

Article

Application of Minnan Folk Light and Shadow Animation in Built Environment in Object Detection Algorithm

Sichao Wu ¹, Xiaoyu Huang ¹, Yiqi Xiong ², Shengzhen Wu ^{3,*}, Enlong Li ^{4,*} and Chen Pan ^{5,6,*}

¹ Xiamen Academy of Arts and Design, Fuzhou University, Xiamen 361000, China; huangxiaoyu@fzu.edu.cn (X.H.)

² School of Business, Guangdong Polytechnic of Science and Technology, Zhuhai 519000, China; yiqicolin@126.com

³ College of Arts and Design, Jimei University, Xiamen 361000, China

⁴ Faculty of International Tourism Management, City University of Macau, Macau 999078, China

⁵ Architecture and Civil Engineering Institute, Guangdong University of Petrochemical Technology, Maoming 525000, China

⁶ Urban Planning and Design, Faculty of Innovation and Design, City University of Macau, Macau 999078, China

* Correspondence: u19092105153@cityu.mo (S.W.); t19092105092@cityu.mo (E.L.); u19092100017@cityu.mo (C.P.)

Abstract: To resolve the problems of deep convolutional neural network models with many parameters and high memory resource consumption, a lightweight network-based algorithm for building detection of Minnan folk light synthetic aperture radar (SAR) images is proposed. Firstly, based on the rotating target detection algorithm R-centernet, the Ghost ResNet network is constructed to reduce the number of model parameters by replacing the traditional convolution in the backbone network with Ghost convolution. Secondly, a channel attention module integrating width and height information is proposed to enhance the network's ability to accurately locate salient regions in folk light images. Content-aware reassembly of features (CARAFE) up-sampling is used to replace the deconvolution module in the network to fully incorporate feature map information during up-sampling to improve target detection. Finally, the constructed dataset of rotated and annotated light and shadow SAR images is trained and tested using the improved R-centernet algorithm. The experimental results show that the improved algorithm improves the accuracy by 3.8%, the recall by 1.2% and the detection speed by 12 frames/second compared with the original R-centernet algorithm.

Keywords: light and shadow animation; convolutional neural network; mechanism of attention; Minnan folk image; Ghost convolution



Citation: Wu, S.; Huang, X.; Xiong, Y.; Wu, S.; Li, E.; Pan, C. Application of Minnan Folk Light and Shadow Animation in Built Environment in Object Detection Algorithm. *Buildings* **2023**, *13*, 1394. <https://doi.org/10.3390/buildings13061394>

Academic Editor: Maxim A. Dulebenets

Received: 1 March 2023

Revised: 18 May 2023

Accepted: 21 May 2023

Published: 27 May 2023



Copyright: © 2023 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

As an active microwave sensing technology, synthetic aperture radar (SAR) has the characteristics of being independent of light and weather conditions and having strong penetration and all-weather detection capabilities. With the development of SAR imaging technology, SAR images have been widely used in military and civilian fields [1,2]. Building target detection in SAR images can quickly obtain building area information, which has important research significance in urban construction planning, military reconnaissance, disaster assessments, target strikes, etc. [3,4].

With the proliferation of SAR image data and the rapid development of computer vision, deep convolutional neural networks are being introduced to solve the target detection problem in Minnan folk light SAR images. The commonly used convolutional neural network-based target detection methods are mainly divided into two-stage detection algorithms and single-stage detection algorithms [5].

Among them, two-stage detection algorithms have two main steps: candidate region extraction and candidate region localization and classification. Thus, the detection speed is

slow. The most representative two-stage detection algorithms include the R-CNN [6] series, among others. The single-stage detection algorithm selects anchor frames directly from the light and shadow images and predicts the location and class of the target; it is both an accurate and fast target detection method. The yolo [7] series, SSD [8] and Centernet [9] are well-known single-stage detection algorithms. Deep convolutional neural networks are also widely used for SAR building detection. The authors of [10] suggest analyzing the Minnan SAR building min-light images with the help of a priori information and using synthetic aperture radar tomography to distinguish the building and non-building areas and produce a dataset. After model training, building detection in the Berlin city area in Minnan folk light SAR images was achieved. The authors of [11] proposed a multiscale convolutional neural network model that extracts multiscale features directly from SAR image blocks to detect buildings and experimentally validated it for high-resolution SAR images of the Beijing area. The authors of [12] analyzed the correlation between neighboring pixels in SAR images, introduced structured prediction into the network, and used multiscale features to classify pixels for building detection in SAR images. The building targets in light and shadow SAR images are often densely arranged in any direction, and the detection with the traditional horizontal rectangular box is easy to affect the adjacent targets. Therefore, a rotating rectangular frame is proposed to detect building objects in SAR images. The target detection algorithm based on a rotating rectangular box has the following advantages: (1) the orientation of the building can be fully considered in the detection, and the detection results represent the orientation information of the target; (2) it is possible to separate a single building from a dense arrangement to reduce the occurrence of missed detection; (3) it is possible to filter out the background information around a single building to avoid affecting the detection effect. Therefore, the object detection algorithm based on the rotating frame has an important research prospect in SAR image-building detection. At present, there are also two detection algorithms based on rotating rectangular boxes, among which DRbox-v2 [13] and SCRDet [14] are typical two-stage detectors, while R-centernet [15], R3Det [16], EAST [17] and FOST [18] are single-stage detectors.

Classical deep convolutional neural network models are usually accompanied by a large number of parameters and calculations, which occupy computer memory during training and reduce the efficiency of detection. Therefore, lightweight convolutional neural networks have promising applications and can be competent for target detection tasks with multiple real-time requirements. Lightweight target detection algorithms include the MobileNet [19] series, GhostNet [20], ShuffleNet [21], and so on.

Based on R-centernet, a single-stage target detection algorithm based on the rotating rectangle framework, this paper proposes a more lightweight, improved building detection algorithm for SAR images. First, Ghost convolution is used to replace the traditional convolution in the original network, and a Ghost ResNet network model is built to reduce the number of parameters. Secondly, a channel attention module integrating width and height information is proposed to improve detection accuracy and ensure a low parameter count. Then, the up-sampling method is improved to further reduce the computational cost of the network. Finally, the improved algorithm is used for training and testing datasets, and the performance of the improved algorithm in SAR image-building detection is verified. With the algorithm in this paper, we can achieve the goal of accurate road extraction for SAR images of buildings.

