

Article

Lightweight Multilevel Feature-Fusion Network for Built-Up Area Mapping from Gaofen-2 Satellite Images

Yixiang Chen ^{1,2}, Feifei Peng ^{2,*}, Shuai Yao ² and Yuxin Xie ²

¹ Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, China; cheniyixiang@njupt.edu.cn

² School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; yaoshuai14@huawei.com (S.Y.); 1021173507@njupt.edu.cn (Y.X.)

* Correspondence: 1022173204@njupt.edu.cn

Abstract: The timely, accurate acquisition of geographic spatial information such as the location, scope, and distribution of built-up areas is of great importance for urban planning, management, and decision-making. Due to the diversity of target features and the complexity of spatial layouts, the large-scale mapping of urban built-up areas using high-resolution (HR) satellite imagery still faces considerable challenges. To address this issue, this study adopted a block-based processing strategy and constructed a lightweight multilevel feature-fusion (FF) convolutional neural network for the feature representation and discrimination of built-up areas in HR images. The proposed network consists of three feature extraction modules composed of lightweight convolutions to extract features at different levels, which are further fused sequentially through two attention-based FF modules. Furthermore, to improve the problem of incorrect discrimination and severe jagged boundaries caused by block-based processing, a majority voting method based on a grid offset is adopted to achieve a refined extraction of built-up areas. The effectiveness of this method is evaluated using Gaofen-2 satellite image data covering Shenzhen, China. Compared with several state-of-the-art algorithms for detecting built-up areas, the proposed method achieves a higher detection accuracy and preserves better shape integrity and boundary smoothness in the extracted results.

Keywords: built-up area; high resolution; satellite image; CNN



Citation: Chen, Y.; Peng, F.; Yao, S.; Xie, Y. Lightweight Multilevel Feature-Fusion Network for Built-Up Area Mapping from Gaofen-2 Satellite Images. *Remote Sens.* **2024**, *16*, 716. <https://doi.org/10.3390/rs16040716>

Academic Editor: Amin Beiranvand Pour

Received: 18 January 2024
Revised: 12 February 2024
Accepted: 15 February 2024
Published: 18 February 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

Built-up areas are the gathering place for human activities. The timely, accurate access to the spatial distribution information of built-up areas can provide indispensable reference and assistance for applications, such as urban planning, construction, management, decision-making, and research [1–3]. The wide availability of high-resolution (HR) satellite data enables the fine-scale mapping of built-up areas [4]. Traditional built-up-area extraction methods mainly use artificially designed algorithms to extract the spectrum, texture, and local features of an image and then identify the built-up area by thresholding the built-up index/saliency map [5–10] or by using a supervised classifier [11]. Built-up areas are a composite target, covering a large geographical range, and their image features have an extremely large spatial heterogeneity that makes the design of a feature extraction (FE) algorithm with strong adaptability and robustness very difficult [4,12]. Thus, these methods can achieve good detection results for low- and medium-resolution satellite images or HR images of simple scenes, but their detection performance is often poor when they are applied to the HR images of large-scale complex scenes [13].

The development of deep learning has brought new opportunities for the recognition of built-up areas in remote sensing images [14,15]. In the field of computer vision and pattern recognition, various deep convolutional neural network (CNN) models, such as VGG-Net [16], GoogleNet [17], ResNet [18], and DenseNet [19], have been developed. However, these networks were originally designed for the multiclassification problem of

ordinary images (e.g., 1000 classes in the ImageNet dataset). The extraction of built-up areas from satellite images is essentially a binary classification problem of built-up areas/non-built-up areas, requiring targeted networks to achieve efficient processing and accurate mapping of built-up areas. Under the paradigm of deep learning, two processing strategies have been developed for extracting built-up areas. The first approach is to consider it as a segmentation problem [20], and semantic segmentation networks (such as FCN [21] and U-Net [22]) are used to achieve the pixel-level category labeling of images [23,24]. Wu et al. [25] constructed a network of the U-Net structure for the semantic segmentation of built-up areas in GF-3 SAR images with a 10 m resolution. Li et al. [26] proposed a dual-attention-based transformer model for built-up area extraction, and the GF-3 and Sentinel-1 SAR image datasets demonstrated the effectiveness of this method in large-scale built-up area mapping. To perform pixel-level dense prediction, semantic segmentation networks based on deep learning require pixel-level labeled data. Accurately creating a sufficient number of such datasets is often time-consuming, laborious, and even difficult, especially in complex scenes [27,28]. To alleviate the pressure of these methods on sample requirements, weakly supervised and unsupervised domain adaptive algorithms have recently been studied to improve the mapping of built-up areas in optical and SAR images [29,30]. In addition to the complexity of collecting labeled samples, pixel-based prediction can lead to high computational costs and salt-and-pepper noises in mapping results, especially for large-scale HR images.

