remote sensing

MDPI

Article

Few Shot Object Detection for SAR Images via Feature Enhancement and Dynamic Relationship Modeling

Shiqi Chen [1], Jun Zhang [1], Ronghui Zhan [1,*], Rongqiang Zhu [2] and Wei Wang [1]

[1] NationalKey Laboratory of Science and Technology on Automatic Target Recognition, National University of Defense Technology, Changsha 410073, China; chenshiqi12@nudt.edu.cn (S.C.); zhangjun@nudt.edu.cn (J.Z.); wwang@nudt.edu.cn (W.W.)

[2] Southwest Electronics and Telecommunication Technology Research Institute, Chengdu 610000, China; r.zhu@nudt.edu.cn

* Correspondence: zhanrh@nudt.edu.cn

Abstract: Current Synthetic Aperture Radar (SAR) image object detection methods require huge amounts of annotated data and can only detect the categories that appears in the training set. Due to the lack of training samples in the real applications, the performance decreases sharply on rare categories, which largely inhibits the detection model from reaching robustness. To tackle this problem, a novel few-shot SAR object detection framework is proposed, which is built upon the meta-learning architecture and aims at detecting objects of unseen classes given only a few annotated examples. Observing the quality of support features determines the performance of the few-shot object detection task, we propose an attention mechanism to highlight class-specific features while softening the irrelevant background information. Considering the variation between different support images, we also employ a support-guided module to enhance query features, thus generating high-qualified proposals more relevant to support images. To further exploit the relevance between support and query images, which is ignored in single class representation, a dynamic relationship learning paradigm is designed via constructing a graph convolutional network and imposing orthogonality constraint in hidden feature space, which both make features from the same category more closer and those from different classes more separable. Comprehensive experiments have been completed on the self-constructed SAR multi-class object detection dataset, which demonstrate the effectiveness of our few-shot object detection framework in learning more generalized features to both enhance the performance on novel classes and maintain the performance on base classes.

Keywords: Synthetic Aperture Radar (SAR); few-shot object detection; attention; support-guided module; graph convolutional network

Citation: Chen, S.; Zhang, J.; Zhan, R.; Zhu, R.; Wang, W. Few Shot Object Detection for SAR Images via Feature Enhancement and Dynamic Relationship Modeling. *Remote Sens.* **2022**, *14*, 3669. https://doi.org/10.3390/rs14153669

Academic Editors: Deliang Xiang, Ying Luo, Xueru Bai, Gangyao Kuang and Xiaolan Qiu

Received: 16 June 2022
Accepted: 26 July 2022
Published: 31 July 2022

1. Introduction

1.1. Background

Due to its distinctive capabilities of all weather and all day imaging, as well as ground-penetrating, Synthetic Aperture Radar (SAR) has become the mainstream active observation system. With the rapid development of SAR imaging technology, a large amount of high-resolution SAR imagery data from various sensors are available, which further promote SAR applications, such as oil spill detection, urban planning and military reconnaissance. In particular, SAR object detection and recognition is a hotspot task widely used in maritime management and battlefield situation perception. Recently, some high-resolution SAR satellites, such as Gaofen-3 and HISEA-1 [1], have been successfully launched, which enables applications, such as radar target recognition, ship target detection, image registration, and so on [2–4].

Given the strong scattering features of the SAR image object and the statistical distribution of background clutter, the Constant False Alarm Rate (CFAR) algorithm [5] calculates

the threshold and determines whether the pixel belongs to the target or clutter. Though numerous methods have been proposed on the basis of CFAR, the model performance becomes unsatisfied under complex scene. In recent years, benefiting from the powerful feature representation capabilities, Convolutional Neural Networks (CNNs) have achieved a superior performance in the field of computer vision and largely boosted the development of data-driven SAR object detection algorithms. The state-of-the-art deep CNN-based detectors can be divided into two-stage detection algorithms and one-stage detection algorithms. The two-stage methods [6–8] dominate detection accuracy in prediction results, whereas one-stage approaches [9–11] perform better in training speed and inference efficiency.

In contrast to large-scale remote sensing datasets, the data volume and object type in SAR ship dataset is very limited. Although superior performance have been gained on SAR object detection task by leveraging a variety of advanced CNN algorithms trained on massive labeled samples, it encounters over fitting problem and poor generalization ability when the training set is not sufficient and new categories emerge.

Furthermore, the object occurrence in SAR interpretation applications follows a long-tailed distribution, which means common objects appear quite often while novel objects are long-tailed distributed. The most straightforward way is retraining a neural network with the aid of a large amount of data with various object types, nevertheless, the collection and annotation of new data require a lot of manpower and material resources. As for SAR image interpretation, the identification of object type becomes even more demanding since only the highly experienced researchers can distinguish the type of target under complex scenarios. The above issues largely hindered the development of SAR ship detection to some extent. Thus, it is indispensably and challenging to devise more robust object detector capable of detecting novel objects with only few labeled samples. In this work, we aim to design a few-shot paradigm which can still perform well on SAR ship detection models when only a limited amount of training data are acquired.

1.2. Related Work

SAR Object Detection Methods. We mainly focus on deep convolutional neural network methods applied for SAR object detection from anchor-based and anchor-free perspective. To effectively detect large-difference-scale targets under large scene images, Tang et al. [12] proposed a new metric revised bhattacharyya distance (RBD) instead of IoU metric and utilized it in label assignment and non-maximum suppression (NMS) of multi-stage detector Cascade RCNN. To mitigate the influence of background clutter and enhance the edge information of targets, Zhao et al. [13] introduced morphological feature-pyramid to preprocess the SAR images and present lightweight YOLOv4-tiny [14] network combined with feature pyramid fusion structure to improve detection accuracy. Moreover, Zhu et al. [15] proposed an optimal high-speed and high-accuracy detector H2Det based on the YOLOv5 detection framework. Chen et al. [16] designed a tiny ship detector with lighter architecture and fast inference speed by network pruning and knowledge distillation. Considering the characteristics of SAR image scenes and ship targets, Hu et al. [17] integrated deformable convolution, context information, and feature pyramid networks to construct BANet for multiscale ship detection. Zhang et al. [18] proposed a high-speed and high-performance anchor-free detector under a deformed complex scene and noise power distribution. Ma et al. [19] applied key-point estimation to eliminate the undetected targets and channel attention mechanism to suppress background noise, which is also established on an anchor-free framework. To enhance the spatial features, Cui et al. [20] added Spatial Group-wise Enhance (SGE) attention module to CenterNet and achieves high accuracy under large scene SAR image. Although the CNN-based method has shown superior performance on the SAR target detection task, the models are all trained with large amounts of labeled samples, which imposes severe limitations in real applications. Therefore, how to effectively use the existing training data and generalize the model to new targets has become a bottleneck.

Few-shot Object Detection Methods. To tackle the issue resulted from data scarcity, we resort to the few-shot learning technique, which has achieved commendable progress in image recognition and segmentation task, while more challenging in an object detection task. Current FSOD methods can be categorized into two branches: meta-learning-based methods and fine-tuning-based methods. The first type of methods usually adopt a parallel structure to extract features from both support images and query images. Meta R-CNN [21] learned the category-agnostic knowledge by the combination of region of interest (RoI) feature and class-attentive vectors to make more accurate object recognition rate. Inspired by category-agnostic transformation, MetaDet [22] defined RCNN head as task-specific modules and learns a meta generator, which guide the query images to detect novel classes. On the other hand, the two-stage fine-tuning methods are mainly based on transfer learning theory, which first pre-trains the model on base classes and then fine-tune on the support set by frozen some network parameters. TFA [23] fixes the feature extractor parameters and applies a straightforward fine-tuning scheme on the last layer of faster RCNN detector. MPSR [24] increases the scales of training sets by inserting object pyramids to refine FPN [8] and augment the scales in the sparse distribution of novel categories. To compensate the negative effect caused by naive data augmentation, Li et al. [25] designed a Transformation Invariant Principle (TIP) to make the encoder invariant to intra-class variations.

Few-shot Learning Applications in Remote Sensing Fields. In the remote sensing field, many researchers have been devoted to introducing few-shot learning in their approach due to its less dependency on training samples. For few-shot-classification methods in optical remote sensing images, Shi et al. [26] presented a metric-based few-shot method to generate prototypes for novel classes. In SAR interpretation applications, Yang et al. [27] proposed a novel few-shot SAR target classification framework by designing mixed loss graph attention network. Fu et al. [28] solved the sample restriction problem in SAR ATR via meta-learning framework. When it comes to few-shot object detection on remote sensing images, much less attention has been paid on this more challenging task since the classification and localization subtasks require the model to locate the object with correct category. Li et al. [29] built FSODM detector, which utilized the YOLOv3 [11] framework as a meta-feature extractor and learns feature adjustment through feature reweighting module proposed in [30]. To tackle the issue of sparse orientation space caused by lack of samples in novel classes, Cheng et al. [31] devised a prototype learning network named prototype-CNN (P-CNN) and further incorporated prototype-guided region proposal network into the whole framework. These methods are based on meta-learning and other studies focus on fine-tuning based methods. Zhao et al. [32] proposed multi-scale few-shot object detection approach by designing a more representative feature extractor, establishing a feature pyramid for multi-scale prediction and increasing the shape bias. Zhou et al. [33] also tackled the scale variation issue in remote sensing images by proposing a context-aware pixel-aggregation module and feature aggregation module. Huang et al. [34] considered the characteristic of remote sensing images and then proposed a shared attention module with a balanced fine tuning strategy to accommodate the few-shot settings.

