remote sensing

MDPI

Article

Multitask Learning-Based for SAR Image Superpixel Generation

Jiafei Liu [1], Qingsong Wang [2,*], Jianda Cheng [1], Deliang Xiang [3,4] and Wenbo Jing [5]

[1] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China; 2020200813@mail.buct.edu.cn (J.L.); 2020400179@mail.buct.edu.cn (J.C.)
[2] School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China
[3] Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China; xiangdeliang@mail.buct.edu.cn
[4] Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing 100029, China
[5] College of Electronic Science, National University of Defense Technology, Changsha 410073, China; wenbo.jing@nudt.edu.cn
* Correspondence: wangqs5@mail.sysu.edu.cn

Abstract: Most of the existing synthetic aperture radar (SAR) image superpixel generation methods are designed based on the raw SAR images or artificially designed features. However, such methods have the following limitations: (1) SAR images are severely affected by speckle noise, resulting in unstable pixel distance estimation. (2) Artificially designed features cannot be well-adapted to complex SAR image scenes, such as the building regions. Aiming to overcome these shortcomings, we propose a multitask learning-based superpixel generation network (ML-SGN) for SAR images. ML-SGN firstly utilizes a multitask feature extractor to extract deep features, and constructs a high-dimensional feature space containing intensity information, deep semantic informantion, and spatial information. Then, we define an effective pixel distance measure based on the high-dimensional feature space. In addition, we design a differentiable soft assignment operation instead of the non-differentiable nearest neighbor operation, so that the differentiable Simple Linear Iterative Clustering (SLIC) and multitask feature extractor can be combined into an end-to-end superpixel generation network. Comprehensive evaluations are performed on two real SAR images with different bands, which demonstrate that our proposed method outperforms other state-of-the-art methods.

Keywords: multitask learning; SAR image superpixel generation; high-dimensional feature space; pixel-superpixel soft assignment

check for updates

Citation: Liu, J.; Wang, Q.; Cheng, J.; Xiang, D.; Jing, W. Multitask Learning-Based for SAR Image Superpixel Generation. *Remote Sens.* **2022**, *14*, 899. https://doi.org/10.3390/rs14040899

Academic Editors: Elise Colin-Koeniguer and Flora Weissgerber

Received: 28 December 2021
Accepted: 7 February 2022
Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Synthetic aperture radar (SAR) has been widely used in many fields due to the capability of providing unique and useful information in all-weather and multi-climate conditions, such as sea monitoring [1], agricultural development [2], urban planning [3]. Due to the severe speckle noise within SAR images, traditional pixel-level SAR image interpretation methods cannot effectively reduce the impact of speckle noise. In addition, the high computational load of traditional pixel-level methods severely restricts the application of large-size SAR images. Superpixel generation is an image over-segmentation technology that has been widely concerned in recent years. Compared with pixels, superpixels can provide the spatial context information of the pixel and the boundary information of the target, both of which play an important role in subsequent tasks [4]. Additionally, through the homogeneous description of the pixels within the superpixel, the impact of speckle noise can be effectively reduced. In addition, for the reason that the number of superpixels is much smaller than the number of pixels, the computational cost of subsequent tasks is greatly reduced, which opens the way for the application of large-size SAR images [5].

The purpose of superpixel generation is to over-segment the image into locally connected areas with perceptual significance. The pixels in the connected areas have similar

attributes. To date, there have been many superpixel generation algorithms in the field of natural images. The existing superpixel generation algorithms can be roughly divided into two categories, graph-based algorithms and gradient-ascent-based algorithms. In the graph-based methods, the nodes of the graph represent pixels, and the edges of the graph, represent the similarity between adjacent pixels. Finally, superpixels are generated by minimizing a cost function defined over the graph [6]. The classic graph-based superpixel generation methods include Normalized Cut (NC) [7], Minimum spanning tree (MST) [8], Entropy Rate Superpixels (ERS) [9], etc. In the gradient-ascent-based methods, the initial clusters are iteratively optimized along the gradient ascent direction until some convergence criteria are met. The classic gradient-ascent-based superpixel generation methods include Mean Shift (MS) [10], Watersheds [11], Turbopixels (TP) [12], Simple Linear Iterative Clustering (SLIC) [13], etc. Among the above methods, SLIC is widely used in the natural image superpixel generation due to its simple operation and easy understanding.

SLIC uses a modified K-means clustering to generate superpixels by measuring the similarity between pixels in a five-dimensional space composed of LAB colors and spatial coordinates. Since the imaging mechanism of SAR images is different from optical images, SLIC cannot be directly applied to the superpixel generation of SAR images. Xiang et al. [14] quantified the similarity between SAR image pixels by defining pixel intensity and location similarity, and for the first time successfully applied SLIC to superpixel generation for SAR images. Zou et al. [15] used the generalized gamma distribution to estimate the likelihood information of SAR image pixel clusters, and combined spatial context and likelihood information to improve the boundary adherence of superpixels. In terms of SAR image pixel similarity measurement, Akyilmaz et al. [16] used Mahalanobis distance as a measure of pixel similarity to achieve superpixel generation for SAR images. Xiang et al. [17] defined the distance to measure both the feature similarity and spatial proximity, and combined local K-means clustering and Ncut to generate SAR image superpixels. Jing et al. [18] proposed a content-sensitive superpixel generation method with edge penalty and the contraction expansion search strategy for SAR images. It is worth noting that the above superpixel generation methods are designed based on the raw SAR data or artificially designed features to measure the similarity between pixels. However, SAR data is severely affected by speckle noise, and artificially designed features have great limitations in complex scenes, resulting in unstable pixel similarity measurements and unsatisfactory superpixel generation results.

Benefiting from the emergence of deep learning represented by convolutional neural networks (CNNs), it is expected to make up for the lack of artificially designed features [19,20]. Chen et al. [21] used a CNN-based method to mine the deep representation of artificially designed features, and obtains better classification results than artificially designed features. Yue et al. [22] proposed a full convolutional network (FCN) of channel attention and spatial attention to enhance the ability of extracting deep features for SAR image segmentation. Therefore, using deep features for pixel similarity measurement is expected to obtain better superpixel generation results. Yang et al. [23] designed a novel method that uses a simple FCN to predict superpixels on a regular image grid. Jampani et al. [24] developed a differentiable model for superpixel sampling that leverages deep networks for learning superpixel generation. However, there is no research result on the superpixel generation algorithm for SAR images based on the deep features. One reason is that the labeled data used to generate superpixels in SAR images is scarce, which is not enough to support the training of deep learning models.

In this paper, we propose a novel SAR image superpixel generation network based on multitask learning (ML-SGN). First, we adopt multitask learning to solve the problem of insufficient training data for SAR image superpixel generation. We utilize the SAR segmentation dataset to train the SAR image segmentation task, and intercept convolutional layers in the segmentation model as the feature extractor of the superpixel generation task. In multitask learning, network parameters of the auxiliary task need to be fine-tuned on the main task [25]. However, SLIC is non-differentiable because it contains

nearest neighbor operations. Therefore, it is impossible to construct an end-to-end network architecture combined with the multi-task feature extractor, thus it is impossible to fine-tune the parameters of the multi-task feature extractor. Inspired by [24,26], we replace the nearest neighbor operation in SLIC with a differentiable soft assignment. In addition, we construct a high-dimensional feature space including deep semantic information, spatial information, and intensity information, and define an effective distance measure between pixels based on the high-dimensional feature space. The experimental results on two real SAR images show that the proposed method is superior to other state-of-the-art superpixel generation methods.

The main contributions of this article are summarized as follows.

(1) We propose a multitask learning-based superpixel generation network (ML-SGN). SAR image segmentation is used as an auxiliary task to extract deep features of SAR images, which solves the problem of insufficient labeled samples for SAR image superpixel generation.
(2) We construct a high-dimensional feature space containing deep semantic information, intensity information and spatial information, and define an effective pixel distance measure based on this high-dimensional feature space.
(3) We design a soft assignment operation instead of the nearest neighbor operation to make SLIC differentiable. It can be used to construct an end-to-end superpixel generation network with the multitask feature extractor, which can be used to fine-tune the parameters of the multitask feature extractor.

The rest of this paper is organized as follows. In Section 2, the proposed method is introduced, including multitask feature extractor, pixel distance measure, and pixel-superpixel soft assignment. Section 3 lists the experimental results of our proposed method on two real SAR images as well as the comparisons with other state-of-the-art methods. Section 4 gives the discussion on the number of superpixels and parameter fine-turning. Finally, conclusions are given in Section 5.

2. Methodology

Figure 1 shows a schematic diagram of the proposed method. The proposed method consists of 3 main parts, including multitask feature extractor, pixel distance measure, and pixel–superpixel soft assignment. First, the multitask feature extractor is trained in auxiliary task. Then, the deep features extracted by the multi-task feature extractor are combined with the intensity information and spatial information into a high-dimensional feature space. Second, we define an effective distance measure between pixels based on the high-dimensional feature space. Finally, we design a pixel–superpixel soft assignment to make SLIC differentiable.