The other approach is to consider it as a scene-based classification problem [31], where CNNs are used to distinguish built-up and non-built-up areas through patch- or object-level classification. Mboga et al. [32] proposed using CNNs to detect informal settlements from VHR images, but they used sliding windows to learn the context information and category labels of each central pixel, resulting in a low processing efficiency. Corbane et al. [33] designed a CNN named GHS-S2Net that contains only four convolutional layers and two flattened layers for the large-scale built-up area mapping of Sentinel-2 images. However, the model uses a patch-based method to label the central pixel (similar to pixel-wise classification). For HR images, a large window size for each pixel is required, which would lead to numerous redundant calculations. Huang et al. [34] combined deep learning with object-oriented methods to extract impervious surfaces in HR satellite images. The spectrum, shape, and CNN features of each segmented object are jointly used to determine its category. However, due to the effect of image segmentation quality, expressing HR complex scenes is difficult for image objects. Recently, block-based deep learning methods have been applied to extract built-up areas from HR images [12,35–37]. An image block typically contains multiple objects and their spatial distribution patterns. An image block, which is a basic processing unit, has strong feature representation and discrimination capabilities, making it very suitable for the mapping of built-up areas in HR images of large-scale, complex scenes [35]. Particularly, creating block-level labeled samples is easier and takes less time than pixel-level-labeled samples. Nevertheless, scene-classification-based methods still face substantial challenges in extracting built-up areas from large-scale HR and VHR images due to the following main reasons:

- (1) With the improvement of spatial resolution and the increase in geographical ranges, the scene blocks as basic units contain richer object details, have a more complex spatial layout, and exhibit remarkable scene heterogeneity in different geographical locations. This approach requires the model to have a strong discriminative ability and adaptability.
- (2) The increased spatial resolution means that larger images are processed. For example, the size of a 1 m resolution image is 100 times that of a 10 m resolution image covering the same area, which greatly increases the computational complexity of the model. Therefore, a lightweight deep learning model would be a better choice.
- (3) Block-based discrimination often overlooks the spatial relationships between blocks, leading to incorrect discrimination and severe jagged boundaries in the extracted results. This method requires the recognition model to consider the contextual information of the block and obtain a complete built-up area target at the pixel level solely through block-level processing.

To address the above issues effectively, this study designed a simple but highly effective model for identifying built-up areas in large-scale HR satellite images. The main contributions are summarized as follows:

(1) A lightweight multilevel feature-fusion convolutional neural network (LMLFF-CNN), which utilizes three FE modules composed of lightweight convolutions, is designed to extract features at different levels. Two attention-based feature-fusion (FF) modules are sequentially used to fuse features at different levels. The use of this network can effectively distinguish between image blocks in built-up and non-built-up areas, and it has a lower computational consumption.

(2) A block-level framework for extracting built-up areas considering contextual information is proposed. This framework uses a set of offset grids to partition images and obtain spatially overlapping image blocks. By integrating classification labels of multiple contextual blocks, a pixel-level mapping of built-up areas is achieved.

(3) Based on the Gaofen-2 satellite images, a block-level sample set of built-up and non-built-up areas, which will be publicly available online with the publication of the paper, is constructed. To our knowledge, proprietary datasets suitable for identifying built-up areas using HR satellite imagery are currently lacking. This paper will be a beneficial supplement to publicly available remote sensing datasets.

(4) The proposed method is used to extract built-up areas from Gaofen-2 satellite images of the entire Shenzhen City. A 1 m resolution distribution map of built-up areas is obtained, demonstrating the potential and advantages of the proposed method in the large-scale, HR mapping of built-up areas.

2. Methods

Built-up areas are large-scale artificial geographic objects that present complex, diverse scenes in HR satellite images. Thus, pixel-based or object-oriented processing is not conducive to the feature representation of built-up areas. In this study, a block-based processing strategy was adopted, and scene classification methods were utilized to achieve the feature representation and discrimination of built-up areas. Figure 1 shows the overall workflow of this method, which mainly consists of three key components: first, the input image is divided into image blocks with a certain overlap ratio through a set of multidirectional, multistep shifting grids. Then, an LMLFF-CNN model is constructed to achieve a binary classification of image blocks. Finally, the refined pixel-level built-up area extraction results are obtained by integrating multiple preliminary prediction maps based on block classification.