1.3. Problems and Motivations

Although existing few-shot object detection methods are mainly designed for optical natural images, only a few approaches are developed for the remote sensing images and even less for SAR images. First, geometric distortion of target in SAR images always exists due to its special imaging mechanism. Furthermore, SAR ship targets under complex scenes are easily confused with speckle noise in the background. These make it difficult for few-shot object detector to extract the most effective features under complex backgrounds. How to select the discriminative support features without background interference and then generate enough high-qualified proposals for further classification and localization are pivotal for the few-shot object detection framework. Secondly, the misclassification of novel targets as base classes is more distinguished in SAR images owing to the similarity between different instances, which further affect the detection accuracy of few-shot detection model.

The SAR ship datasets comprise objects with arbitrary orientation and various scales, these intra-class difference also increase the difficultly of classification subtask in few-shot object detection.

To address the aforementioned challenges, we develop a novel few-shot detection framework, named the Relational graph convolutional network (RelationGCN), on the basis of the typical two-stage object detection structure, Faster Region-based CNN (Faster RCNN), which usually contains a region proposal network (RPN) and the Region of Interest (RoI) head. Correspondingly, the proposed detector is mainly composed of the support-guided region proposal network, the detection head and the relationship modeling graph structure between them.

To eliminate the background noise and ensure the high quality of support features, an attention module is attached at the extracted support feature to undermine the irrelevant information while strengthening the most discriminative features. As the randomly selected support objects appear in different views, shapes, and illuminations, the support-guided module is proposed, which generates dynamically changed kernels according to the support feature, and thus make full use of support information to enhance query features and further filter support-irrelevant proposals. In conventional meta-learning-based methods, only the single-class support data are used for generating class-attentive vectors, leading the relationship between support data and query data remain unexploited. Considering that the limited number of novel class samples in the few-shot fine-tuning stage leads to class imbalance and further degrades the detection accuracy on both base and novel classes, we establish the relationship between support and query features in a graph structure to make better knowledge transfer. In addition, consistent with other few-shot object detection methods, a two-stage training strategy is used to enable the proposed detector to quickly learn the knowledge of novel classes.

1.4. Contributions

With the proposed method, our detection framework can quickly adapt to novel classes with only few annotated training samples, which largely alleviate the intensive labeling cost in SAR images and perform well under data-constrained conditions. To sum up, our contributions are listed as following: (1) first, we propose a novel RelationGCN framework designed for few-shot object detection in SAR images. (2) Second, support feature guidance via dynamic convolutions is leveraged to better utilize support images and then provide more representative class-aware prototypes for adaptively enhancing query features. Additionally, a lightweight channel attention mechanism is also introduced to suppress the background noise in SAR images, thus elevating the quality of support images. (3) Thirdly, to make the model can generalize to novel classes in the few-shot settings, correlations between support and query features are captured by graph structure to facilitate feature transfer learning process. (4) Fourth, a new SAR multi-class ship dataset for few-shot object detection is constructed, several novel and base split settings are randomly selected to verify the effectiveness of the method, and some baseline results are exhibited for further research.

As for the organization of this article, Section 2 briefly introduces the preliminaries of the meta-learning-based few-shot object detection method. In Section 3, the overall architecture of our proposed few-shot detector and detailed description of each module are illustrated. Next, experimental results and analyses on the SAR multi-class ship detection dataset are presented in Section 4. Finally, we conclude this paper in Section 5.

2. Preliminaries

2.1. Problem Definition

In this section, we describe the dominant FSOD framework which aims at detecting novel objects given only few annotated instances. Formally, FSOD partitions the objects into two disjoint sets of categories: base classes C_{base}, the categories for which we have access to abundant training examples in large-scale dataset D_{base}; and novel classes C_{novel}, for which

only few instances are available in small-scale dataset D_{novel}. Given the support dataset S composed of K annotated samples for each category and $|C_{novel}| = N$, this is denoted as an N-way K-shot few-shot object detection problem. The performance is evaluated to jointly detect both base and novel categories from the test set. The evaluation metrics are reported separately for base and novel classes.

2.2. Meta-Learning Structure

One of the most successful approach for FSOD is the expanded upon meta-learning method, which trains a meta-learner and teaches the model to transfer previously learnt knowledge on common classes to object detection tasks on rare classes.

The meta-learning method for FSOD define a series of detection tasks on the base dataset to train the model. Typically, the initial model equipped with pretrained weight is first trained on base dataset, then an additional fine-tuning stage is attached to fine-tune the model on novel dataset, which adopts the similar meta-learning setting as in the base dataset.

Distinguished from general object detection, the training and testing dataset are organized in the episodic paradigm. In each episode, class c is randomly selected and K support objects from several support images are involved. Each task in an episode can be formulated as $T_i = (S^1, S^2, \ldots S^n, Q)$ where S^m means the mth support image and Q denotes a query image. Then, the detector is trained to detect all objects of class c in a query image with K supports via an episode meta-training manner. The difference between meta-training and meta-testing lies in that the ground truths of each class in the query dataset are only available in the meta-training stage.

3. Proposed Methods

The overall framework of the proposed method is shown in Figure 1, and the specific modules and implementations are illustrated in Sections 3.2–3.4. First, we elaborate on the construction of attention module and support-guided feature enhance module. Then, the graph convolutional network is introduced into the few-shot detection framework.

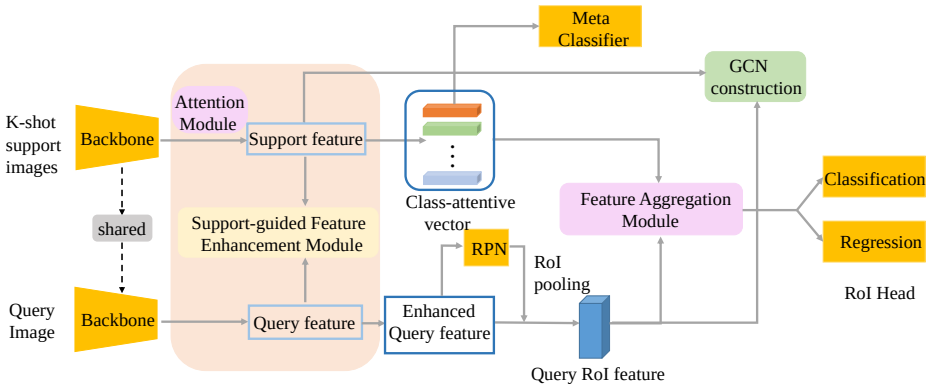


Figure 1. Architecture of the proposed RelationGCN for few-shot object detection on SAR images, which mainly consists of three components: a lightweight attention module to extract more relevant support information, a support-guided feature enhance module that injects different support images as guidance for better generation of higher-qualified object proposals, and a graph convolutional network which models the relationship between support features and query RoI features to further implicitly boost class representation.

3.1. Overall Architecture

The proposal-based few-shot detector is defined as $f(I, \theta)$, where I refers to the input data information and θ means the model parameters. Our approach follows a typical two-stage training scheme—*base training* and *few-shot adaptation*. $f(I, \theta)$ adopts the episodic training strategy in both stages, in which each episode is organized as the way illustrated in

Section 2.2. The final objective of our few-shot detector is to learn generalized features from abundant training samples of base classes during the first base training stage, and then rapidly adapt to novel classes under a balanced small scale dataset with equal number of annotations in base and novel classes during the few-shot fine-tuning stage. Our proposed few-shot detector is developed from Faster R-CNN, where the region proposal network is responsible for generating proposals and RCNN trains the classification and regression head. The input images in both the support set and query set share the same backbone as the feature extractor. With the consideration of the characteristic of SAR images, a lightweight attention module is designed and attached at the output support feature maps, which highlights the most relevant support features and attenuates the background noise. Since support feature vectors are diversified from each other due to the various characteristic of objects in support images, dynamic kernels are incorporated to adaptively enhance the query features according to different support features. Finally, we propose a graph convolutional network-based correlation learning mechanism to further model the relevance between support features and query RoI features.

3.2. Attention Mechanism and Support-Guided Feature Enhancement

The performance of the few-shot object detection model heavily depends on the support information from small amounts of samples to detect novel objects. Nevertheless, the ambiguous outline, side-lobes, and shadow caused by SAR imaging make feature extraction in SAR images more tricky. Therefore, how to guarantee the quality of support features and make them more representative becomes especially important in the SAR ship detection task under a few-shot setting. In this paper, we devise a lightweight attention module comprised of local context branch and global context branch to impair the influence of background noise and highlight features that are more relevant to the task. The detailed structure of efficient double-branch channel attention (E-DCA) module is illustrated in Figure 2.