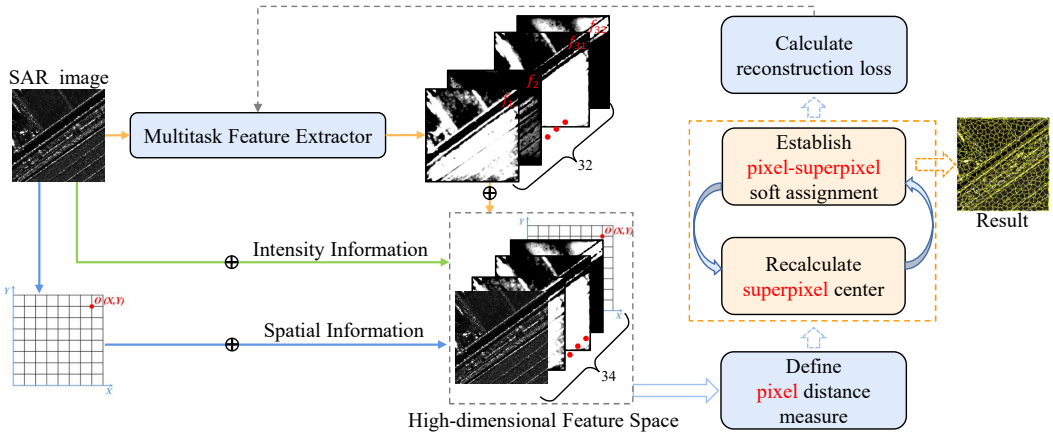


Figure 1. The architecture of the proposed classification method.

2.1. Multitask Feature Extractor

To solve the problem of insufficient training samples, we adopt multitask learning to train the feature extractor. Multitask learning is a deduced transfer learning method that puts multiple related tasks together for learning. During the learning process, auxiliary task promotes the main task learning through the network layer and parameter sharing. In this paper, we use the U-Net [27,28] image segmentation network as an auxiliary task, and the network structure and parameters are shown in Figure 2.

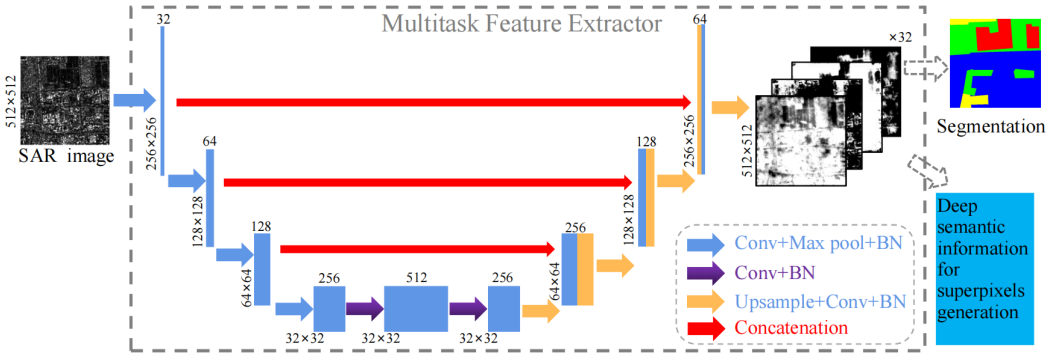


Figure 2. The network structure of multitask feature extractor.

SAR image superpixel generation cannot only consider deep semantic information, but also need edge information, which plays an important role in the superpixel generation [29]. As we all know, with the deepening of the convolutional layer, the salient information is retained, whereas the edge information is gradually lost. UNet adopts skip connection to fuse the features of the encoder and decoder, so that the extracted deep features can contain both low-level and high-level semantic information. At the same time, through four times of upsampling, the edge information recovered in deep features is more refined.

The SAR image segmentation dataset is selected from the 2020 Challenge on Automated High-Resolution Earth Observation Image Interpretation acquired by Gaofen 3 [30,31]. We select 100 SAR images with HV polarization, as shown in Figure 3. The size of each SAR image is 512 × 512. The ground truth map contains five classical land covers, including grass, water, industrial area, bare soil, and urban area.

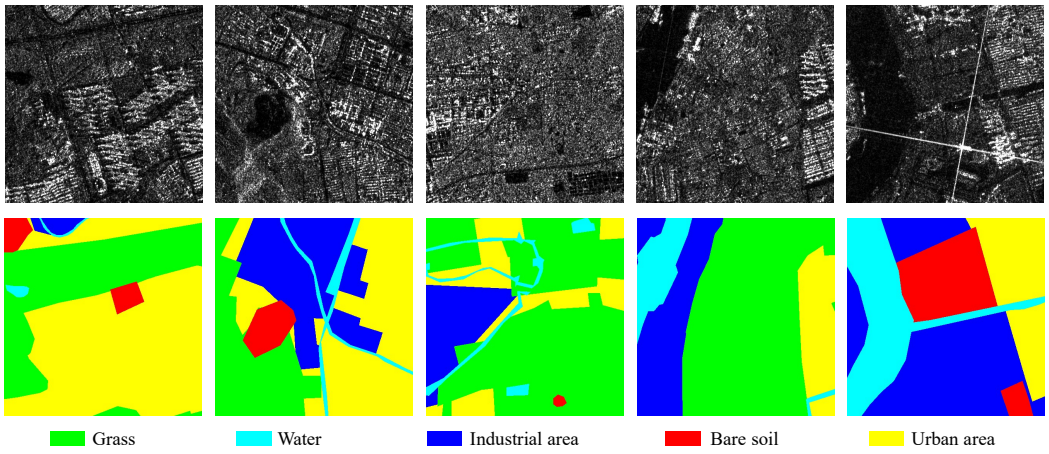


Figure 3. Partial training dataset of multi-task feature extractor.

2.2. Pixel Distance Measure

Although the deep features extracted by UNet contain both low-level and high-level semantic information, but the original intensity information and spatial information are dis-

carded. Therefore, in this paper, we construct a high-dimensional feature space containing deep semantic information (F), spatial information (O), and intensity information (I).

Classic SLIC uses Euclidean distance as the distance metric between pixels [32,33]. However, a large amount of multiplicative speckle is introduced in the SAR imaging. Euclidean distance is robust to additive noise, but sensitive to multiplicative noise. Therefore, we use a ratio distance as the measure of intensity information, which has been proven to be robust to multiplicative noise [34]. Let i and j denote two pixels, I_{N_i} and I_{N_j} denote patches with the same size and shape centered on i and j in the intensity image, respectively. The ratio distance is then defined as

$$d_I(i,j) = \left\| \frac{I_{N_i}}{I_{N_j}} \right\|_{2,G} = \sqrt{\sum_{k=1}^{M} G(k) \left| \frac{I_{N_i}(k)}{I_{N_j}(k)} \right|^2}, \tag{1}$$

where M represents the number of pixels contained in the patch. Function $G(\bullet)$ represents the standard Gaussian kernel. In this paper, we set the size of patch as 3×3 [14,34].

Let $F = \{f_1, f_2, \ldots, f_q\}$ denote the q deep features extracted by the multitask feature extractor, and the distance between the pixel deep features is defined as

$$d_F(i,j) = \sqrt{\sum_{k=1}^{q} \left(F_{N_i}(k) - F_{N_j}(k)\right)^2}. \tag{2}$$

Let the positions of pixels i and j be $O(x_i, y_i)$ and $O(x_j, y_j)$, the pixel location distance is defined as

$$d_O(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{3}$$

Utilizing the definition of (1)–(3), the proposed pixel distance measure based on the high-dimensional feature space is defined as

$$D(i,j) = \sqrt{d_I^2(i,j) + d_F^2(i,j) + \lambda d_O^2(i,j)}, \tag{4}$$

where the parameter λ is used to measure the relative importance of intensity information, spatial information, and deep semantic information.

2.3. Pixel-Superpixel Soft Assignment

The basic idea of classic SLIC is to iteratively assign pixels to neighboring superpixels with similar intensity values and spatial locations through the local K-means clustering. The classic SLIC mainly includes three steps: (1) initialization of the cluster centers, (2) pixel-superpixel assignment, and (3) recalculation of the superpixel center.

(1) Initialization of the cluster centers. Let N be the number of pixels and K be the number of superpixels. Initially, the image is divided into K grids with the same size, and K initial cluster centers are obtained, as shown in Figure 4a. The step length of adjacent initial cluster centers is approximately calculated as $S = \sqrt{\frac{N}{K}}$, and then calculating the point with the smallest gradient in the 3×3 neighborhood of the initial cluster center as the new cluster center.

