance among all the studied approaches.

A comparison of the proposed method and other methods with both datasets is shown in Fig. 6. A detailed analysis of the results is provided next. For the building dataset, the comparison algorithm and the proposed method both achieve good

TABLE III
COMPARISON OF EXPERIMENTAL RESULTS ON THE BUILDING DATASET AND THE PLANE DATASET

Method	$AP_{s\,building50:95}$	$Time_{building}/ms$	$AP_{s\,plane50:95}$	$Time_{plane}/ms$
AlignPS [1]	63.71%	25.24	45.31%	25.68
AlignPS+ [1]	64.69%	26.37	46.52%	26.83
Roi-AlignPS [51]	70.31%	39.56	49.27%	40.02
SeqNet [47]	69.47%	40.43	48.65%	40.97
Our Method	**74.58%**	30.23	**51.65%**	30.71

performance. However, the many irrelevant objects in the dataset still pose challenges to the searching task. Regarding the comparison algorithms, the features obtained only through the backbone network have difficulty managing many similar objects, resulting in a wide range of false detections and missed detections. Therefore, the MSFA module is proposed to improve the ability of feature description, which can obtain more shallow details. These details will greatly enhance the identification ability in the reidentification subtask. Therefore, the proposed method yields the highest search accuracy. The plane searching tasks need to locate more numbers and dense objects, which makes searching markedly more difficult. Therefore, the comparison algorithm performs poorly in the plane searching task. On the one hand, the lack of sufficient detailed descriptive information leads to missed and false detections. On the other hand, the inaccurate localization of dense objects leads to low accuracy of the bounding box. However, the proposed method achieves good adaptability to these problems. The proposed MSFA module incorporates more shallow detail information, which can better distinguish the objects and improve detection accuracy. In addition, the proposed DAOE module enhances the focus of the object and has better landmark results for dense objects. Thus, both the quantitative and qualitative results show that the proposed method outperforms all the comparison algorithms.

V. CONCLUSION

In this article, we proposed a new deep learning framework for object searching in remote sensing images. Two modules were proposed to enhance the representation of effective features at different levels. Specifically, due to the difficulty of distinguishing too many feature descriptions of objects in remote sensing scenes, we proposed an MSFA module with top-down fused feature mapping to enhance the representation of object details. Then, a DAOE module was proposed to enhance object features from channel and spatial dimensions, which greatly improved the accuracy of localization of dense objects. Finally, the proposed method was tested on two challenging self-manually labeled datasets, and experimental results demonstrated the improved performance of the proposed method.

ACKNOWLEDGMENT

The authors would like to thank all the anonymous reviewers' insightful comments that greatly helped to improve the quality and presentation of this article.

REFERENCES

[1] Y. Yan et al., "Anchor-free person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 7690–7699.

[2] D. Chen, S. Zhang, J. Yang, and B. Schiele, "Norm-aware embedding for efficient person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12615–12624.

[3] Y. Jing, C. Si, J. Wang, W. Wang, L. Wang, and T. Tan, "Pose-guided multi-granularity attention network for text-based person search," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 11189–11196.

[4] Y. Yan, Q. Zhang, B. Ni, W. Zhang, M. Xu, and X. Yang, "Learning context graph for person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2158–2167.

[5] J. Xiao, Y. Xie, T. Tillo, K. Huang, Y. Wei, and J. Feng, "IAN: The individual aggregation network for person search," *Pattern Recognit.*, vol. 87, pp. 332–340, 2019.

[6] C. Liu, H. Yang, Q. Zhou, and S. Zheng, "Making person search enjoy the merits of person re-identification," *Pattern Recognit.*, vol. 127, 2022, Art. no. 108654.

[7] Y. Zhong, X. Wang, and S. Zhang, "Robust partial matching for person search in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6827–6835.

[8] S. Hou, C. Zhao, Z. Chen, J. Wu, Z. Wei, and D. Miao, "Improved instance discrimination and feature compactness for end-to-end person search," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2079–2090, Apr. 2022.

[9] G.-S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.

[10] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 296–307, 2020.

[11] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, Jan. 2012.

[12] 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.

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

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[15] 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.

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

[17] J. Redmon, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[18] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

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

[20] J. Kang et al., "DisOptNet: Distilling semantic knowledge from optical images for weather-independent building segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4706315.

