remote sensing

MDPI

Article

Multi-Scale Similarity Guidance Few-Shot Network for Ship Segmentation in SAR Images

Ruimin Li [1], Jichao Li [2], Shuiping Gou [2,*], Haofan Lu [2], Shasha Mao [2] and Zhang Guo [1]

[1] Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710071, China; rmli@xidian.edu.cn (R.L.); guozhang@xidian.edu.cn (Z.G.)
[2] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China; jcli_14@stu.xidian.edu.cn (J.L.); hflu@stu.xidian.edu.cn (H.L.); ssmao@xidian.edu.cn (S.M.)
* Correspondence: shpgou@mail.xidian.edu.cn

Abstract: Target detection and segmentation in synthetic aperture radar (SAR) images are vital steps for many remote sensing applications. In the era of data-driven deep learning, this task is extremely challenging due to the limited labeled data. Few-shot learning has the ability to learn quickly from a few samples with supervised information. Inspired by this, a few-shot learning framework named MSG-FN is proposed to solve the segmentation of ship targets in heterologous SAR images with few annotated samples. The proposed MSG-FN adopts a dual-branch network consisting of a support branch and a query branch. The support branch is used to extract features with an encoder, and the query branch uses a U-shaped encoder–decoder structure to segment the target in the query image. The encoder of each branch is composed of well-designed residual blocks combined with filter response normalization to capture robust and domain-independent features. A multi-scale similarity guidance module is proposed to improve the scale adaptability of detection by applying hand-on-hand guidance of support features to query features of various scales. In addition, a SAR dataset named SARShip-4i is built to evaluate the proposed MSG-FN, and the experimental results show that the proposed method achieves superior segmentation results compared with the state-of-the-art.

Keywords: SAR image; ship segmentation; few-shot learning; multi-scale similarity guidance

Citation: Li, R.; Li, J.; Gou, S.; Lu, H.; Mao, S.; Guo, Z. Multi-Scale Similarity Guidance Few-Shot Network for Ship Segmentation in SAR Images. *Remote Sens.* **2023**, *15*, 3304. https://doi.org/10.3390/rs15133304

Academic Editor: Ali Khenchaf

Received: 24 May 2023
Revised: 25 June 2023
Accepted: 26 June 2023
Published: 27 June 2023

1. Introduction

Synthetic aperture radar (SAR) is an imaging radar with high range and azimuth resolution, which is widely used in military and civilian fields due to its all-day and all-weather imaging capabilities. Target detection and segmentation are important parts of SAR image understanding and analysis. As the main transport carrier and effective combat weapon, automatic ship detection and segmentation provide important support for protecting inviolable maritime rights and maintaining maritime military security. Therefore, it is of great significance to carry out research on ship detection and segmentation in SAR images.

Most of the current ship detection methods [1–3] are based on the conventional object detection framework to achieve ship detection. These methods provide the position information of the bounding box covering the target, but they do not provide detailed contour information on the target. Target segmentation refers to segmenting the target of interest in images at the pixel level, which simultaneously provides position information and contour information of the target. Hence, ship segmentation is treated as a more accurate and comprehensive means to achieve ship detection.

The segmentation algorithms based on active contour are popular in the field of image segmentation, including the improved K-means active contour model [4,5], the

entropy-based active contour segmentation model [6], and the Chan–Vese model [7]. The Hidden Markov model is a commonly used method for image segmentation, which is a two-level structure model consisting of an unobservable hidden layer and an observable upper layer. Clustering analysis technology is also widely used to solve this issue, such as multi-center clustering algorithm [8], fast fuzzy segmentation [9], adaptive fuzzy C-means algorithm [10], and the bias correction fuzzy C-means algorithm [11]. As for SAR image segmentation, the most representative methods are the image segmentation algorithms [12–14] based on the constant false alarm rate (CFAR) detector [15], in which a threshold is determined based on the statistical characteristics of each image, and the image is segmented by comparing the gray level value of each pixel against the threshold value. CFAR-based methods consider pixel contrast information while ignoring the structural features of the target, which leads to speckle noise in the segmentation results, incorrect target localization, and a large number of false alarms.

With the rapid development of deep-learning technology, the convolutional neural network (CNN) has achieved excellent performance in the field of image processing [16–19], such as image classification [16], object detection [17], and object segmentation [18,19]. Many mature deep-learning methods have also been put forward in the field of SAR image processing. For example, Henry et al. [20] presented a fully convolutional neural network for road segmentation in SAR images and enhanced the sensitivity toward thin objects by adding spatial tolerance rules. Bianchi et al. [21] explored the capability of deep-learning models in segmenting the snow avalanches in SAR images at a pixel granularity for the first time. Deep-learning-based methods effectively improve the performance in data-intensive tasks, where a large amount of data is required to train the deep models. However, the performance of the deep-learning methods is limited or even ineffective when the available training dataset is relatively small. Moreover, the deep-learning-based methods have insufficient generalization ability in the task of SAR image processing due to the large imaging area and various imaging characteristics of the SAR images. Specifically, most of these methods have superior performance in the source domain data, but their performance is degraded in the target domain. Therefore, how to solve the problem of SAR ship segmentation on the cross-domain small dataset is still an extremely challenging task.

Transfer learning is a commonly used strategy in cross-domain tasks by transferring knowledge learned from a source domain with sufficient training data to a target domain lacking training data. However, a certain scale of target domain data is still required to achieve better results. Such a requirement is still a burden in SAR image processing because it is expensive and time-consuming to collect SAR images and provide the label, especially in the task of SAR segmentation, where the pixel-level ground truth is needed. Few-shot learning (FSL) has the ability to learn and generalize from a small number (one or several) of samples, which provides a feasible solution to the above problem. As a typical few-shot learning framework, meta-learning is borrowed from the way humans learn a new task. Humans rarely learn from scratch but learn based on the experience gained from the learning process of the related tasks when they learn a new skill. Meta-learning, also known as learning to learn, is proposed based on this learning mechanism of the human brain. The purpose of meta-learning is to learn from previous learning tasks in a systematic and data-driven way to obtain a learning method or meta-knowledge so as to accelerate the learning process of new tasks [22]. Therefore, the meta-learning framework is applied to solve the problem of SAR ship segmentation on the small cross-domain dataset.

The distribution of ship data in different regions is quite different due to various imaging modes, imaging resolutions, and imaging satellites. Ship segmentation in SAR images in different regions is considered a task originating from different domains. In this paper, a multi-scale similarity guidance few-shot network titled MSG-FN is proposed for ship segmentation in heterogeneous SAR images with few labeled samples in the target domain. The proposed MSG-FN adopts a dual-branch network structure, including a

support branch and a query branch. The support branch is used to extract the features of a specific domain target with a single encoder structure, while the query branch utilizes a U-shaped encoder–decoder structure to segment the target in the query image. These two branches share the same parameters in the encoder part, and the encoder is composed of well-designed residual blocks combined with filter response normalization (FRN). A similarity guidance module is designed to guide the segmentation process of the query branch by incorporating pixel-wise similarities between the features of support objects and query images. Four similarity guidance modules are deployed between the support branch and the query branch at various scales to enhance the detection adaptability of targets with different scales. In addition, a challenging ship target segmentation dataset named SARShip-4i is built by us to evaluate the proposed ship segmentation network, which includes both offshore and inshore ships.

The key contributions of this paper are as follows:

- A multi-scale similarity guidance few-shot learning framework with a dual-branch structure is proposed to implement ship segmentation in heterogeneous SAR images with few annotated samples;
- A residual block combined with FRN is designed to improve the generalization capability in the target domain, which forms the encoder of the support and query branches for domain-independent feature extraction;
- A similarity guidance module is proposed and inserted between two branches at various scales to perform hand-on-hand segmentation guidance of the query branch by pixel-wise similarity measurement;
- A ship segmentation dataset named SARShip-4i is built, and the experiment results on this dataset demonstrate that the proposed MSG-FN has superior ship segmentation performance.

The remainder of this paper is organized as follows. The previous related works are briefly described in Section 2. The proposed MSG-FN is presented in detail in Section 3. Experimental results and analysis are demonstrated in Section 4. Finally, the conclusion is made in Section 5.

2. Related Work

2.1. Semantic Segmentation

Semantic segmentation is a classic problem in the field of computer vision, which aims at the pixel-level classification of images, providing a foundation for subsequent tasks of image scene understanding and environment perception. The deep-learning method initially applied to image semantic segmentation is patch classification [23], in which the image is cut into blocks and fed into the depth model, and then the pixels are classified. Subsequently, the fully convolutional network (FCN) [24] was developed, which removes the original fully connected layer and converts the network to a fully convolutional model. The speed of FCN is much faster than that of the patch classification method, and the FCN method does not require the fixed size of the input image. However, the linear interpolation decoding method in the FCN leads to the loss of structure information, and the obtained boundary is relatively coarse despite the fact that some skipping connections are used. SegNet [25] is proposed to solve this problem by introducing more skipping connections and replicating the maximum pooled index. Another issue of the FCN model for semantic segmentation is the unbalance between the scale of the receptive field and the resolution of the feature map. The pooling layer enlarges the receptive field, but the resolution is reduced due to the down-sampling operation of the pooling layer, thus weakening the position information that semantic segmentation needs to preserve.