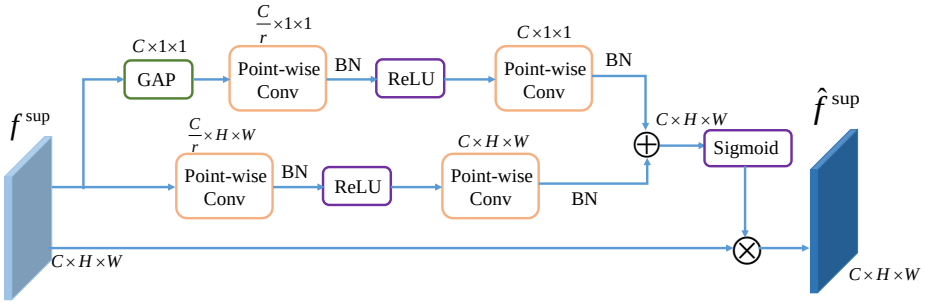


Figure 2. Diagram of our efficient double-branch channel attention (E-DCA) module.

Unlike general channel attention structure, which squeezes feature maps by a global max pooling operator, we add the local context branch to maintain the support information of small objects in SAR images. Both branches apply point-wise convolution for context aggregation. Given the support feature map f^{sup}, the local channel context is calculated via a similar bottleneck structure as in SENet [35], which can be represented as follows:

$$f^{sup}_{local} = BN(Conv^2_{pw}(\delta(BN(Conv^1_{pw}(f^{sup}))))) \quad (1)$$

where BN means Batch Normalization (BN) and δ indicates the Rectified Linear Unit (ReLU). The conventional fully-connected layers are replaced by two efficient point-wise convolutions $Conv^1_{pw}$ and $Conv^2_{pw}$, whose kernel sizes are $\frac{C}{r}\times C\times 1\times 1$ and $C\times\frac{C}{r}\times 1\times 1$, respectively. C denotes the channel number of support features, and r means the channel reduction ratio. For another global channel branch, the global average pooling (GAP)

is first employed for each channel independently, resulting in the channel-wise feature $G(f^{sup}) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} f^{sup}$. Similarly, the global channel context can be represented as:

$$f^{sup}_{global} = BN(Conv^{2}_{pw}(\delta(BN(Conv^{1}_{pw}(G(f^{sup})))))) \quad (2)$$

Given the local and global context information, the final refined support feature can be obtained as follows:

$$\hat{f}^{sup} = f^{sup} \otimes \sigma(f^{sup}_{local} \oplus f^{sup}_{global}) \quad (3)$$

where $\otimes$ means the broadcasting addition and $\oplus$ is the element-wise multiplication. σ denotes the sigmoid function.

Thus, the proposed attention module can benefit the few-shot detection network by highlighting class-specific features and softening redundant information for the task. Furthermore, this block is more efficient since it abandons the fully-connected layers to generate channel weights.

Some meta-learning-based methods, such as MetaRCNN [21] and Attention-RPN [36], perform channel-wise multiplication to reweight the query features, which filters a large number of support-irrelevant proposals. Nevertheless, the object view, size, or even occlusion by objects of other categories in SAR images are various in both support images and query images, which results in the unequal amount of information from different images. To resolve this problem, a support-guided query feature enhance module is devised and each support feature serves as an individual prototype, which allows for better aggregation between diverse support data and query images and provides enhanced query features for the RPN stage. The support-guided feature enhancement makes up for loss of information for simply averaging the information of support data for obtaining class-wise representative prototype in the next section. Traditional convolution operations achieve feature fusion by fixed kernels, while dynamic convolution [37] generates various kernels which are input-dependent and has more representation power. Motivated by this, we propose to generate dynamic kernels from support features to sufficiently interact with query features. The schematic of kernel generation is depicted in Figure 3.

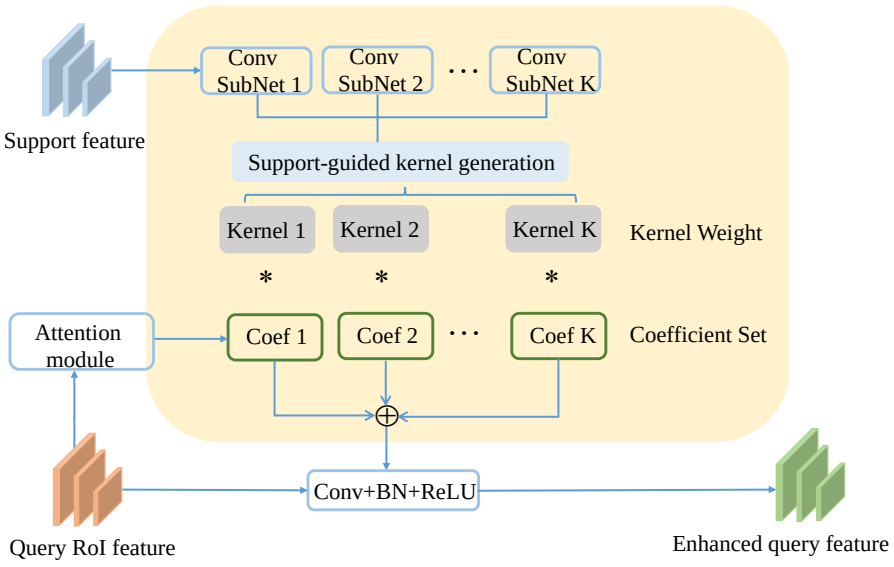


Figure 3. The structure of generating support-related kernel in support-guided query feature enhancement module. $*$ and $\oplus$ means the aggregation operation of different convolutional kernels.

Specifically, different support features are input to a kernel generator G_{ker}, which consists of several convolution subnetworks with the same number as the shot number. We denote the number of channels of both support feature and query feature as C_{in}. The shot

number is given as K, the number of kernels and its kernel size as K and c. The input and output of each subnetwork both have C_{in} channels and the kernel size of each subnetwork is $C_{in} - c + 1$. $\mathcal{K}_\theta^i$ serves as the kernel weight, which can be calculated as:

$$\mathcal{K}_\theta^i = G_{ker}(f_i^s) \tag{4}$$

where f_i^s denotes the feature of the ith support image refined by the proposed attention module. Then, K convolution operations are carried out over the query features using the support-related dynamic kernels to highlight the support-related regions. The resulting enhanced query features can be expressed as $\hat{v}_q \in \mathbb{R}^{C \times H \times W}$:

$$\hat{v}_q = \mathcal{K}_\theta^i \odot v_q = \sigma(c_k^i(\mathcal{K}_w^i v_q + \mathcal{K}_b^i)) \tag{5}$$

where $\odot$ denotes the convolutional operation, $\hat{v}_q$ is the strengthened query feature. $\mathcal{K}_w^i$ and $\mathcal{K}_b^i$ represent the weights and bias the ith kernel. Specifically, the coefficient of each kernel weight c_k^i is obtained via transformations of the input query feature, which can be formulated as:

$$c_k^i = Softmax(FC_2(Relu(FC_1(Avgpool(v_q))))) \tag{6}$$

where c_k^i means the attention weight for the ith kernel weight.

Assembling K multiple kernel functions is computationally efficient due to the small kernel size, and the way of aggregation via coefficient calculated by query feature related attention further boosts the feature representation capability.

3.3. Class Attentive Vector and Feature Aggregation

Most meta-learning based methods aggregate the RoI feature with support feature to obtain class-specific RoI features for classification and regression tasks. Incorporated with class-specific soft-attention vectors to achieve feature selection on RoI feature, meta-learning paradigm achieves model predictions on novel classes. Given the support image $x_{c,k}^{sup}$ and the query image x_i^q, these datasets are sent to the same detection network in a parallel way. Both region proposal network and RoIAlign operator are utilized to generate RoI features from a query feature map. In most few-shot object detection methods, such as FSRW [30] and Meta R-CNN [21], class-specific RoI features are aggregated by simple element-wise or channel-wise multiplication between the class-attentive feature vector and RoI feature vector generated from RPN. The aggregation combined features are used for further classification and regression in the second stage of Faster RCNN. The way of aggregation between class prototype with the query features determines the performance of FSOD methods based on meta-learning. Generally, the average value of K sample features from support dataset is taken as the class representation. The class attentive feature of class c is calculated as,

$$a_s^c = \sigma(\frac{1}{K}\sum_{k=1}^{K} Att(F(x_{c,k}^{sup}))) \tag{7}$$

where F means the feature extraction network, Att denotes the proposed attention module in Section 3.2. In this paper, a more complex RoI feature aggregation proposed in [38] is adopted to obtain more accurate RoI feature. Given the enhanced query feature v_q from support-guided RPN, and the class attentive feature vector a_s, the aggregated feature vector v^{fuse} is represented as:

$$v^{fuse} = [FC(v_q \otimes a_s^c), FC(v_q - a_s^c), v_q] \tag{8}$$

where FC denotes the fully-connected layer, which unifies the channel dimension generated after multiplication and subtraction. By aggregating the query features and class attentive

features, the classification probability of query RoI features can be obtained using a fully-connected sub-network as the predictor:

$$p_j = softmax(F_{cls}(v^{fuse})) \tag{9}$$

where F_{cls} denotes the classification head implemented by a fully-connected layer. The classification loss and bounding box regression loss are derived from the RPN training stage of Faster RCNN framework.