2. Building Detection Algorithm in SAR Images

2.1. Basic Principle of R-Centernet Algorithm

Currently, most common target detection algorithms use horizontal rectangular frames to extract anchor frames from images and then classify and localize them, which is inefficient and slow in detection. Centernet proposes a target detection method based on key point estimation. In the detection process, the key point is found to estimate the centroid of the target, and then other attributes are regressed. R-centernet, an improved version of centernet, is a rotating target detection algorithm. Based on the horizontal framework,

angular parameters are introduced for training and prediction. Finally, four features of the target are regressed: the heat map feature, the centroid, the size and the angle, and the exact position and orientation of the target in the image are determined. The algorithm structure is shown in Figure 1.

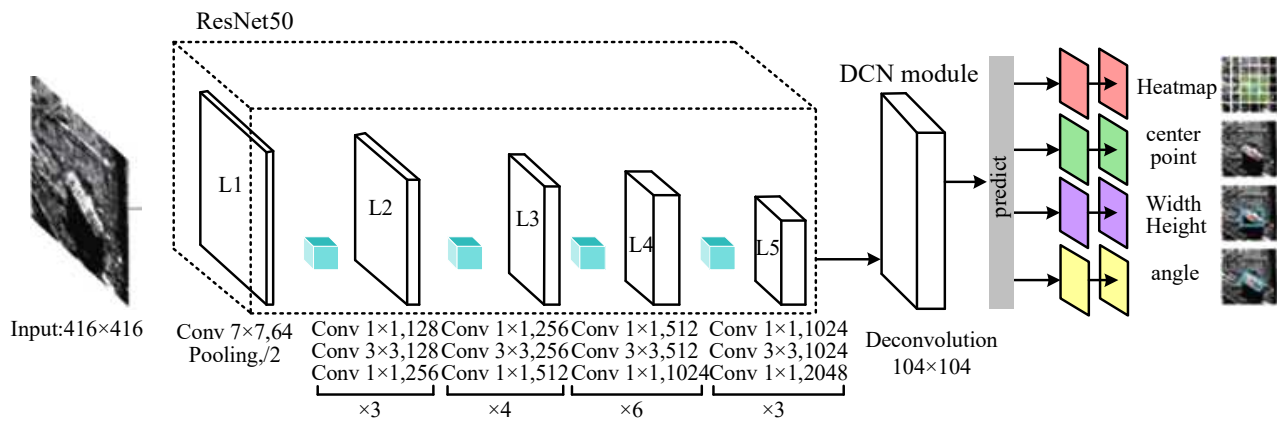


Figure 1. Structure of R-centernet algorithm.

As shown in Figure 1, after the image is fed to the R-centernet, it first undergoes feature extraction through the backbone network ResNet50 [22] and then is fed to the inverse folded product (DCN) module, which is mainly used for up-sampling. Finally, four predictions are obtained: the heat map, the centroid coordinates, the width and height, and the angle. Therefore, the loss function consists of three components: the thermal characteristic loss, the position loss (the centroid offset loss and the width and height prediction loss), and the rotation angle loss. The calculation equation is as follows.

$$L_{\text{det}} = L_{hm} + \lambda_{off} L_{off} + \lambda_{wh} L_{wh} + \lambda_{\theta} L_{\theta} \quad (1)$$

where L_{det} represents the total loss value; L_{hm} represents the thermal characteristic loss; L_{off} is the offset error of the center point of the prediction box; L_{wh} is the error of the width and the height of the prediction box; L_{θ} is the rotation angle error of the prediction box; λ_{off} , λ_{wh} , λ_{θ} is the corresponding weight.

2.2. Improved R-Centernet Algorithm

In the process of building detection in SAR images, the original R-centernet algorithm requires a large number of parameters and floating-point algorithms to obtain the benefit of detection accuracy, leading to a decrease in detection speed. In this paper, the R-centernet algorithm is improved to reduce the number of network parameters while ensuring detection accuracy, thus making building detection in SAR images efficient and effective.

The network structure of the improved R-centernet algorithm is shown in Figure 2, and the parts marked by red dashed boxes in the figure are the improved parts, which mainly include: (1) using Ghost convolution [23] instead of traditional convolution, with the backbone network improved from ResNet50 to Ghost ResNet; (2) adding the channel integrating width and height information in the Ghost ResNet attention mechanism, as shown in WH-ECA in Figure 2; (3) improving the DCN module in the original algorithm with an up-sampling method more suitable for lightweight networks.

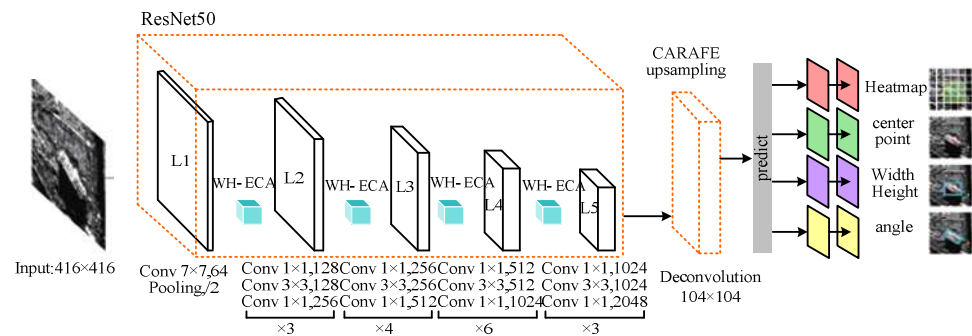


Figure 2. Improved R-centernet algorithm structure.

2.2.1. Application of Ghost Convolution in Residual Network

Due to the limitations of computing performance, storage space and detection speed, the target detection network should be as lightweight as possible to ensure high accuracy. Based on the single-stage detection algorithm R-centernet, this paper takes ResNet50 as the main backbone network and adopts Ghost convolution instead of the traditional convolution method in the network, which not only achieves the approximation effect but also reduces the number of parameters in the network.