To keep the trade-off between the scale of the receptive field and the resolution of the feature map, the dilated convolutional structure and the encoder–decoder structure were proposed. Fisher et al. [26] designed a dilated convolutional network to realize semantic segmentation, which increases the respective field without decreasing the spatial

dimension. U-Net [27] is a typical encoder–decoder structure; the encoder gradually reduces the spatial dimension of the pooling layer, and the decoder recovers the details and spatial dimension of the target step by step. Moreover, there is a skip connection between the encoder and the decoder so that shallow features can assist in recovering the details of the target. Furthermore, RefineNet [28] was proposed based on U-Net, which exploited all the information available along the down-sampling process and used long-range residual connections to enable high-resolution prediction. In this way, the fine-grained features in the early convolution are used to refine the high-level semantic features captured by the deeper layers.

Accurate segmentation of targets with different scales is the focal and difficult issue of semantic segmentation. In order to achieve this goal, semantic segmentation methods need to integrate the spatial features of different scales to achieve an accurate description of multi-scale objects. The simple idea is to use the image pyramid [29], in which the input image is scaled into different sizes, and then the final segmentation result is obtained in an integrated way. In addition to the image pyramid, most of the current methods focus on how to make effective use of low-level features and high-level features. It is believed that the low-level features include rich location information, which is particularly important for accurate positioning, while the high-level features contain abundant semantic information, which is of great benefit to fine classification. In [30], a multi-scale context-aggregated module called the pyramid pooling module (PPM) was introduced, which uses different large-scale pooling kernels to capture global context information. On the basis of this work, Chen et al. [31] proposed an atrous spatial pyramid pooling (ASPP) module by replacing the pooling and convolution in PPM with the atrous convolution. Subsequently, the DenseASPP [32] was proposed to generate features with more various scales in a larger range by combining the advantages of parallel and cascade expansion convolution.

The above methods work well on large-scale natural images, but the performance of these algorithms decreases when the amount of training data is small. As for SAR images, the number of SAR images collected in a scene is limited due to the special imaging mode of the SAR images. Likewise, the amount of labeled SAR images that can be used to train the segmentation model is small because the pixel-level labeling of SAR images is time-consuming and laborious. Therefore, how to use the knowledge learned in other scenes to make predictions with few training data is an urgent problem worthy of consideration.

2.2. Few-Shot Learning

Few-shot learning is a learning paradigm proposed to solve the problem of small-scale training data, which refers to learning from a limited number of instance samples with supervised information. The proposal of few-shot learning has drawn lessons from the rapid learning mechanism of the human brain; that is, human beings quickly learn new tasks by using what they have learned in the past. The amount of training data determines the upper limit of the algorithm's performance. If a small-scale dataset is used to train a complex deep neural network in the traditional way, the over-fitting problem is inevitable. Due to the little demand for well-annotated training data, FSL has attracted wide attention and has been adopted in various image processing tasks, such as image classification [33–35], semantic segmentation [36–38], and object detection [39,40].

FSL aims at obtaining good learning performance given limited training samples, specifically, given a learning task and a dataset that consists of a training set and a test set. The number of training samples in the training set is relatively small, usually less than or equal to 5. The training set is also called the support set, and the test set is also called the query set. Suppose there is a theoretical mapping function satisfied between the input and the corresponding label. The purpose of few-shot learning is to find an approximate optimal mapping function in the mapping space by learning from other similar tasks so as to achieve accurate prediction on the test set. Few-shot learning is mainly reflected in the

number of samples in the support set, which is the number of well-annotated samples required when learning a new task.

Taking the most classic task of image classification as an example. The training set contains training data belonging to different categories. N is the number of categories contained in the training set; K is the number of images corresponding to each category, and the number of the training samples is $I = N \times K$. This kind of few-shot learning is called N-way K-shot learning. In particular, it is called one-shot learning, when $K = 1$.

2.3. Few-Shot Semantic Segmentation

Currently, there are some researchers trying to use few-shot learning to achieve image semantic segmentation. The most widely adopted technical route is to use the guidance information in the support set and guide the segmentation of the target in the query set by cleverly designing the network structure. The generally adopted network is a double-branch structure, as shown in Figure 1. The support image and its corresponding label are fed into the support branch to provide guidance for the query branch, and then the prediction result of the query image is obtained. From the perspective of the way to achieve guidance, the existing few-shot segmentation methods can be divided into three types [36], namely, matching-based methods [37,38], prototype-based methods [41], and optimization-based methods [42].

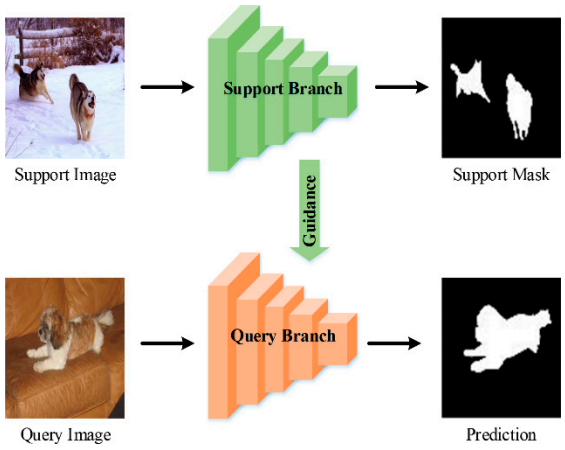


Figure 1. Typical structure of the few-shot semantic segmentation.

The typical matching-based method is SG-One [37], in which a similarity-guided one-shot semantic segmentation network was proposed. SG-One uses the dense pairwise feature to measure the similarity and a specific decoding network to generate segmentation results. On this basis, CANet [38] adds a multi-level feature comparison module to the dual-branch network structure and improves the segmentation performance through multiple iterations of optimization.

The prototype-based methods extract the global context information to represent each semantic category and use the overall prototype of the semantic category to match the query image at the pixel level. PANet [41] learns class-specific prototype representations by introducing prototype alignment regularization between the support branch and the query branch. Both the prototype-based methods and the matching-based methods use a metric-based meta-learning framework to compare the similarity between the support image and query image.

The optimization-based methods regard the few-shot semantic segmentation problem as a pixel classification problem. There are a few related works. Among them, MetaSegNet [42], which uses global and local feature extraction branches to extract meta-knowledge and integrates linear classifiers into the network to deal with pixel classification problems. MetaSegNet mainly focuses on N-way K-shot (N > 1) problem to realize the multi-objective segmentation problem.

The above methods mainly focus on the few-shot semantic segmentation of natural images. In the field of SAR image processing, it is not feasible to directly use the few-shot segmentation method in natural image scenes because of the large distribution difference of SAR images under different imaging conditions. Therefore, we remodel the few-shot segmentation method of SAR images and propose a multi-scale similarity guidance network to achieve ship segmentation in heterogeneous SAR images with limited annotation data.

3. Method

3.1. Problem Setup

In the few-shot semantic segmentation of natural images, segmenting targets of different classes is considered a different segmentation task. Different from this setup, ship segmentations in SAR images collected in various scenarios are treated as different segmentation tasks because there are large differences in the data distribution of SAR images due to the different imaging satellites, various imaging resolutions, and so on. Therefore, in this paper, the problem of SAR ship segmentation on the small cross-domain dataset is described as follows: the ship segmentation model is trained on SAR images collected in several regions, which is called a meta-training set, and our goal is to use the trained model to predict SAR images in the meta-testing set, i.e., SAR images collected from the target region with few annotated samples. There is no intersection between the regions of SAR image data used in the meta-training set and the meta-testing set.

For better understanding, there is an example to illustrate the definition of meta-training and meta-testing, as shown in Figure 2. The SAR images in the meta-training set are collected from Aswan Dam, Barcelona, Houston, and Singapore, while the SAR images in the meta-testing set are collected from Qingdao and Strait Gibraltar. Both the meta-training set $D_{train} = \{(S_i, Q_i)\}_{i=1}^{N_{train}}$ and the meta-testing set $D_{test} = \{(S_i, Q_i)\}_{i=1}^{N_{test}}$ consist of several episodes. The episode (S_i, Q_i) is the sample unit, which is composed of a support set S_i and a query set Q_i. The support set consists of several support images I_{S_i} and their corresponding segmentation annotation mask M_{S_i}.The query set consists of the input query images I_Q and their labels M_Q. In the training phase, the support–query pair (S_i, Q_i) in the meta-training set is used to train the model. In the test phase, $\{S, I_Q\} = \{I_{S_1}, M_{S_1}, I_{S_2}, M_{S_2}, \ldots, I_{S_k}, M_{S_k}, I_Q\}$ forms the input batch to the model, and the ground truth M_Q is used to evaluate the segmentation performance on the query image in each episode.

