



Article

Small Object Detection in Medium–Low-Resolution Remote Sensing Images Based on Degradation Reconstruction

Yongxian Zhao ^{1,2}, Haijiang Sun ^{1,*} and Shuai Wang ¹

¹ Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; zhaoyongxian20@mails.ucas.ac.cn (Y.Z.); wangshuai@ciomp.ac.cn (S.W.)

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: sunhj@ciomp.ac.cn

Abstract: With the continuous development of space remote sensing technology, the spatial resolution of visible remote sensing images has been continuously improved, which has promoted the progress of remote sensing target detection. However, due to the limitation of sensor lattice size, it is still challenging to obtain a large range of high-resolution (HR) remote sensing images in practical applications, which makes it difficult to carry out target monitoring in a large range of areas. At present, many object detection methods focus on the detection and positioning technology of HR remote sensing images, but there are relatively few studies on object detection methods using medium- and low-resolution (M-LR) remote sensing images. Because of its wide coverage area and short observation period, M-LR remote sensing imagery is of great significance for obtaining information quickly in space applications. However, the small amount of fine-texture information on objects in M-LR images brings great challenges to detection and recognition tasks. Therefore, we propose a small target detection method based on degradation reconstruction, named DRADNet. Different from the previous methods that use super resolution as a pre-processing step and then directly input the image into the detector, we have designed an additional degenerate reconstruction-assisted framework to effectively improve the detector's performance in detection tasks with M-LR remote sensing images. In addition, we introduce a hybrid parallel-attention feature fusion module in the detector to achieve focused attention on target features and suppress redundant complex backgrounds, thus improving the accuracy of the model in small target localization. The experimental results are based on the widely used VEDAI dataset and Airbus-Ships dataset, and verify the effectiveness of our method in the detection of small- and medium-sized targets in M-LR remote sensing images.

Keywords: remote sensing; deep learning; medium–low-resolution images; object detection; super resolution



Citation: Zhao, Y.; Sun, H.; Wang, S. Small Object Detection in Medium–Low-Resolution Remote Sensing Images Based on Degradation Reconstruction. *Remote Sens.* **2024**, *16*, 2645. <https://doi.org/10.3390/rs16142645>

Academic Editors: Pedram Ghamisi and Yonghao Xu

Received: 5 June 2024

Revised: 11 July 2024

Accepted: 17 July 2024

Published: 19 July 2024



Copyright: © 2024 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

Space remote sensing technology obtains image data through Earth detection via satellite remote sensors, and is widely used in national defense and civilian fields [1–6]. How to obtain small target information from massive remote sensing data is an important research question in the field of computer vision [7]. In recent years, with the development of space remote sensing technology and semiconductor optical devices, the spatial resolution of remote sensing images has been continuously improved. HR remote sensing images can present rich details and clear texture features of objects. Therefore, most current studies are based on HR remote sensing images [8–11]. However, due to the limitation of sensor lattice size, HR remote sensing images render it difficult to obtain a large range of remote sensing images, which is not conducive to real-time monitoring in wide-area space. In contrast, M-LR remote sensing images have the characteristics of long detection distances, wide coverage, and strong data timeliness. With the same amount of data, M-LR remote sensing images can cover a wider area, so they are more practical for large-scale

target monitoring, rapid search tasks, and positioning. However, M-LR images contains a small number of pixels, low clarity, and fuzzy imaging, which makes the appearance and texture information presented by the target extremely small. This brings great challenges to the detection task [12]. In Figure 1, different appearance characteristics of vehicle and ship targets in HR remote sensing images are shown. In M-LR images, a large amount of detailed information is lost, and the boundary between the target and the background is blurred, which greatly increases the difficulty of detection. These images were taken from a SPOT satellite.



Figure 1. Examples of objects in remote sensing images with different resolutions. The image on the left is an HR image, and the image on the right is an M-LR image.

At present, there are few studies devoted to target detection within M-LR remote sensing images. Due to the low spatial resolution of these images, the information and detailed characteristics of the target are not obvious, and it is easy for complex backgrounds to interfere, which increases the difficulty of small target detection. Some existing traditional algorithms for small object detection in remote sensing images are susceptible to clutter in the background and need to rely on fine-feature and texture information for detection,

such as gray statistical features [13], shape texture feature extraction [14], and template matching [15], etc. It is difficult to achieve the detection effect with strong robustness and versatility. With the development of object detection technology, deep learning networks have made great breakthroughs in object detection within remote sensing images by virtue of their powerful capabilities of automatic feature extraction and excellent portability, such as NWD [16], RFLA [17], SuperYOLO [18], etc. However, there is still room for improvement in the detection in M-LR remote sensing images. At its core, the use of deep learning networks to detect small targets in M-LR remote sensing images mainly covers two aspects: on the one hand, they are used to enhance the feature information of the target and strengthen the network's ability to extract and learn the target features, and on the other hand, they are used to suppress the interference of complex backgrounds. The combination of the two can make the target and background more recognizable, and this is more conducive to the network's ability to learn and detect. Therefore, we focus on enhancing the network's learning ability to target features and suppress the background, having designed a method for small target detection within M-LR remote sensing images based on deep learning networks, named the degradation reconstruction-assisted detection network method (DRADNet). We use the degradation reconstruction module to degrade the HR remote sensing images acquired from the sensor into M-LR remote sensing images with different sizes and different blur degrees, so that they are closer to the real detection scene. We use the assisted enhancement module to learn different degraded images, so as to improve the detection network's ability to cope with the change in scale and clarity. This is helpful for obtaining a more robust detection effect and achieving higher precision for small target detection. The main contributions of this paper are as follows:

1. We have designed a degradation reconstruction-assisted enhancement branch for learning M-LR remote sensing images with different degenerated degrees, and used it as a branch structure of the network to guide the model to capture feature expressions at different resolutions, so as to improve the feature learning ability of the network in M-LR images.
2. To capture target details effectively and suppress the interference of redundant background information, we propose a feature fusion module based on hybrid parallel attention. This module can effectively use the feature information to help the network screen out the target against a complex background. It can make the network focus on the target features and improve the ability of feature extraction and model detection accuracy at the same time.
3. We propose a method for small target detection within M-LR remote sensing images based on a degradation reconstruction-assisted detection network (DRADNet). We describe the results of extensive ablation experiments to demonstrate its effectiveness. In addition, we compare the results of several mainstream methods on public datasets. The results show that our method can improve the accuracy of detection significantly and has competitive performance.

2. Related Works

2.1. Image Super Resolution

Image super resolution (SR) refers to the process of restoring an HR image from an M-LR image or image sequence. Super-resolution reconstruction technology has played a very important role in many fields such as satellite remote sensing, astronomical observation, security, biomedicine, and historical image restoration [19]. Super-resolution reconstruction methods based on deep learning have made breakthroughs and achieved significant results on various benchmarks. Dong et al. [20] first applied a convolutional neural network to super-resolution reconstruction technology and proposed the SRCNN, a super-resolution reconstruction convolutional neural network. Although this neural network has only three layers, it achieved remarkable results compared with classical super-resolution algorithms used in the past. The emergence of the SRCNN has opened up new research ideas for the development of image super-resolution reconstruction technology and attracted domestic

and foreign scholars to apply various variants of neural networks to image super-resolution research, including convolutional neural networks [21], adversarial neural networks [22], and combinations of the two [23]. Shi et al. [24] proposed a more efficient network, the ESPCN, which directly extracts features from LR small maps and greatly reduces training time. Kim et al. [25] proposed the use of a deeper network to learn the high-frequency residuals between HR images and LR images, effectively solving the problem of difficult convergence. Ledig et al. [26] proposed a generative adversarial network [27] to solve the problem of super resolution for the first time, in which perceptual loss and adversarial loss are used to enhance the visual realism of pictures. In addition, more image super-resolution algorithms based on deep learning have been derived, such as the DRCN [28], WDSR [29], EDSR [30], and SCNet [31] methods, etc.