(2) Pixel–superpixel assignment. In the neighborhood of each cluster center, a certain superpixel label is assigned to each pixel, as shown in Figure 4b. Given an SAR image (H), with high-dimensional features ($[I, F, O]$) at N pixels. The task of pixel–superpixel assignment is to assign each pixel to one of the m superpixels. The assignment option can be expressed as

$$L_p^t = \underset{i \in \{0,\ldots,m-1\}}{\arg\min} \{D(H_p, C_i^{t-1})\}, \tag{5}$$

where H_p represents pixel p. C_i^{t-1} represents the cluster center of the i-th superpixel in the $(t-1)$-th iteration. L_p^t represent the superpixel label of pixel p at the t-th iteration.

(3) Recalculation of the superpixel center. For each superpixel i, SLIC calculates its feature average values as the superpixel center.

$$C_i^t = \frac{1}{M_i^t} \sum_{p|L_p^t=i} H_p, \tag{6}$$

where, M_i^t represents the number of pixels in the i-th superpixel in the t-th iteration.

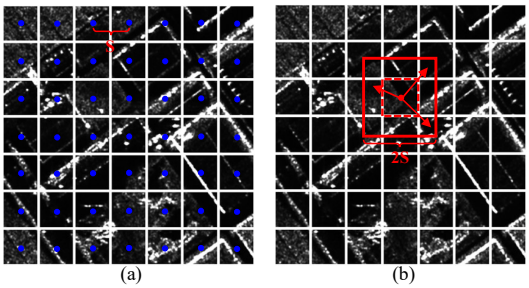


Figure 4. SLIC program diagram. (**a**) Initialization of the cluster centers. (**b**) Pixel–superpixel assignment.

As mentioned earlier, the parameters of the multitask feature extractor need to be fine-tuned in the superpixel generation network. Therefore, it is necessary to design an end-to-end superpixel generation framework. However, because SLIC introduces a non-differentiable nearest neighbor operation in the pixel assignation stage, as shown in (5), SLIC cannot be added to the end-to-end framework. In this paper, we learn from the idea of "*softmax*" [24,35,36] and replace the non-differential "*argmin*" with the differential "*softmin*" in (5), as shown in Figure 5. The pixel–superpixel soft assignment is defined as

$$L_{pi}^{'t} = \underset{i\in\{0,\ldots,m-1\}}{soft\min} \{D(H_p, C_i^{'t-1})\} = \frac{e^{-D(H_p,C_i^{'t-1})}}{\sum_{p=1}^{N} e^{-D(H_p,C_i^{'t-1})}}, \tag{7}$$

where $L_{pi}^{'t}$ represents the probability that the pixel p belongs to the superpixel i.

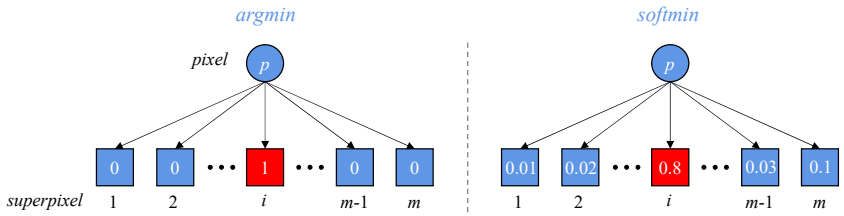


Figure 5. Illustration of the different between argmin and softmin. In the argmin strategy, a pixel is assigned to a superpixel. In the softmin strategy, pixels are assigned to surrounding superpixels in the form of probability.

Correspondingly, the superpixel center update is rewritten as

$$C_i^{'t} = \sum_{p=1}^{N} L_{pi}^{'t} H_p. \tag{8}$$

2.4. Algorithm

In detail, the proposed method is shown in Algorithm 1. First, we use multitask feature extractor to extract deep features. Then, we construct a high-dimensional feature space including deep semantic information, intensity information and spatial information, and we

define an effective pixel distance measure based on the high-dimensional feature space. Next, we use softmin instead of argmin to implement the pixel–superpixel soft assignment, which makes the SLIC differentiable. In the training process, we iteratively update the parameters of the multitask feature extractor by calculating the reconstruction loss [37,38]. Equation (8) describes the mapping from pixel to superpixel, which is $C = L'^{T}H$. Therefore, the inverse mapping from superpixel to pixel can be calculated as

$$\widehat{H} = \widetilde{L}'C, \widetilde{L}' = \frac{e^{-D(H_p, C_i'^{t-1})}}{\sum_{i=0}^{m-1} e^{-D(H_p, C_i'^{t-1})}}, \tag{9}$$

where $\widehat{H}$ represents the pixel representation of the superpixel mapped back to the pixel.

The reconstruction loss can be written as

$$Loss = \mathcal{L}(H, \widehat{H}) = \mathcal{L}(H, \widetilde{L}'L'^{T}H), \tag{10}$$

In this paper, we use the cross-entropy loss function for the superpixel generation task.

Algorithm 1 Our Proposed Method.

Input: SAR image, the number of superpixels K.

while *iteration* $<$ *epochs* **do**
  deep features $F = \{f_1, f_2, \ldots, f_q\} \leftarrow$ multitask feature extractor
  High-dimensional feature space $[H] \leftarrow H = [I, F, O]$
  Initialization of the cluster centers.
  for $t = 1 \rightarrow v$ **do**
    Pixel-superpixel soft assignment by Equation (7).
    Recalculate superpixel center by Equation (8).
  end for
  Calculate reconstruction loss by Equation (10).
  Update parameters of the multitask feature extractor.
end while
function D(i, j)
  $d_I(i, j) \leftarrow$ Equation (1)
  $d_F(i, j) \leftarrow$ Equation (2)
  $d_O(i, j) \leftarrow$ Equation (3)
  $D(i, j) \leftarrow$ Equation (4)
end function

Output: Superpixel generation result.

3. Experimental Results and Analysis

In this section, we evaluate the performance of the proposed method for SAR image superpixel generation. First, we briefly describe the used SAR images, and parameter settings. Then, the selection of hyperparameters λ and v are given. Finally, four different SAR image superpixel generation methods are conducted on two real SAR images to investigate the effectiveness of our proposed method.

3.1. Data Description and Parameter Settings

We conduct experiments on two real SAR images with different bands and different resolutions. Figure 6a shows the X-band SAR image with 0.5 m resolution. The size of the airborne X-band SAR image is 1063×1165. Figure 6c shows the C-band SAR image acquired by Gaofen-3 with 8m resolution. The size of the C-band SAR image is 512×512. The manually extracted ground truths of two real SAR images are shown in Figure 6b,d, which can be used to quantitatively evaluate the different methods equitably to a great extent.

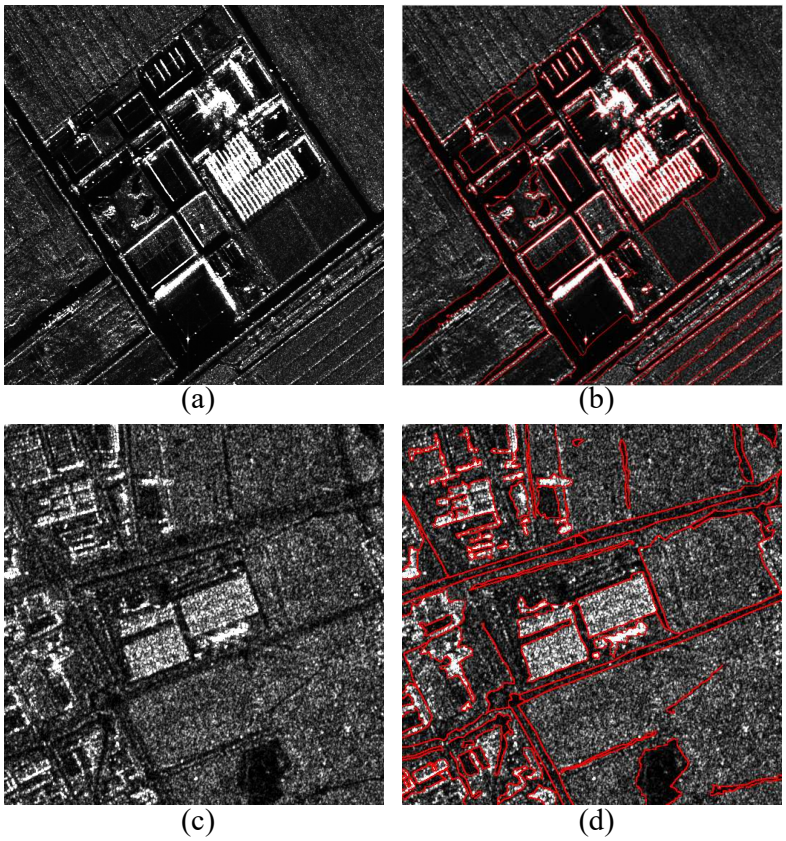


Figure 6. Real SAR images. (**a**) X-band. (**c**) C-band. (**b**) Ground truth of (**a**). (**d**) Ground truth of (**c**).