3.4. GCN

Graph Convolution Neural Networks (GCNs) were first proposed by Kipf et al. [39] to represent the data with a graph structure. Motivated by [40], which constructs a meta-graph where each edge represents the correlation between two classes, we propose a graph CNN-aided structure to effectively represent the interactions among support and query RoI features. This learning paradigm can better build appropriate connections between support and RoI features, thus guides more discriminative feature representation learning in a implicit manner.

As illustrated in Figure 4, we construct graph $G = (X, E)$ which is composed of the sets of nodes X and edges E. The two types of nodes in X are the support nodes X_s, namely the support features, and the query nodes X_q, namely the query RoI features. S denotes the adjacency relationship matrix of G, indicating the interaction between each node. The correlations between support features and query RoI features are adopted as the adjacency matrix of GCN, of which each element can be calculated by cosine similarity metric,

$$s_{ij} = \frac{exp^{\cos(s_i, q_j)}}{\sum_{k \in n_q} exp^{\cos(s_i, q_k)}} \tag{10}$$

where s_i and q_j denotes the support feature vector and the query RoI feature, respectively. n_q means the number of query images. Since the support images are randomly selected during the training process, the graph nodes and edges are dynamically updated at each training iteration, resulting in a dynamic correlation matrix S.

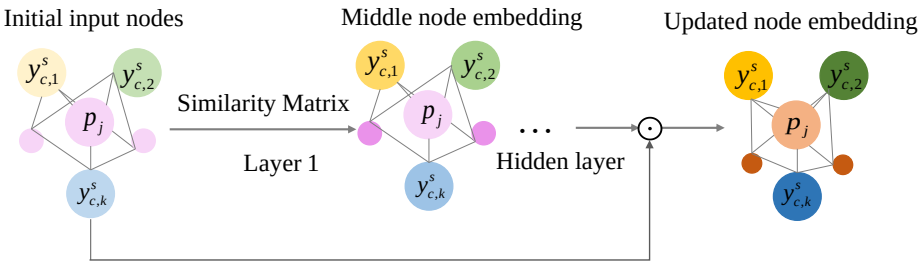


Figure 4. The construction of graph convolutional network (GCN) designed for few-shot object detection task ⊙ means multiplication operator.

The initialization of the proposed GCN can be represented as:

$$X_{init} = \{p_j\}_{j=1}^{n_q} \cup \left\{ y_{c,k}^{\text{support}} \right\}_{c=1;\, k=1}^{n_{\text{meta}}\, K} \tag{11}$$

where n_{meta} means the category number in support image set, and K denotes the shot number. Generally, the graph convolution is defined as:

$$A(l) = sigmoid(SX(l)W(l)) \tag{12}$$

where l denotes the layer index of GCN, X_l is the input of layer l, $W(l) \in \mathbb{R}^{n_{\text{meta}} \times n_{\text{meta}}}$ means the dynamically learned weight matrix of layer l. $A(l)$ is the output of layer l, which

can be considered as the relevance attention to represent the confidence of each category. Then, these weights are multiplied by the initial input to obtain the node embeddings of each category:

$$G(L) = A(L) \odot X_{init} \tag{13}$$

where L is the total number of GCN layers. To this end, the implicit relationship between support and query RoI features can be reflected in the updated node embeddings.

The relationships between the RoI and support features are implicitly constrained by GCN loss calculated through cross-entropy function, which is calculated by

$$L_{gcn} = -\frac{1}{n_q} \sum_{i=1}^{n_q} \mathbf{y}_i^{query} \log(\hat{\mathbf{g}}_i(L)) \tag{14}$$

where n_q means the number of RoI features, $\hat{\mathbf{g}}_i(L)$ denotes the i-th node embedding of the graph structure, which represents the enhanced probability that current RoI features are classified as the ground truth class. If the high similarity is captured between pairs of support and query RoI features, these two nodes share the strong relationship and in turn makes the predicted class probability more close to the support label. Otherwise, more penalties are attached on weak correlations to increase the separability between different classes in the embedding representation.

Although the application of GCN helps in forming strong relevance between RoI and support features, the standard cross-entropy loss may fail to ensure sufficient margin between heterogeneous classes which show high visual similarities. Inspired from metric-learning based FSOD techniques, we adopt a novel loss function to further reduce the intra-class variance and inter-class bias among target classes. The orthogonality constraint is first presented in [41] for the classification task, and we make modifications to suit for few-shot object detection task. This is achieved by enforcing all classes to be orthogonal to each other in the intermediate feature space. The computation of orthogonality constraint loss is formulated in Equation (15), where the angular distances between feature vectors is calculated using cosine similarity operator.

$$L_{oc}^{fea} = 1 - \sum_{\substack{i,j \in (N \times K) \\ y_i = y_j}} \left\langle \hat{f}_i^q, \hat{f}_j^q \right\rangle + \left| \sum_{\substack{i,j \in (N \times K) \\ y_i \neq y_j}} \left\langle \hat{f}_i^q, \hat{f}_j^q \right\rangle \right| \tag{15}$$

where $\langle \, , \cdot, \cdot \rangle$ is the cosine similarity operator applied on two features vectors, $|\cdot|$ means the absolute value operator, $\hat{f}_i^q$ denotes the support features of the i-th support image. Note that the features are first normalized when computing the similarity.

3.5. Training Strategy

Given the updated RoI feature, to ensure the diversity of attentive vectors from different classes, meta loss L_{meta} is proposed to obtain the optimal support representation in the meta-learning stage [21]. This is implemented by cross-entropy loss to diversify the inferred object attentive vector, however, it only works well under a sufficient number of training samples. Therefore, the meta loss is only participant in the base training stage while absent in the fine-tuning stage.

Finally, the total loss in the base training stage can be defined as:

$$L_{\text{base}} = L_{\text{rpn}} + L_{\text{cls}} + L_{\text{reg}} + L_{\text{meta}} + L_{\text{gcn}} \tag{16}$$

During the few-shot adaptation stage, the orthogonality constraint is introduced in the final loss, as described in Equation (17).

$$L_{\text{finetune}} = L_{\text{rpn}} + L_{oc}^{fea} + L_{\text{reg}} + L_{\text{meta}} + L_{\text{gcn}} \tag{17}$$

In the inference stage, since the discriminative feature representation is well learned in the training stages, we only use the original class probability prediction for classification without the participation of GCN and features constraint between classes.

4. Experiments

In this section, we describe our experimental setup and benchmark the performance of the proposed method on the multi-class SAR ship detection dataset.

4.1. Datasets

The high-resolution FUSAR-Ship dataset [42] is collected from the Gaofen-3 satellite, it contains 15 ship categories and used as an benchmark for SAR ship target recognition. To use this dataset for few-shot object detection task, 11 types of ships were selected to build a multi-type SAR ship detection dataset. Since the ship slices in FUSAR ship dataset are quite large, we remove the redundant background in chip images and insert these targets into several background images from both inshore and offshore scenarios selected mainly from AIRSARShip dataset [43]. Finally, 224 large-scale images of size 3000×3000 are constructed and the synthetic dataset is named as FUSAR-GEN. There are 11 categories in FUSAR-GEN, as displayed in Figure 5. These large-scale images are cropped to 500×500 size and the whole dataset is comprised of 2009 chips. Each chip image contains at least one ship target to be detected. The number of each category is analyzed in Table 1.

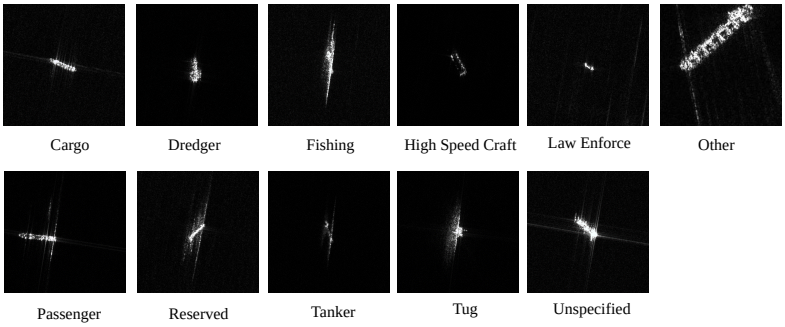


Figure 5. Different categories of ships in FUSAR-GEN dataset.

Table 1. Number of objects in each category of FUSAR-GEN dataset.

Class Name	Cargo	Dredger	Fishing	HighSpeedCraft	LawEnforce	Other	Passenger	Reserved	Tanker	Tug	Unspecified
Class Index	C01	C03	C04	C05	C06	C07	C08	C10	C12	C13	C14
Object Number	867	138	243	36	68	448	61	71	214	51	111

4.2. Implementation Details

Our few-shot object detection architecture is based on the two-stage Faster RCNN framework with ResNet-101 backbone. The input of the network consists of a batch of 4 query images which are resized to 800×800 and K-shot support images resized to 224×224 for each query image. As for the experimental parameter settings, the initial learning rate is set as 0.005 for the base training stage and a constant learning rate of 0.001 is used for fine-tuning stage. We train 18,000 iterations for base training stage and 1000 iterations for few-shot fine-tuning stage. Stochastic gradient descent (SGD) is used as the optimizer with 0.9 momentum and 0.0001 weight decay. All benchmark experiments are conducted using Pytorch framework on Ubuntu 16.04 system and supported by GeForce RTX 2080Ti GPU with 11G memory.