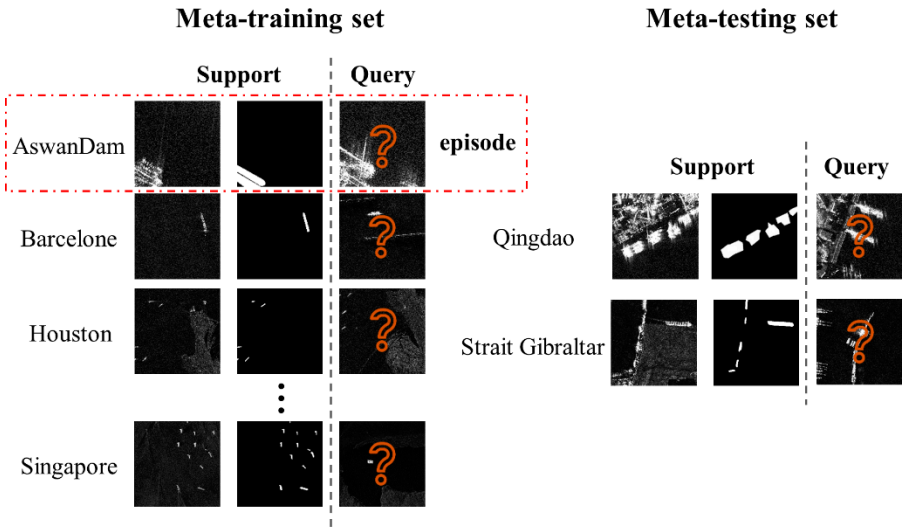


Figure 2. Schematic diagram of the meta-training set and the meta-testing set defined in the task of one-shot SAR ship segmentation. The red dotted box indicates the sample unit, *i.e.*, an episode.

3.2. MSG-FN Architecture

A multi-scale similarity guidance network (MSG-FN) is proposed to perform ship segmentation in heterogeneous SAR images with few labeled data in the target domain. The MSG-FN is a matching-based few-shot learning framework with two main modules, i.e., residual block combined with FRN (Res-FRN) and multi-scale similarity guidance module (SGM), which are introduced in the following Sections 3.3 and 3.4. The well-designed Res-FRN Block is proposed to capture robust and domain-independent features, and the multi-scale SGM is designed to conduct multi-scale hand-on-hand segmentation guidance.

The proposed MSG-FN is a dual-branch network, and its diagram is shown in Figure 3, where the support branch contains a convolutional layer and four Res-FRN blocks, and the query branch contains an input convolutional layer, four Res-FRN blocks, three deconvolution blocks, and an output convolutional layer. The support input is obtained by directly multiplying the support image with the support mask, which effectively removes the background and retains the target feature of the ship, avoiding interference caused by the complex and changeable background. Four SGMs are embedded between these two branches at various scales to better play the guiding role of the support branch to the query branch. The parameter-sharing mechanism is used between the support branch and the query branch. The support branch extracts the multi-scale features of the ship target from the limited support images and their corresponding support masks. Then, the features of the target are fused with the query features obtained by the query branch through the multi-scale SGM. The similarity map is obtained and used to realize the segmentation of the ship target in the query image by multiplying it with query features.

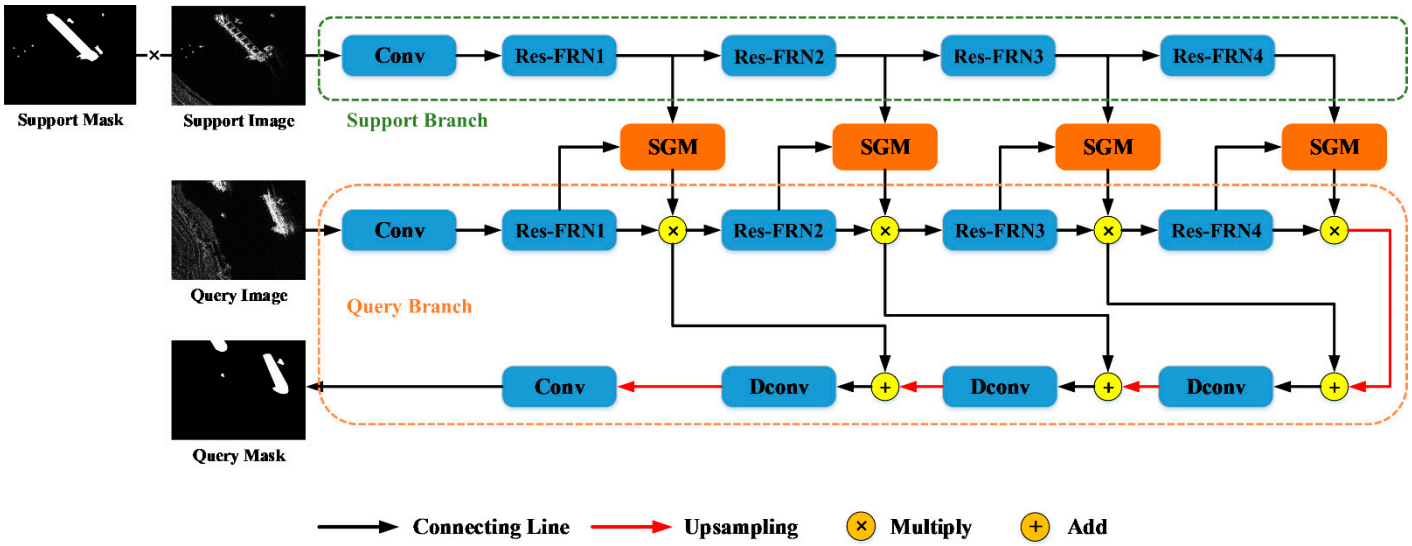


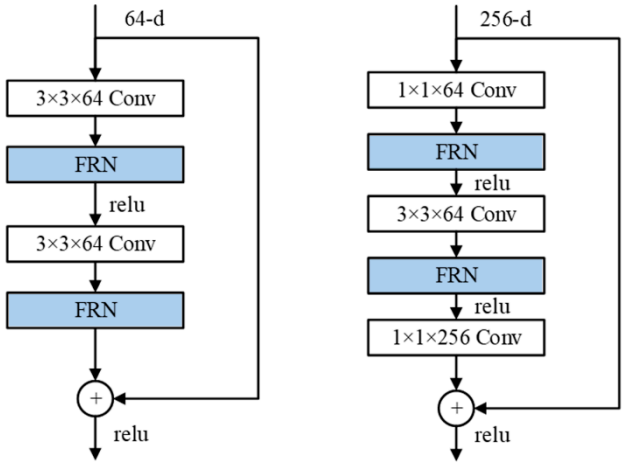
Figure 3. Schematic diagram of the proposed MSG-FN.

3.3. Residual Block Combined with FRN

The residual network (ResNet) is the most commonly used feature extraction network, which effectively reduces the difficulty of deep network training, making it possible to train networks with hundreds or even thousands of layers. There are two typical residual modules of ResNet: the basic block and the bottleneck. Batch normalization (BN) is a commonly used normalization method both in the basic block and the bottleneck, which is designed to alleviate the problem of internal covariate shifting. However, BN introduces dependence among samples, which leads to performance degradation when the source domain and the target domain have different distributions. More specifically, BN first standardizes each feature in a mini-batch and learns a common slope and bias for each mini-batch in the training phase, and then the global statistics of all training samples are

used to normalize each mini-batch of test data during the test. The statistics of the BN layer contain the traits of the source domain [43]. If the global statistics obtained from the training samples are used to normalize the test data on the new domain, performance degradation occurs due to differences in distribution between the source domain and the target domain. Moreover, the batch size set in BN may cause correlation among samples, affecting the training process. In the task of few-shot segmentation, the batch size during training is often set relatively small due to its special problem setting, so the network may have poor convergence if BN is used.

Filter response normalization (FRN) [44] is another normalization method, which operates on each activation map of each batch sample independently, eliminating the dependency of samples in the same batch. Therefore, we propose a residual block combined with FRN, namely, the Res-FRN block, to extract the domain-independent features for stronger generalization performance in target domains. As shown in Figure 4, FRN layers replace the BN layers in the designed Res-FRN block.



(a) Res-FRN block (basic) (b) Res-FRN block (bottleneck)

Figure 4. Schematic diagram of the designed Res-FRN block. (**a**) Res-FRN block (basic). (**b**) Res-FRN block (bottleneck).

In the FRN layer, assuming that the shape of the input tensor X is $[B, C, W, H]$, in which B is the batch size during training; C is the number of channels, and W and H represent the width and height of the input tensor, respectively. Let $v^2 = \frac{1}{M}\sum_i x_i^2$ be the mean squared norm of x, where $M = W \times H$. The FRN is calculated as follows:

$$\hat{x} = \frac{x}{\sqrt{v^2 + \varepsilon}} \tag{1}$$

where $x = X_{b,c,:,:} \in \mathbb{R}^N$; ε is a small constant to avoid the denominator being zero. Then, FRN performs affine transformation after normalization, as computed in Formula (2).

$$y = \gamma\hat{x} + \beta \tag{2}$$

where γ and β are both learnable parameters. This transformation guarantees that the input distribution of each layer remains unchanged across different mini-batches.

The bias term is fixed as a constant in the commonly used activation function rectified linear unit (ReLU), making the output value of the network lack some flexibility. Here, the threshold linear unit (TLU) activation function is selected as the activation function in the FRN layer, as computed in Formula (3), where a threshold τ is set as an optimizable parameter, which increases the flexibility of the network.

$$z = \max(y, \tau) \tag{3}$$

In addition, FRN is carried out on a per-channel basis, which ensures that all convolution kernel parameters have the same relative importance in the final model.