The above non-blind image super-resolution algorithms based on deep learning usually synthesize HR and LR image pairs as training sets through predefined downsampling methods, which makes the process of reconstructing low-resolution images into high-resolution images become standardized. However, the degree of blur degradation of remote sensing images in the real world is different. Therefore, only using predefined downsampling methods cannot well reflect the super-resolution reconstruction effect of remote sensing images in real scenes. Moreover, the gap between the predefined degradation method and the actual degradation will largely undermine the performance of the SR algorithm in practical applications [32–34]. In order to make super-resolution reconstruction algorithms more reliable and further solve the challenges faced in real application scenarios, some super-resolution methods based on blind degradation [35–38] have been proposed. Chen et al. [39] proposed a nonlinear degradation model to narrow the gap between the real scene and the trained image pairs. Zhang et al. [40] proposed a probabilistic degradation model to model random factors in the degradation process and make it automatically learn the degradation distribution of SR images. Zhang et al. [41] proposed a unified gated degradation model to avoid ignoring common corner cases in the real world and generate more extensive degradation cases. Wang et al. [42] used high-level degradation modeling to better simulate the complex real-world degradation process.

These blind SR methods based on deep learning degrade the input images reconstructed at high resolution to make them closer to the real remote sensing images, thus increasing the reliability of image super-resolution reconstruction. Many research works have confirmed that modeling the approximate coverage degradation space is effective. Inspired by the blind degradation model, we designed an image degradation model before super-resolution reconstruction to enhance the robustness and reliability of the super-resolution reconstruction branch, and combined the two to form a degradation reconstruction-assisted enhancement branch as a super-resolution-assisted branch. Different from the method of preprocessing the image to obtain a high-resolution image, the purpose of this assisted branch is to guide the model to learn high-quality feature representations during the super-resolution reconstruction process, thereby enhancing the feature response to small targets and improving the detection performance for small targets in medium and low-resolution remote sensing images.

2.2. Object Detection

Due to the limitation of spatial resolution, the objects in M-LR remote sensing images lack texture information and the proportion of pixels is very small, which increases the difficulty of detection. There are relatively few algorithms for small target detection in M-LR remote sensing images, and the common ones are based on gray distribution, visual saliency, and deep learning. The method based on gray distribution finds the position of the target contour according to the discontinuity of the image gray value, and the segmentation between the target and the background is completed by designing the gray threshold. Yang et al. [43] proposed a function for selecting ship targets by using intensity and texture information, which can automatically assign weights and effectively shield the background area to optimize the detection performance. This method

is suitable for cases where the contrast between the target and the background is large, but the detection effect is not ideal for images with similar colors. The method based on visual saliency simulates the characteristics of the human visual system, and extracts the significant regions in the image through the spatial domain or a transform domain algorithm. Song et al. [44] constructed pixel intensity features and noise edge features from the spatial domain, used the contrast sensitivity function to weight different feature maps, and then realized small target detection in optical remote sensing images with noise through a difference algorithm. This verified the feasibility of visual saliency algorithms for small target detection in M-LR remote sensing images. This method has a good effect on suppressing background noise, but it is prone to producing false alarms when the image is affected by illumination and weather. In contrast, the feature learning capability of deep learning methods can effectively adapt to small target detection in M-LR remote sensing images. Rabbi et al. [45] proposed an Edge-Enhanced GAN (EESRGAN) combined with super-resolution technology: they combined it with different detection networks to complete end-to-end target detection, effectively improving the detection accuracy of small- and medium-sized targets in M-LR remote sensing images. Zou et al. [46] designed a method that uses image scaling to guide image super-resolution training (RASR), embedding more information into M-LR remote sensing images in order to achieve better detection performance. Chen et al. [47] designed a degradation reconstruction enhancer and combined it with an improved YOLOv5 network, then obtained the degradation reconstruction enhancement network (DRENet). This network can effectively realize ship detection in M-LR remote sensing images. He et al. [48] designed a feature distillation framework to make full use of the information in HR remote sensing images to process ship targets in M-LR images, and combined the framework with mainstream detectors to achieve relatively high-precision target detection. Although network models based on deep learning have made some achievements in small target detection, their detection accuracy still has great room for improvement.

2.3. Feature Fusion

Feature fusion is an image processing method that combines multiple different features to obtain richer feature information. It can also reduce the data dimension and improve the robustness of the algorithm. Since small targets have less feature information, it is necessary to make more full use of the detail information in the image. Therefore, feature fusion is widely used in the detection of small targets. Among them, multi-scale feature fusion uses low-level features to help the network obtain discriminative feature information and accurate positioning information, which greatly enhances the detection and recognition of small targets. Lin et al. [49] proposed the RetinaNet, which draws on the idea of multi-scale target detection, and adopted an improved loss function to solve the problem of class imbalance between positive and negative samples in the model training process, thus improving the detection effect on small targets. Lin et al. [50] proposed feature pyramid networks (FPNs) in order to combine M-LR's strong semantic features with HR's weak semantic features, using top-down paths and horizontal connections in multi-scale feature maps to enhance the information of feature maps. Although the FPN structure can make full use of the semantic information of feature maps of different scales, shallow feature information is seriously lost after the bottom-up multilayer transmission. Later, Liu et al. [51] proposed a path aggregation Network (PANet) based on the FPN, which added a bottom-up path enhancement structure to achieve the purpose of better preserving shallow feature information. This structure not only achieves good results in instance segmentation tasks, but also plays a certain role in object detection tasks. Liu et al. [52] also put forward an adaptive spatial feature fusion method, called Adaptively Spatial Feature Fusion (ASFF), where through adaptively learning the fused spatial weights of the feature maps at each scale, the feature inconsistency problem caused by the simultaneous existence of feature maps with different scales in the feature pyramid is effectively solved.

For target detection tasks, high-level and low-level information are complementary. The above multi-scale feature fusion method can achieve the fusion of low-level semantic information and high-level feature information through bottom-up path enhancement and lateral connection. However, in each fusion process, the feature information of small targets will be lost or diluted to a certain extent. In order to reduce the adverse effects of feature fusion and retain the detailed features of the target as much as possible, we have introduced an attention mechanism module into the feature fusion model. The attention mechanism module suppresses irrelevant information and enhances semantic information, reduces information loss in the feature fusion process, and further improves the network's effect in small target detection.

3. Method

3.1. Overall Structure

The overall structure of DRADNet is shown in Figure 2. It consists of a degraded reconstruction-assisted enhancement branch and a detection branch. Among them, target detection is the main task, and the introduced super-resolution branch only serves as an auxiliary guide to enhance feature learning. In the training and inference stages of our method, the model adopts different structures. The part connected by the dotted line is the degraded reconstruction auxiliary enhancement branch we designed, which consists of an image degradation module and a super-resolution reconstruction branch. As an auxiliary branch, this branch plays an important role in the training process. It supervises the model to extract high-quality target feature representation by sharing the backbone network part with the target detection branch, thereby avoiding information loss caused by the network in the downsampling process, to a certain extent, and better learning target feature information. The part connected by the solid line is the detection branch we designed, which is used to detect the target. We introduced a hybrid parallel-attention mechanism into the feature fusion network, which can guide the network to pay more attention to the feature information of small targets and effectively suppress the interference of redundant background information. The detection branch and SISR branch share the same backbone network to share optimization weights, helping the detection network learn the ability to extract high-quality features in the super-resolution reconstruction process, thereby guiding the detection model to enhance its feature response to small targets and improve its performance in the detection of small targets in medium- and low-resolution remote sensing images.