In our experiments, the epoches is set to 100, and the Adam [39] is employed as an optimizer with a learning rate of 0.001. The selection of hyperparameter λ and the number of iterations v will be explained in detail in the next section. In this paper, we adopt two normally used metrics, boundary recall (BR) and undersegmentation error (UE), to evaluate the performance of superpixel generation quantitatively.

BR denotes the ratio of the superpixel boundaries falling on the ground truth boundaries. The higher BR indicates the better boundary adhesion of superpixel. The BR can be calculated as

$$BR = \frac{\sum\limits_{p \in B_G, q \in B_S} TP(p,q)}{\sum\limits_{p \in B_G, q \in B_S} [TP(p,q) + FN(p,q)]}, \tag{11}$$

where B_G and B_S represent the pixel set of the ground truth boundaries and the superpixel boundaries, respectively. $TP(p,q)$ indicates that the superpixel boundary pixel p falls on the neighborhood of the ground truth boundary pixel q. The size of the neighborhood is commonly set to 2×2 [40]. On the contrary, $FN(p,q)$ indicates the neighborhood of the ground truth boundary pixel q without the superpixel boundary pixel p.

UE measures the degree that the superpixels spill over the ground truth segmentation. A lower UE indicates that fewer superpixels cross multiple objects. The UE can be calculated as

$$UE = \frac{\sum_i \sum_j \min\{|S_j \cap G_i|, |S_j - G_i|\}}{\sum_i |G_i|}, \tag{12}$$

where G_i and S_j represent a ground truth segmentation and a superpixel, respectively. $|\bullet|$ indicates the number of pixels.

3.2. Hyperparameter Selection

Here, we give the influence of λ on the superpixel generation results through experiments in a slice of the X-band image, as shown in Figure 7. We can observe that the compactness of superpixels is poor, when $\lambda = 0.2$. With the increase of λ, the compactness of superpixels gradually increases, whereas the boundary adherence of superpixels gradually decreases, as shown in Figure 7b–f. When $\lambda = 1.0$, a large number of superpixels spill over the ground truth segmentation. To select the appropriate hyperparameter λ, we give the quantitative results, as shown in Figure 8. We can observe that when $\lambda = 0.4$, the UE of the superpixel generation result is the smallest and the BR remains at a high value. Therefore, the λ is set to 0.4, which is used to balance the proportion of spatial information in the pixel distance measure.

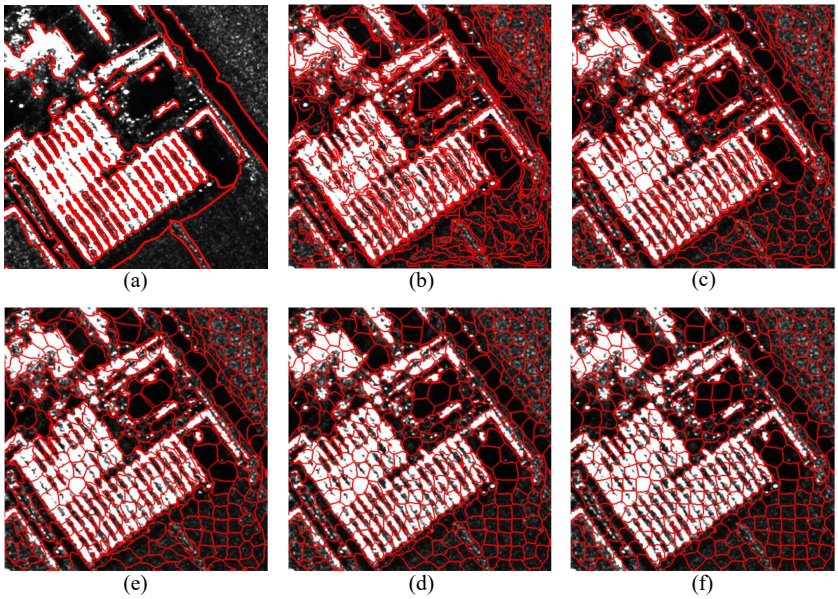


Figure 7. Superpixel generation results with different λ. (**a**) Ground truth. (**b**) $\lambda = 0.2$. (**c**) $\lambda = 0.4$. (**d**) $\lambda = 0.6$. (**e**) $\lambda = 0.8$. (**f**) $\lambda = 1.0$. (The number of superpixel K is 400 and $v = 10$).

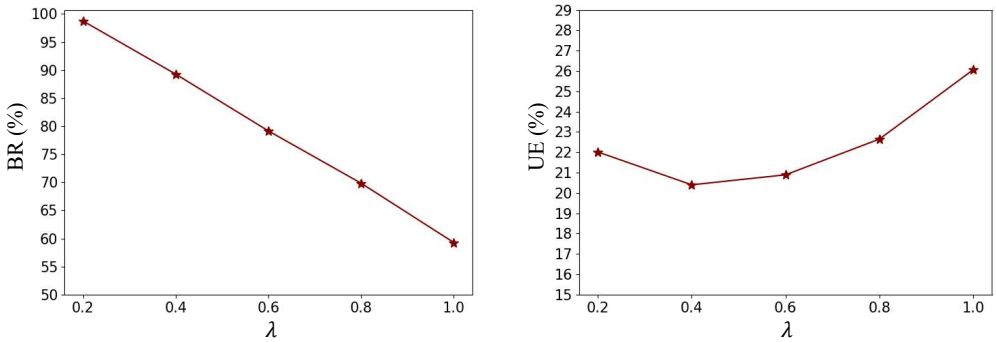


Figure 8. Quantitative comparison of superpixel generation results with different λ. (**left**) BR curves. (**right**) UE covers. ($K = 400$ and $v = 10$).

Next, we give the influence of the number of iterations v on the superpixel generation results, as shown in Figure 9. We can find that the number of iterations of SLIC has little effect on the compactness and boundary adherence of superpixels. Correspondingly, we give the quantitative results of the above experiments, as shown in Figure 10. We can find that the BR and UE of the superpixel generation results tend to stable after ten iterations. Therefore, the v is set to 10.

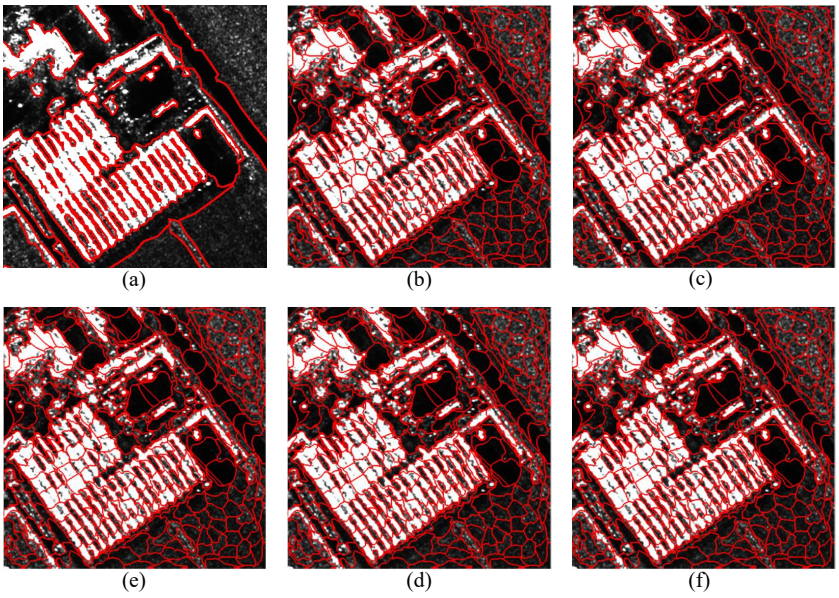


Figure 9. Superpixel generation results with different v. (**a**) Ground truth. (**b**) $v = 5$. (**c**) $v = 10$. (**d**) $v = 15$. (**e**) $v = 20$. (**f**) $v = 25$. ($K = 400$ and $\lambda = 0.4$).

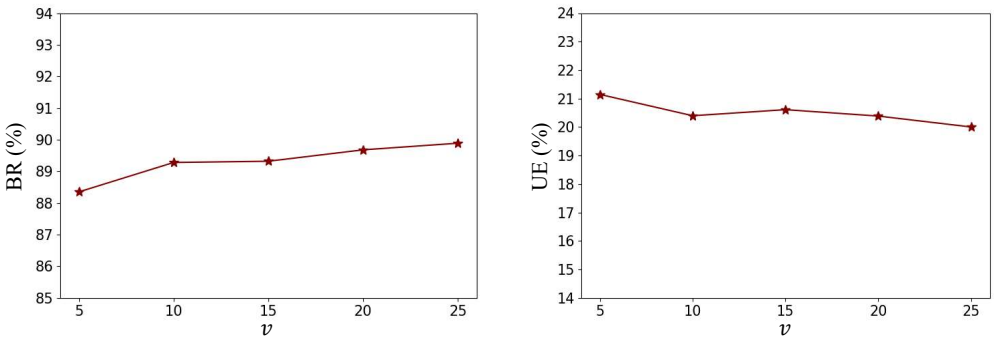


Figure 10. Quantitative comparison of superpixel generation results with different v. (**left**) BR curves. (**right**) UE covers. ($K = 400$ and $\lambda = 0.4$).