3.4. Multi-Scale Similarity Guidance Module

Most of the existing few-shot semantic segmentation methods [37,38] fuse the extracted support features with the query features at a single scale to achieve guidance. The features extracted from the support branch are precious, as they determine the final category the network will segment. However, inefficient support feature utilization has occurred in the above methods. Specifically, one single-scale guidance only considers the single output from the end of the network, which does not take full advantage of the multi-scale context features. Furthermore, ship targets have the characteristics of multiple scales, and the sizes of ships in the SAR images have significant differences. The fusion of support features and query features at a single scale is not enough to achieve the segmentation of ships of various scales. For example, if only the deepest features of the network are used to perform guidance, the segmentation of small targets will be greatly affected and even lead to a complete loss of information because the small-scale targets may lose location information as the depth of the network increases. Therefore, a multi-scale similarity guidance module is proposed to apply sufficient hand-on-hand guidance of support features to query features of different scales, which also enhances the adaptability of the algorithm to ship targets of different scales.

The architecture of the designed multi-scale SGM embedded in the proposed MSG-FN is illustrated in Figure 3. There are four residual blocks (Res-FRN1, Res-FRN2, Res-FRN3, Res-FRN4) in the encoder part of both the support branch and the query branch. Four SGMs are embedded between the support feature and the corresponding query feature with multi-scale sizes. The internal structure of the SGM is shown in Figure 5. The inputs are the support feature and the query feature, which are extracted by the residual blocks from the support branch and the query branch. The support feature contains semantic category information of the target, and the feature vector of the target is obtained through global average pooling, which contains global context semantic features of the ship target. Then, the cosine function is used to measure the similarity between the target feature vector obtained from the support image and the feature vector at each pixel in the query feature. Finally, a similarity matrix is generated as the guidance map to activate the target ship area in the query image by using the prior information in the support feature.

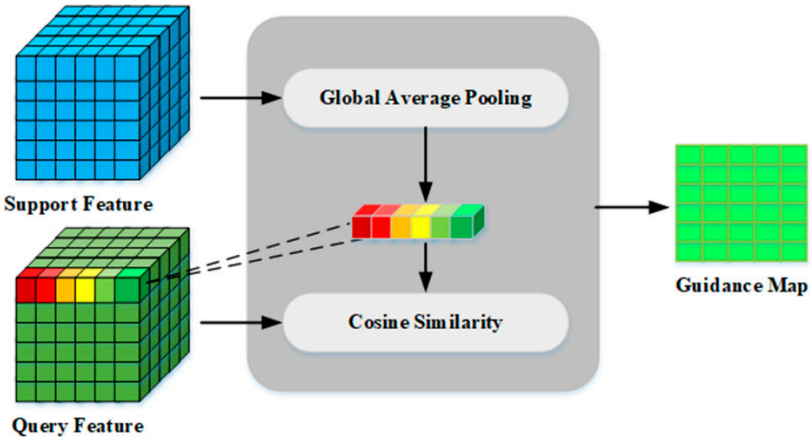


Figure 5. The internal structure of the similarity guidance module.

The guidance map, which is the output of the SGM, is calculated as follows:

$$s_{x,y} = \frac{u \cdot F_{x,y}^{query}}{\|u\|_2 \cdot \left\|F_{x,y}^{query}\right\|_2} \quad (4)$$

where $s_{x,y}$ represents the similarity value between the query feature and the feature vector at the pixel position of (x, y), and the value range is [−1,1]. $F_{x,y}^{support}$ and $F_{x,y}^{query}$ are the support feature vector and query feature vector at the pixel position of (x, y), respectively. u is the target feature vector, and u_i is the i^{th} element of u. u_i is calculated as follows, which is the average of all the pixels on the i^{th} support feature map:

$$u_i = \frac{\sum_{x=0,y=0}^{w,h} F_{i,x,y}^{support}}{\sum_{x=0,y=0}^{w,h} F_{x,y}^{support}} \quad (5)$$

The multi-scale guidance maps are visualized for better understanding of the designed multi-scale similarity guidance module, as shown in Figure 6. The first row is the support image, query image, and ground truth of ship segmentation. The second to fourth rows are support features, query features, and the generated guidance maps, respectively. Each column in the second to fourth rows is feature maps at different scales with the size of 256 × 256,128 × 128,64 × 64, and 32 × 32, and they are zoomed into a uniform scale for display. It is observed that the guidance maps contain the contour semantic information of the ship target learned from the support features, in addition to query features. Guidance maps at different scales contain complementary information, with shallower ones being able to focus on small targets, and deeper ones capturing more precise target location information. The multi-scale guidance maps provide sufficient hand-on-hand guidance, which allows the proposed MSG-FN method to be adaptable to ship targets at various scales.

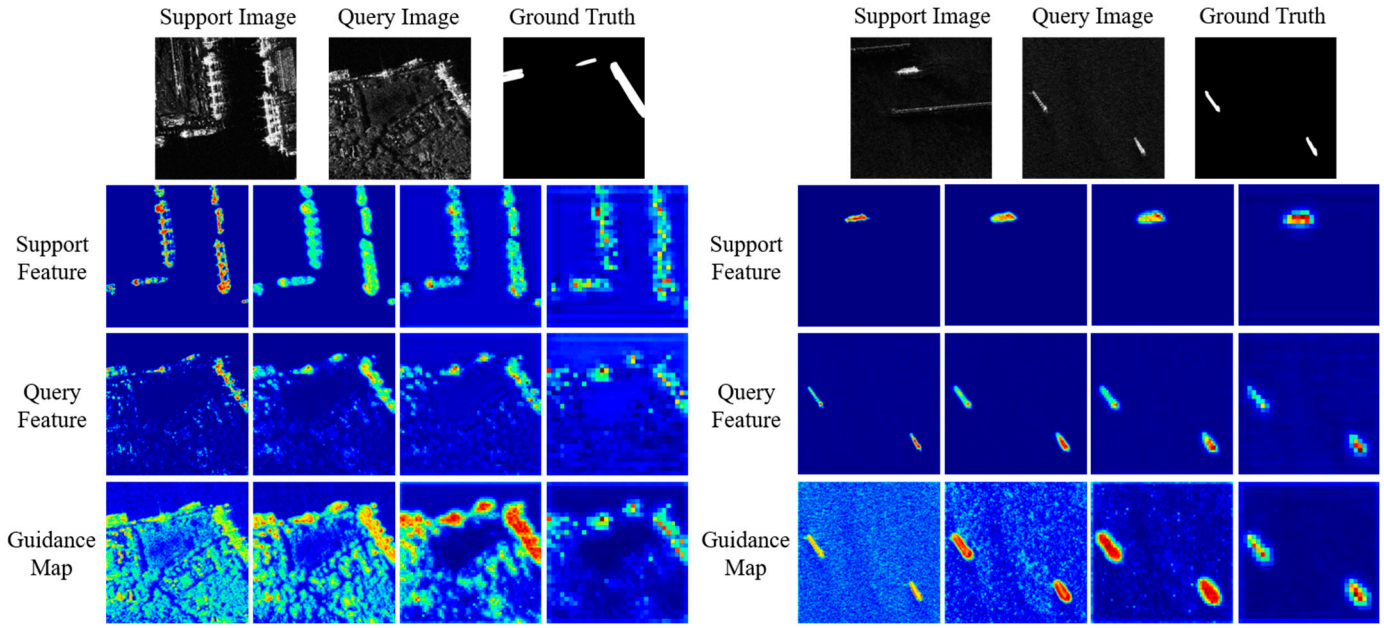


Figure 6. Visualization of multi-scale guidance maps (zoomed into a uniform scale for display).

3.5. Training and Inference

The whole training and inference procedures of the proposed MSG-FN on few-shot ship segmentation are summarized in Algorithm 1.

Algorithm 1 The Training and Test Procedures of the Proposed MSG-FN

Input: Meta-training set D_{train} and meta-testing set D_{test}.

Output: Network parameters θ.

Initialization: Initialize MSG-FN with Kaiming uniform

for each episode (S_i, Q_i) in D_{train} **do**

1. Extract feature from the support branch and query branch to obtain support features $F_1^s, F_2^s, F_3^s, F_4^s$ and query features $F_1^q, F_2^q, F_3^q, F_4^q$;
2. Get the similarity guide map s_1, s_2, s_3, s_4 for the query image;
3. Obtain the guided query features $F_1^{qs}, F_2^{qs}, F_3^{qs}, F_4^{qs}$ by multiplying s_1 and F_1^q, s_2 and F_2^q, s_3 and F_3^q, s_4 and F_4^q, respectively;
4. Fuse feature at different scale:
   (1) Concatenate the F_3^{qs} with up-sampled F_4^{qs} and then feed it into the Res-FRN block to get $F_3^{'qs}$;
   (2) Concatenate the F_2^{qs} with up-sampled $F_3^{'qs}$ and then feed it into the Res-FRN block to get $F_2^{'qs}$;
   (3) Concatenate the F_1^{qs} with up-sampled $F_2^{'qs}$ and then feed it into the Res-FRN block to get $F_1^{'qs}$;
5. Predict the segmentation mask of query image by feeding the $F_1^{'qs}$ to a Convout layer;
6. Update θ to minimize the cross-entropy loss via SGD.

End

for each episode (S_i, Q_i) in D_{test} **do**

1. Put forward the S_i and Q_i into the well-trained MSG-FN;
2. Predict the segmentation mask of the query image.

End

4. Experiment

4.1. SARShip-4i Dataset

This paper aims at ship segmentation in SAR images under the condition of a few annotated samples in the target domain, and the proposed MSG-FN should be evaluated on the few-shot ship segmentation dataset consisting of SAR images. However, there is no SAR dataset available so far to evaluate the performance of the few-shot ship segmentation algorithms. Therefore, we built a SAR dataset named SARShip-4i with reference to current COCO-20i [45] and Pascal-5i [46] datasets used for few-shot natural image segmentation to evaluate the proposed MSG-FN method for few-shot ship segmentation.