[21] J. Kang, Z. Wang, R. Zhu, X. Sun, R. Fernandez-Beltran, and A. Plaza, "PiCoCo: Pixelwise contrast and consistency learning for semisupervised building footprint segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10548–10559, 2021.

[22] P. Duan, Z. Xie, X. Kang, and S. Li, "Self-supervised learning-based oil spill detection of hyperspectral images," *Sci. China Technol. Sci.*, vol. 65, no. 4, pp. 793–801, 2022.

[23] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, and R. Gloaguen, "Fusion of dual spatial information for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7726–7738, Sep. 2021.

[24] F. Shi, T. Zhang, and T. Zhang, "Orientation-aware vehicle detection in aerial images via an anchor-free object detection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5221–5233, Jun. 2021.

[25] Y. Yu, X. Yang, J. Li, and X. Gao, "A cascade rotated anchor-aided detector for ship detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5600514.

[26] J. Han, J. Ding, J. Li, and G.-S. Xia, "Align deep features for oriented object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602511.

[27] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and Fisher discriminative convolutional neural networks for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.

[28] K. Zhou, Z. Zhang, C. Gao, and J. Liu, "Rotated feature network for multiorientation object detection of remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 33–37, Jan. 2021.

[29] Z. Chen et al., "PIoU loss: Towards accurate oriented object detection in complex environments," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 195–211.

[30] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented R-CNN for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3520–3529.

[31] T.-Y. Lin, P. Dollar, 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. 936–944.

[32] X. Lu, H. Sun, and X. Zheng, "A feature aggregation convolutional neural network for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7894–7906, Oct. 2019.

[33] J. Chen, L. Wan, J. Zhu, G. Xu, and M. Deng, "Multi-scale spatial and channel-wise attention for improving object detection in remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 681–685, Apr. 2020.

[34] P. Wang, X. Sun, W. Diao, and K. Fu, "FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3377–3390, May 2020.

[35] X. Yang et al., "SCRDet: Towards more robust detection for small, cluttered and rotated objects," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8232–8241.

[36] Y. Liu, Q. Li, Y. Yuan, Q. Du, and Q. Wang, "ABNet: Adaptive balanced network for multi-scale object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5614914.

[37] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and T. S. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6111–6120.

[38] Y. Zhang, Q. Zhong, L. Ma, D. Xie, and S. Pu, "Learning incremental triplet margin for person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, pp. 9243–9250.

[39] X. Zhu, J. Qian, H. Wang, and P. Liu, "Curriculum enhanced supervised attention network for person re-identification," *IEEE Signal Process. Lett.*, vol. 27, pp. 1665–1669, 2020.

[40] T. Si, F. He, H. Wu, and Y. Duan, "Spatial-driven features based on image dependencies for person re-identification," *Pattern Recognit.*, vol. 124, 2022, Art. no. 108462.

[41] W. Huang, M. Luo, P. Zhang, and Y. Zha, "Full-scaled deep metric learning for pedestrian re-identification," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 5945–5975, 2021.

[42] C. Han et al., "Re-ID driven localization refinement for person search," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9814–9823.

[43] B. Munjal, S. Amin, F. Tombari, and F. Galasso, "Query-guided end-to-end person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 811–820.

[44] W. Dong, Z. Zhang, C. Song, and T. Tan, "Instance guided proposal network for person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2582–2591.

[45] B.-J. Han, K. Ko, and J.-Y. Sim, "End-to-end trainable trident person search network using adaptive gradient propagation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 925–933.

[46] Z. Li and D. Miao, "Sequential end-to-end network for efficient person search," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 3, pp. 2011–2019.

[47] X. Yang et al., "Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 132.

[48] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722.

[49] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[50] Y. Yan, J. Li, J. Qin, S. Liao, and X. Yang, "Efficient person search: An anchor-free approach," 2021, *arXiv:2109.00211*.

Haolong Fu received the B.S. degree in vehicle engineering from the Jiangsu University, Zhenjiang, China, in 2017, and the M.S. degree in vehicle engineering from Guizhou University, Guiyang, China, in 2020. He is currently working toward the Ph.D. degree with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.

His research interests include computer vision, deep learning, and object detection.