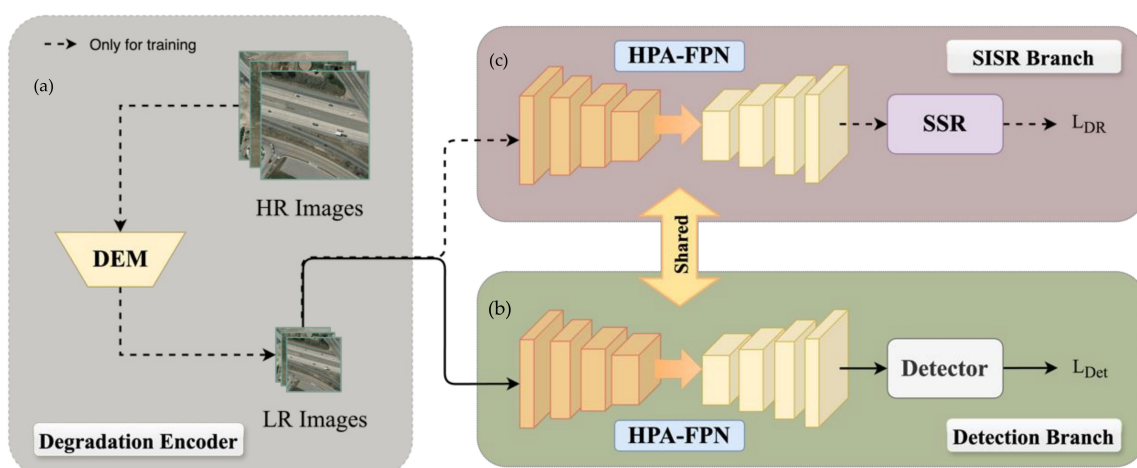


Figure 2. Illustrations of the architecture of the proposed method DRADNet: (a) the degradation encoder module for building a real-world degradation space (yellow part); (b) the object detection branch for achieving accurate acquisition of small targets (green part); (c) the single-image super-resolution reconstruction branch for guiding the detection network to better learn the target features (pink part).

During the training process, one inputs the high-resolution image into the degradation coding module to obtain low-resolution images of different sizes through random degradation, and then guides the feature extraction network of the target detection branch to learn how to effectively extract high-quality target features through the supervision of the super-resolution branch. Among them, the degradation coding module is used to obtain LR images of different sizes that are closer to the actual degradation, so as to cover the unknown degradation conditions in the real low-resolution remote sensing images as much as possible. Then, each obtained LR image is sent to the SISR branch and the detection branch for model training. The SISR branch reconstructs the degraded LR image with super resolution, thereby enhancing the feature information of the target. Through the SISR branch, the model is guided to have the ability to learn high-quality feature information of the target. The detection branch uses the CenterNet [53] detector without anchor boxes, which has a simple structure and higher versatility. The two branches share the same backbone network. Through continuous training and loss optimization, the detection network can have the ability to achieve super-resolution enhancement of the target features when detecting medium- and low-resolution images of the original dataset without going through a super-resolution reconstruction module. This intentionally guides the detection branch to better extract prominent features similar to those after high-resolution reconstruction, thereby ensuring that the network achieves better detection results without increasing the amount of additional computation.

In the training process of the entire network, the super-resolution reconstruction branch plays an important role. However, in the actual reasoning process, the degraded reconstruction auxiliary branch will be removed and will not participate in the reasoning process of the entire model, thereby effectively improving the performance of the detector without increasing the computational cost of model reasoning. Specifically, the medium- and low-resolution remote sensing images in the selected dataset will be directly entered into the detection branch for detection tasks without going through the SISR branch. At this time, the trained detection network has the ability to extract high-quality target feature information after super-resolution reconstruction. In addition, we introduced an attention enhancement mechanism in the feature fusion network, which can guide the network to pay more attention to the feature information of small targets, suppress redundant complex backgrounds, and improve the robustness and stability of detection. The proposed work is introduced in detail below.

3.2. Degradation Reconstruction-Assisted Enhancement Branch

Firstly, we designed a degradation reconstruction-assisted enhancement branch, which has two functions of image degradation and super-resolution reconstruction. By using the degraded model, M-LR images of different sizes and resolutions can be obtained and sent to the SISR branch for training, so that the network has the ability to learn images of different degradation degrees and extract objects. The degradation reconstruction-assisted enhancement branch is used as an assisted training branch in the whole network to guide the model to capture feature expressions at different resolutions, so as to improve the feature learning ability of the network in M-LR images.

3.2.1. Degradation Reconstruction-Assisted Enhancement Branch

The traditional degradation model assumes that HR images are affected by various degradation types, which usually include three key factors: blur, noise, and downsampling. Mathematically, the degradation process is given by Equation (1):

$$D(x) = (x \otimes k) \downarrow_s + n \quad (1)$$

where $D(x)$ represents the degradation process of the image, k represents the fuzzy kernel, x represents the original HR image, $\downarrow_s$ represents an s -times downsampling operation, and n represents the noise. The order of degradation in existing methods is fixed. First, the HR image is blurred by convolution with the fuzzy kernel. Then, it is downsampled to

reduce the image resolution. Finally, a degraded M-LR image is obtained by superimposing additive noise. This method uses a fixed fuzzy kernel and noise on the input HR image, and then obtains the M-LR image with the same degree of degradation after a fixed degradation sequence and a fixed downsampling size. However, in practical applications, remote sensing images are usually affected by satellite platforms, imaging sensors, atmospheric interference, and other factors, resulting in unknown complex degradation. A single degradation method cannot well simulate actual degradation. Therefore, in order to cover the unknown degradation in real M-LR remote sensing images as much as possible, we introduced a practical degradation representation module to achieve the goal of better approximating real degradation.

The overall structure of the degradation module is shown in Figure 3, which is mainly realized through the mixed random downsampling mechanism and the gated random selection mechanism. The mixed random downsampling mechanism obtains degraded images of different sizes by randomly placing them at different locations in the degraded model, unlike the traditional degradation model, which can only obtain images of the same size by fixing the downsampling multiple, so as to have a more effective representation of real image degradation. The gated random selection mechanism creates more subsets of degradation types through the randomness of the gated switch. It can randomly process images with different downsampling degrees for fuzzy processing or noise processing, or a combination of both, unlike the traditional method which is limited by fixed degradation types and degradation order, thus forming a more powerful degradation representation model.

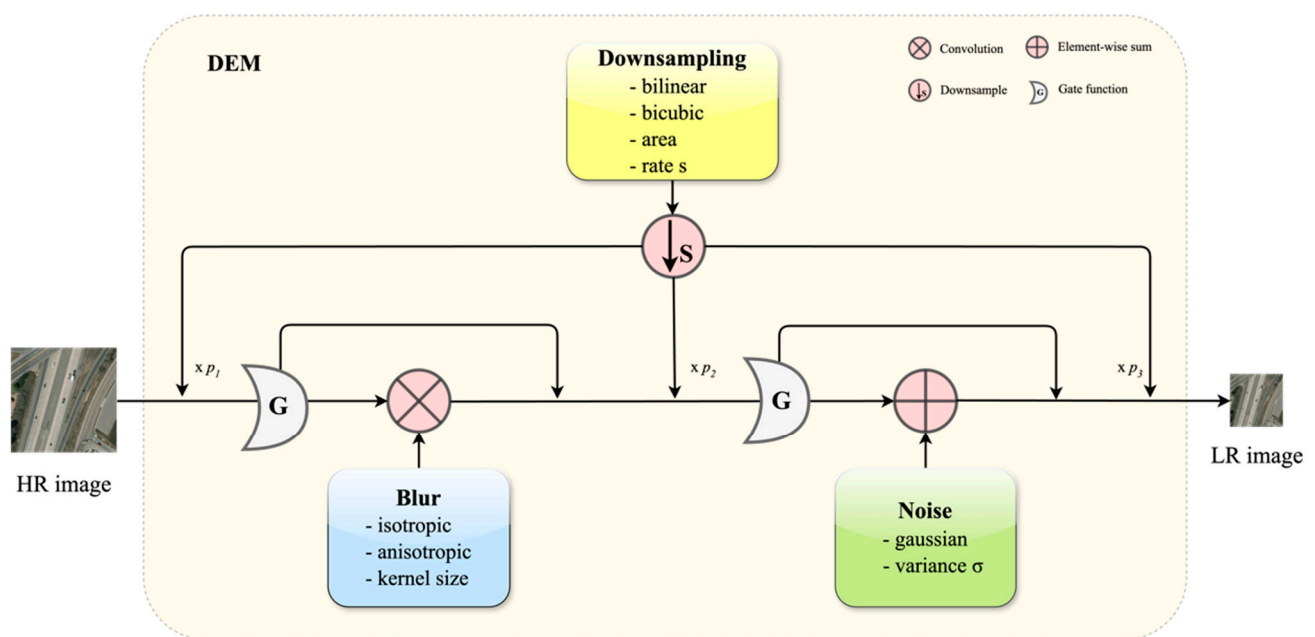


Figure 3. The detailed structure of the proposed degradation module.