The SAR images in SARShip-4i dataset consist of two parts, one is the self-collected SAR images, whose segmentation labels are provided by the pixel-by-pixel manual annotation, and the other is the SAR images in the dataset HRSID [47], whose segmentation labels are generated based on the segmentation polygons provided in HRSID. SARShip-4i dataset contains 139 high-resolution real-world SAR images with resolutions ranging from 0.3 m to 3 m. There are several examples in the SARShip-4i dataset, as shown in Figure 7. It is obvious that all samples have different noise levels, especially since the first SAR image in the first row has severe noise. There are occlusions in the second and third SAR images in the first row, which are caused by the close proximity of multiple ships. The SAR images in the SARShip-4i dataset are captured under different conditions, such

as satellite, resolution, imaging mode, polarization, and incident angle. The detailed information is shown in Table 1. Note that the resolution of SAR images mentioned in Table 1 refers to the actual distance represented by one pixel in the SAR images. Furthermore, in the SARShip-4i dataset, the distortion introduced by incidence angles is relatively small, and most of the ship target information is abundant.

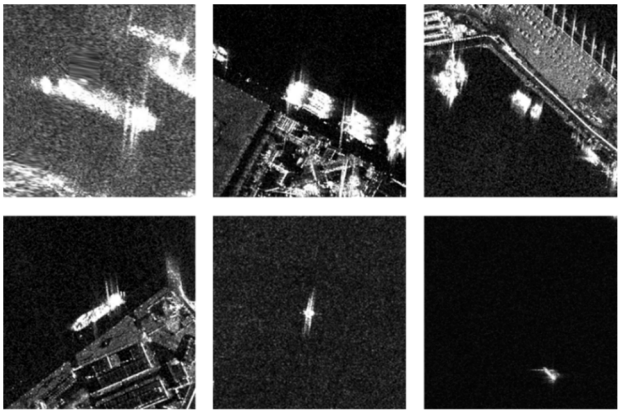


Figure 7. SAR images in the SARShip-4i dataset (image patches cropped from the original high-resolution SAR images are displayed here due to the limited space).

Table 1. The detailed information of SAR images in the SARShip-4i dataset.

Region	Imaging Satellite	Resolution (m)	Num. of Images	Imaging Mode	Polarization	Incident Angle (°)	Min. Size of Ships (Pixel)
Qingdao	TanDEM-X	0.3	1	ST	HH	-	68
Shanghai	TanDEM-X	0.3	1	ST	HH	-	66
Hong Kong	TerraSAR-X	1.0	1	HS	HH	-	761
Istanbul	TerraSAR-X	0.3	1	ST	VV	-	54
Houston	Sentinel-1B	3	40	S3-SM	HH	27.6~34.8	11
Sao Paulo	Sentinel-1B	3	21	S3-SM	HH	27.6~34.8	24
Sao Paulo	Sentinel-1B	3	20	S3-SM	HV	27.6~34.8	15
Barcelona	TerraSAR-X	3	23	SM	VV	20~45	26
Chittagong	Sentinel-1B	3	18	S3-SM	VV	27.6~34.8	23
Aswan Dam	TerraSAR-X	0.5	2	ST	HH	20~60	3478
Shanghai	TerraSAR-X	0.5	2	ST	HH	20~60	167
Panama Canal	TanDEM	1	1	HS	HH	20~55	86
Visakhapatnam	TerraSAR-X	1	1	HS	VV	20~55	182
Singapore	TerraSAR-X	3	4	SM	HH	20~45	47
Strait Gibraltar	TerraSAR-X	3	2	SM	HH	20~45	179
Bay Plenty	TerraSAR-X	3	1	SM	VV	20~45	43

The size of the ship targets in the SARShip-4i varies greatly, which poses a challenge for the design of the segmentation algorithm. Moreover, the appearance of ship targets with similar actual sizes in SAR images of different resolutions and different modes looks different, as shown in the second row in Figure 7. The pixel size of the ship targets in these three SAR images are 4685, 1344, and 475, respectively, and the resolution of these three SAR images are 0.3 m, 1 m, and 3 m, respectively. The actual size, which equals the pixel size multiplied by the image resolution of these three ship targets, is similar, but the ships in the three images look quite different due to the different SAR image resolutions and imaging modes.

The high-resolution SAR images are cropped to several image patches and rescaled to the same size of 512 × 512, and there is a total of 6961 image patches in the SARShip-4i

dataset. As mentioned in Section 3.1, SAR ship segmentations in different regions are treated as different segmentation tasks. The meta-training set and meta-testing set are set as SAR data from different regions considering different imaging modes and regional factors, and there is no intersection between the regions of SAR data used in the meta-training set and those predicted in the meta-testing set. The cross-validation strategy is applied here to evaluate the proposed MSG-FN. The SAR image patches in the SARShip-4i dataset are divided into four folds according to imaging regions, as shown in Table 2. In each fold, the SAR image patches in a fold form the meta-testing set, and the SAR image patches in the other three folds form the meta-training set. To the best of our knowledge, SARShip-4i is the first dataset that can be used to evaluate the few-shot ship segmentation methods in the SAR images.

Table 2. Details of the fold partition used for cross-validation in the SARShip-4i dataset.

Fold	Test Regions
SARShip-4^0	Visakhapatnam, Hong Kong, Barcelona, Chittagong
SARShip-4^1	Shanghai-HH, Singapore, Shanghai, Sao Paulo-HV
SARShip-4^2	Panama Canal, Bay Plenty, Istanbul, Sao Paulo-HH
SARShip-4^3	Aswan Dam, Strait Gibraltar, Qingdao, Houston

4.2. Implementation Details

In the setting of few-shot ship segmentation in the SAR images, the training process is carried out in a meta-learning manner, and the fundamental unit for training and testing is the episode. Each episode is composed of a support set and a query set. Each support set consists of several image patches; for example, the support set contains five image patches in the 1-way 5-shot, and the query set contains one image patch in this paper. Before training and testing, the image patch in the dataset should be organized into episode-based data. That is, an episode is generated by randomly selecting several image patches as support–query pair, and it is necessary to ensure that there are no duplicate image patches between the support set and the query set in an episode.

The backbone of the proposed MSG-FN is selected as the lightweight ResNet-18. Because of the large difference between the SAR images and natural images, the parameter of pre-trained on large-scale natural image datasets, such as ImageNet or COCO, cannot be used to initialize our model, and our model is trained from scratch. In the training phase, the network is optimized with stochastic gradient descent (SGD); the batch size is set as 3, and the momentum and weight decay are set as 0.9 and 0.0001, respectively. The learning rate linearly increases from 0 to 0.001 in the first 2000 steps and then decays exponentially to 300,000 steps with a decay rate of 0.9. The network is implemented using PyTorch, and all networks are trained and tested on an NVIDIA GTX 1080 GPU with 8 GB Memory.

4.3. Evaluation Metrics

There are four evaluation metrics used to evaluate the performance of the proposed MSG-FN, i.e., Precision, Recall, F1, and intersection over union (IoU). Precision and Recall are a pair of contradictory evaluation matrices, neither of which can fully measure the segmentation performance. F1 is a more comprehensive evaluation criterion, which maintains a trade-off between Precision and Recall. IoU is used to measure the degree of overlap between the segmentation result and the ground truth. These four evaluation metrics are calculated as follows:

$$Precision = \frac{p_{jj}}{\sum_{i=0}^{k} p_{ij}},\ Recall = \frac{p_{jj}}{\sum_{i=0}^{k} p_{ji}}, \tag{6}$$

$$F1 = 2 \times \frac{Precison_j \times Recall_j}{Precison_j + Recall_j},\ IoU = \frac{p_{jj}}{\sum_{i=0}^{k} p_{ij} + \sum_{i=0}^{k} p_{ji} - p_{jj}}. \quad (7)$$

where k is the number of categories of the target to be segmented, and k is set as 1 here because only that ship is the target in this paper. p_{ij} is the number of pixels that are inferred to belong to class j with the ground truth of class i. In other words, p_{ii}, p_{ij}, and p_{ji} represent the numbers of true positives, false positives, and false negatives, respectively.

4.4. Comparison with the State-of-the-Art

The proposed MSG-FN is evaluated against some state-of-the-art few-shot semantic segmentation methods under two experimental settings, namely, 1-way 1-shot and 1-way 5-shot. 1-way 1-shot means that only one annotated support image is used to guide the ship segmentation when making predictions on the query image of the unseen test data, and 1-way 5-shot refers to using five support images to guide the segmentation of the query image. In the setting of 1-way 5-shot, the final segmentation result Y is the average ensemble of the predicted masks generated with the guidance from the five support images, which is calculated as follows:

$$Y_{m,n} = avg(Y^1_{m,n}, Y^2_{m,n}, \ldots, Y^5_{m,n}) \quad (8)$$

where $Y^i_{m,n}, i = \{1,2,\ldots,5\}$ is the predicted semantic label of the pixel at (m,n), corresponding to the support image S_i.