3.3. Comparison with Other Methods

To evaluate the performance of the proposed method, four superpixel generation methods are employed for comparison, including Linear Spectral Clustering (LSC) [41], Similarity Ratio and Mahalanobis Proximity (SRMP) [42], Mixture-based Superpixel (MISP) [43], and Edge-aware Superpixel Generation with One Iteration Merging (ESOM) [29]. For the sake of fair comparison, all these methods use the same number of superpixels. The superpixel generation results of X-band SAR image with 2000 superpixels are shown in Figure 11. The superpixel generation results of C-band image with 1200 superpixels are shown in Figure 12.

It can be seen from Figure 11b that the compactness and boundary adherence of superpixels are very poor, and a large number of mis-segmented pixels are generated in homogeneous areas. To compare the results in detail, six areas are marked with frames of different colors in Figure 11b–f. We can observe that, compared with other methods, our proposed method can generate better superpixel results, where the details of the building are clearer, as shown in the building region marked by red rectangles in Figure 11c–f. Although ESOM considers the edge information in the superpixel generation process, the edges of the man-made region are not good, as shown in three regions marked by green rectangles in Figure 11e. Similarly, our proposed method can preserve more accurate building edges, as shown in three regions marked by green rectangles in Figure 11f. We use the yellow triangle frame to mark the superpixel generation result in the homogeneous

area. It can be seen that the boundary adherence of superpixels generated by SRMP, MIPS, and ESOM in homogeneous regions is poor, while LSC and ML-SGN can retain better terrain edges.

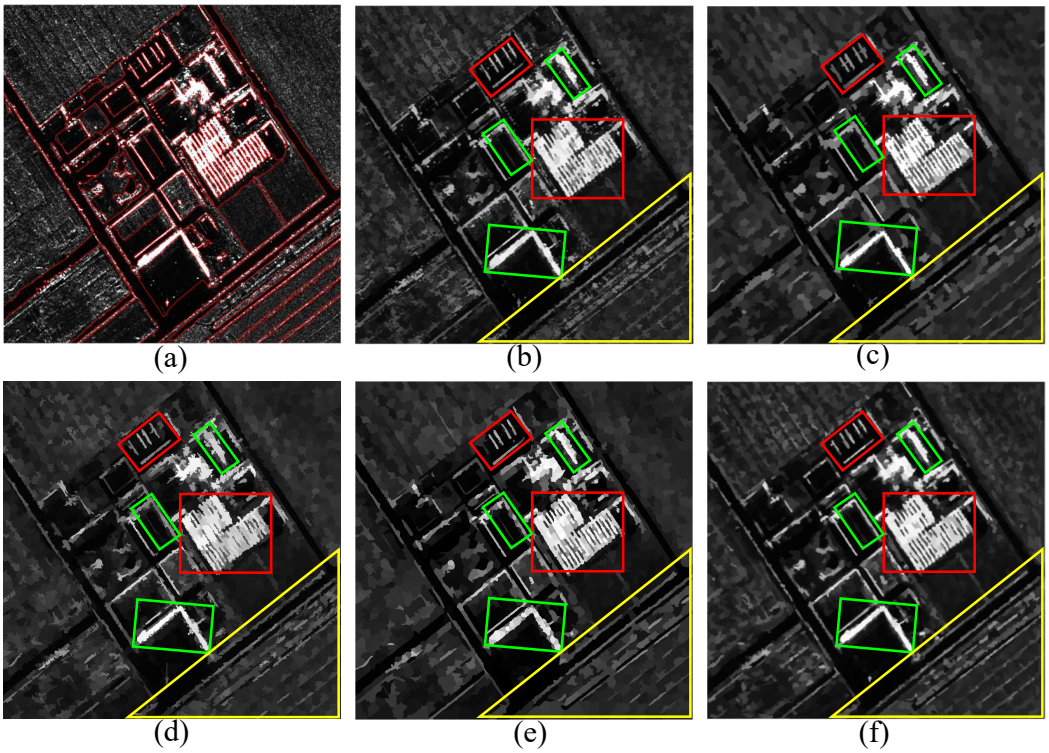


Figure 11. Superpixel generation results of the five methods for the X-band SAR image. (**a**) Ground truth. (**b**) SRMP. (**c**) MIPS. (**d**) LSC. (**e**) ESOM. (**f**) ML-SGN.

To explore the reason why the superpixels generated by the proposed method can retain better details and edges, we visualize some deep features extracted by the multtask feature extractor, as shown in Figure 13. We can find that some deep features extracted by the multitask feature extractor can clearly describe the details of complex man-made targets, as shown in Figure 13a–d. In addition, some deep features can reflect the edges of homogeneous natural regions, as shown in Figure 13e–h. This fact indicates that our proposed method is suitable for the superpixel generation not only in heterogeneous building regions, but also in homogeneous natural regions.

To verify the effectiveness of the method, we compared the superpixel generation results on a low-resolution SAR image, as shown in Figure 12. It can be seen from Figure 12b that the compactness of the superpixels generated by SRMP is extremely poor. We can find that the details of buildings and targets are not well preserved, as shown in the heterogenous areas marked by red rectangle in Figure 12c,d. It can be seen from Figure 12e that ESOM can retain the internal details of complex regions. However, it is not effective in segmenting homogeneous areas such as small paths, as shown in the areas marked by green rectangle in Figure 12d. From Figure 12f, it can be observed that the superpixel generated by the proposed method can precisely adheres to the boundaries even though in low-resolution SAR images. Correspondingly, we visualize some deep features extracted by the multitask feature extractor, as shown in Figure 14. We can observe that the details and edges of buildings can be clearly distinguished with deep features.

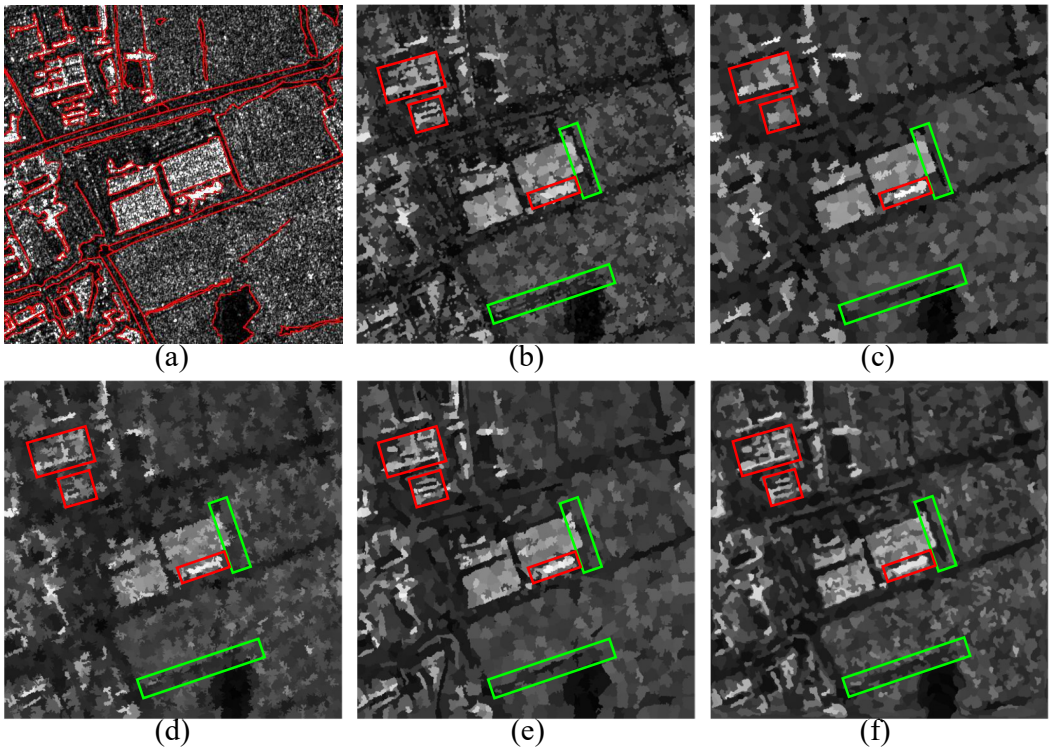


Figure 12. Superpixel generation results of the five methods for the C-band SAR image. (**a**) Ground truth. (**b**) SRMP. (**c**) MISP. (**d**) LSC. (**e**) ESOM. (**f**) ML-SGN.