2.1. Image Partitioning Using Multi-Directional and Multi-Step Offset Grids

Block-based extraction strategies typically use a regular grid to partition images to produce non-overlapping image blocks. This partitioning method may split the spatial context of the target, resulting in incorrect discrimination and severe jagged boundaries in the extraction results, especially when the block size is large. To alleviate this problem, the input image is divided several times through a set of multidirectional, multistep offset grids to generate spatially overlapping image blocks, which are then input into the trained LMLFF-Net model to determine their category labels.

To illustrate this process intuitively, assuming that the input image is divided into a series of image blocks of size $L \times L$ using a predefined regular grid, Figure 2 shows the result of moving the original grid to the right and down by $L/2$. The size of the generated image block is reduced to $1/4$ of the original size. Each newly generated block is covered by three original image blocks, which provide additional contextual information for that block. The final category of the block can be obtained by integrating the contextual labels of the original blocks that cover it. In this way, compared with directly dividing into smaller blocks for classification, the accuracy and computational efficiency of built-up area mapping can be greatly improved, and the jagged boundaries can be substantially refined.

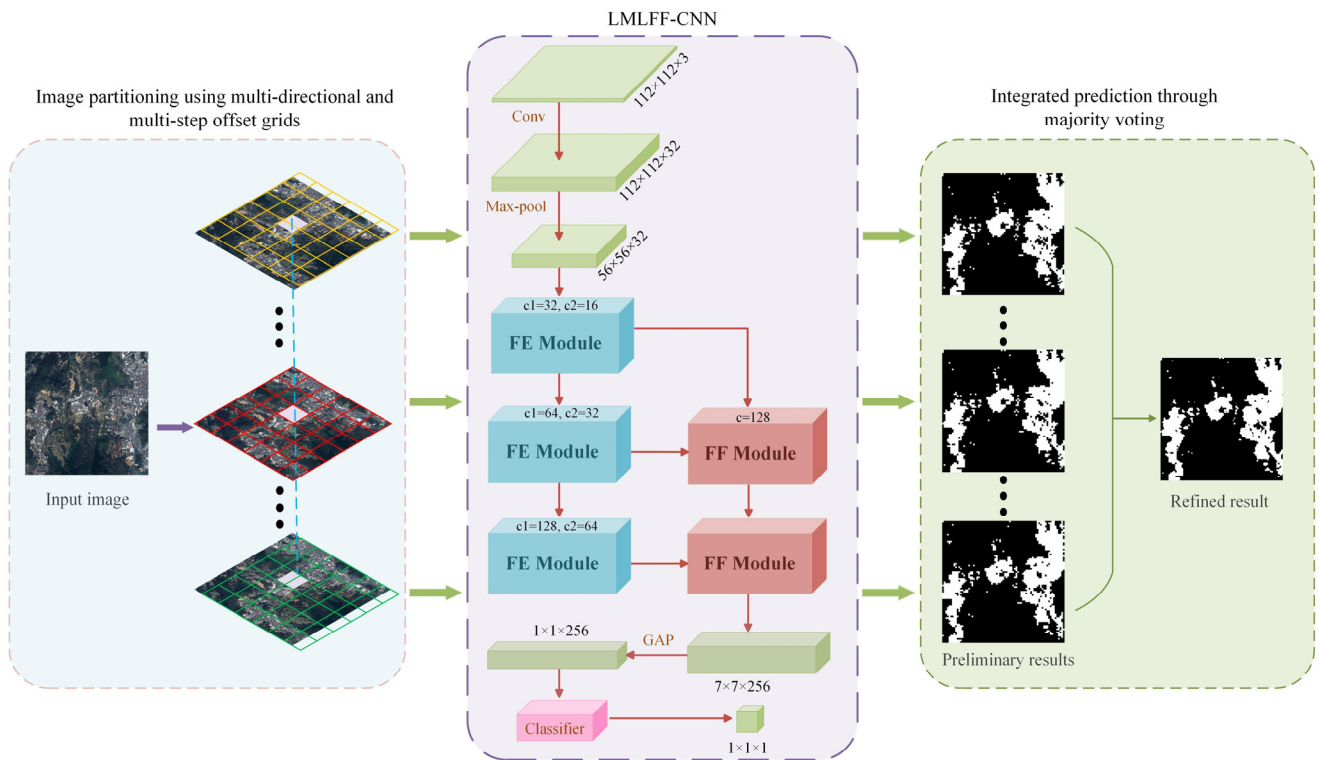


Figure 1. Flowchart of the proposed method.