Blur is a very common degradation type in remote sensing images. Here, we modeled the blur degradation as a convolution operation between the image and the Gaussian filter. In order to include a variety of Gaussian blur types, we chose two kernel forms of the Gaussian filter for our degenerate model: they are isotropic and anisotropic, respectively. Their dimensions range from $\{7 \times 7, 9 \times 9, \dots, 21 \times 21\}$, and are taken randomly from this range. Among them, the kernel width of the isotropic Gaussian kernel is uniformly sampled in the range of $[0.1, 2.4]$, the longer kernel width of the anisotropic Gaussian kernel is uniformly sampled in the range of $[0.5, 6]$, and the kernel angle is uniformly sampled in the range of $[0, \pi]$.

Downsampling is the basic operation of conversion from high resolution to low resolution. Different downsampling methods will bring different effects. There are many common downsampling methods, including nearest-neighbor interpolation, bilinear interpolation, bicubic interpolation, area adjustment, and so on. Because the nearest-neighbor interpolation downsampling causes the space deviation error, we discarded it. In order to achieve complex and diversified downsampling degradation, we randomly chose the downsampling order, i.e., through the three parameters p_1 , p_2 , and p_3 , where the downsampling rate is obtained from the uniform sampling of $[1.0, 4.0]$.

Noise is widespread in remote sensing images and is caused by various external factors. Gaussian noise is the most widely used noise type. Considering that noise information was unknown, we conservatively adopted the Gaussian noise model to simulate sensor noise and other external noise. Here, our method uses the widely used channel-independent additive white Gaussian noise (AWGN) model $n \sim N(0, \sigma)$ in Equation (1). Variance σ is obtained from a uniform sampling of $\{1/255, 2/255, \dots, 25/255\}$.

The HR remote sensing image is randomly processed by the above fuzzy, downsampling, and noise techniques, and the M-LR remote sensing image is closer to the real degradation. Downsampling must be experienced; p_1 , p_2 , and p_3 represent different probabilities, which determine the position of the image to be downsampled—here, we set these to 10%, 10%, and 80%, while blur and noise are randomly experienced and are controlled by the gate switch. For example, after the image enters the degradation model, it may be randomly downsampled first, followed by blur and noise processing, or it may be blurred first, followed by downsampling and noise processing, or it may only go through downsampling plus one of either the blur or noise processes. All in all, downsampling must be experienced, noise and blur can be randomly selected through gating, and the positions of the three are also randomly selected. This degradation method can obtain M-LR images with different degradation degrees and sizes, which are closer to the images obtained in the actual scene.

3.2.2. Single-Image Super-Resolution Branch

In convolutional neural networks, with the increase in the number of backbone feature extraction network layers, the downsampling multiple increases, the feature map size decreases, and the receptive field increases. The model can extract more semantic information through hierarchical downsampling, but it is prone to losing some important details in the process. Especially for M-LR remote sensing images with low resolution and weak target texture information, more feature information will be lost after the extraction of the backbone. In order to solve this problem, we introduced an SISR branch, which aims to enhance the feature information of the target through super-resolution reconstruction of the feature map after feature extraction, strengthening the ability of the network to extract and learn the target features. The overall structure of the SISR branch is shown in Figure 2, which is composed of the backbone network and SRR module. The SISR branch shares the feature extraction backbone network with the detection branch, so that the detection network can learn the capability of super-resolution reconstruction. M-LR images are firstly aggregated by the feature fusion module with different scale features from the backbone, and then the aggregated features are sent to the SRR module for super-resolution reconstruction.

As shown in Figure 4, the deep features after HPA-FPN feature fusion are input into the SRR module and converted into corresponding SR images. The obtained SR images and the original HR images are used to minimize the learning error. Specifically, given a low-resolution deep feature, the feature is first continuously upsampled by an alternating cascade of Spatial Shift-Conv Residual Block (SSCBlock) and upsampling operations. After that, the upsampled features generated by skip connections are aggregated. Finally, the reconstructed SR image is obtained through a pixel-shuffle operation with the superimposed features. Among them, SSCBlock mainly includes 1×1 convolution, 3×3 convolution, a spatial shift operation, and a ReLU activation function. The specific structure is shown in Figure 4. SSCBlock can align features of its neighbors and aggregate local features by

means of the spatial shift operation. It can further extend the extraction of long-distance feature relations by controlling the selection of shifted feature points.

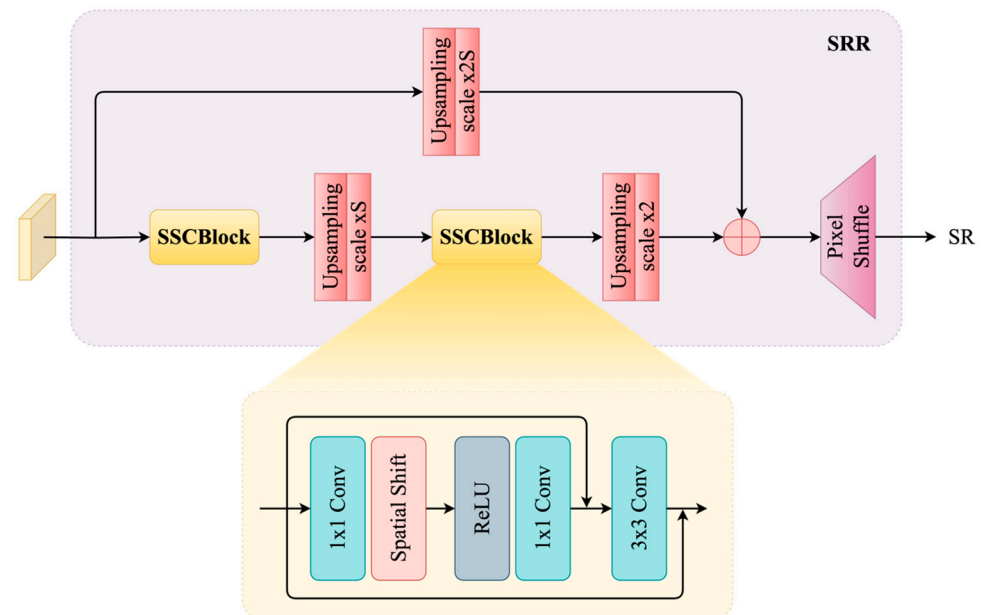


Figure 4. The overall structure of the SRR module.

3.3. Feature Fusion Module HPA-FPN

The traditional feature pyramid structure can alleviate the problem of information loss to a certain extent by directly blending M-LR and HR feature maps, but the direct fusion of features at different stages limits the expression of multi-scale features, which easily causes small targets to be submerged in a large amount of information. Finding a way to suppress the interference in the complex background information and make the network pay more attention to the detection of small targets is the key to enhancing the feature extraction capability of the backbone network. Therefore, we introduced a hybrid parallel-attention mechanism into the backbone network and propose a composite-structure FPN, the principle of which is shown in Figure 5. In the feature fusion stage, HPA-FPN takes the features generated by the backbone network with different resolutions $\{c1, c2, c3, c4\}$ as inputs, extracts the features through the hybrid parallel-attention mechanism, and fuses them to obtain $\{p1, p2, p3, p4\}$.