4.3. Evaluation Metrics

Considering the randomness of selecting base and novel sets and to comprehensively evaluate the detection performance on novel classes, three different divisions of base and novel classes are set randomly in the following experiment. Each split is comprised of 7 base categories and 4 novel categories ($N = 4$), and each set contains at least two categories with small number of samples:

- Set1: $C_{\text{novel}} = \{\text{Law Enforce, Passenger, Reserved, Tug}\}$.
- Set2: $C_{\text{novel}} = \{\text{Fishing, High Speed Craft, Reserved, Tanker}\}$.
- Set3: $C_{\text{novel}} = \{\text{Dredger, High Speed Craft, Law Enforce, Tug}\}$.

The detailed information of each setting can be found in Table 2.

Table 2. Three different base/novel classes split settings in our experiments.

Split	Novel Classes				Base Classes
1	Law Enforce(C06)	Passenger(C08)	Reserved(C10)	Tug(C13)	rest
2	Fishing(C04)	Tanker(C12)	Reserved(C10)	High Speed Craft(C05)	rest
3	Dredger(C03)	Tug(C13)	Law Enforce(C06)	High Speed Craft(C05)	rest

In the base training stage, the information of novel classes in all training images are removed. For the few-shot fine-tuning stage, a small training set which involves k annotated ground truths of both base and novel classes are randomly selected. We follow the widely-used protocol in [30] and compare our results with other methods adopting the same setting. The conventional evaluation metrics for detection is adopted as the K-shot evaluation metrics, which can be expressed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

where TP, FP, and FN denote the number of correctly predicted objects, false positives, and false negatives, respectively. The mAPs under K shot value can be calculated as:

$$mAP^K = \frac{\sum_{n=1}^{N} \int p^K(r) \mathrm{d}r}{N} \tag{20}$$

where N denotes the number of novel classes, p^k and r denotes the precision and recall of K-shot model. Although few-shot object detection algorithms focus on the performance of the novel class, catastrophic forgetting of the base class is also worthy of attention. Apart from the performance on novel classes, we also record the performance on base classes to make an integral evaluation.

5. Results

5.1. Comparison of Results

To verify the effectiveness of our model, we make comprehensive comparisons with state-of-the-art FSOD methods, including both fine-tune based approaches, such as TFA, FSCE, MPSR, and meta-learning-based methods, such as MetaRCNN and FsdetView. The average precision on novel classes are listed in Table 3. All these methods use ResNet101 as their backbone network.

Table 3. Few-shot detection results for the novel classes on FUSAR-GEN for three different splits. We tabulate results for $K = 3, 5, 10$ shots under different methods.

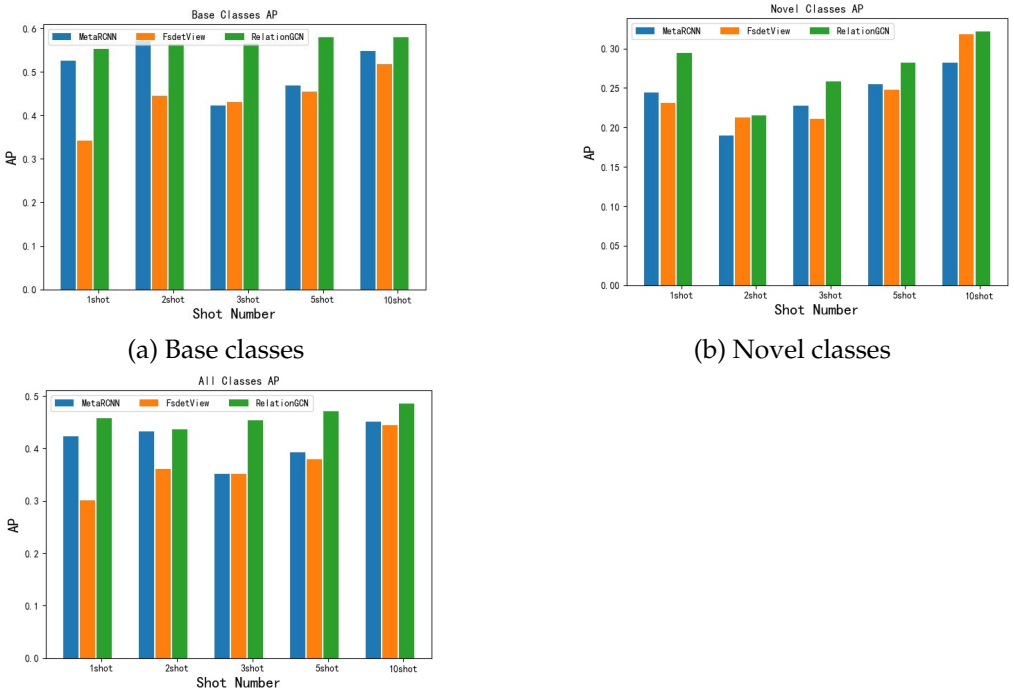
Methods	Set1			Set2			Set3		
	10	**5**	**3**	**10**	**5**	**3**	**10**	**5**	**3**
MetaRCNN	0.350	0.337	0.266	0.282	0.255	0.228	0.435	0.293	0.291
FsdetView	0.317	0.272	0.235	0.297	0.248	0.211	0.393	0.325	0.318
TFA	0.254	0.188	0.199	0.235	0.256	0.194	0.401	0.230	0.277
FSCE	0.355	0.229	0.260	0.276	0.236	0.185	0.413	0.272	0.283
MPSR	0.363	0.341	0.353	0.211	0.241	0.196	0.433	0.378	0.332
Ours	0.396	0.336	0.300	0.322	0.282	0.258	0.484	0.349	0.339

Table 3 lists the few-shot object detection performance of our method and the comparison methods on three different sets of novel classes. From Table 3, we have the following observations. The fine-tuning-based method TFA exhibits the lowest AP performance, since the inter-class differences means that the knowledge learned in the base training stage cannot be well transferred to the novel classes. MPSR achieves the best overall performance among three fine-tuning based methods owing to the augmented scales compensated for the missing scales in the sparse distribution of novel classes. The advantage of our method is not obvious in class split1 under shot number of 5, but the AP under 3-shot is still 3.4% higher than MetaRCNN while still inferior to MPSR. When the shot number of samples decreases from 10 to 3, the performance of our method for novel classes stills outperforms the best AP under other methods by 4.9% and 3.0% in split3 and split2, respectively. In comparison with a strong competitor, such as MetaRCNN and FsdetView, our method exhibits better performance under all shot numbers.

Furthermore, to compare the results of each meta-learning based FSOD approaches more intuitively, we visualize the mAP results on base classes, novel classes, and all classes at different shots K in a bar chart. As shown in Figure 6, the performance boost on novel classes can be as large as 6.3% under one shot and improvement of our method is still distinctive under other shot numbers. For base classes, FsdetView exhibits significant drop of performance compared with MetaRCNN especially under low shots. Nevertheless, the base AP performance on MetaRCNN decreases a lot when more samples of novel classes participate in the fine-tuning stage while our method maintains the high-level base AP at all shots. The overall performance is also outstanding from the perspective of both base and novel classes, which verifies the superiority of our method in adaptation to few-shot object detection task on SAR images.

5.2. Ablation Study

To analyze the effectiveness of different module in the proposed method more concretely, we implement ablation studies using the Set1 of FUSAR-GEN as novel classes. Our few-shot object detection framework can be decomposed into three major components: support-related and attention mechanism-guided query feature enhancement, graph structure relationship modeling, and orthogonal constraint loss. From the analysis of ablation studies, we can create a comprehensive understanding of the impact of different modules on the whole model performance. The implementation details are as follows: (1) for models without query feature enhancement, the original query features is directly feed into RPN; (2) for models without GCN, the relationship between support feature and query RoI feature is not considered and then the corresponding loss function is skipped; and (3) for models without orthogonal constraint in the hidden feature space, the loss function keeps the same as that in MetaRCNN.



(a) Base classes (b) Novel classes

(c) All classes

Figure 6. Performanceof different method on three type of classes under different shots. (**a**) Base classes. (**b**) Novel classes. (**c**) All classes.

After thorough investigation into the effects of different modules, the results are summarized in Table 4. We can see from the table that attaching either module achieves better results than the baseline model FsdetView. With the enhanced query features, the performance on novel class becomes 4.3% and 6.6% improvement when adding attention mechanism and support-related feature refinement, respectively, which indicates the effectiveness of support-related information in generating more representative proposals. However, the performance on base classes drops a little. After the introduction of relevance learning between support and query RoI features, the overall performance on all classes can increase up to 2.9%, which means better feature representation can be obtained and further leads to more stable RoI learning stage. As for the loss constraint, we can observe a more significant boost of 2.4% in novel classes than 1.5% performance lift in base classes. Orthogonal constraint in the feature space can improve the precision on both base and novel classes due to it guarantees the inter-class and intra-class separability in the feature embedding space.

Table 4. Ablation study of different components of the proposed few-shot detection architecture.