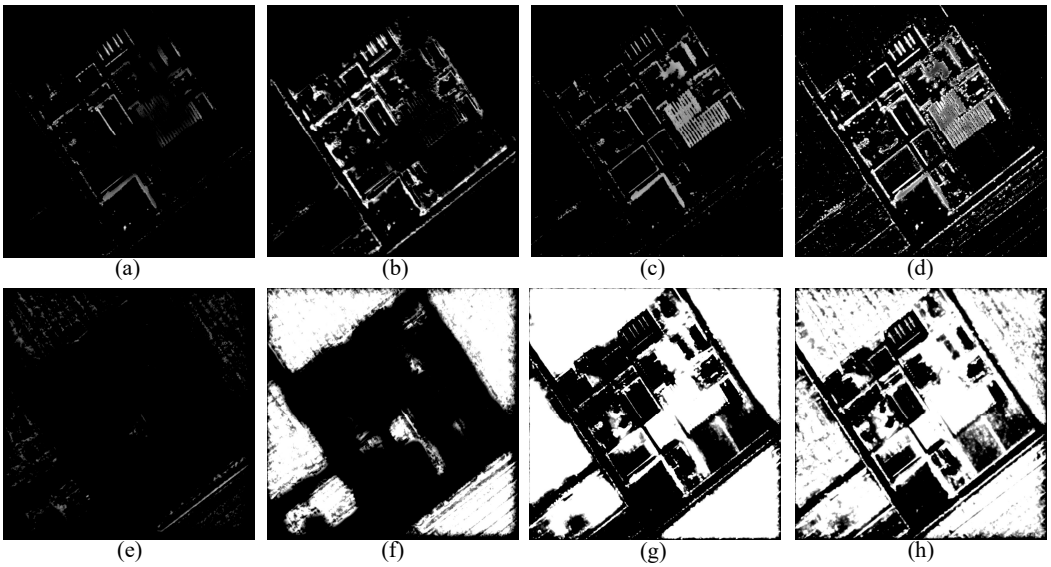


Figure 13. Visualization of deep features extracted by the multitask feature extractor on the X-band SAR image. (**a**–**d**) describe the details of heterogeneous building regions. (**e**–**h**) describe the edges of homogeneous natural regions.

Table 1 gives the quantitative evaluation indices of the superpixel generation results in two SAR images with different methods. We can find that the BR and UE of SRMP are the worst among the comparison methods. LSC obtains good boundary adherence, but its undersegmentation error is high. In contrast, ESOM achieves a small under-segmentation error, but has relative poor boundary adherence. Similar to the conclusions obtained in the above images, ML-SGN outperforms the other methods in terms of BR and UE. As a preprocessing step of many methods, the running time of the superpixel generation algorithm is an important indicator of its performance. All of the experiments are implemented

on a desktop with an Intel Core i7-10870H CPU with frequency of 2.2 GHz, and 32 GB RAM, and RTX 3060 GPU with 12 GB video memory. From Table 1, we can find that the efficiencies of SRMP, MISP and LSC are too poor, especially for large-scale SAR images. Although ESOM has a short running time, its boundary recall is poor. In contrast, our proposed method achieves high BR and UE with a higher speed.

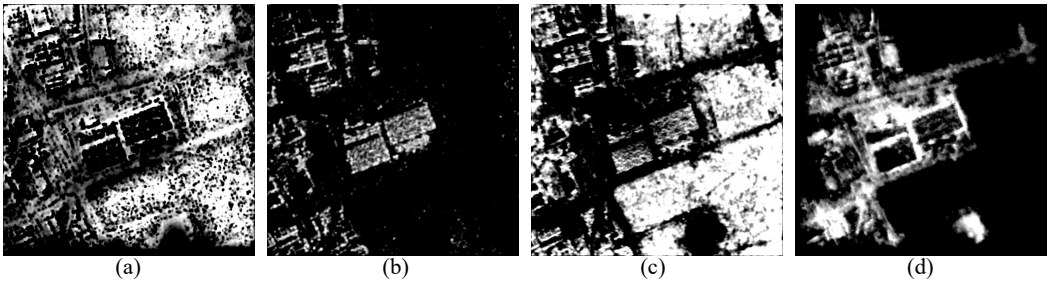


Figure 14. Visualization of deep features extracted by the multitask feature extractor on the C-band SAR image. (**a**–**d**) are different levels of deep features generated by the proposed method.

Table 1. Quantitative evaluation measures of the superpixel generation results using different methods.

Method	X-Band			C-Band		
	BR	**UE**	**Time**	**BR**	**UE**	**Time**
SRMP [42]	0.6280	0.1770	157.70	0.6113	0.1010	31.59
MISP [43]	0.6889	0.1583	1380	0.7708	0.0659	87.18
LSC [41]	0.9274	0.1584	91.53	0.9071	0.0812	17.56
ESOM [29]	0.6756	0.1195	2.55	0.8523	0.0620	1.52
ML-SGN	**0.9571**	**0.1165**	3.37	**0.9460**	**0.0549**	2.35

4. Discussion

4.1. The Impact of the Number of Superpixels

As we all know, the number of superpixels seriously affects the performance of superpixels. To verify the robustness of the proposed method to the number of superpixels, we give superpixel generation results on complex building scenes, as shown in Figure 15. We can observe that as the number of superpixels increases, the details of the building become clearer.

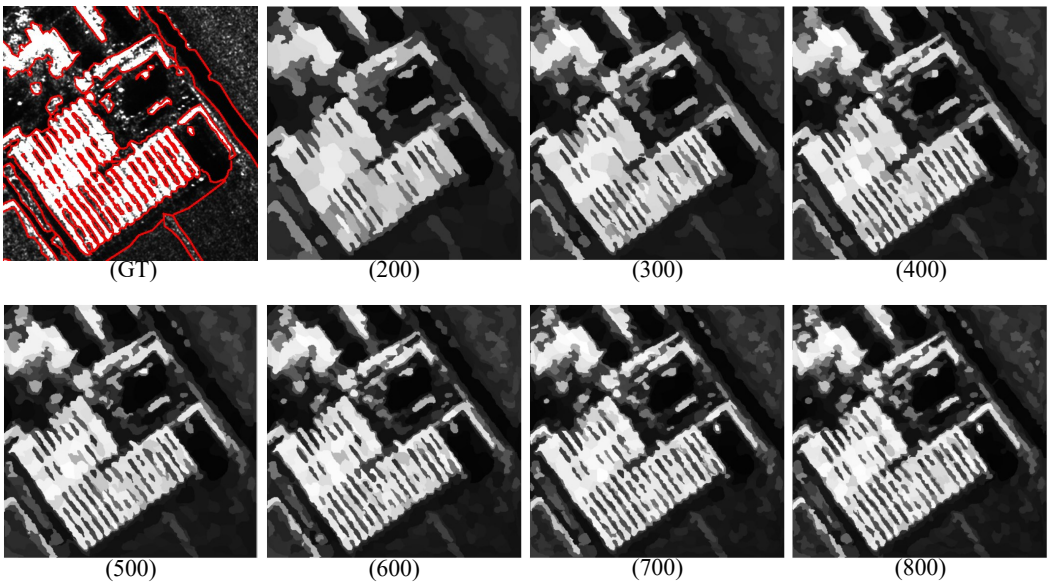


Figure 15. Superpixel results generated by the proposed method with different numbers of superpixels.

Figure 16 gives the performance of different methods with different numbers of superpixels. We can find that increasing the number of superpixels will improve the performance in BR and UE. In terms of boundary recall, ML-SGN and LSC obtain competitive results, and other methods cannot completely retain the edges of the complex building regions. In terms of undersegmentation error, LSC obtain the highest UE value, whereas ML-SGN consistently achieves lower UE values than other methods with different numbers of superpixels.

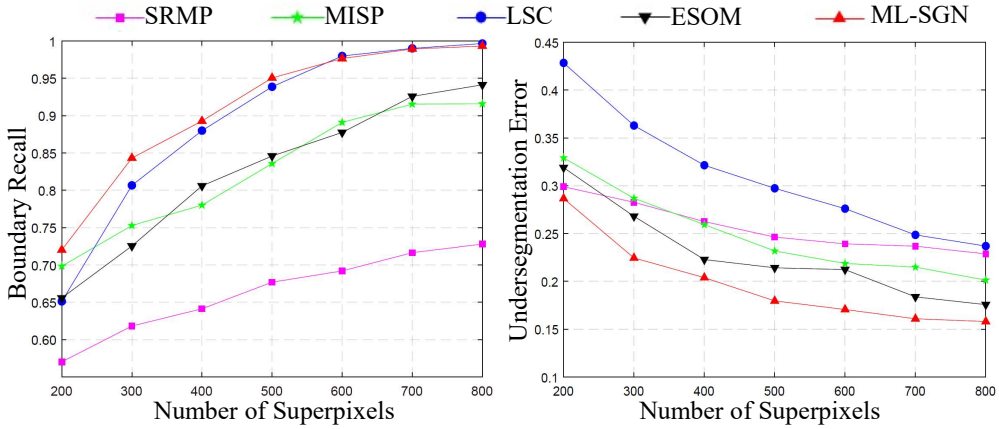


Figure 16. Comparison of the performances of different superpixel generation algorithms with different numbers of superpixels.