The detailed structure of the proposed HPA is shown in Figure 5, which we define as a two-branch parallel module. For the given input feature $Z \in \mathbb{R}^{C \times H \times W}$, our method passes it to the branches separately and outputs a thinning feature of the same shape. For the left branch, our method aims to aggregate the feature maps in each channel to fully learn the specific relationships between channels. First, the input feature map is globally pooled to obtain a feature map with shape $C \times 1 \times 1$. The channel is then scaled r times by the 2D convolution of 1×1 and the nonlinear activation function ReLU, and its characteristic shape is transformed to $C/r \times 1 \times 1$. The channel is then restored to C by a 2D convolution of 1×1 . Finally, the weight W_C of each feature channel is obtained through the mapping operation of the Sigmoid activation function. For the right branch, our purpose is to learn the spatial information of features, so as to better focus the effective information on the feature map. First, the input feature maps are average-pooled in the X and Y dimensions, respectively, and feature maps with shapes of $C \times H \times 1$ and $C \times 1 \times W$ are obtained, respectively. Then, these two feature maps are spliced in the spatial dimension and a 2D convolution of 1×1 is used to obtain a feature map with the shape $C/r \times 1 \times (W + H)$. Then, BatchNorm and the nonlinear function are used to encode the spatial information in the X and Y directions. Next, the channel is restored to C through a split operation and a 2D convolution of 1×1 , respectively. Finally, the weight W_S is also mapped by the Sigmoid

activation function. We can then multiply the attention weights W_C and W_S , the values obtained from the two branches, element by element with the original input feature map to obtain the refined features.

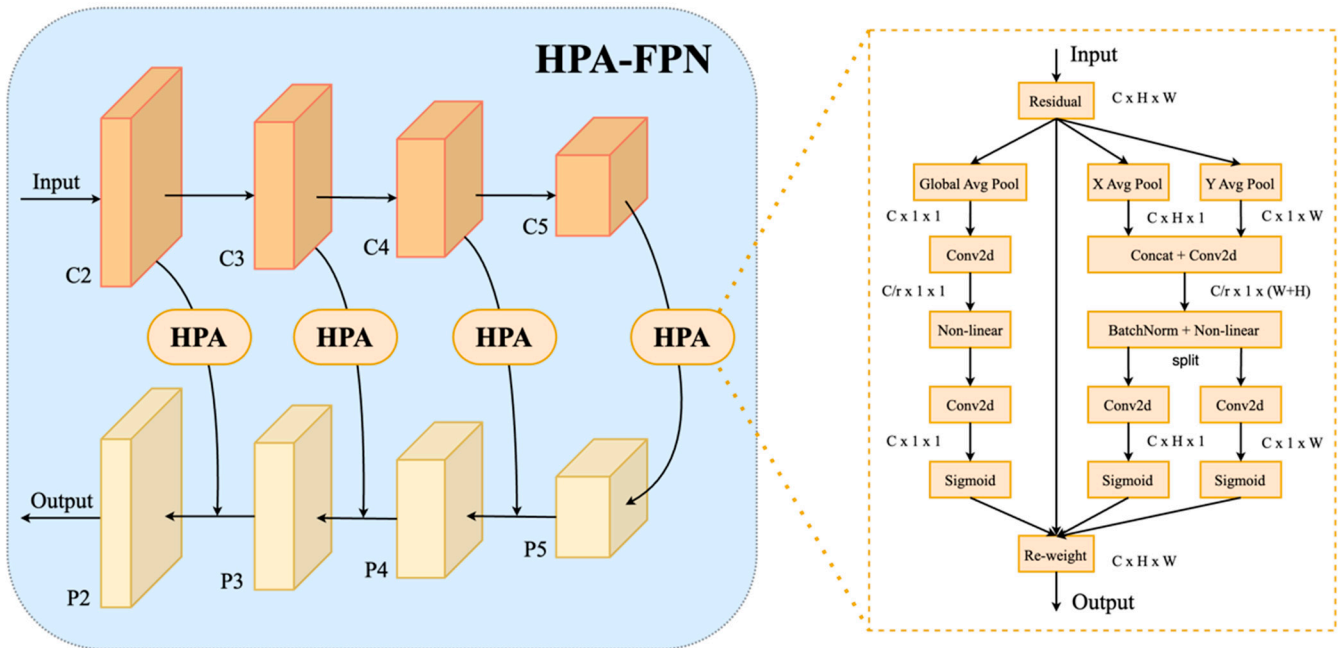


Figure 5. Proposed structure of HPA-FPN.

3.4. Loss Function

The loss of our network mainly consists of two parts, namely degradation reconstruction loss and target detection loss. These two types of losses are described as follows:

(1) Degradation Reconstruction Loss

The aim of the image super-resolution reconstruction branch is to generate clear HR images from degraded input M-LR images. Only one degenerate reconstruction loss L_{DR} is included in this branch. Here, we define L_{DR} as the L_1 distance between the reconstructed image I_{SR} after degradation and the original HR image I_{HR} , expressed mathematically as Formula (2).

$$L_{DR} = \| I_{SR} - I_{HR} \|_1 \quad (2)$$

(2) Target Detection Loss

The loss of the target detection branch L_{Det} includes the dimension prediction branch loss L_S , the center-point displacement loss L_O , and the center-point prediction loss L_C . Gaussian Focal loss is used for the center-point prediction loss, and L_1 loss is used for the dimension prediction branch and center-point displacement loss. The total loss of the target detection branch is defined as shown in Formula (3). In this formula, the hyperparameters λ_1 and λ_2 are set to 0.1 and 1, respectively.

$$L_{Det} = \lambda_1 L_S + \lambda_2 L_O + L_C \quad (3)$$

(3) Total Loss

When we jointly train the super-resolution reconstruction subnetwork and the object detection subnetwork, the mathematical expression of the total loss is shown in Formula (4), where α and β are balance factors.

$$L_{Total} = \alpha L_{DR} + \beta L_{Det} \quad (4)$$

The total loss is the key to guiding the network training and optimization. The detection branch can continuously optimize its learning ability of super-resolution reconstruction through the guidance of total loss to achieve better detection results.

4. Experiments

4.1. Dataset and Experimental Details

To evaluate the effectiveness of our proposed method, two challenging datasets containing a large number of small objects were selected for experimentation, namely the VEDAI dataset [54] and the Airbus-Ships dataset [55]. In order to better verify the method of this paper, we conducted some processing on the datasets, respectively, and the specific processing details are described as follows.

4.1.1. VEDAI Dataset

VEDAI is a small and medium vehicle object detection dataset for visible-light remote sensing images published by Razakarivony et al. It uses cropped images obtained from the Utah AGRC. The background and distractors of the targets in this dataset are diverse, covering complex scenes such as cities, fields, forests, and mountains. After eliminating the very few targets and erroneous data in the dataset, the data contain 1245 images of nine common types of vehicles: boats, campers, cars, pickups, planes, tractors, trucks, vans, and others. There are 3750 targets in this dataset, each with a size of 1024 pixels $\times$ 1024 pixels and 512 pixels $\times$ 512 pixels, respectively, and a spatial resolution of 12.5 cm $\times$ 12.5 cm per pixel and 25 cm $\times$ 25 cm per pixel, respectively. The training set and the test set were composed of 996 and 249 images, respectively. The dataset has strong robustness not only because its targets are very small, but also because it has the condition of illumination change and occlusion. The statistics of the VEDAI dataset are shown in Table 1, which shows the data distribution of the training and testing samples.

Table 1. Statistics of the VEDAI dataset.