Method	Attention	Dynamic Conv	GCN	Loss Constraint	mAP_{base}	mAP_{novel}	mAP_{all}
FsdetView	-	-	-	-	0.5562	0.3168	0.4692
RelationGCN(Ours)	✓				0.5489	0.3598	0.4802
	✓	✓			0.5496	0.3829	0.4890
			✓		0.5652	0.3616	0.4912
	✓	✓	✓		0.5710	0.3719	0.4986
	✓	✓	✓	✓	0.5858	0.3959	0.5168

The number of layers in GCN structure is also an important hyper-parameter. In the proposed method, the dynamic GCN structure is implemented in a more complicated way than the static GCN, thus, the effect of different layers for our GCN structure should be

investigated. The changes of novel AP (nAP), base AP (bAP), and overall AP (AllAP) in the 10-shot setting of split1 under different numbers of GCN layers are reflected in Figure 7. It can be concluded that the overall performance on all classes is not very sensitive to the number of GCN layers and attaching more layers does not necessarily leads to the performance boost. The performance of base AP slightly increases with more GCN layers are involved in the model architecture. Although the model attains the best base AP when the number of GCN layer is 4, the performance on novel classes drops a lot. To make a comprise between overall performance and computational complexity, we choose the final GCN layer number as 2, which guarantees good performance on both novel and base classes.

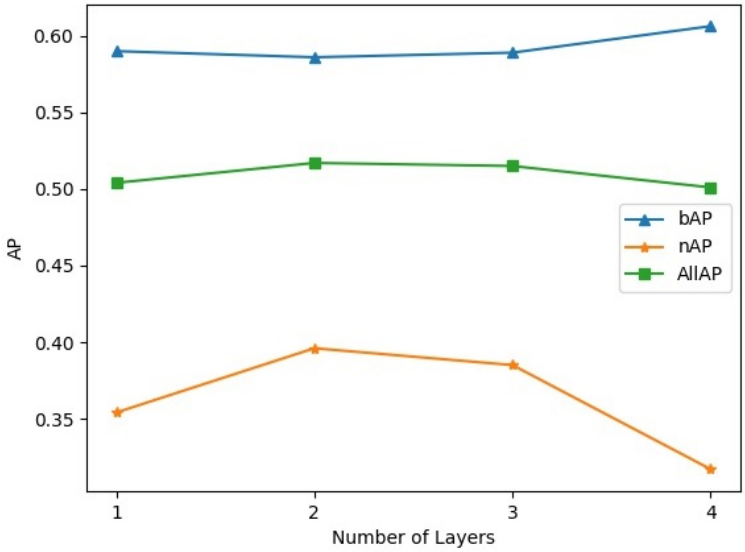


Figure 7. The effect of different GCN layer number on performance of base classes, novel classes, and all classes. The average precision on these three types of class sets are denoted as bAP, nAP, and AllAP, respectively.

5.3. Visual Analysis

Qualitatively, Figure 8 visualizes some results of 10-shot model in the first split of novel classes.

Most of the ships of novel classes can be successfully detected on both offshore and inshore scenario. Although scale distribution varies in different classes and some objects are too blurred to be distinguished from the background, there exists less missing detections and false alarms, demonstrating the effectiveness of our model. We also select a demanding situation in which both novel and base classes appear in chip images to display the overall performance of the proposed few-shot detector. The proposed method is capable of simultaneously detecting both base and novel targets. In addition, it can be seen that the base classes can also be accurately detected with high prediction scores, indicating the knowledge learned from abundant samples in the base training stage can be maintained without catastrophic forgetting.

As shown in Figure 9, we also summarize some typical failure results, which can be roughly divided into: (1) confusion between similar classes and (2) objects with low illumination. For example, Cargo is misclassified as Fishing or Other class, and Unspecified is mistaken as Cargo since they share similar characteristic of size and aspect ratio. For novel classes, Reserved appears in lower scattering intensity and tends to be misclassified as Tanker. Similar class confusion phenomena occurs in LawEnforce, which is incorrectly predicted as Passenger and Tug. The aforementioned failure cases deserves further research.

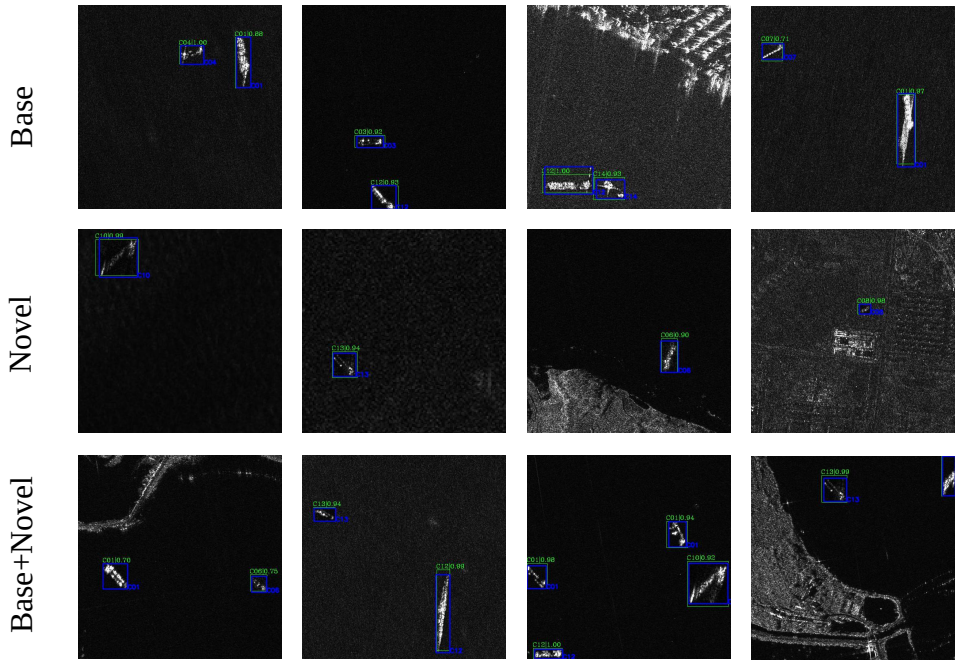


Figure 8. Visualization of detection results on base, novel, and base+novel classes. The blue and green boxes denote the ground truth and detection results with confidence scores, respectively.

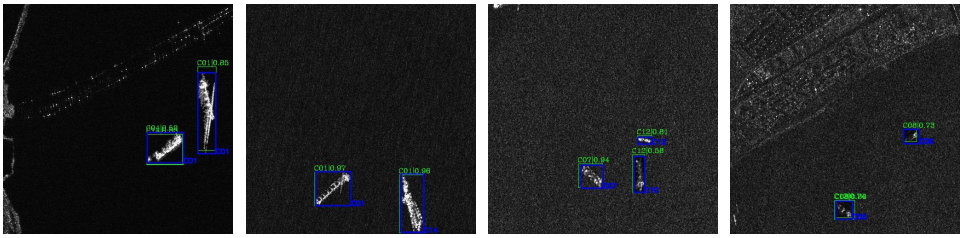
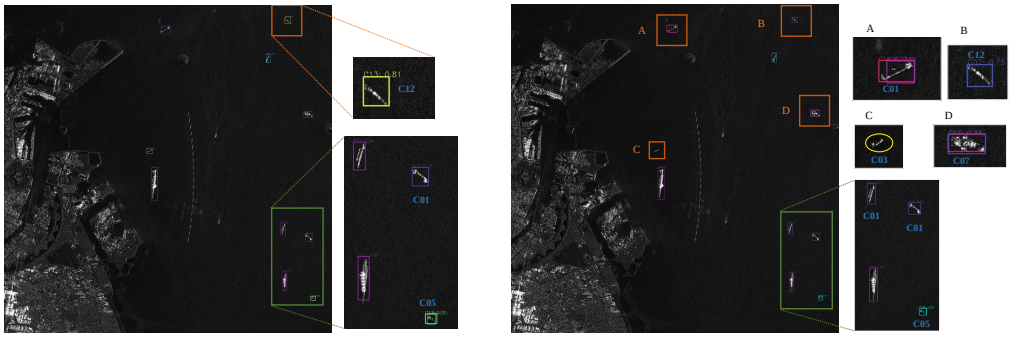


Figure 9. Typical failure detection results. The first two columns and the last two columns present the results for base classes and novel classes, respectively.

To demonstrate the holistic performance on a large demo image, we also display qualitative results in Figure 10. The performance of our model is contrasted against the state-of-the-art approach for dataset split1 under 10-shot setting. We choose MetaRCNN for comparison owing to its capability of elevating the novel class performance without sacrificing base class performance. The prediction results of different classes are visualized in different colors. We zoom in on specific areas to display the predicted category and corresponding confidence score for more intuitive comparison. Owing to the small scale, as well as low contrast compared with the background, Dredger (C03) becomes a missing detection in MetaRCNN but can be detected with 0.75 confidence score in our method, which is lower than other type of targets. C01 (Cargo) and C07 (Other) are easily confused with C14 (Unspecified), as shown in Figure 10b while our method can be well adapted to the inter-class variations. Some false alarms, such as Tanker (C12), are not correctly detected in both methods due to their blurred scattering characteristic and especially narrowed width. As shown in the enlarged area of the selected green area, although High Speed Craft (C05) is both misclassified as novel category Passenger (C08) in two comparing methods, the proposed few-shot object detector outperforms the other in detecting base class Cargo. The class confusion between Cargo and Other is slightly alleviated in our method since two ships of Cargo were rightly detected with a high confidence score, while Meta RCNN treats them as the Other type with a high confidence score. The three instances are distinctive from each other in appearance, view, and pose, so the intra-class variation

makes it difficult for both methods to distinguish them from each other. Overall, our model suffers from less catastrophic forgetting on base classes, such as Cargo and Other, compared with MetaRCNN.