4.2. The Necessity of End-to-End Network Construction

For the reason that the training objectives of the multitask feature extractor and the superpixel generation task are inconsistent, the objective function of the multitask feature extractor is deviated from the objective function of the main task, and it is difficult for the two-step model to achieve the optimal performance. The end-to-end network structure can fine-tune the network parameters of the multitask feature extractor, making the deep features extracted by the multitask feature extractor more suitable for the superpixel generation task. Figure 17 gives the deep features extracted by the multitask feature extractor without parameter fine-tuning. Compared with Figure 13, Figure 17a–d can only reflect the difference between homogeneous regions and heterogeneous regions. Compared with Figure 14, Figure 17e–h can only extract the edges of different terrains, and cannot obtain internal details.

In addition, the parameter fine-tuning of the multi-task feature extractor improves the generalization performance of the network, making the proposed method suitable for SAR images with different scenes. Figure 18 gives the superpixel generation results in different scenes, including water, grass, urban area, and industrial area. It can be seen from Figure 18a that natural regions (such as rivers) can be segmented by regular and compact superpixels using our proposed method. Similarly, our proposed method can generate superpixels with good performance in grass and road regions, as shown in Figure 18b. It is worth noting that our proposed method can retain the edges and details of man-made regions, as shown in Figure 18c,d. In summary, the end-to-end network can extract deep features well by fine-tuning the parameters of the multitask feature extractor, and perform well on SAR images with different scenes.

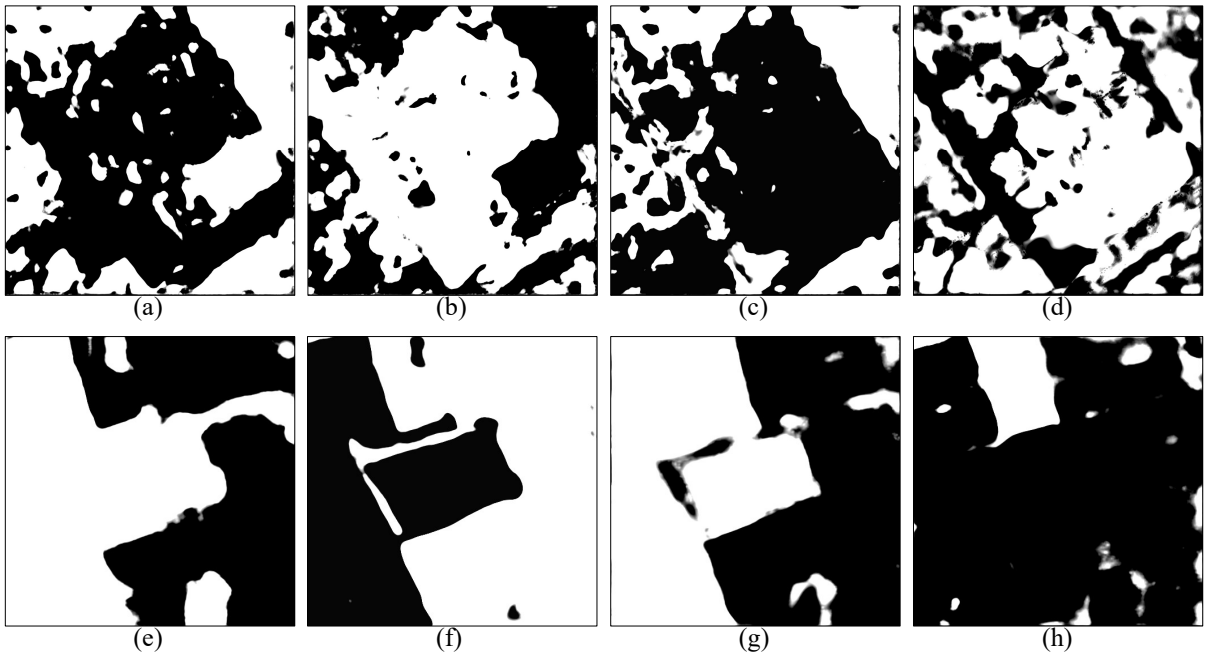


Figure 17. Visualization of deep features extracted by the multitask feature extractor without parameter fine-tuning. (**a**–**d**) Deep features of X-band SAR images. (**e**–**h**) Deep features of C-band SAR images.

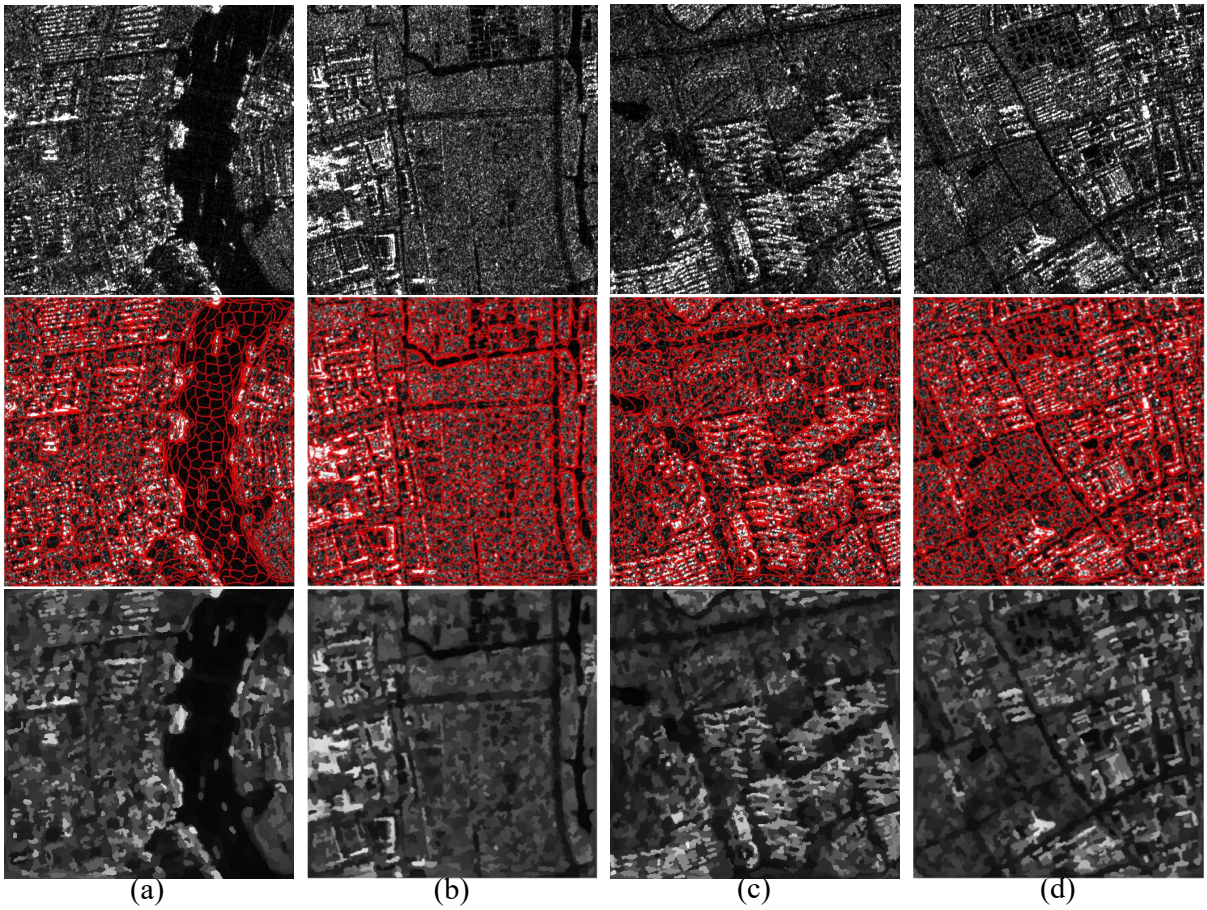


Figure 18. Superpixel generation results of the proposed method for Gaofen-3 SAR images. (**a**) Water. (**b**) Grass. (**c**) Urban area. (**d**) Industrial area.