The traditional convolution process and the Ghost convolution process are shown in Figure 3. In Figure 3, C , H and W are the number, quotient and width of the input feature graph, respectively; C' , H' and W' are the number, height and width of the output feature graph, respectively; the size of the convolution kernel adopted is $k \times k$. Ghost convolution is mainly divided into three parts: (1) The input feature map is generated by traditional convolution, and the number of channels of the feature map is smaller than that of the output feature map; (2) The Ghost feature map is obtained by Depthwise convolution operation. The Depthwise convolution Φ is run on each channel, the number of convolution kernels is the same as the number of channels in the previous layer, and the computational burden is far less than that of traditional convolution. (3) The final output result is obtained by stitching the eigenfeature map and Ghost feature map.

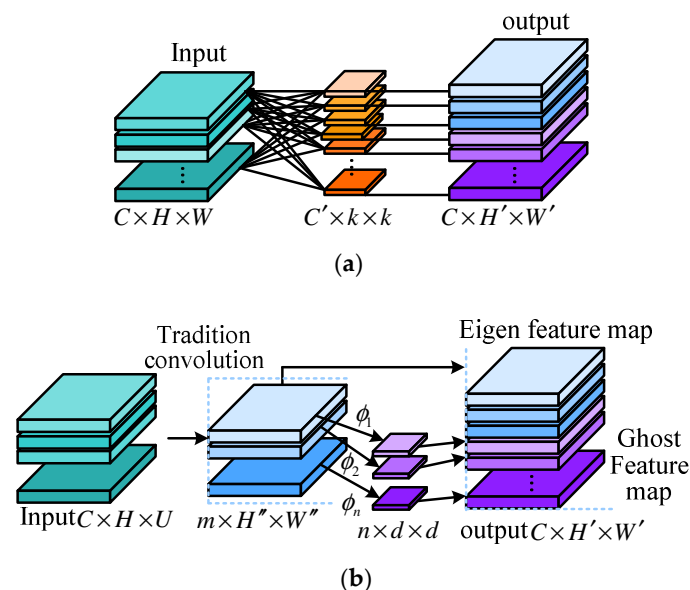


Figure 3. Traditional convolution structure and Ghost convolution structure. (a) Traditional convolution; (b) Ghost convolution structure.

Ghost convolution does not completely abandon the traditional convolution part; it first adopts the traditional convolution to generate a small number of channel feature maps and then generates the Ghost feature map. This method is an efficient and effective

convolutional method that can reduce the computational burden and ensure the recognition performance of features. When the size of the input feature map is $C \times H \times W$, the size of the convolution kernel used is $k \times k$ and the size of the output feature map is $C' \times H' \times W'$, the number of parameters required in the process of traditional convolution and Ghost convolution are shown in Equations (2) and (3), respectively.

$$P_{conv} = C \times C' \times k \times k \quad (2)$$

$$P_{ghost} = C \times m \times k \times k + m \times n \times d \times d \quad (3)$$

where m is the number of channels of the intrinsic feature map, n is the number of cores for linear operation, $d \times d$ is the size of the linear kernel and $d \times d \leq k \times k$. Therefore, the ratio of the number of parameters between the traditional convolution and Ghost convolution is

$$R_P = \frac{P_{conv}}{P_{ghost}} = \frac{C \times C' \times k \times k}{C \times m \times k \times k + m \times n \times d \times d} \approx \frac{C'}{m} \quad (4)$$

Through theoretical analysis, it is found that the ratio of the number of parameters obtained by traditional convolution and Ghost convolution is $\frac{C'}{m}$, and, as the number of eigenmap channels decreases, the number of parameters of Ghost convolution is less than that of traditional convolution. When the traditional convolution step is skipped and the Ghost feature map is generated directly by linear operation, the number of parameters reaches the minimum.

In this paper, ResNet50 is used as the main backbone network, and Ghost convolution is used to replace the traditional convolution, constituting the Ghost ResNet network structure. The composition of ResNet50 and Ghost ResNet is shown in Figure 4.

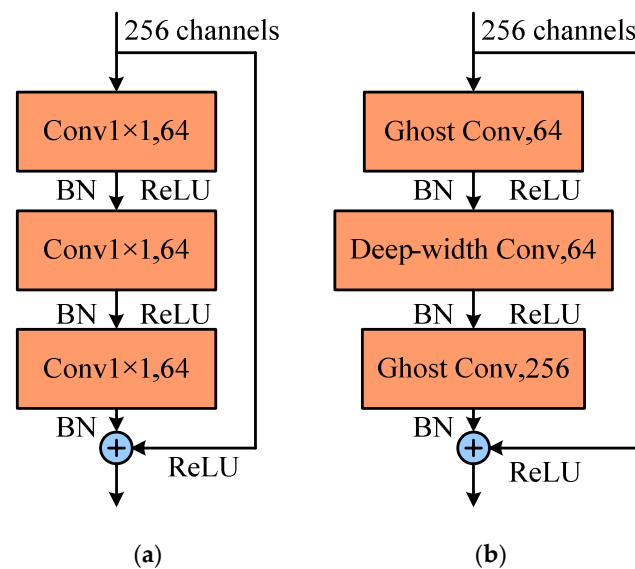


Figure 4. Components of ResNet50 and Ghost ResNet. (a) ResNet50 component module; (b) Ghost ResNet component module.

As shown in Figure 4, the structure of the two modules is similar. Ghost ResNet effectively combines the Ghost volume product and the depth volume product (DW-Conv) by adding a depth convolution between the two Ghost convolutions and reducing the size of the feature map to 1/2 of the input features. In this way, the down-sampling effect in ResNet50 synthesis is achieved.

2.2.2. Attention Mechanism

Due to the dihedral angle reflection effect, the buildings in SAR images mainly appear as L-shaped or as linear bright lines, which are distinctly different from the surrounding background. Therefore, in this paper, a focus mechanism is added to R-centernet to enhance the feature extraction capability of the network for building regions with strong salient features in the images.