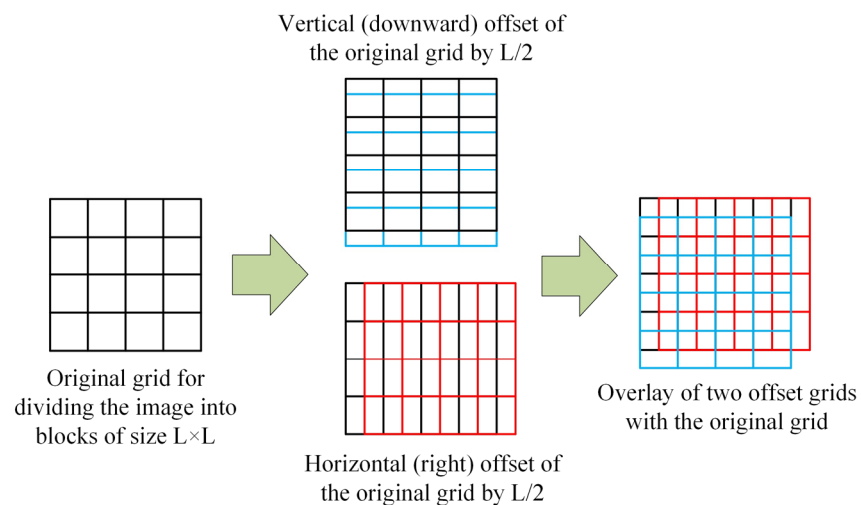


Figure 2. Example of grid offsets and their integration.

For this overlapping partitioning strategy, the direction and step size of the grid offset are two key parameters that determine the position and ratio of overlap between image blocks, respectively. As described in Section 4.2, the influence of these two parameters on the extraction results was explored in detail through experiments.

2.2. LMLFF-CNN Model

In this study, a CNN-based method was used to obtain the category labels for each image block. Instead of using existing CNN models, an LMLFF-CNN model was constructed to characterize the distinctive features of built-up areas in HR images more effectively, thereby achieving more robust and efficient discrimination. In the design process of the model, considering that the discrimination between built and non-built blocks is a binary classification problem, a lightweight network is a more ideal choice. It can reduce the

parameter size of the model, thereby lowering the requirement for sample size and the consumption of computational costs for large-scale image processing. Furthermore, to improve the discriminant performance of the model, the multiscale features of the built-up areas, which play a crucial role in their correct recognition, are fully utilized. Therefore, unlike commonly used CNN models that mainly utilize high-level abstract features from the end outputs of the network, a multiscale FF module is designed to fuse the low-level details and high-level semantic features of built-up areas, which can provide valuable information for distinguishing built-up areas.

The proposed LMLFF-CNN has a lightweight network architecture with several typical characteristics: (1) the structure is simple, and its key components only include three FE modules and two attention-based FF modules. (2) The FE module adopts a dual-branch structure to reduce the depth of the network, and each branch uses a depthwise separable convolution (DSC) and 1×1 convolution to reduce the number of parameters and computational burden of the network. (3) At the end of the network, instead of the commonly used fully connected layer, a global average pooling layer is adopted to reduce the number of parameters and computational costs further.

Specifically, Figure 1 shows that the input image first passes through a 3×3 convolutional layer and a 2×2 maximum pooling layer to extract the initial low-level features and then enters the FE and FF modules.

(1) Feature Extraction Module

Each FE module (Figure 3) contains two branches composed of different convolutional layers to obtain features of different scales and types. The final output feature maps of the two branches are concatenated and then enter a 2×2 max-pooling layer. The lower branch only contains a 1×1 convolutional layer that is used to undertake the low-level features of the previous convolutional module/layer. The upper branch consists of three (including one 1×1 and two 3×3) DSC [38] layers that have fewer parameters but a higher computational efficiency than the standard convolution. Here, the 1×1 DSC layer is used to compress the number of channels of the feature and reduce the amount of calculation for subsequent operations. The first 3×3 DSC layer is used to obtain the initial semantic information, whereas the second 3×3 DSC layer is used to enhance and obtain higher-level semantic features. In addition, all convolutional and DSC layers are followed by batch normalization and a rectified linear unit.