Qingpeng Li (Member, IEEE) received the B.S. degree in mechanical engineering and automation from Beijing Jiaotong University, Beijing, China, in 2015, and the Ph.D. degree in computer science and technology from Beihang University, Beijing, in 2019.

He is currently an Associate Professor with the State Laboratory of Robot Visual Perception and Control Technology School of Robotics, Hunan University, Changsha, China. His research interests include computer vision and machine/deep learning applications in remote sensing, especially object detection in remote sensing images and videos.

Puhong Duan (Member, IEEE) received the B.S. degree in mathematics and applied mathematics from Suzhou University, Suzhou, China, in 2014, and the Ph.D. degree in control science and engineering from Guizhou University, Changsha, China, in 2021.

From October 2019 to December 2019, he was a Visiting Ph.D. Student with the Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany. From January 2020 to October 2020, he was a Visiting Ph.D. Student with the Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany. He is currently an Associate Researcher with the College of Electrical and Information Engineering, Hunan University. His research interests include image restoration, multimodal image fusion, and classification.

Dr. Duan was the Finalist for the Best Student Paper Award at the 2020 International Geoscience and Remote Sensing Symposium and was selected as the Best Reviewer for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2021. He is an Associate Editor for *Frontiers in Remote Sensing*.

Jiacheng Lin received the B.S. degree in mechanical design manufacture and automation major from the Lanzhou University of Technology, Lanzhou, China, in 2018, and the M.S. degree in mechatronic engineering from Guizhou University, Guiyang, China, in 2021. He is currently working toward the Ph.D. degree with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.

His research interests include computer vision, deep learning, data security, and privacy protection.

Renwei Dian (Member, IEEE) received the B.S. degree in automation from the Wuhan University of Science and Technology, Wuhan, China, in 2015, and the Ph.D. degree in control science and engineering from Guizhou University, Changsha, China, in 2020.

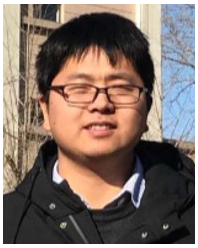
He is currently an Associate Professor with the School of Robotics, Hunan University. From November 2017 to November 2018, he was a visiting Ph.D. Student with the University of Lisbon, Lisbon, Portugal, supported by the China Scholarship Council. From August 2020 to July 2022, he was a Postdoctoral Researcher with the College of Electrical and Information Engineering, Hunan University. His research interests include hyperspectral image super-resolution, image fusion, tensor decomposition, and deep learning.

Dr. Dian was a finalist for the Best Student Paper Award at the International Geoscience and Remote Sensing Symposium 2018. He was awarded the Fellowship of China National Postdoctoral Program for Innovative Talents in 2020 and Excellent Doctoral Dissertation by the China Society of Image and Graphics in 2020.

Shutao Li (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he was with the College of Electrical and Information Engineering, Hunan University. In 2001, he was also a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong. From 2002 to 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, Surrey, U.K. In 2005, he was a Visiting Professor with the Department of Computer Science, The Hong Kong University of Science and Technology. In 2013, he was granted the National Science Fund for Distinguished Young Scholars in China. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored or coauthored more than 180 refereed papers. His research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li was a recipient of two Second-Grade National Awards with the Science and Technology Progress of China in 2004 and 2006. He is a Member of the Editorial Board of *Information Fusion* and *Sensing and Imaging*.



Xudong Kang (Senior Member, IEEE) received the B.S. degree in automation from Northeast University, Shenyang, China, in 2007, and the Ph.D. degree in control science and engineering from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the National Natural Science Award of China (Second Class and Rank as Third) and the Second Prize in the Student Paper Competition in 2014 International Geoscience and Remote Sensing Symposium. He was also selected as the Best Reviewer for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. From 2018 to 2019, he was an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS.

Zhiyong Li (Member, IEEE) received the M.Sc. degree in system engineering from the National University of Defense Technology, Changsha, China, in 1996, and the Ph.D. degree in control theory and control engineering from Hunan University, Changsha, in 2004.

In 2004, he joined the College of Computer Science and Electronic Engineering, Hunan University, where he is currently a Full Professor. He has authored or coauthored more than 100 papers in international journals and conferences. His research interests include intelligent perception and autonomous moving body, machine learning and industrial big data, and intelligent optimization algorithms with applications.

Dr. Li is a Member of the China Computer Federation.