There is no work specifically designed for few-shot SAR ship segmentation, and thus, we modify the state-of-the-art few-shot semantic segmentation approaches [37,48] on natural images to fit our settings for algorithm comparison. SG-One [37] predicts the segmentation mask of a query image by referring to one densely labeled support image of the same category, where only the deepest support features are used to provide guidance for the segmentation. PMMs [48] correlate various object parts with multiple prototypes estimated via an expectation–maximization algorithm to enhance few-shot semantic segmentation. RPMMs [48] are assembled by multi-PMMs using a residual structure. In the experiments, the training and testing settings specified in Section 4.1 are adopted. The experimental results of the proposed MSG-FN and three comparison methods on the settings of 1-way 1-shot and 1-way 5-shot are shown in Table 3 and Table 4, respectively.

Table 3. Segmentation results of the proposed MSG-FN and three state-of-the-art methods under the setting of 1-way 1-shot.

Metric	Method	SARShip-4^0	SARShip-4^1	SARShip-4^2	SARShip-4^3	Mean
Precision	SG-One [37]	0.4075	0.5777	0.5632	0.5507	0.5248
	PMMs [48]	0.6018	**0.8717**	0.6827	0.7973	0.7384
	RPMMs [48]	0.6023	0.7252	0.7477	0.7805	0.7139
	MSG-FN (ours)	**0.6822**	0.6890	**0.8453**	**0.8221**	**0.7597**
Recall	SG-One [37]	0.5208	0.6013	0.7093	0.6673	0.6247
	PMMs [48]	**0.7512**	0.6469	**0.8667**	**0.8526**	**0.7794**
	RPMMs [48]	0.6940	0.7692	0.8246	0.8295	0.7793
	MSG-FN (ours)	0.6699	**0.7939**	0.7889	0.8471	0.7750
F1	SG-One [37]	0.4204	0.5266	0.5865	0.5710	0.5261
	PMMs [48]	0.6264	0.7265	0.7232	0.8132	0.7223
	RPMMs [48]	0.6129	**0.7297**	0.7538	0.7930	0.7224
	MSG-FN (ours)	**0.6422**	0.7026	**0.8011**	**0.8282**	**0.7435**
IoU	SG-One [37]	0.3038	0.3790	0.4359	0.4287	0.3869
	PMMs [48]	0.5081	0.5853	0.6035	0.7068	0.6009

	RPMMs [48]	0.4897	**0.6027**	0.6320	0.6784	0.6007
	MSG-FN (ours)	**0.5314**	0.5962	**0.6927**	**0.7236**	**0.6360**

Table 4. Segmentation results of the proposed MSG-FN and three state-of-the-art methods under the setting of 1-way 5-shot.

Metric	Method	SARShip-4^0	SARShip-4^1	SARShip-4^2	SARShip-4^3	Mean
Precision	SG-One [37]	0.4135	0.6830	0.6175	0.5722	0.5716
	PMMs [48]	0.6066	0.8731	0.6840	0.7967	0.7401
	RPMMs [48]	0.6264	0.7544	0.7528	0.7980	0.7329
	MSG-FN (ours)	**0.6821**	0.6891	0.8451	0.8225	**0.7597**
Recall	SG-One [37]	0.5191	0.5741	0.6926	0.6594	0.6113
	PMMs [48]	**0.7494**	0.6456	**0.8674**	**0.8526**	**0.7788**
	RPMMs [48]	0.5938	0.6664	0.7246	0.7023	0.6718
	MSG-FN (ours)	0.6705	**0.7938**	0.7892	0.8469	0.7751
F1	SG-One [37]	0.4234	0.5748	0.6186	0.5822	0.5498
	PMMs [48]	0.6292	**0.7263**	0.7234	0.8131	0.7230
	RPMMs [48]	0.5783	0.6922	0.7028	0.7360	0.6773
	MSG-FN (ours)	**0.6425**	0.7027	**0.8012**	**0.8282**	**0.7437**
IoU	SG-One [37]	0.3065	0.4214	0.4661	0.4390	0.4083
	PMMs [48]	0.5106	0.5849	0.6037	0.7067	0.6015
	RPMMs [48]	0.4418	0.5497	0.5590	0.5983	0.5372
	MSG-FN (ours)	**0.5319**	**0.5963**	**0.6929**	**0.7237**	**0.6362**

The segmentation results on the four folds, as well as the mean results, are given in Tables 3 and 4. We pay more attention to the mean results, which comprehensively evaluate the performance of the segmentation algorithm. The proposed method has superior performance than the SG-One method [37] in terms of precision, recall, F1, and IoU under the settings of both 1-way 1-shot and 1-way 5-shot. This is because only the deepest semantic features of the supporting image are used for segmentation guidance in the SG-One method, and this single-scale feature guidance tends to ignore small ship targets, resulting in poor segmentation performance. The performance of the proposed MSG-FN is better than that of the PMMs and RPMMs methods [48] in terms of precision, F1, and IoU under both settings. The recall of the proposed MSG-FN is a little bit lower than that of the PMMs [48]. Overall, the proposed method achieves the best results for SAR image ship segmentation. In particular, the F1 and IoU of our method are 74.35% and 63.60% on the setting of 1-way 1-shot and 74.37% and 63.62% on the setting of 1-way 5-shot. The results on the setting of 1-way 5-shot are better than those on the setting of 1-way 1-shot because there are more images provided in the target domain on the setting of 1-way 5-shot.

The segmentation results on several samples are presented in Figure 8 to visually illustrate the superiority of the proposed MSG-FN. The first three rows are samples of ship segmentation in the off-shore scenes, and the last three rows are samples in the inshore scenes. The first and second columns are the SAR image and the ground truth of ship segmentation, and the third to sixth columns are the segmentation results of the SG-One [37], PMMs [48], RPMMs [48], and the proposed MSG-FN methods, respectively. It is obvious that the segmentation results of the proposed MSG-FN are more consistent with the ground truth compared with the other three methods. SG-One [37] has missing segmentation for some small-scale ship targets, as shown by the dashed yellow circles in the third and fifth rows. Meanwhile, there are many false alarms in the segmentation result of SG-One in complex inshore scenes, as shown by the dashed red circles in the fourth and sixth rows. The reason for the above phenomenon is that SG-One uses only a single-scale guidance module, and its applicability to ship targets of various scales is inferior to our

method. The segmentation results of the PMMs [48] and RPMMs [48] in the off-shore scene are similar to our method because the background in the off-shore scene is relatively simple. As for the inshore ship segmentation with a more complex and changeable background, there are many false alarms appearing in the segmentation results of PMMs and RPMMs, as shown by the dashed red circles in the fourth and sixth rows. This is because PMMs and RPMMs use a simple up-sampling interpolation method in the decoder part, while the proposed MSG-FN utilizes a U-shaped encoder–decoder structure. In conclusion, the experiments demonstrate that the segmentation results of the proposed MSG-FN are superior to other state-of-the-arts in terms of both quantitative metrics and qualitative visualization.