The most representative channel attention module is the squeeze-and-excitation network (SENet). As a modified algorithm of SENet, the efficient channel attention mechanism (ECANet) can effectively reduce the computational effort while maintaining the ability to extract salient features of the network, making it more suitable for lightweight networks. The model structures of SENet and ECANet are shown in Figure 5. In Figure 5, σ is the sigmoid mapping and r is the dimensionality reduction ratio.

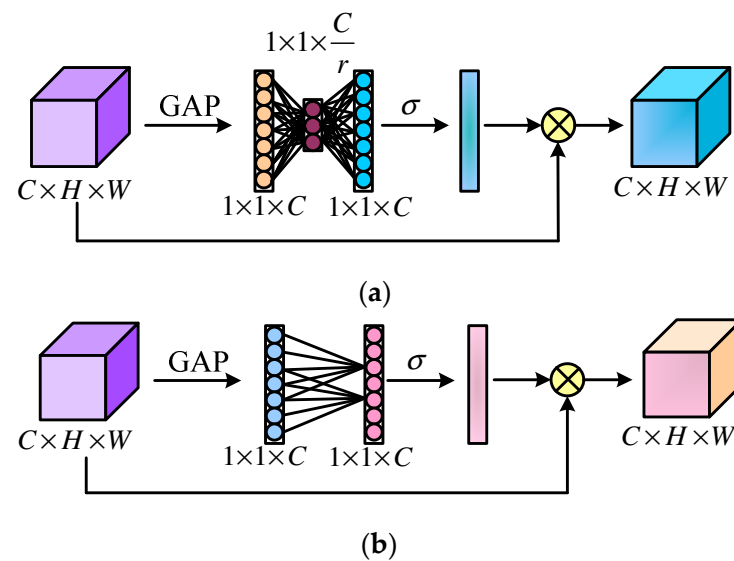


Figure 5. Structure diagram of SENet and ECANet. (a) SENet structure; (b) ECANet structure.

SENet first performs global average pooling (GAP) on the input image channels and then generates channel weights using a fully connected layer with a reduced-dimensional structure and a nonlinear sigmoid function. ECANet differs from SENet in that it uses a non-dimensionalized k-nearest neighbor operation rather than a fully connected layer to capture the relationships between different channels and regenerate channel weights. However, both consider only the relationships between channels and place emphasis on generating channel weights, ignoring the importance of significant target location information.

In this paper, we improved on the basis of ECANet, borrowed the idea of coordinated attention (CA), and proposed a channel attention mechanism (WH-ECA) that integrates width and height information. Its model structure is shown in Figure 6.

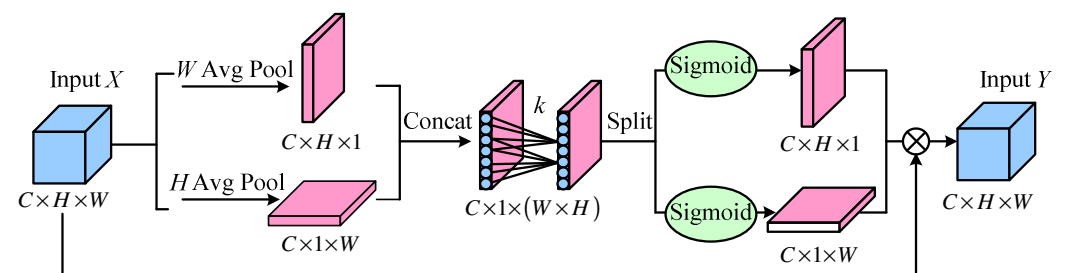


Figure 6. Structure of WH-ECA model.

As shown in Figure 6, the input image with size $C \times H \times W$, $x(i,j)$ is the input pixel value. Firstly, pooling kernels with size $(H,1)$ and size $(1,W)$ are used to perform average pooling operations on each channel along the width and height directions of the image, respectively, and then the output results are obtained z_w, z_h as shown in Equations (5) and (6), respectively:

$$z_w(\omega) = \frac{1}{H} \sum_{0 \leq i < H} x(i, \omega) \quad (5)$$

$$z_h(h) = \frac{1}{W} \sum_{0 \leq j < W} x(h, j) \quad (6)$$

When the input is $C \times W \times H$, the computation amount of the global pooling operation is

$$P_{GAP} = C \times W \times H \quad (7)$$

and the calculation amount of the pooling operation along the width and height directions is

$$P_{WH-P} = C \times W \times 1 + C \times 1 \times H \quad (8)$$

According to Equations (7) and (8), the ratio of the calculation amounts of the two methods is $(W \times H) / (W + H)$. Therefore, the directional pooling method adopted in this paper can not only realize the channel coding from width to height so as to determine the location information of the salient region but also occupy fewer parameters.

Tensor concatenation was performed on the two feature maps generated above, and then the k-nearest neighbor operation in ECANet was used to capture channel relations and re-encode each channel. The formula was as follows:

$$f = F_k[z_w, z_h] \quad (9)$$

where f represents the output of this part and F_k represents the k-nearest neighbor operation.

The above results were then decomposed into individual tensors along the spatial dimension f_w and f_h , and the sigmoid function was used to generate the weight of each channel in the direction of width and height, as shown in Equations (10) and (11):

$$g_w = \sigma(f_w) \quad (10)$$

$$g_h = \sigma(f_h) \quad (11)$$

In this case, the final output result of WH-ECA is shown in Equation (12):

$$y(i, j) = x(i, j) \times g_w(i) \times g_h(j) \quad (12)$$

where $y(i, j)$ is the output pixel value; $x(i, j)$ is the pixel value of the input; g_w and g_h is the weight of each channel in the direction of width and height.

In this paper, the feature maps obtained from the images after extracting the two-layer network were compared and analyzed, as shown in Figure 7. Figure 7a shows the input raw SAR image data, and the output of the L2 layer thermal map after extraction from the original R-centered network is shown in Figure 7b. The thermal features obtained after adding the ECANet attention module to the original network are shown in Figure 7c, and Figure 7d shows the output heat map after the fused WH-ECA attention module.