(a) Our method. (b) MetaRCNN.

Figure 10. Comparison of detection results on large-scale SAR image. The orange and green area are shown enlarged on the right side of the sub-image. (**a**) Detection results of our method. (**b**) Detection results of MetaRCNN.

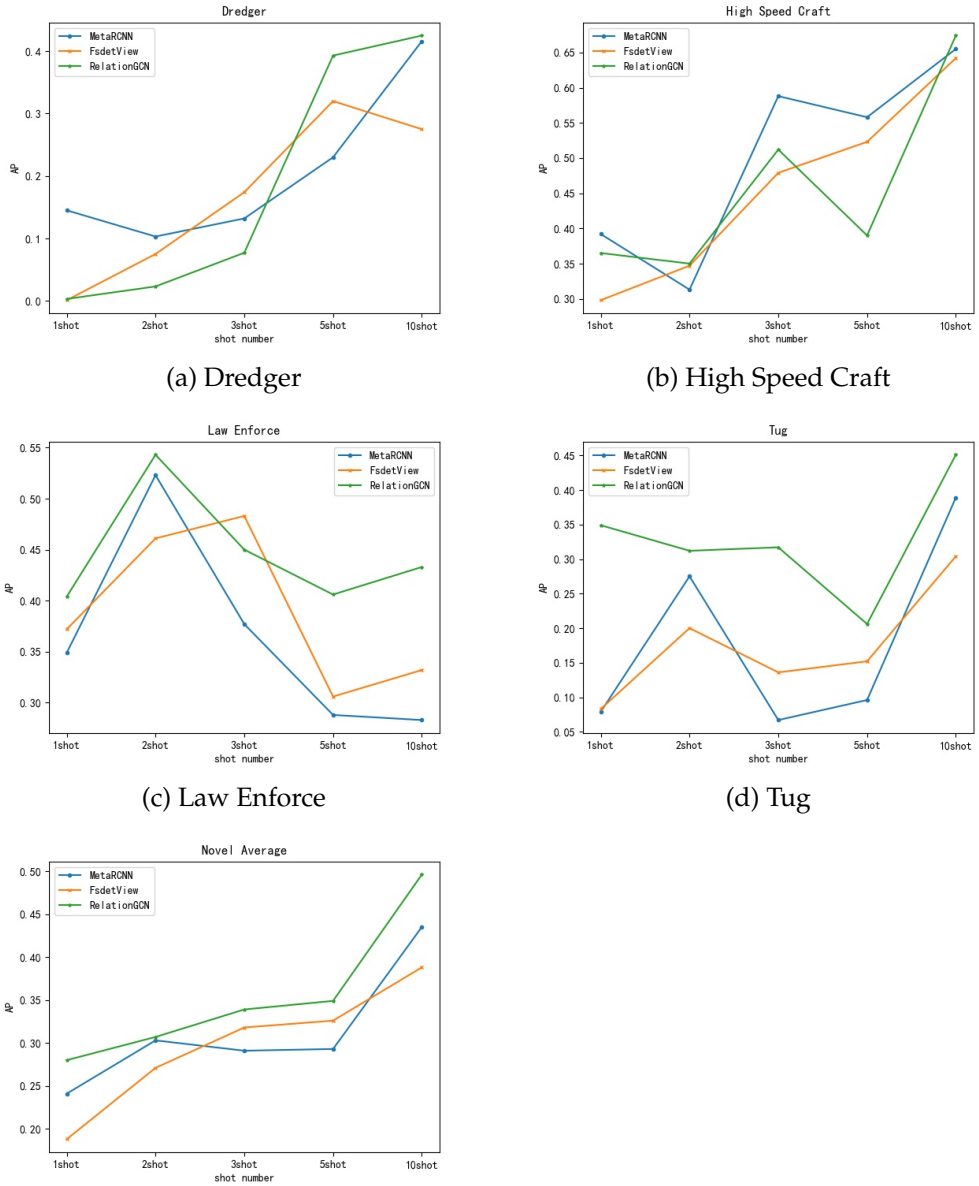
6. Discussions

6.1. Performance on Novel Classes

Since our method is developed on the meta-learning-based few-shot object detection methods, we specifically evaluate the performance of MetaRCNN, FsdetView, and our method on novel class. The accuracy trend of each category and the average novel class AP under different shot numbers are visualized in Figure 11. As for Dredger, the performance is poorer than the others under low shot while it increases a lot under 5-shot and reaches the peak point under 10-shot. Compared with MetaRCNN, smaller samples can guarantee almost equivalent precision on this class. This can be attributed to the well-designed graph structure modeling of the relationship between query RoI and support features. As for Law Enforce, our method achieves the best AP under 2-shot, and more novel samples do not bring a performance boost, which is consistent with other methods. Even under the condition of an only 1-shot sample, the precision of Tug can start at a high initial point and the performance under all shots always stays ahead of others. Nevertheless, for High Speed Craft, the performance fluctuates a lot except for FsdetView. This is due to the fact that the ambiguous internal structure of the target and the scattering information is only highlighted at the outlines, thus making it susceptible to be interfered with by other similar classes. In general, the precision of High Speed Craft outperforms other classes and can perform higher than 0.65 under 10-shot, while the precision on Tug and Dredger has much room for accuracy improvement which requires further research. From the perspective of novel average precision, the proposed method remains higher than others under all shots, demonstrating the comprehensive effectiveness of the proposed modules.

6.2. Adaptation Speed of Different Methods

To verify the efficiency of the proposed method, we also compare the number of iterations required for model convergence. Concretely, we evaluate a model every 100 iterations and update the best AP at each test iterations. For fair comparison, we set the default fine-tuning iterations as 1000. If the current AP no longer surpasses the best recorded on before 1000 iterations, we regard that the model has converged and report the iteration which exhibits the best AP as adaptation speed. The fewer iterations required to achieve the best AP means faster adaptation speed.



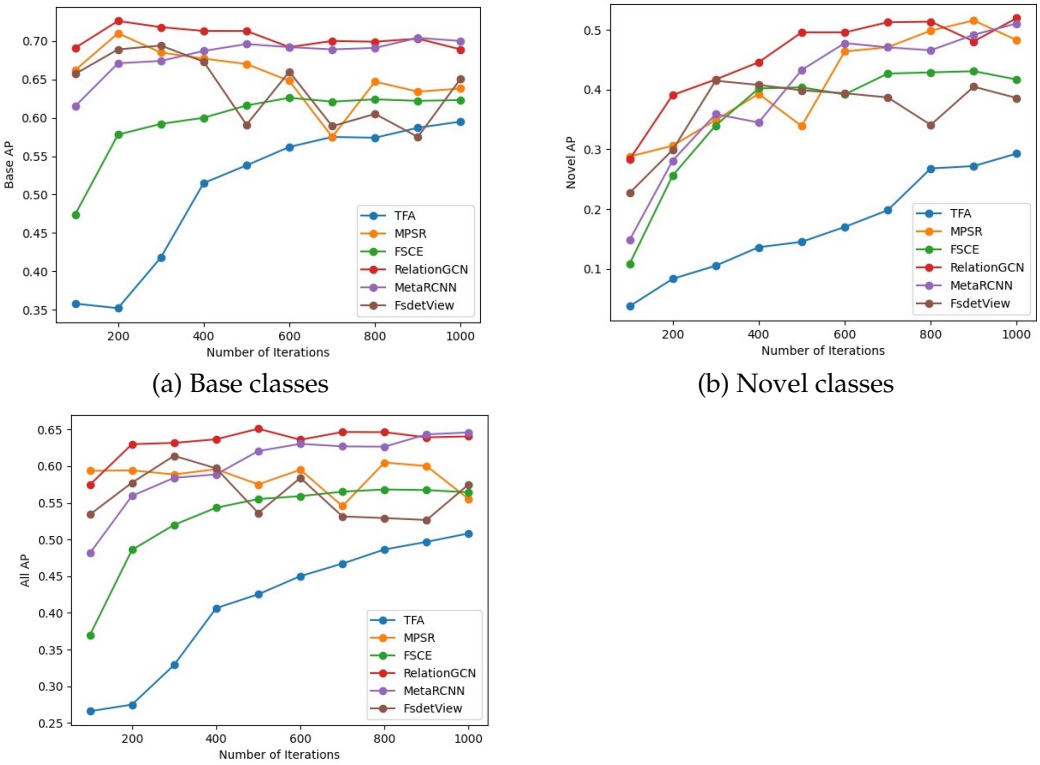
(a) Dredger (b) High Speed Craft

(c) Law Enforce (d) Tug

(e) Novel Average

Figure 11. Comparison of precision on each novel class and average result.