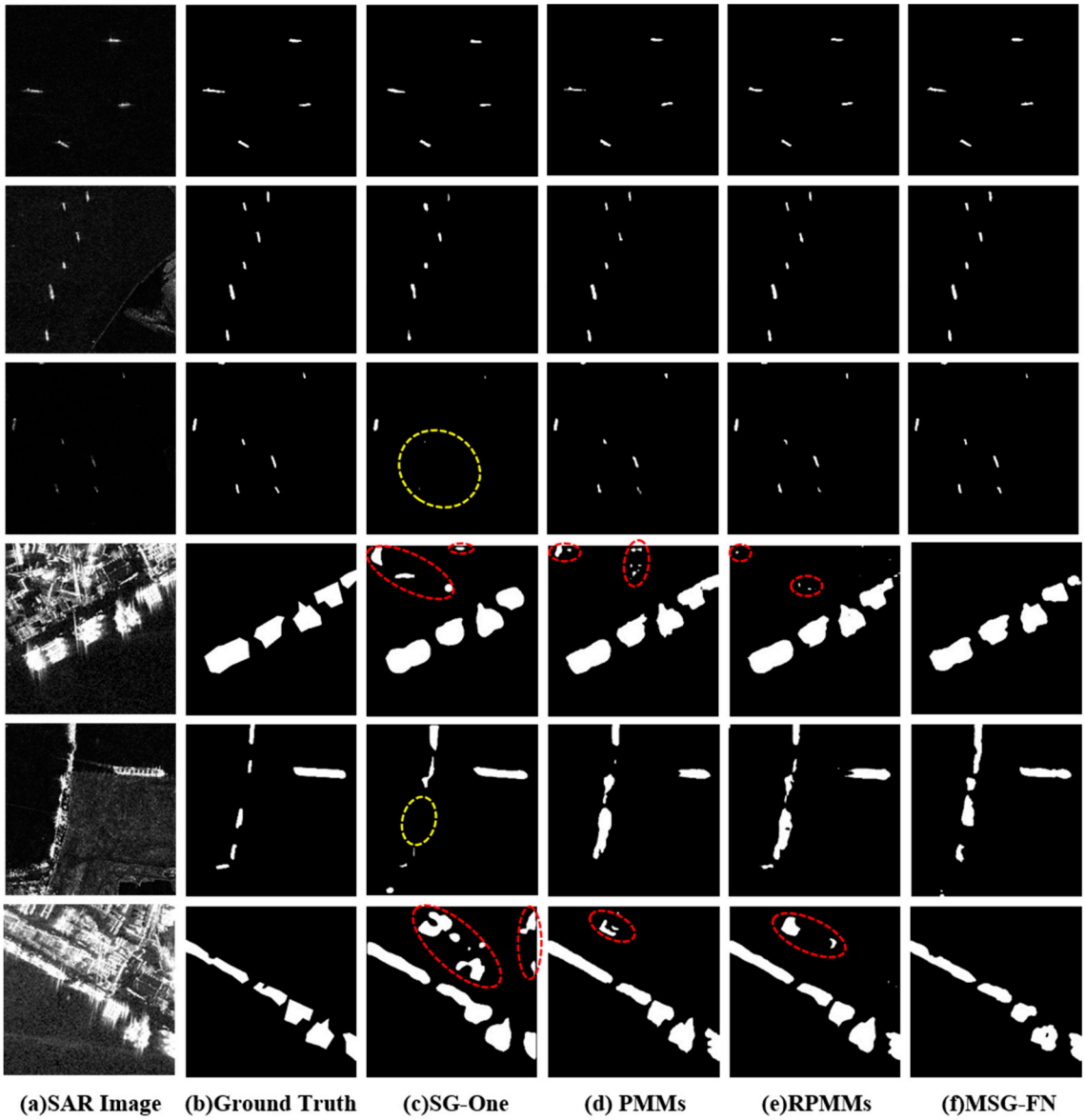


Figure 8. Visualization of segmentation results of the proposed MSG-FN and three comparison methods. The yellow circle represents a missed detection and the red circle represents a false alarm. (**a**) Original SAR image. (**b**) Ground truth. (**c**) Prediction results of SG-One [37]. (**d**) Prediction results of PMMs [48]. (**e**) Prediction results of RPMMs[48]. (**f**) Prediction results of MSG-FN.

4.5. Analysis of the Learning Strategy

In this section, we analyze three kinds of learning strategies, which are typically used to migrate models from the source domain to the target domain and validate the effectiveness of the few-shot strategy used in the proposed MSG-FN. The comparison results are reported in Table 5. U-Net [27] and PSPNet [30] are two classic segmentation methods, where the model trained on the source domain is directly used to perform inference on the target domain. U-Net (TL) and PSPNet (TL) are applied with a transfer learning strategy, which utilizes 40% of data from the target domain to fine-tune the model trained on the source domain. MSG-FN (1-shot) and MSG-FN (5-shot) are the proposed few-shot methods on the settings of the 1-way 1-shot and 1-way 5-shot.

Table 5. Comparison of different learning strategies used in the ship segmentation methods.

Method	IoU	Precision	Recall	F1-Score
U-Net [27]	0.5085	0.6461	0.7527	0.6411
PSPNet [30]	0.4481	0.5086	0.8009	0.5659
U-Net (TL)	0.5562	0.7129	0.7516	0.6886
PSPNet (TL)	0.6071	0.7704	0.7380	0.7301
MSG-FN (1-shot)	0.7236	0.8221	**0.8471**	**0.8282**
MSG-FN (5-shot)	**0.7237**	**0.8225**	0.8469	**0.8282**

As reported in Table 5, the performance of segmentation is poor when the trained model is directly used for prediction in the target domain. The performance has improved after utilizing the transfer learning strategy because a large amount of target information is learned to narrow the gap between the source domain and the target domain. Although transfer learning brings performance improvement, this strategy requires a certain number of annotated samples to be available for training in the target domain, which is not feasible in practical applications. It is noted that the proposed few-shot MSG-FN method has achieved the best performance. This is because the few-shot MSG-FN obtains meta-information about each domain data over a series of episodes of training, and it utilizes meta-information for the prediction of unseen data. Furthermore, the amount of the required labeled data in the target domain has been greatly reduced compared with the transfer learning methods. The experiment results have verified that the few-shot learning strategy in the proposed MSG-FN is effective in solving the problem of semantic segmentation of SAR images with few labeled training data available in the target domain.

4.6. Ablation Study

In this section, ablation experiments are carried out to verify the effectiveness of the two main modules in the proposed MSG-FN. The fourth fold, SARShip-4^{3,} defined in Table 2, is selected randomly to perform the ablation study under the setting of a 1-way 1-shot. The results of the ablation experiments are shown in Table 6. W/o Res-FRN represents a simplified version of MSG-FN, in which the Res-FRN block is replaced by the plain residual block. W/o multi-scale SGM represents another simplified version of MSG-FN, in which the multi-scale similarity guidance module is removed, and only a single similarity guidance module is deployed at the end of the encoder part.

Table 6. Results of ablation experiments.

Method	Precision	Recall	F1-Score	IoU
W/o Res-FRN	0.6925	0.8357	0.7269	0.6127
W/o Multi-scale SGM	0.8201	**0.8654**	0.8319	0.7337
MSG-FN (ours)	**0.8400**	0.8513	**0.8353**	**0.7398**

Res-FRN block. As shown in Table 6, the experiments demonstrate that MSG-FN with the proposed Res-FRN block achieves significant improvements by 10.84% and 12.71% in terms of F1 and IoU over the W/o Res-FRN method, which indicates that the Res-FRN block extracts the domain-independent features and has stronger generalization performance in target domains than the plain residual block.

Multi-scale SGM. In the proposed MSG-FN, a multi-scale similarity guidance module is used to perform hand-on-hand guidance of support features to query features of various scales. Its performance has been improved by 0.34% and 0.61% in terms of F1 and IoU compared to using a single similarity guidance module, which illustrates the adequacy of the hand-on-hand guidance, especially for the segmentation targets with various scales.

4.7. Robustness Analysis

The samples contained in the SARShip-4i dataset are real-world SAR images with different noise levels and actual occlusions. Several representative samples in the SARShip-4i dataset and their corresponding segmentation results are shown in Figure 9 to qualitatively illustrate the robustness of the proposed MSG-FN method for real-world SAR images with noise or occlusion. In Figure 9, the first row is the original SAR images in the SARShip-4i dataset. The second and third rows are the corresponding ground truth and segmentation results, respectively. It is obvious that all five SAR images have different noise levels; the first two SAR images especially have severe noise. There are occlusions in the last three SAR images, which are caused by the close proximity of multiple ships. The proposed method has good segmentation performance even under these real-world SAR images with different noise levels and actual occlusions, which illustrate the robustness of the proposed MSG-FN model.

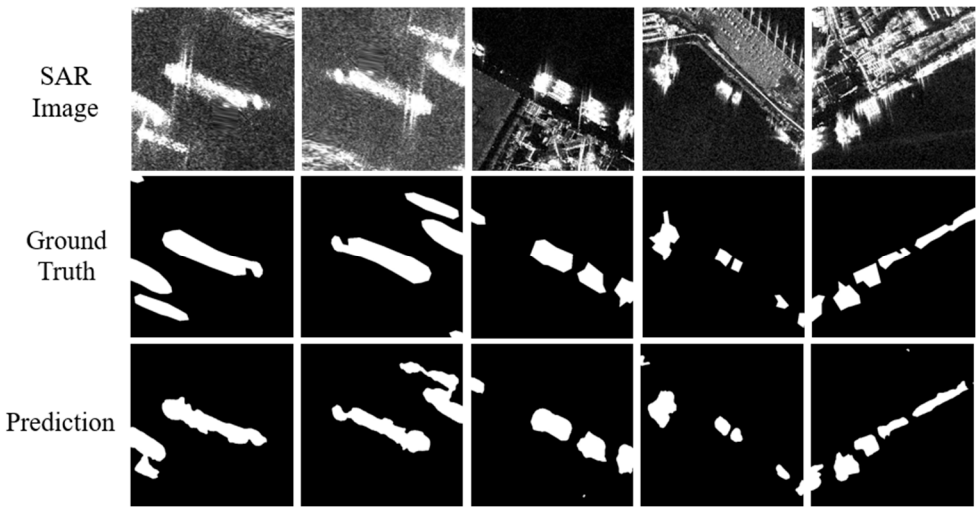


Figure 9. Real-word SAR images with noise or occlusions in the SARShip-4i dataset and their corresponding segmentation results of the proposed MSG-FN method.

4.8. Running time

In this section, the number of parameters, training time, and test time of the proposed method are reported. There are 15.82 M parameters in the proposed MSG-FN model. Taking the example of testing on the first fold (SARShip-40) and training on the other three folds (SARShip-41, SARShip-42, SARShip-43), the training time of the proposed method is 62.11 h, with 300,000 steps. The average test time for a single sample is 0.31 s on an NVIDIA GTX 1080 GPU with 8 GB Memory.

5. Conclusions

In this paper, a multi-scale similarity guidance network (MSG-FN) is proposed to perform the segmentation of ship targets in heterologous SAR images with few labeled data in the target domain. The proposed MSG-FN is a matching-based few-shot learning framework that has two main innovations. The first one is the well-designed Res-FRN block, which is proposed to capture robust and domain-independent features. The second one is a multi-scale similarity guidance module, which is proposed to provide sufficient hand-on-hand guidance of support features to query features of different scales, enhancing the adaptability of the algorithm to ship targets of different scales. In addition, a SAR ship segmentation dataset named SARShip-4i was built by us to evaluate the performance of the few-shot ship segmentation methods in SAR images. The proposed MSG-FN achieves superior segmentation results compared with other methods with only a few annotated data in the target domain, which provides a practical and feasible solution to the ship segmentation task in heterologous SAR images. Meanwhile, the proposed method establishes a new baseline for few-shot segmentation models, which can be applied to other few-shot segmentation problems after targeted improvement. At present, the inference speed of the proposed MSG-FN is less satisfactory, which will be addressed in future work.