5. Conclusions

In this paper, we propose an end-to-end SAR image superpixel generation method, named the multitask learning-based superpixel generation network (ML-SGN). ML-SGN adopts a segmentation task to train the multitask feature extractor, which solves the problem of insufficient labeled samples in the SAR superpixel generation task. We replace the nearest neighbor operation in SLIC with a soft assignment operation, so that SLIC can be integrated into an end-to-end network architecture. In addition, we construct a high-dimensional feature space containing deep semantic information, spatial information and intensity information, and define an effective distance metric based on the high-level feature space. The superiority of the proposed method is quantitatively evaluated on two real SAR images, and the generalization performance of the proposed method is verified on SAR images with different scenes. Future work will further focus on the combination of traditional features and deep features for the SAR image superpixel generation, which is beneficial to the field of intelligent interpretation of SAR images.

Author Contributions: Conceptualization, D.X. and J.C.; methodology, Q.W.; software, J.L. and W.J.; resources, Q.W.; writing, J.L. 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 62171015, and in part by the Fundamental Research Funds for the Central Universities under Grant buctrc202121, and in part by the funding provided by the Alexander von Humboldt Foundation.

Acknowledgments: The authors acknowledge the National High-resolution Earth Observation System Major Science and Technology Project and the Institute of Aerospace Information Innovation of the Chinese Academy of Sciences for providing the high-resolution fully polarimetric SAR dataset.

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

References

1. Renga, A.; Graziano, M.D.; Moccia, A. Segmentation of marine SAR images by sublook analysis and application to sea traffic monitoring. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1463–1477. [CrossRef]
2. Cheng, J.; Zhang, F.; Xiang, D.; Yin, Q.; Zhou, Y.; Wang, W. PolSAR Image Land Cover Classification Based on Hierarchical Capsule Network. *Remote Sens.* **2021**, *13*, 3132. [CrossRef]
3. Quan, S.; Xiong, B.; Xiang, D.; Zhao, L.; Zhang, S.; Kuang, G. Eigenvalue-based urban area extraction using polarimetric SAR data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 458–471. [CrossRef]
4. Cheng, J.; Zhang, F.; Xiang, D.; Yin, Q.; Zhou, Y. PolSAR Image Classification with Multiscale Superpixel-Based Graph Convolutional Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60* , 1–14. [CrossRef]
5. Guan, D.; Xiang, D.; Dong, G.; Tang, T.; Tang, X.; Kuang, G. SAR image classification by exploiting adaptive contextual information and composite kernels. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1035–1039. [CrossRef]
6. Subudhi, S.; Patro, R.N.; Biswal, P.K.; Dell'Acqua, F. A Survey on Superpixel Segmentation as a Preprocessing Step in Hyperspectral Image Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5015–5035. [CrossRef]
7. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
8. Zhang, W.; Xiang, D.; Su, Y. Fast multiscale superpixel segmentation for SAR imagery. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19* , 1–5. [CrossRef]
9. Liu, M.Y.; Tuzel, O.; Ramalingam, S.; Chellappa, R. Entropy rate superpixel segmentation. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR 2011), Colorado Springs, CO, USA, 20–25 June 2011; pp. 2097–2104.
10. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
11. Vincent, L.; Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 583–598. [CrossRef]
12. Levinshtein, A.; Stere, A.; Kutulakos, K.N.; Fleet, D.J.; Dickinson, S.J.; Siddiqi, K. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2290–2297. [CrossRef]
13. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. Slic Superpixels. Technical Report. 2010. Available online: https://infoscience.epfl.ch/record/149300 (accessed on 20 December 2021).
14. Xiang, D.; Tang, T.; Zhao, L.; Su, Y. Superpixel generating algorithm based on pixel intensity and location similarity for SAR image classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1414–1418. [CrossRef]

15. Zou, H.; Qin, X.; Zhou, S.; Ji, K. A likelihood-based SLIC superpixel algorithm for SAR images using generalized Gamma distribution. *Sensors* **2016**, *16*, 1107. [CrossRef]
16. Akyilmaz, E.; Leloglu, U.M. Segmentation of SAR images using similarity ratios for generating and clustering superpixels. *Electron. Lett.* **2016**, *52*, 654–656. [CrossRef]
17. Xiang, D.; Tang, T.; Quan, S.; Guan, D.; Su, Y. Adaptive superpixel generation for SAR images with linear feature clustering and edge constraint. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3873–3889. [CrossRef]
18. Jing, W.; Jin, T.; Xiang, D. Content-Sensitive Superpixel Generation for SAR Images with Edge Penalty and Contraction-Expansion Search Strategy. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60* , 1–15. [CrossRef]
19. Wang, Y.; Cheng, J.; Zhou, Y.; Zhang, F.; Yin, Q. A Multichannel Fusion Convolutional Neural Network Based on Scattering Mechanism for PolSAR Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19* , 1–5. [CrossRef]
20. Bi, H.; Sun, J.; Xu, Z. A graph-based semisupervised deep learning model for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2116–2132. [CrossRef]
21. Chen, S.W.; Tao, C.S. PolSAR image classification using polarimetric-feature-driven deep convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 627–631. [CrossRef]
22. Yue, Z.; Gao, F.; Xiong, Q.; Wang, J.; Hussain, A.; Zhou, H. A novel attention fully convolutional network method for synthetic aperture radar image segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4585–4598. [CrossRef]
23. Yang, F.; Sun, Q.; Jin, H.; Zhou, Z. Superpixel segmentation with fully convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 13964–13973.
24. Jampani, V.; Sun, D.; Liu, M.Y.; Yang, M.H.; Kautz, J. Superpixel sampling networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 352–368.
25. Du, W.; Zhang, F.; Ma, F.; Yin, Q.; Zhou, Y. Improving SAR Target Recognition with Multi-Task Learning. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 284–287.
26. Liu, W.; Wen, Y.; Yu, Z.; Yang, M. Large-margin softmax loss for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML), New York, NY, USA, 20–22 June 2016; Volume 2, p. 7.
27. Shamsolmoali, P.; Zareapoor, M.; Wang, R.; Zhou, H.; Yang, J. A novel deep structure U-Net for sea-land segmentation in remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3219–3232. [CrossRef]
28. Nagi, A.S.; Kumar, D.; Sola, D.; Scott, K.A. RUF: Effective Sea Ice Floe Segmentation Using End-to-End RES-UNET-CRF with Dual Loss. *Remote Sens.* **2021**, *13*, 2460. [CrossRef]
29. Jing, W.; Jin, T.; Xiang, D. Edge-Aware superpixel generation for SAR imagery with one iteration merging. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1600–1604. [CrossRef]
30. Sun, X.; Shi, A.; Huang, H.; Mayer, H. BAS^4 Net: Boundary-Aware Semi-Supervised Semantic Segmentation Network for Very High Resolution Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5398–5413. [CrossRef]
31. Sun, X.; Wang, P.; Yan, Z.; Diao, W.; Lu, X.; Yang, Z.; Zhang, Y.; Xiang, D.; Yan, C.; Guo, J.; et al. Automated High-Resolution Earth Observation Image Interpretation: Outcome of the 2020 Gaofen Challenge. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8922–8940. [CrossRef]
32. Csillik, O. Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. *Remote Sens.* **2017**, *9*, 243. [CrossRef]
33. Yin, J.; Wang, T.; Du, Y.; Liu, X.; Zhou, L.; Yang, J. SLIC Superpixel Segmentation for Polarimetric SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60* , 1–17. [CrossRef]
34. Feng, H.; Hou, B.; Gong, M. SAR image despeckling based on local homogeneous-region segmentation by using pixel-relativity measurement. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2724–2737. [CrossRef]
35. Gold, S.; Rangarajan, A. Softmax to softassign: Neural network algorithms for combinatorial optimization. *J. Artif. Neural Netw.* **1996**, *2*, 381–399.
36. Inthumathi, V.; Chitra, V.; Jayasree, S. Fuzzy soft min-max decision making and its applications. *J. Inform. Math. Sci.* **2017**, *9*, 827–834.
37. Ma, F.; Zhang, F.; Yin, Q.; Xiang, D.; Zhou, Y. Fast SAR Image Segmentation with Deep Task-Specific Superpixel Sampling and Soft Graph Convolution. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60* , 1–16. [CrossRef]
38. Liu, R.; Sisman, B.; Gao, G.; Li, H. Expressive tts training with frame and style reconstruction loss. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 1806–1818. [CrossRef]
39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
40. Xiang, D.; Wang, W.; Tang, T.; Guan, D.; Quan, S.; Liu, T.; Su, Y. Adaptive statistical superpixel merging with edge penalty for PolSAR image segmentation. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2412–2429. [CrossRef]
41. Li, Z.; Chen, J. Superpixel segmentation using linear spectral clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1356–1363.
42. Akyilmaz, E.; Leloglu, U.M. Similarity ratio based adaptive Mahalanobis distance algorithm to generate SAR superpixels. *Can. J. Remote Sens.* **2017**, *43*, 569–581. [CrossRef]
43. Arisoy, S.; Kayabol, K. Mixture-based superpixel segmentation and classification of SAR images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1721–1725. [CrossRef]