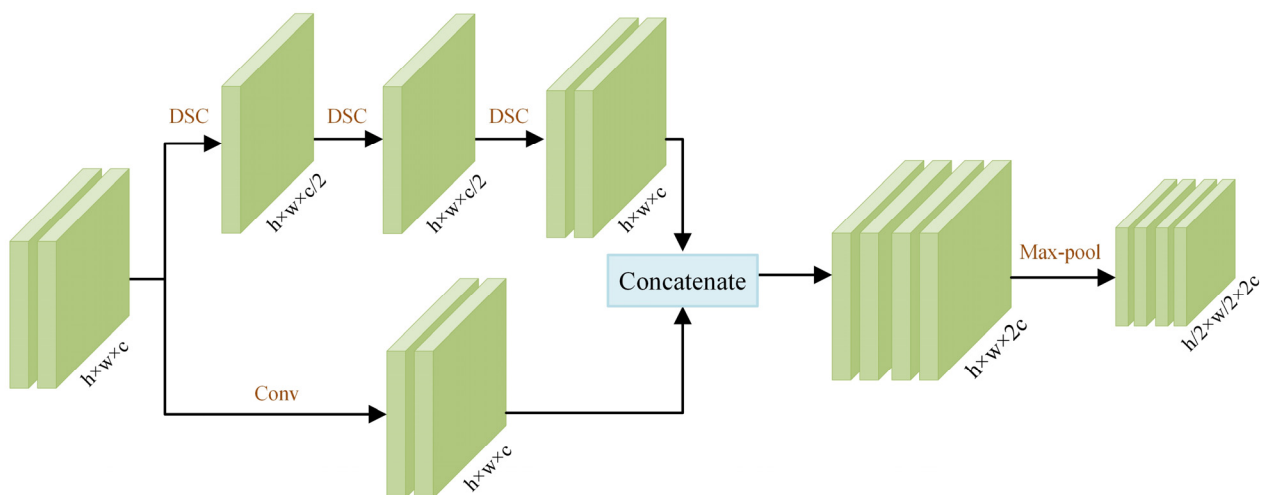


Figure 3. Feature extraction module.

(2) Multilevel Feature-Fusion Module

Existing deep-learning-based methods mainly use the high-level semantic features output by the full connection (FC) layer at the end of CNNs [39]. However, the built-up

area has multiscale characteristics. In addition to the high-level semantic information, the structure and details of the middle and low levels play a very important role in the discrimination of built-up areas. In our network, the three FE modules above output different levels of features that can provide complementary information for the discrimination of built-up areas. To make full use of these features, two FF modules are designed to integrate this information with different degrees of abstraction. FF is conducted module by module. The structure of the FF module is shown in Figure 4. Its design is inspired by the human visual attention mechanism, which enables the human visual system to focus on useful information in the scene consciously while suppressing unnecessary information. The attention-based FF module can achieve multilevel FF of different channels. Specifically, the input low-level feature maps (F_{low}) is first transformed to generate a weight vector that is then fused with the input high-level feature maps (F_{high}) by multiplication. This process can be expressed as follows:

$$F_{fused} = F_{high} \times T(F_{low}), \quad (1)$$

where F_{fused} is the fused feature maps; T is a composite transformation, including four consecutive operations: a 1×1 convolution, a global max-pooling, and an FC, followed by a sigmoid activation function.