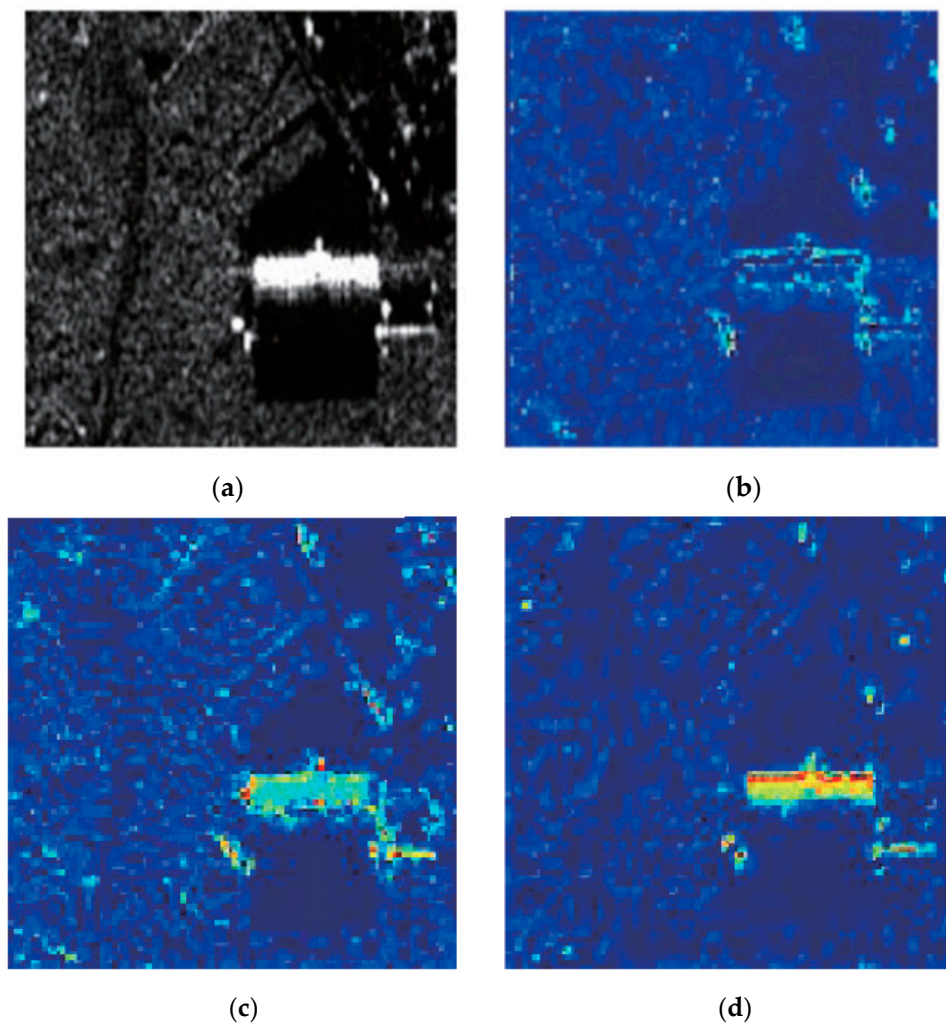


Figure 7. Thermal diagram after feature extraction. (a) Raw SAR image; (b) L2 layer thermal map; (c) Thermal map of fused ECANet; (d) Thermal map of fused WH-ECA.

In the heat map, the red color represents the region of high importance. After comparative analysis, it was found that the difference between the target and the surrounding background is not significant in the heat map of the L2 layer. Although the heat map obtained by integrating ECANet shows the importance of the target, there are some problems, such as the blurred target boundary, which is not conducive to the precise location of the target. However, the background information in the heat map obtained by integrating the WH-ECA attention module is suppressed, and the target features are highlighted. In addition, the clear boundaries of the targets on the map facilitate locating the coordinates of the center point and return accurate width and height information.

The WH-ECA proposed in this paper improves the global pool in the original ECANet into an average pool in the width and height directions, respectively, with the following three advantages: (1) it can effectively analyze the relationships between channels and essentially acts as a channel focus; (2) it uses the width and height information of the feature map to accurately locate important regions; and (3) WH-ECA is simple and efficient and retains a small number of parameters, making it suitable for lightweight networks.

2.2.3. Up-Sampling Improvement

R-centernet uses ResNet50 to extract features from images and then feeds the feature images to the deconvolution (DCN) module for up-sampling. Deconvolution is the inverse process of convolution. After learning the parameters in the network, it inserts the pixel values into the feature map for up-sampling purposes. Inverse convolution has the follow-

ing drawbacks: (1) it uses the same up-sampling kernel for each position of the feature map, which does not capture the feature map information; (2) a large number of parameters are introduced, thus increasing the computational effort.

In order to effectively solve the above problems, this paper uses CARAFE attribute-aware feature recombination up-sampling (whose structure is shown in Figure 8) instead of an inverse fold product. The CARAFE up-sampling process is mainly divided into two parts: kernel prediction and feature recombination.

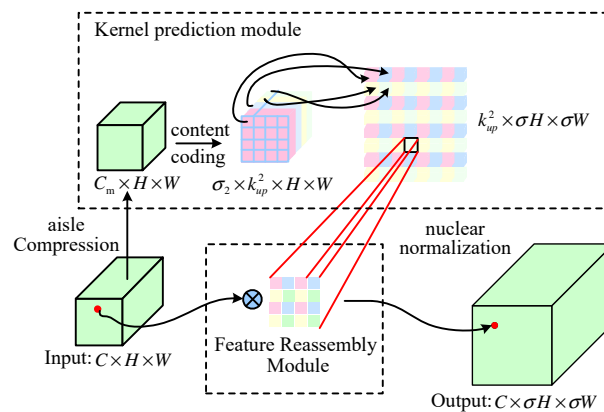


Figure 8. CARAFE up-sampling structure.

The algorithmic process of CARAFE up-sampling is as follows:

Step 1: Channel compression is performed on the input feature map, whose size is $C \times H \times W$, and the result, whose size is $C_m \times H \times W$, is obtained. This step reduces the subsequent computation. C_m is the number of channels after channel compression.

Step 2: The usage size is $k_{en} \times k_{en}$ ($k_{en} \times k_{en}$ convolution kernel size encoded for content); the convolution is obtained by the content encoding of the above compressed feature map $\sigma^2 \times k_{up}^2$ ($k_{up} \times k_{up}$ is the predicted up-sampled kernel size) $\times H \times W$ and then expanded in the channel dimension, where the size becomes $k_{up}^2 \times \sigma H \times \sigma W$.