Class Name	Targets in the Training Set	Targets in the Test Set	Total
Car	1110	267	1377
Truck	242	65	307
Pickup	766	189	955
Tractor	154	36	190
Camping Car	326	71	397
Boat	139	32	171
Van	80	21	101
Other Vehicle	153	51	204
Plane	40	8	48
	3010	740	3750

4.1.2. Airbus-Ships Dataset

The Airbus-Ships dataset is provided by Airbus in the Ship Detection Challenge launched in Kaggle. Airbus collects satellite remote sensing images containing ships. This dataset comprises remote sensing images that contain a single ship or several ships, located in the open sea, at docks, on marinas, etc. There are 2730 images in this dataset, of which 2184 were used for training and 546 were used for testing. Each image has two specifications of 768 pixels $\times$ 768 pixels and 384 pixels $\times$ 384 pixels, respectively.

4.1.3. Implementation Details

All experiments were built on the Ubuntu platform and based on the PyTorch deep learning framework for model training and inference on the NVIDIA3090 graphics card in this paper. In these experiments, we adopted the Stochastic Gradient Descent (SGD) method for training, the initial learning rate was 0.02, and the momentum and weight

decay parameters were set at 0.0001 and 0.9, respectively. For the evaluation of the experiment, we used the commonly used average accuracy (map) as the main index of method performance comparison.

4.2. Ablation Study

In this section, we performed a series of ablation experiments on the VEDAI dataset in order to verify the validity of key modules in our proposed method. Compared with the Airbus-Ships dataset, the VEDAI dataset has better richness and complexity, so it can better evaluate the actual effect produced by the method in this paper. In the experiments, we used CenterNet as the baseline model and gradually added the different proposed components to show the impacts they bring. The results of our ablation experiments on the VEDAI dataset are shown in Table 2.

Table 2. Ablation experiments performed on the proposed modules using the VEDAI dataset.

Method	SISR Branch	HPA-FPN	mAP ₅₀
Baseline	-	-	64.4
DRADNet	✓	-	72.3
	-	✓	69.8
	✓	✓	77.5

Three different combinations were used to enrich our experimental effect, namely adding the SISR branch separately, adding HPA-FPN separately, and adding the SISR branch and HPA-FPN simultaneously. The experimental results are shown in Table 2. When the SISR branch structure was added separately, the mAP index increased by 7.9%, indicating that the super-resolution reconstruction module plays a good role in enhancing the target features and can help the network to better detect the target. When the HPA-FPN structure was added separately, the mAP index increased by 5.4%, indicating that when the hybrid parallel-attention mechanism is introduced into the backbone network, the interference of redundant background information can be effectively suppressed, and the network's ability to extract and recognize small target features can be improved. When these two structures were added to the network, the overall detection performance of the network improved by 13.1%, indicating that the SISR branch and detection branch share the backbone network, which can well enable the detection branch to learn the ability of super-resolution reconstruction and enhance the feature extraction ability of the network in relation to small targets while suppressing background interference. Thus, the detection performance is effectively improved. The results of the ablation experiments clearly show the effectiveness of our proposed method, which brings benefits to the detection of small- and medium-sized targets in M-LR remote sensing images.

4.3. Experimental Results and Discussions

To further evaluate the performance of our method for small target detection within M-LR remote sensing images, we compared our method with several commonly used advanced detection methods, including common general target detection methods, such as Faster R-CNN [56], Cascade R-CNN [57], CenterNet [53], FOCS [58], and RetinaNet [49]. At the same time, we also compared small target detection algorithms for remote sensing images, such as SuperYOLO [18], RTMDet [59], NWD [16], and RFLA [17]. In order to ensure the validity and accuracy of the comparison results, we ensured that all of the compared algorithms were carried out on the same platform and under the same training reasoning conditions. The experimental results of our method on the VEDAI dataset and Airbus-Ships dataset are described in detail below.

4.3.1. Results on the VEDAI Dataset

The VEDAI dataset covers small targets in different scenarios, which exhibit a large amount of similar-target interference. Here, we compare the test results achieved using DRADNet on the VEDAI dataset with those achieved with some other advanced detection methods. As shown in Table 3, the mAP evaluation index of our proposed method in the VEDAI dataset reaches 77.5%; compared with the common general object detection algorithms Faster R-CNN, Cascade R-CNN, CenterNet, and RetinaNet, the mAP value is increased by 11.4%, 9.7%, 13.1% and 26.3%, respectively, and the detection accuracy of various types of targets is greatly improved. As can be seen from Figure 6, compared with the general algorithm, DRADNet is more suitable for small target detection in complex backgrounds. In Figure 6c, due to more distractors around the target, the Cascade R-CNN algorithm shows incorrectly detected targets, and there is the phenomenon of misclassification of target categories in Figure 6b. The robustness and adaptability are not as good as those of DRADNet. Compared with the small object detection algorithms SuperYOLO, RTMDet, NWD, and RFLA for remote sensing images, the mAP value is increased with DRADNet by 4.4%, 10.4%, 9.2%, and 5.9%, respectively, which also shows great improvement. DRADNet also has the highest mAP in the car, truck, pickup, tractor, camping car, boat, and van categories. As shown in Figure 6a, the RTMDet algorithm has the problems of misdetection and object classification error. In Figure 6b, the RTMDet and RFLA algorithms have confused the car and pickup targets to varying degrees, and do not show accurate discrimination. In the dense case shown in Figure 6d, the target boundary in the M-LR image is blurred, which causes a great hindrance to the accurate identification of the target. However, the method we propose shows good performance in all of the above cases, and alleviates the missed detection and difficult identification of small targets to a certain extent.



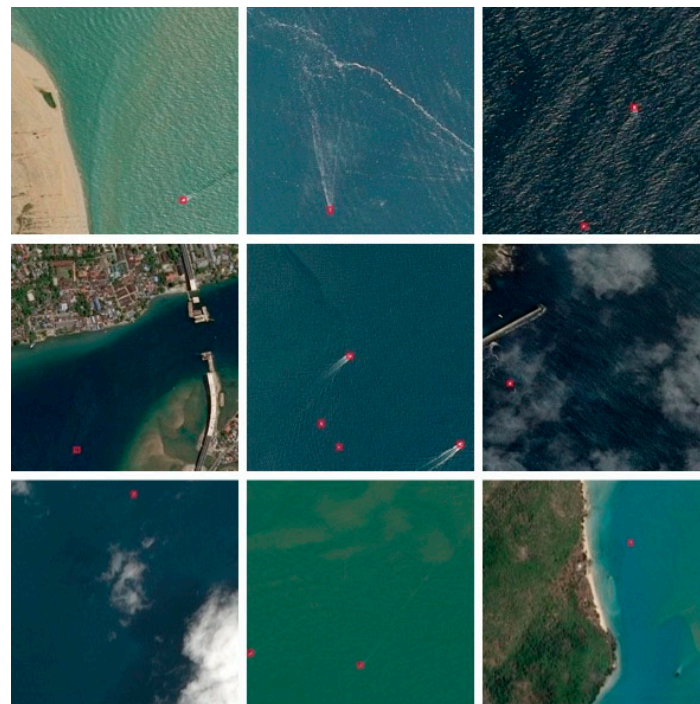
Figure 6. Detection results for the VEDAI dataset achieved with different algorithms. In the figure (a–d) are examples of different scenarios.

Table 3. Comparison with other state-of-the-art methods performed on the VEDAI dataset.