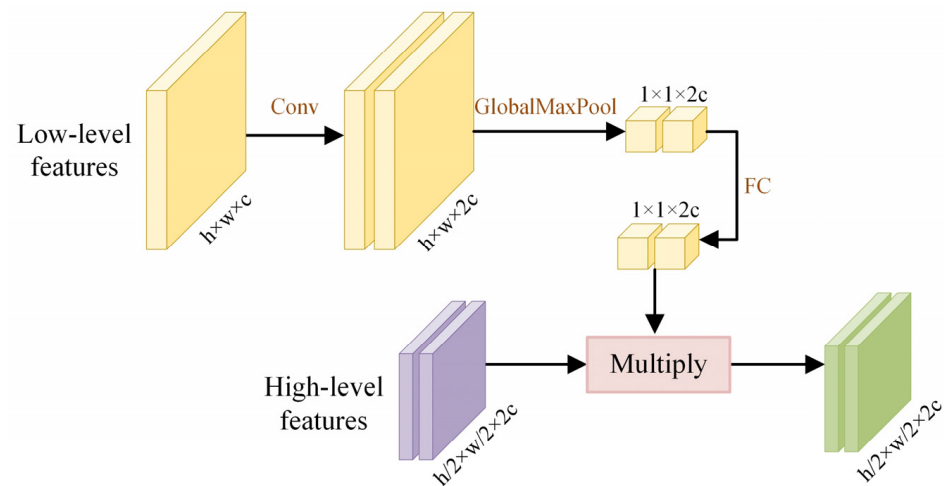


Figure 4. Feature-fusion module.

In Section 4.1, the effectiveness and advantages of the FF module is discussed through ablation experiments.

2.3. Integrated Prediction through Majority Voting

For each input image block, using the trained LMLFF-CNN, a label can be obtained: 1 or 0, representing the built class or non-built class, respectively. If all pixels contained in each grid cell are assigned the same label value, each grid division of the original image generates a binary map similar to a mosaic puzzle through block-level classification. To further improve the recognition error and rough jagged boundaries caused by block-based discrimination, this study used the said set of offset grids to divide the original image multiple times to generate image blocks with a certain degree of spatial overlap. By integrating the classification results of these image blocks, a refined extraction of built-up areas can be achieved. Specifically, each pixel is covered by image blocks from different grids. These blocks may have different labels, providing contextual label information for the discrimination of that pixel. By conducting a majority vote on these labels, the final label value for that pixel is determined. In this way, the extraction results of built-up areas can be greatly improved by utilizing contextual information, and their boundaries can be remarkably refined.

3. Results

3.1. Study Area and Dataset

The selected study area is in Shenzhen, southern China's Pearl River Delta region (Figure 5). As a window city for China's reform and opening up, Shenzhen has rapidly developed from a small fishing village in the 1980s to a modern metropolis, creating a world-renowned "Shenzhen speed" and is known as the "Silicon Valley of China". It is also one of the four central cities in the Guangdong–Hong Kong–Macao Greater Bay Area. As of 2022, the city had nine districts under its jurisdiction, with a total area of 1997.47 square kilometers and a built-up area of 927.96 square kilometers. The terrain of Shenzhen is high in the southeast and low in the northwest; most of it is low hilly areas, interspersed with gentle terraces; and the western part is a coastal plain.