Step 3: Use the Softmax function for normalization so that the sum of the weights of the up-sampling kernels is 1.

Step 4: Convolve the input feature map with the predicted up-sampling kernel to obtain the final up-sampling result. The number of parameters in the CARAFE up-sampling process is shown in Equation (13):

$$P = C \times C_m + \sigma^2 k_{up}^2 (\sigma^2 k_{en}^2 C_m + 1) + \sigma^2 k_{up}^2 C \quad (13)$$

3. Methods

3.1. Datasets

In this paper, based on the SBD (SAR Building Dataset) [24], rotated rectangular boxes were used for relabeling. After filtering, the dataset contained 1087 images of 416×416 pixels and 512×512 pixels, including 12,001 buildings. The data sources mainly consist of synthetic aperture radar (SAR) images acquired from airborne and satellite-based platforms, such as TerraSAR [25], the Gaofen-3 satellite [26] and Sandia National Laboratories. The signal bands include X-band, C-band and Ku-band; the image resolution is 0.5–5 m; the polarization modes include HH, HV, VH and VV. The datasets were randomly assigned as training and test sets according to a ratio of 8:2. The data volume of the dataset contained 25,000 building area SAR sample slices and 25,000 building area labeled image slices, with a total data volume of 2.01 GB. The size of the images was 256×256 pixels, and the primary source of the original SAR image imaging time for the dataset acquisition was 2019.

3.2. Model Training

In this paper, R-centernet was used as the base algorithm, and Ghost convolution, the WH-ECA attention module and CARAFE up-sampling were used to improve, train and test it, respectively. In this paper, the loss function values (average loss) of each algorithm during training were recorded, and the change curves of the loss functions were plotted, as shown in Figure 9. From Figure 9, it can be seen that the loss values of the original algorithm and the three improved algorithms can converge with an increase in the number of iterations. Among them, the algorithm combining Ghost convolution, the WH-ECA attention module and CARAFE up-sampling decreases the fastest and has the least loss after convergence.

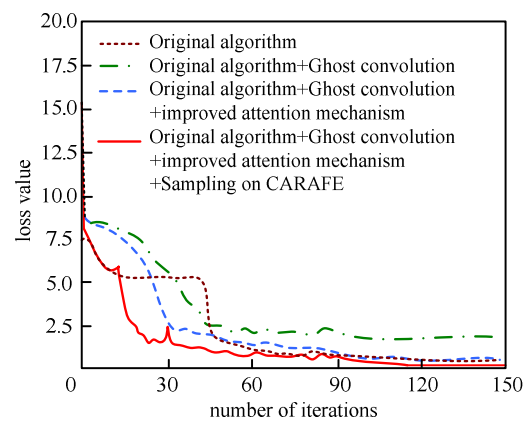


Figure 9. Change of loss function curve.

4. Analysis of Test Results

To quantify the detection performance of each improved algorithm for constructing targets in SAR images, this paper uses precision accuracy, recall and F1 score for quantitative analysis during testing, defined as follows.

$$P = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (14)$$

$$R = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (15)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (16)$$

where P is the accuracy rate; R is recall rate; F_1 represents the F_1 score, and the higher the F_1 score, the more balanced the accuracy and recall, and the better the detection effect; n_{TP} is the number of targets for which the truth box is a positive example and the prediction box is a positive example; n_{FP} is the number of targets for which the truth box is a negative example and the prediction box is a positive example; n_{FN} is the number of targets for which the truth box is a negative example and the prediction box is a negative example. At the same time, the number of parameters and floating points and the detection speed are introduced to measure the size of the network model.

The experimental results of each detection algorithm [27,28] are shown in Table 1. As can be seen from the table, when the backbone network is replaced by Ghost ResNet, the number of parameters and floating points decrease, and the F_1 score also decreases, which proves that Ghost convolution reduces the network burden but affects the algorithm's ability to detect SAR buildings. WH-ECA, a lightweight attention module, increases the F_1 score by 3.5% with only a small increase in weight. Using CARAFE up-sampling instead of the DCN module reduces the number of parameters and increases the F_1 score by 1. Finally, by comparing our experiments with the lightweight networks SqueezeNet and

Yolov8s, using SqueezeNet and Yolov8s parameters and FLOPs directly, we found that our algorithm worked better than SqueezeNet, while Yolov8s showed overfitting on our dataset, and therefore is not suitable to be applied in the method of this paper.

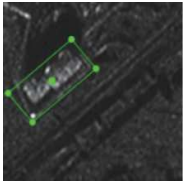
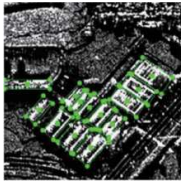
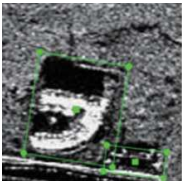
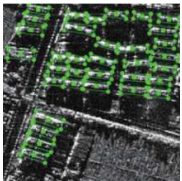
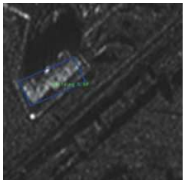
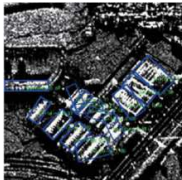
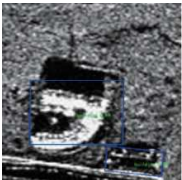
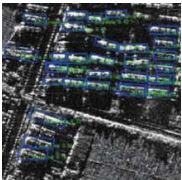
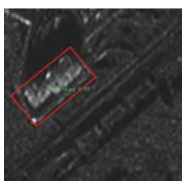
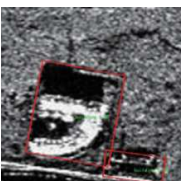
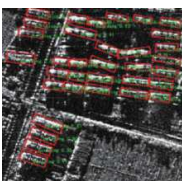
Table 1. Comparison of experimental results.