Author Contributions: Conceptualization, R.L. and S.G.; methodology, R.L. and J.L.; software, J.L. and H.L.; writing—original draft, J.L. and R.L.; writing—review and editing, S.G., S.M. and Z.G.; visualization, R.L.; supervision, R.L. and S.G.; funding acquisition, R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 62102296) and the Fundamental Research Funds for the Central Universities (Grant No. XJS222221, XJS222215).

Data Availability Statement: The SARShip-4i datasets can be obtained from https://drive.google.com/file/d/15Q-5A_GQBAQTOp9rHXedbU7S1Dti-tR9/view?usp=share_link (accessed on 25 June 2023). The code is available at https://github.com/imiuupload/MSG-FN (accessed on 25 June 2023).

Acknowledgments: The authors would like to express their gratitude to the editors and the anonymous reviewers for their insightful comments.

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

Abbreviations

The following abbreviations and mathematical symbols are used in this manuscript:

SAR	Synthetic aperture radar;
MSG-FN	Multi-scale similarity guidance few-shot network;
CFAR	Constant false alarm rate;
CNN	Convolutional neural network;
FSL	Few-shot learning;
FRN	Filter response normalization;
FCN	Fully convolutional network;
PPM	Pyramid pooling module;
ASPP	Atrous spatial pyramid pooling;
Res-FRN	Residual block combined with FRN;
SGM	Similarity guidance module;
ResNet	Residual network;
BN	Batch normalization;
TLU	Threshold linear unit;
ReLU	Rectified linear unit;
SGD	Stochastic gradient descent;
IoU	Intersection over union;
N	The number of categories in the training set;

K	The number of samples of each category in the training set;
I	The number of training samples;
D_{train}, D_{test}	The meta-training set and the meta-testing set;
S_i, Q_i	The support set and the query set;
I_{S_i}, M_{S_i}	The support image and its segmentation label;
I_Q, M_Q	The query image and its segmentation label;
X	The input tensor of the FRN layer;
B	The batch size during training;
C	The number of channels;
W, H	The width and height of X;
ε	A small constant;
γ, β	Learnable parameters;
$s_{x,y}$	The similarity value;
$F_{x,y}^{support}$	The support feature vector;
$F_{x,y}^{query}$	The query feature vector;
u	The target feature vector;
p_{ii}	The number of true positives;
p_{ij}	The number of false positives;
p_{ji}	The number of false negatives;
$Y_{m,n}$	The predicted semantic label of pixel at (m, n).

References

1. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *16*, 751–755.
2. Zhang, X.; Wang, H.; Xu, C.; Lv, Y.; Fu, C.; Xiao, H.; He, Y. A lightweight feature optimizing network for ship detection in SAR image. *IEEE Access* **2019**, *7*, 141662–141678.
3. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote. Sens.* **2019**, *57*, 8983–8997.
4. Song, Y.; Peng, G.; Sun, D.; Xie. X. Active contours driven by Gaussian function and adaptive-scale local correntropy-based K-means clustering for fast image segmentation. *Signal Process.* 2020, 174, 107625.
5. Xia, X.; Lin, T.; Chen, Z.; Xu, H.. Salient object segmentation based on active contouring. *PLoS ONE* **2018**, *12*, e0188118.
6. Zong, J.; Qiu, T.; Li, W.; Guo, D. Automatic ultrasound image segmentation based on local entropy and active contour model. *Comput. Math. With Appl.* 2019, 78(3), 929-943.
7. Xu, L.; Xiao, J.; Yi, B.; Lou, L.An Improved C-V Image Segmentation Method Based on Level Set Model. In Proceedings of the International Conference on Intelligent Networks and Intelligent Systems, Wuhan, China, 1–3 November 2008; pp. 507–510.
8. Visalakshi, N.K.; Suguna, J. K-means clustering using Max-min distance measure. In Proceedings of the Nafips 2009 Meeting of the North American, Cincinnati, OH, USA, 14–17 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6.
9. Dubey, Y.K.; Mushrif, M.M. FCM Clustering Algorithms for Segmentation of Brain MR Images. *Adv. Fuzzy Syst.* **2016**, *2016*, 3406406.
10. Pham, D.L.; Prince, J.L. An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recognit. Lett.* **1999**, *20*, 57–68.
11. Salem, W.S.; Ali, H.F.; Seddik, A.F. Spatial Fuzzy C-Means Algorithm for Bias Correction and Segmentation of Brain MRI Data. *Int. Conf. Biomed. Eng. Sci.* **2015, 57-65**.
12. Cui, Y.; Yang, J.; Zhang, X. New CFAR target detector for SAR images based on kernel density estimation and mean square error distance. *J. Syst. Eng. Electron.* **2012**, *23*, 40–46.
13. Huang, S.; Huang, W.; Zhang, T. A New SAR Image Segmentation Algorithm for the Detection of Target and Shadow Regions. *Sci. Rep.* **2016**, *6*, 38596.
14. Hou, B.; Chen, X.; Jiao, L. Multilayer CFAR Detection of Ship Targets in Very High Resolution SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 811–815.
15. Crisp, D.J. The state-of-the-art in ship detection in Synthetic Aperture Radar imagery. Defence Science and Technology Organisation Salisbury (Australia) Info Sciences Lab,**2004**.
16. Tan, M.; Le, Q.V.; EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
17. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *99*, 2999–3007.
18. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H.*Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*; Springer: Cham, Switzerland, 2018.

19. Ding, H.; Jiang, X.; Shuai, B.; Liu, A.; Wang, G.Semantic Correlation Promoted Shape-Variant Context for Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; IEEE: Piscataway, NJ, USA, 2019.
20. Henry, C.; Azimi, S.M.; Merkle, N. Road Segmentation in SAR Satellite Images with Deep Fully Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1867–1871.
21. Bianchi, F.M.; Grahn, J.; Eckerstorfer, M.; Malnes, E.; Vickers, H.Snow Avalanche Segmentation in SAR Images with Fully Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 75–82.
22. Vanschoren, J. Meta-learning: A survey. *arXiv* **2018**, arXiv:1810.03548.
23. Dan, C.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J.Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 2852–2860.
24. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference Oil Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3432–3440.
25. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495.
26. Fisher, Y.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2015**, arXiv:1511.07122.
27. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
28. Lin, G.; Milan, A.; Shen, C.; Reid, I.Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
29. Rezaee, M.R.; van der Zwet, P.M.J.; Lelieveldt, B.P.E.; van der Geest, R.J.; Reiber, J.H.C. A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. *IEEE Trans. Image Process.* **2000**, *9*, 1238–1248. https://doi.org/10.1109/83.847836.
30. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239. https://doi.org/10.1109/CVPR.2017.660.
31. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H.Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
32. Yang, M.; Yu, K.; Zhang, C.; Li, Z.; Yang, K. DenseASPP for Semantic Segmentation in Street Scenes. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3684–3692. https://doi.org/10.1109/CVPR.2018.00388.
33. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. *arXiv* **2017**, arXiv:1703.03400.
34. Nichol, A.; Schulman, J. Reptile: A scalable metalearning algorithm. *arXiv* **2018**, arXiv:1803.02999.
35. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4077–4087.
36. Liu, Y.; Zhang, X.; Zhang, S.; He, X.Part-aware prototype network for few-shot semantic segmentation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 142–158.
37. Zhang, X.; Wei, Y.; Yang, Y.; Huang, T.Sg-one: Similarity guidance network for one-shot semantic segmentation. *IEEE Trans. Cybern.* **2020**, *50*, 3855–3865.
38. Zhang, C.; Lin, G.; Liu, F.; Yao, R.; Shen, C.Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 5217–5226.
39. Perez-Rua, J.-M.; Zhu, X.; Hospedales, T.; Xiang, T. Incremental Few-Shot Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
40. 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, Seoul, Republic of Korea, 27 October–2 November 2019.
41. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J.Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9197–9206.
42. Tian, P.; Wu, Z.; Qi, L.; Wang, L.; Shi, Y.; Gao, Y.Differentiable meta-learning model for few-shot semantic segmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7-12 February 2020; Volume 34, pp. 12087–12094.
43. Li, Y.; Wang, N.; Shi, J.; Liu, J.; Hou, X.Revisiting Batch Normalization for Practical Domain Adaptation. *arXiv* 2016, arXiv:1603.04779.
44. Singh, S.; Krishnan, S. Filter Response Normalization Layer: Eliminating Batch Dependence in the Training of Deep Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020.
45. Shaban, A.; Bansal, S.; Liu, Z.; Essa, I.; Boots, B.One-Shot Learning for Semantic Segmentation. In Proceedings of the British Machine Vision Conference 2017, London, UK, 4–7 September 2017.
46. Nguyen, K.; Todorovic, S. Feature weighting and boosting for few-shot segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 622–631.

47. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J.HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access* **2020**, *8*, 120234–120254.
48. Yang, B.; Liu, C.; Li, B.; Jiao, J.; Ye, Q.Prototype Mixture Models for Few-Shot Semantic Segmentation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 763–778.