Method	Car	Truck	Pickup	Tractor	Camping Car	Boat	Van	Other Vehicle	Plane	mAP (%)
Faster R-CNN [56]	83.2	60.4	73.7	56.4	76.2	54.6	50.0	51.2	80.4	65.1
Cascade R-CNN [57]	81.6	65.4	75.1	67.7	77.3	42.8	60.2	47.0	92.8	67.8
SuperYOLO [18]	84.5	75.6	79.4	72.6	78.1	62.3	69.8	52.4	83.3	73.1
RTMDet [59]	64.9	68.1	77.5	74.5	76.0	52.2	65.4	37.0	88.4	67.1
CenterNet [53]	77.8	58.6	72.3	67.6	75.9	42.8	57.5	45.7	81.7	64.4
FOCS [58]	77.5	42.4	69.5	62.5	74.9	36.5	42.7	38.3	75.2	57.7
RetinaNet [49]	68.7	34.6	57.7	53.6	59.2	37.7	23.5	42.2	83.6	51.2
NWD [16]	81.8	70.2	74.5	69.2	77.2	43.1	61.3	46.7	90.6	68.3
RFLA [17]	83.1	68.2	78.4	72.1	74.9	63.4	67.6	50.7	85.6	71.6
DRADNet	89.0	76.0	80.2	82.8	82.0	73.7	71.9	51.1	90.3	77.5

4.3.2. Results on the Airbus-Ships Dataset

There are many extremely small ship targets in the Airbus-Ships Dataset, which are usually disturbed by waves and thin clouds. Compared with other algorithms, the DRADNet algorithm we propose has better performance on the Airbus-Ships Dataset, achieving 82.7% mAP. Compared with the target detection algorithms SuperYOLO, RTMDet, NWD, and RFLA for remote sensing images, the mAP is increased by 2.3%, 5.9%, 3.9%, and 3.6%, respectively, which is a significant improvement. The specific situation is shown in Table 4. In the face of the serious problem of misdetection encountered by other algorithms on this dataset, with our algorithm, this is well improved. Figure 7 shows the visualization effect of the proposed algorithm for the Airbus-Ships dataset. Even in the face of interference from thin clouds, sea clutter, or small islands and reefs, small ship targets in the images can be well detected.

**Figure 7.** Examples of detection results using the proposed method DRADNet on the Airbus-Ships dataset.**Table 4.** Comparison with other state-of-the-art methods performed on the Airbus-Ships dataset.

Method	mAP (%)
Faster R-CNN [56]	75.2
Cascade R-CNN [57]	78.5

Table 4. Cont.

Method	mAP (%)
SuperYOLO [18]	80.4
RTMDet [59]	76.8
CenterNet [53]	70.9
FOCS [58]	65.1
RetinaNet [49]	62.3
NWD [16]	78.8
RFLA [17]	79.1
DRADNet	82.7

4.3.3. Visual Analysis

To further analyze the improvement of the detection effect brought by the optimization algorithm in this paper compared with the baseline algorithm, we use a heat map to visualize the features in the network. A heat map can intuitively reflect the intensity of the response of the network to the target area through the depth of color, making the detection effect more intuitive. In Figure 8, the first row is the original image, the second row is the feature visualization result obtained using the baseline algorithm, the third row is the feature visualization effect of the optimization algorithm proposed by us, and the fourth row is the ground truth. As can be seen from this figure, the baseline algorithm is not ideal for vehicle detection in complex backgrounds, the identified area is not accurate, and the entire shape of the vehicle target is not fully identified. At the same time, part of the background is also identified as the target, which greatly affects the accuracy of network detection; our optimization algorithm introduces the degraded reconstruction-assisted enhancement branch as an assisted branch before detection to enhance the model's learning ability for high-quality target features, and introduces the attention enhancement mechanism in the feature fusion network, which improves the feature information while suppressing the interference of redundant background, effectively solving the problem of the baseline algorithm not being comprehensive and accurate in detecting small targets in complex backgrounds. As can be seen from the effect comparison diagram, the algorithm proposed by us has a higher recognition degree for the target and a more accurate response, which fully verifies that the algorithm in this paper has good adaptability and robustness.

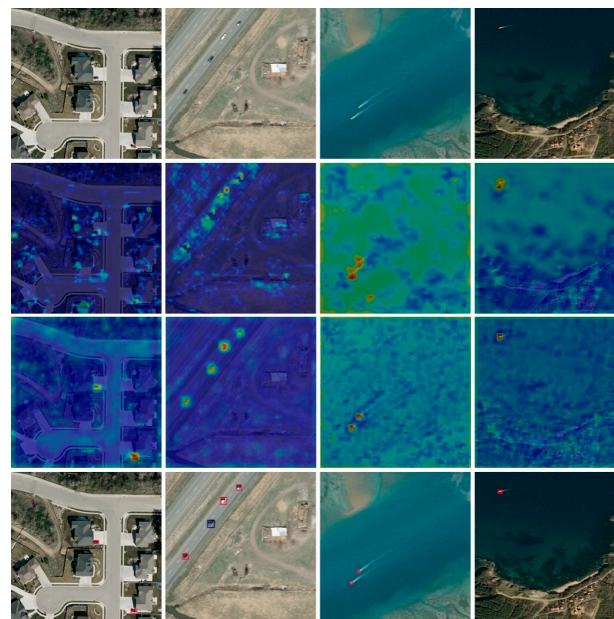


Figure 8. Examples of feature visualization. The first row is the original image, the second row is the feature visualization result obtained using the baseline algorithm, the third row is the feature visualization effect of the DRADNet algorithm, and the fourth row is the ground truth.

5. Conclusions

In this paper, we propose a method for small target detection in M-LR remote sensing images based on degraded reconstruction. Different from the previous method of using image super resolution as a preprocessing step and then directly inputting data into the detector, we designed an additional degraded reconstruction auxiliary framework, which can effectively improve the performance of the detector in M-LR remote sensing image detection tasks without increasing the computational cost of model inference. Specifically, it consists of a degraded reconstruction auxiliary enhancement branch and a detection branch. Among these, target detection is the main task, and the introduced super-resolution branch only serves as an auxiliary guide to enhance the features. In the training and inference stages of our method, the model adopts different structures. In the training process of the entire network, the super-resolution reconstruction branch plays an important role. It supervises the model in extracting high-quality target feature representations by sharing the backbone network part with the target detection branch, thereby avoiding the information loss caused by the network in the downsampling process to a certain extent. In addition, we introduced a hybrid parallel-attention mechanism into the feature fusion network, which can guide the network to pay more attention to the feature information of small targets and effectively suppress the interference of redundant background information. In the actual inference stage, the degraded reconstruction auxiliary enhancement branch is removed; that is, it will not participate in the entire inference process of the model. In this way, the detection of small targets in M-LR remote sensing images can be effectively improved without changing the computational complexity of the original target detection model. We have demonstrated the robustness and effectiveness of this algorithm through a large number of experiments on the VEDAI dataset and the Airbus-Ships dataset, which fully demonstrates that DRADNet has good adaptability in different complex scenarios and can be effectively applied to small target detection in M-LR remote sensing images.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z. and H.S.; formal analysis, Y.Z.; software, Y.Z.; validation, S.W. and H.S.; writing—original draft preparation, Y.Z.; investigation, Y.Z. and S.W.; resources, H.S.; project administration, H.S. and S.W.; funding acquisition, H.S.; data curation, S.W.; writing—review and editing, H.S. and S.W.; visualization, Y.Z.; supervision, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Changchun Science and Technology Bureau, grant number E31093U7T010.

Data Availability Statement: The data of experimental images used to support the findings of this research are available from the corresponding author upon reasonable request.

Acknowledgments: This research was supported by the Jilin Provincial Key Laboratory of Machine Vision Intelligent Equipment.