As shown in Figure 12, meta-learning-based methods can adapt faster than fine-tuning-based methods. Specifically, TFA requires much more iterations toward convergence due to the random initialization for novel classes. In contrast, our model shows comparable adaptation speed with FsdetView, while achieving 11.2% better novel AP and 3.7% better all AP performance. More surprisingly, only half of the total iterations can achieve 65.1% of the peak overall performance, which further indicates the fast adaptation ability of our method in the early fine-tuning stage. In addition, RelationGCN can obtain a higher initial point than most of the methods in terms of novel AP without any iteration. In conclusion, our method can achieve satisfactory adaptation speed and maintain the high AP performance with less few-shot transfer time consumption.



(a) Base classes (b) Novel classes

(c) All classes

Figure 12. Comparison of few-shot fine-tuning speed of different methods on three types of classes under 10-shot setting.

7. Conclusions

In this paper, a well-adapted two-stage detector based on meta-learning is proposed to address the challenging few-shot object detection task in SAR images. To fully exploit the most discriminative features of support images, a lightweight double-branch channel attention module is incorporated to reduce background interference while strengthening the most representative information of support images. Considering the variety between different support objects, a support-guided module is proposed to enhance query features with weighted support features. This dynamic convolution incorporates the importance of each support image, which not only strengthens the shared information between support feature and query feature but also maintains intrinsic representation of support data. In addition, we design a correlation learning mechanism via a graph structure to further model the relevance between support images and query images, making the features from the same category more close while from different categories far apart. With all these modules integrated into the conventional Faster R-CNN detector, a novel few-shot detection framework RelationGCN is developed. Comprehensive experiments have been conducted on the self-constructed FUSAR-GEN dataset which contains various types of ship objects, and the results fully verify the feasibility of applying meta-learning based few-shot learning method into SAR ship detection under few-shot scenario. In the future, we will also make further research into some fine-tuning-based few-shot detection methods and focus on the scale variation issue in SAR images.

Author Contributions: Conceptualization, S.C.; methodology, S.C.; software, S.C.; validation, S.C.; data curation, S.C., R.Z. (Rongqiang Zhu); writing—original draft preparation, S.C.; writing—review and editing, S.C., R.Z. (Ronghui Zhan); visualization, S.C.; supervision, J.Z.; funding acquisition, R.Z. (Ronghui Zhan), W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61901500, and the Natural Science Foundation of Hunan Province under Grant 2020JJ5674.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
RPN	Region proposal network
RoI	Region of interest
mAP	Mean average precision
GCN	Graph convolution network

References

1. Xia, R.; Chen, J.; Huang, Z.; Wan, H.; Wu, B.; Sun, L.; Yao, B.; Xiang, H.; Xing, M. CRTransSar: A Visual Transformer Based on Contextual Joint Representation Learning for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 1488. [CrossRef]
2. Zhang, F.; Tianying, M.; Xiang, D.; Ma, F.; Sun, X.; Zhou, Y. Adversarial deception against SAR target recognition network. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4507–4520. doi: 10. [CrossRef]
3. Ao, W.; Xu, F.; Li, Y.; Wang, H. Detection and Discrimination of Ship Targets in Complex Background From Spaceborne ALOS-2 SAR Images. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 536–550. doi: 10. [CrossRef]
4. Zhang, H.; Lei, L.; Ni, W.; Tang, T.; Wu, J.; Xiang, D.; Kuang, G. Explore Better Network Framework for High-Resolution Optical and SAR Image Matching. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. doi: 10. [CrossRef]
5. Robey, F.C.; Fuhrmann, D.R.; Kelly, E.J.; Nitzberg, R. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [CrossRef]
6. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 7–12 December 2015, Volume *28*, pp. 91–99.
7. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
8. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
9. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
10. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
11. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
12. Tang, J.; Cheng, J.; Xiang, D.; Hu, C. Large-Difference-Scale Target Detection Using a Revised Bhattacharyya Distance in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
13. Zhao, C.; Fu, X.; Dong, J.; Qin, R.; Chang, J.; Lang, P. SAR Ship Detection Based on End-to-End Morphological Feature Pyramid Network. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4599–4611. doi: 10. [CrossRef]
14. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
15. Zhu, M.; Hu, G.; Zhou, H.; Wang, S. H2Det: A High-Speed and High-Accurate Ship Detector in SAR Images. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 12455–12466. doi: 10. [CrossRef]
16. Chen, S.; Zhan, R.; Wang, W.; Zhang, J. Learning Slimming SAR Ship Object Detector Through Network Pruning and Knowledge Distillation. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1267–1282. doi: 10. [CrossRef]
17. Hu, Q.; Hu, S.; Liu, S. BANet: A Balance Attention Network for Anchor-Free Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
18. Zhang, Y.; Cao, Y.; Feng, X.; Xie, M.; Li, X.; Xue, Y.; Qian, X. SAR Object Detection Encounters Deformed Complex Scenes and Aliased Scattered Power Distribution. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4482-4495. doi: 10. [CrossRef]
19. Ma, X.; Hou, S.; Wang, Y.; Wang, J.; Wang, H. Multiscale and Dense Ship Detection in SAR Images Based on Key-Point Estimation and Attention Mechanism. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. doi: 10. [CrossRef]
20. Cui, Z.; Wang, X.; Liu, N.; Cao, Z.; Yang, J. Ship Detection in Large-Scale SAR Images Via Spatial Shuffle-Group Enhance Attention. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 379–391. doi: 10. [CrossRef]
21. Yan, X.; Chen, Z.; Xu, A.; Wang, X.; Liang, X.; Lin, L. Meta r-cnn: Towards general solver for instance-level low-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9577–9586.
22. Wang, Y.X.; Ramanan, D.; Hebert, M. Meta-Learning to Detect Rare Objects. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 9924–9933. doi: 10. [CrossRef]

23. Wang, X.; Huang, T.E.; Darrell, T.; Gonzalez, J.E.; Yu, F. Frustratingly simple few-shot object detection. *arXiv* **2020**, arXiv:2003.06957.
24. Wu, J.; Liu, S.; Huang, D.; Wang, Y. Multi-scale positive sample refinement for few-shot object detection. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, pp. 456–472.
25. Li, A.; Li, Z. Transformation Invariant Few-Shot Object Detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online, 20–25 June 2021; pp. 3093–3101. doi: 10. [CrossRef]
26. Shi, J.; Jiang, Z.; Zhang, H. Few-Shot Ship Classification in Optical Remote Sensing Images Using Nearest Neighbor Prototype Representation. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3581–3590. [CrossRef]
27. Yang, M.; Bai, X.; Wang, L.; Zhou, F. Mixed Loss Graph Attention Network for Few-Shot SAR Target Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. doi: 10. [CrossRef]
28. Fu, K.; Zhang, T.; Zhang, Y.; Wang, Z.; Sun, X. Few-shot SAR target classification via metalearning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
29. Li, X.; Deng, J.; Fang, Y. Few-Shot Object Detection on Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. doi: 10. [CrossRef]
30. Kang, B.; Liu, Z.; Wang, X.; Yu, F.; Feng, J.; Darrell, T. Few-Shot Object Detection via Feature Reweighting. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 8419–8428. doi: 10. [CrossRef]
31. Cheng, G.; Yan, B.; Shi, P.; Li, K.; Yao, X.; Guo, L.; Han, J. Prototype-CNN for few-shot object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–10. [CrossRef]
32. Zhao, Z.; Tang, P.; Zhao, L.; Zhang, Z. Few-Shot Object Detection of Remote Sensing Images via Two-Stage Fine-Tuning. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
33. Zhou, Y.; Hu, H.; Zhao, J.; Zhu, H.; Yao, R.; Du, W. Few-shot Object Detection via Context-aware Aggregation for Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. doi: 10. [CrossRef]
34. Huang, X.; He, B.; Tong, M.; Wang, D.; He, C. Few-Shot Object Detection on Remote Sensing Images via Shared Attention Module and Balanced Fine-Tuning Strategy. *Remote Sens.* **2021**, *13*, 3816. [CrossRef]
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
36. Fan, Q.; Zhuo, W.; Tang, C.K.; Tai, Y.W. Few-shot object detection with attention-RPN and multi-relation detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 16–18 June 2020; pp. 4013–4022.
37. Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic Convolution: Attention Over Convolution Kernels. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 16–18 June 2020; pp. 11027–11036. doi: 10. [CrossRef]
38. Xiao, Y.; Marlet, R. Few-shot object detection and viewpoint estimation for objects in the wild. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 192–210.
39. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
40. Kim, G.; Jung, H.G.; Lee, S.W. Few-Shot Object Detection via Knowledge Transfer. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Singapore, 8–10 February 2020; pp. 3564–3569. doi: 10. [CrossRef]
41. Ranasinghe, K.; Naseer, M.; Hayat, M.; Khan, S.; Khan, F.S. Orthogonal projection loss. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Online, 20–25 June 2021; pp. 12333–12343.
42. Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; Xu, F. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inf. Sci.* **2020**, *63*, 1–19. [CrossRef]
43. Xian, S.; Zhirui, W.; Yuanrui, S.; Wenhui, D.; Yue, Z.; Kun, F. AIR-SARShip-1.0: High-resolution SAR ship detection dataset. *J. Radars* **2019**, *8*, 852–862.