Number	Detection Algorithm	Backbone	Accuracy/%	Recall/%	F ₁ Score	Parameter/10 ⁶	FLOPs/10 ⁶	FPS
1	R-centernet	ResNet50	85.8	80.6	82.3	36.4	234.9	32.2
2	R-centernet	SqueezeNet [29]	80.0	52.3	71.5	33.6	236.9	45.0
3	R-centernet	Ghost ResNet	85.5	76.2	80.6	30.3	209.7	39.0
4	R-centernet + WH-ECA	Ghost ResNet	87.4	81.0	84.1	31.6	214.4	37.8
5	R3Det	MobileNetV2-FPN	83.2	79.9	81.5	35.7	222.4	28.5
6	Yolov8s	CSPDarkNet	32.8	52.9	62.6	11.2	28.6	68.8
7	Proposed method	Ghost ResNet	89.6	81.8	85.5	28.1	200.1	44.2

As shown in the fourth row of Table 1, this paper uses a single-stage rotation detector and R3Det to detect architectural datasets of SAR images. Using the lightweight MobileNetV2 FPN as the backbone network, the test results showed an accuracy of 83.2%, a recall of 79.9%, an F1 score of 81.5 and a detection speed of 28.5 frames/second. Because R3Det adds the boundary optimization module to the prediction process, the detection speed is slower despite the smaller number of network parameters and floating points in the detector.

In conclusion, the light SAR building detection algorithm proposed in this paper achieves 89.6% accuracy and 81.8% recall with a detection speed of 44.2 frames/sec. Compared with the original R-centernet and R3Det algorithms, the accuracy, recall and detection speed are significantly improved. Table 2 shows the detection results of some images in the test set. In order to reflect the detection performance of the algorithm for building targets in different scenes, this paper selects the following four arrangements of buildings for comparative experiments: independent buildings, buildings in complex scenes, buildings with special shapes and buildings in dense arrangements. As can be seen from the detection results in Table 2, although the original algorithm can detect the target, the location of the target boundary is relatively blurred, resulting in a large angular deviation of the rectangular box, which cannot completely surround the target. However, the algorithm in this paper adds a channel attention module that integrates width and height information and averages pooling in the width and height directions of the image, which can accurately capture the location information and boundary information of the target and can accurately regress the width and height values and the rotation angle of the target so that the rectangular box in the detection result can accurately surround the target. The original algorithm fails to detect and incorrectly detects images of densely arranged buildings because the DCN module in the original algorithm uses the same core for each location during up-sampling and does not consider the information in the feature map. The improved algorithm in this paper uses CARAFE up-sampling instead of the DCN module, combined with generating different up-sampling cores for different locations, to fully capture the feature map information. In the process of improving the resolution, no information is missed, and false detection and missed detection occurrences can be significantly reduced. The results show that the algorithm is able to detect irregularly shaped buildings in SAR images, and the rotation angle is more accurate than in the original algorithm. Since our data come from real-life data, its application will have good practical value in displaying real-life situations.

Table 2. Comparison of SAR Image Building Test Results.

Method	Independent Buildings	Complex Buildings	Special Buildings	Dense Buildings
tagging				
R-centernet				
our				

5. Discussion and Conclusions

In this paper, a lightweight R-centernet algorithm is proposed to resolve the problem of a convolutional neural network that has many parameters and occupies a lot of computational resources, and it is applied to the field of building detection in Minnan SAR civil light images. The main conclusions are as follows: (1) Ghost convolution is used instead of traditional convolution in the backbone network to form a new network; Ghost ResNet greatly reduces the number of parameters of the model and improves the detection efficiency at the expense of detection accuracy. (2) The proposed channel attention module that fuses width and height information can better capture the spatial information of important regions in the image, which is beneficial to accurately locating target locations and improves detection accuracy while adding only a small number of network parameters; (3) CARAFE up-sampling is used to replace the DCN module and generate different up-sampling kernels for different locations during the up-sampling process to fully integrate feature map information and enhance the feature extraction capability of the network. Compared with the DCN module, CARAFE up-sampling introduces fewer parameters and has a lower network burden. (4) The detection results in the rotating annotated SAR light image construction dataset show that, compared with the original algorithm, the improved algorithm proposed in this paper improves the detection accuracy and detection speed of Minnan folk light animation and proves the feasibility of a lightweight network for building detection directions in SAR images. In the future, the use of more advanced algorithms for the extraction of features from specific parts of the building will be a worthwhile area of research and could include hybrid heuristics [30], metaheuristics, adaptive algorithms [31], self-adaptive algorithms [32] and island algorithms. We also found that there are many different fields that are applying advanced algorithms to resolve their respective problems, such as online learning [33], scheduling [34], multi-objective optimization [35], transportation, medicine [36], data classification, etc. Meanwhile, as SAR technology continues to mature, the use of SAR technology to achieve delicate object segmentation and detection is also a very valuable area of research.

Author Contributions: Writing—original draft, S.W. (Sichao Wu); Writing—review & editing S.W. (Sichao Wu), S.W. (Shengzhen Wu), E.L. and C.P.; Methodology, S.W. (Sichao Wu) and C.P.; Data curation, X.H.; Funding acquisition, X.H.; Investigation, Y.X.; Project administration, Y.X. and E.L.;

Formal analysis, S.W. (Shengzhen Wu); Software, S.W. (Shengzhen Wu); Visualization, E.L. and C.P. All authors have read and agreed to the published version of the manuscript.

Funding: Fujian Provincial Federation of Social Sciences (FJ2021B195).

Data Availability Statement: We declare that the data is unavailable due to privacy or ethical restrictions.