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

References

1. Audebert, N.; Le Saux, B.; Lefèvre, S. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 20–32. [\[CrossRef\]](#)
2. Hird, J.N.; Montaghi, A.; McDermid, G.J.; Kariyeva, J.; Moorman, B.J.; Nielsen, S.E.; McIntosh, A.C. Use of unmanned aerial vehicles for monitoring recovery of forest vegetation on petroleum well sites. *Remote Sens.* **2017**, *9*, 413. [\[CrossRef\]](#)
3. Li, L.; Zhou, Z.; Wang, B.; Miao, L.; Zong, H. A novel CNN-based method for accurate ship detection in HR optical remote sensing images via rotated bounding box. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 686–699. [\[CrossRef\]](#)
4. Zhang, X.; Wang, G.; Zhu, P.; Zhang, T.; Li, C.; Jiao, L. GRS-Det: An anchor-free rotation ship detector based on Gaussian-mask in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3518–3531. [\[CrossRef\]](#)
5. Xiong, G.; Wang, F.; Yu, W.X.; Truong, T.K. Spatial Singularity-Exponent-Domain Multiresolution Imaging-Based SAR Ship Target Detection Method. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 12. [\[CrossRef\]](#)
6. Xiong, G.; Wang, F.; Zhu, L.Y.; Li, J.Y.; Yu, W.X. SAR Target Detection in Complex Scene Based on 2-D Singularity Power Spectrum Analysis. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9993–10003. [\[CrossRef\]](#)

7. Wang, X.; Wang, A.; Yi, J.; Song, Y.; Chehri, A. Small Object Detection Based on Deep Learning for Remote Sensing: A Comprehensive Review. *Remote Sens.* **2023**, *15*, 3265. [\[CrossRef\]](#)
8. Li, Y.; Zhou, Z.; Qi, G.; Hu, G.; Zhu, Z.; Huang, X. Remote Sensing Micro-Object Detection under Global and Local Attention Mechanism. *Remote Sens.* **2024**, *16*, 644. [\[CrossRef\]](#)
9. Shi, T.; Gong, J.; Hu, J.; Zhi, X.; Zhu, G.; Yuan, B.; Sun, Y.; Zhang, W. Adaptive Feature Fusion with Attention-Guided Small Target Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 16. [\[CrossRef\]](#)
10. Yu, L.; Hu, H.; Zhong, Z.; Wu, H.; Deng, Q. GLF-Net: A target detection method based on global and local multiscale feature fusion of remote sensing aircraft images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4021505. [\[CrossRef\]](#)
11. Zhou, L.; Zheng, C.; Yan, H.; Zuo, X.; Liu, Y.; Qiao, B.; Yang, Y. RepDarkNet: A multi-branched detector for small-target detection in remote sensing images. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 158. [\[CrossRef\]](#)
12. Courtrai, L.; Pham, M.-T.; Lefèvre, S. Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks. *Remote Sens.* **2020**, *12*, 3152. [\[CrossRef\]](#)
13. Zhang, L.; Wang, Y.; Sun, Y. Salient target detection based on the combination of super-pixel and statistical saliency feature analysis for remote sensing images. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2336–2340.
14. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3446–3456. [\[CrossRef\]](#)
15. Hou, B.; Ren, Z.; Zhao, W.; Wu, Q.; Jiao, L. Object detection in high-resolution panchromatic images using deep models and spatial template matching. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 956–970. [\[CrossRef\]](#)
16. Wang, J.; Xu, C.; Yang, W.; Yu, L. A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv* **2021**, arXiv:2110.13389.
17. Xu, C.; Wang, J.; Yang, W.; Yu, H.; Yu, L.; Xia, G.-S. RFLA: Gaussian receptive field based label assignment for tiny object detection. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 526–543.
18. Zhang, J.; Lei, J.; Xie, W.; Fang, Z.; Li, Y.; Du, Q. SuperYOLO: Super resolution assisted object detection in multimodal remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5605415. [\[CrossRef\]](#)
19. Lepcha, D.C.; Goyal, B.; Dogra, A.; Goyal, V. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Inf. Fusion* **2023**, *91*, 230–260. [\[CrossRef\]](#)
20. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
22. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 606–615.
23. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L.W. Weakly supervised photo enhancer for digital cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 691–700.
24. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
25. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
26. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
27. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
28. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
29. Yu, J.; Fan, Y.; Yang, J.; Xu, N.; Wang, Z.; Wang, X.; Huang, T. Wide activation for efficient and accurate image super-resolution. *arXiv* **2018**, arXiv:1808.08718.
30. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
31. Wu, G.; Jiang, J.; Jiang, K.; Liu, X. Fully 1X1 Convolutional Network for Lightweight Image Super-Resolution. *arXiv* **2023**, arXiv:2307.16140.
32. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 2223–2232.
33. Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F.; IEEE Communications Society. Real-World Super-Resolution via Kernel Estimation and Noise Injection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Electr Network, Seattle, WA, USA, 14–19 June 2020; pp. 1914–1923.

34. Cui, Z.; Zhu, Y.; Gu, L.; Qi, G.-J.; Li, X.; Zhang, R.; Zhang, Z.; Harada, T. Exploring resolution and degradation clues as self-supervised signal for low quality object detection. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 473–491.
35. Bell-Kligler, S.; Shocher, A.; Irani, M. Blind super-resolution kernel estimation using an internal-gan. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 284–393.
36. Fritsche, M.; Gu, S.; Timofte, R. Frequency separation for real-world super-resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3599–3608.
37. Li, S.; Zhang, G.; Luo, Z.; Liu, J.; Zeng, Z.; Zhang, S. From general to specific: Online updating for blind super-resolution. *Pattern Recognit.* **2022**, *127*, 108613. [\[CrossRef\]](#)
38. Wang, B.; Yang, F.; Yu, X.; Zhang, C.; Zhao, H. APISR: Anime Production Inspired Real-World Anime Super-Resolution. *arXiv* **2024**, arXiv:2403.01598.
39. Chen, S.; Han, Z.; Dai, E.; Jia, X.; Liu, Z.; Xing, L.; Zou, X.; Xu, C.; Liu, J.; Tian, Q. Unsupervised image super-resolution with an indirect supervised path. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 468–469.
40. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4791–4800.
41. Zhang, W.; Shi, G.; Liu, Y.; Dong, C.; Wu, X.-M. A closer look at blind super-resolution: Degradation models, baselines, and performance upper bounds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 527–536.
42. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1905–1914.
43. Yang, G.; Li, B.; Ji, S.; Gao, F.; Xu, Q. Ship detection from optical satellite images based on sea surface analysis. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 641–645. [\[CrossRef\]](#)
44. Song, M.; Qu, H.; Jin, G. Weak ShipTarget Detection of NoisyOptical Remote SensingImage on Sea Surface. *Acta Opt. Sin.* **2017**, *37*, 1011004-1–1011004-8. [\[CrossRef\]](#)
45. Rabbi, J.; Ray, N.; Schubert, M.; Chowdhury, S.; Chao, D. Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network. *Remote Sens.* **2020**, *12*, 1432. [\[CrossRef\]](#)
46. Zou, H.; He, S.; Cao, X.; Sun, L.; Wei, J.; Liu, S.; Liu, J. Rescaling-Assisted Super-Resolution for Medium-Low Resolution Remote Sensing Ship Detection. *Remote Sens.* **2022**, *14*, 2566. [\[CrossRef\]](#)
47. Chen, J.; Chen, K.; Chen, H.; Zou, Z.; Shi, Z. A Degraded Reconstruction Enhancement-Based Method for Tiny Ship Detection in Remote Sensing Images with a New Large-Scale Dataset. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 14. [\[CrossRef\]](#)
48. He, S.; Zou, H.; Wang, Y.; Li, R.; Cheng, F.; Cao, X.; Li, M. Enhancing Mid-Low-Resolution Ship Detection with High-Resolution Feature Distillation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5. [\[CrossRef\]](#)
49. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
51. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
52. Liu, S.; Huang, D.; Wang, Y. Learning spatial fusion for single-shot object detection. *arXiv* **2019**, arXiv:1911.09516.
53. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
54. Razakarivony, S.; Jurie, F. Vehicle detection in aerial imagery: A small target detection benchmark. *J. Vis. Commun. Image Represent.* **2016**, *34*, 187–203. [\[CrossRef\]](#)
55. Inversion, M.J.F. Airbus Ship Detection Challenge. Available online: <https://kaggle.com/competitions/airbus-ship-detection> (accessed on 31 July 2018).
56. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Cai, Z.; Vascon, N. Cascade R-CNN: High Qual. Object Detect. Instance Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1483–1498. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully convolutional one-stage object detection. *arXiv* **2019**, arXiv:1904.01355.
59. Lyu, C.; Zhang, W.; Huang, H.; Zhou, Y.; Wang, Y.; Liu, Y.; Zhang, S.; Chen, K. Rtmddet: An empirical study of designing real-time object detectors. *arXiv* **2022**, arXiv:2212.07784.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.