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

References

1. Poulain, V.; Inglada, J.; Spigai, M.; Tournet, J.-Y.; Marthon, P. High-resolution optical and SAR image fusion for building database updating. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2900–2910. [\[CrossRef\]](#)
2. Alshboul, O.; Shehadeh, A.; Almasabha, G.; Mamlook, R.E.A.; Almuflih, A.S. Evaluating the impact of external support on green building construction cost: A hybrid mathematical and machine learning prediction approach. *Buildings* **2022**, *12*, 1256. [\[CrossRef\]](#)
3. Bui, N.; Merschbrock, C.; Munkvold, B.E. A review of Building Information Modelling for construction in developing countries. *Procedia Eng.* **2016**, *164*, 487–494. [\[CrossRef\]](#)
4. Lippitt, C.D.; Zhang, S. The impact of small unmanned airborne platforms on passive optical remote sensing: A conceptual perspective. *Int. J. Remote Sens.* **2018**, *39*, 4852–4868. [\[CrossRef\]](#)
5. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
6. Jiang, D.; Li, G.; Tan, C.; Huang, L.; Sun, Y.; Kong, J. Semantic segmentation for multiscale target based on object recognition using the improved Faster-RCNN model. *Future Gener. Comput. Syst.* **2021**, *123*, 94–104. [\[CrossRef\]](#)
7. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [\[CrossRef\]](#)
8. Zhai, S.; Shang, D.; Wang, S.; Dong, S. DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion. *IEEE Access* **2020**, *8*, 24344–24357. [\[CrossRef\]](#)
9. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
10. Zhu, M.; Wan, X.; Fei, B.; Qiao, Z.; Ge, C.; Minati, F.; Vecchioli, F.; Li, J.; Costantini, M. Detection of building and infrastructure instabilities by automatic spatiotemporal analysis of satellite SAR interferometry measurements. *Remote Sens.* **2018**, *10*, 1816. [\[CrossRef\]](#)
11. Yang, C.; Zhou, X.; Zhu, W.; Xiang, D.; Chen, Z.; Yuan, J.; Chen, X.; Shi, F. Multi-discriminator adversarial convolutional network for nerve fiber segmentation in confocal corneal microscopy images. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 648–659. [\[CrossRef\]](#)
12. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
13. An, Q.; Pan, Z.; Liu, L.; You, H. DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8333–8349. [\[CrossRef\]](#)
14. Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; Fu, K. Srdet: Towards more robust detection for small, cluttered and rotated objects. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8232–8241.
15. Jiang, Y.; Li, W.; Liu, L. R-CenterNet+: Anchor-Free Detector for Ship Detection in SAR Images. *Sensors* **2021**, *21*, 5693. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Yang, X.; Yan, J.; Feng, Z.; He, T. R3det: Refined single-stage detector with feature refinement for rotating object. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, No. 4. pp. 3163–3171.
17. Zhou, X.; Yao, C.; Wen, H.; Wang, Y.; Zhou, S.; He, W.; Liang, J. East: An efficient and accurate scene text detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5551–5560.
18. Alam, M.S.; Kim, I.J.; Ling, Z.; Mahmood, A.H.; O’Neill, J.J.; Severini, H.; Sun, C.R.; Wappler, F.; Crawford, G.; Daubenmier, C.M.; et al. First measurement of the rate for the inclusive radiative penguin decay $b \rightarrow s \gamma$. *Phys. Rev. Lett.* **1995**, *74*, 2885. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Sinha, D.; El-Sharkawy, M. Thin mobilenet: An enhanced mobilenet architecture. In Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 10–12 October 2019; pp. 0280–0285.
20. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
21. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
22. Thekedath, D.; Sedamkar, R.R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **2020**, *1*, 79. [\[CrossRef\]](#)

23. Briggs, J. The ghost story. In *A New Companion to the Gothic*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2012; Volume 2, pp. 176–185.
24. Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; Xu, F. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inf. Sci.* **2020**, *63*, 140303. [[CrossRef](#)]
25. Prats-Iraola, P.; Scheiber, R.; Marotti, L.; Wollstadt, S.; Reigber, A. TOPS interferometry with TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3179–3188. [[CrossRef](#)]
26. Dong, H.; Xu, X.; Wang, L.; Pu, F. Gaofen-3 PolSAR image classification via XGBoost and polarimetric spatial information. *Sensors* **2018**, *18*, 611. [[CrossRef](#)]
27. Yuan, J.; Ding, X.; Liu, F.; Cai, X. Disaster classification net: A disaster classification algorithm on remote sensing imagery. *Front. Environ. Sci.* **2023**, *10*, 2690. [[CrossRef](#)]
28. An, D.; Chen, L.; Zhou, Z. Clustering Algorithm Improvement in SAR Target Detection. *IEEE Access* **2019**, *7*, 113398–113403. [[CrossRef](#)]
29. Huang, Q. Weight-quantized squeezenet for resource-constrained robot vacuums for indoor obstacle classification. *AI* **2022**, *3*, 180–193. [[CrossRef](#)]
30. Tsao, Y.C.; Delicia, M.; Vu, T.L. Marker planning problem in the apparel industry: Hybrid pso-based heuristics. *Appl. Soft Comput.* **2022**, *123*, 108928. [[CrossRef](#)]
31. Luo, Z.; Zhou, J.; Pu, Y.F.; Li, L. A class of augmented complex-value FLANN adaptive algorithms for nonlinear systems. *Neurocomputing* **2023**, *520*, 331–341. [[CrossRef](#)]
32. Kavoosi, M.; Dulebenets, M.A.; Abioye, O.F.; Pasha, J.; Wang, H.; Chi, H. An augmented self-adaptive parameter control in evolutionary computation: A case study for the berth scheduling problem. *Adv. Eng. Inform.* **2019**, *42*, 100972. [[CrossRef](#)]
33. Zhao, H. An online-learning-based evolutionary many-objective algorithm. *Inf. Sci. Int. J.* **2020**, *509*, 1–21. [[CrossRef](#)]
34. Dulebenets, M.A. An adaptive polyploid memetic algorithm for scheduling trucks at a cross-docking terminal. *Inf. Sci.* **2021**, *565*, 390–421. [[CrossRef](#)]
35. Pasha, J.; Nwodu, A.L.; Fathollahi-Fard, A.M.; Tian, G.; Li, Z.; Wang, H.; Dulebenets, M.A. Exact and metaheuristic algorithms for the vehicle routing problem with a factory-in-a-box in multi-objective settings. *Adv. Eng. Inform.* **2022**, *52*, 101623. [[CrossRef](#)]
36. Rabbani, M.; Oladzad-Abbasabady, N.; Akbarian-Saravi, N. Ambulance routing in disaster response considering variable patient condition: Nsga-ii and mopso algorithms. *J. Ind. Manag. Optim.* **2017**, *13*, 1035. [[CrossRef](#)]

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.