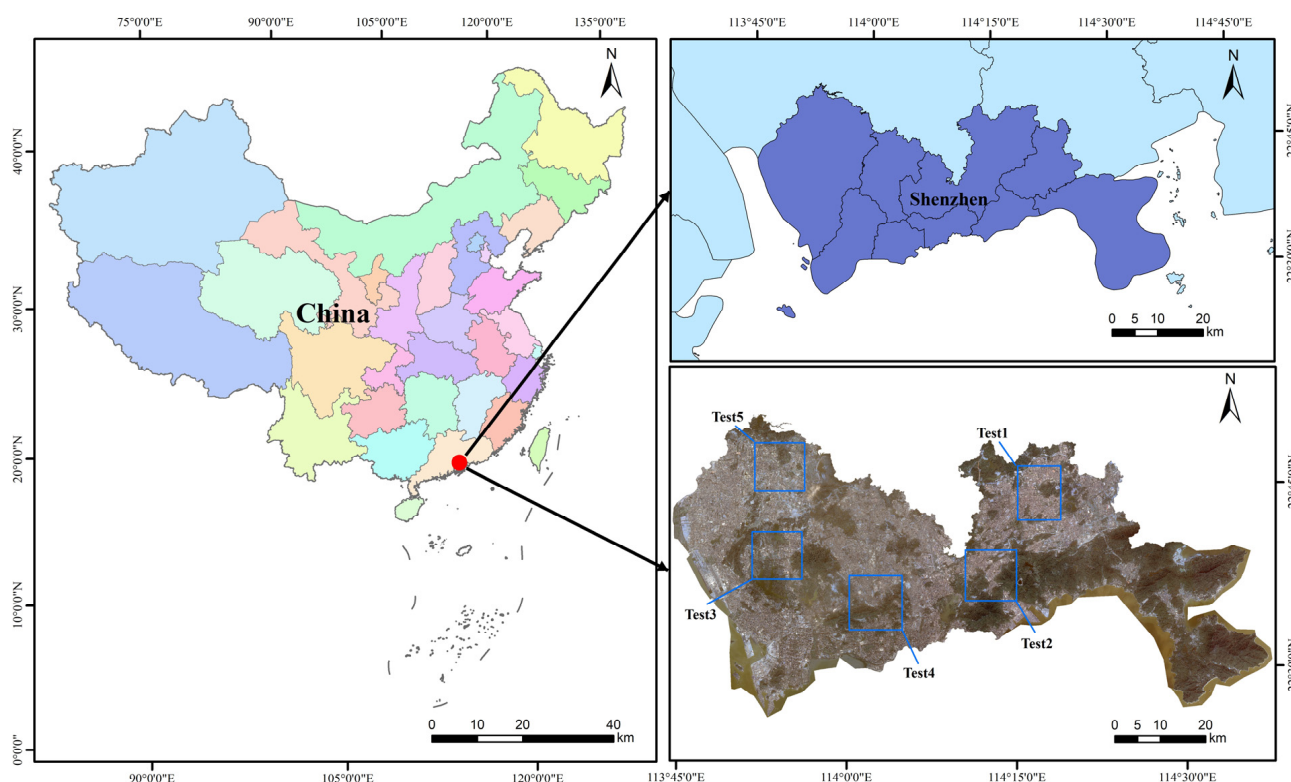
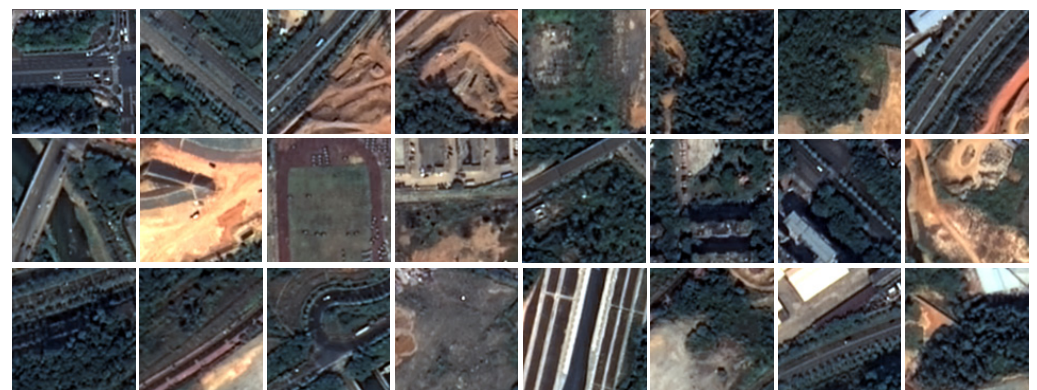


Figure 5. Geographical location of the study area and the Gaofen-2 imagery covering it. Tests 1–5 correspond to five selected test areas.

To verify the effectiveness of the proposed method, an evaluation dataset was constructed using Gaofen-2 satellite images covering the study area. The Gaofen-2 satellite was successfully launched on 19 August 2014. It is China's first civilian optical remote sensing satellite with a spatial resolution better than 1 m (nadir: 0.8 m), equipped with two HR 1 m panchromatic and 4 m multispectral cameras. The used images include three channels of RGB with panchromatic sharpening and a spatial resolution of 1 m. The training set consists of 9808 sample images, with 4904 samples in the built-up and non-built-up areas, and each sample image has a size of 112×112 . Figure 6 shows some sample images from these two classes, each containing scenes that are as diverse as possible. The test set includes satellite images of five sub-regions in Shenzhen, as shown in Figure 5, with sizes of 8960×7840 (Test 1), $10,080 \times 8960$ (Test 2), 8400×8400 (Test 3), 8960×8960 (Test 4), and 8400×8400 (Test 5). The basic processing of these data in the study area was completed through PIE and ArcGIS.



(a)



(b)

Figure 6. Examples of built-up (a) and non-built-up (b) areas included in the sample set.

3.2. Experimental Setup

The hardware platform for the experiment was a Dell workstation equipped with an Intel XeonE5-2620 V3 CPU, 32 GB of memory, and NVIDIA Quadro K620 graphics card. The software platforms included a Windows operating system, Python 3.6, TensorFlow 1.4, and Keras. In the training phase, the binary cross-entropy loss function and the Adam optimizer were selected to train the network. The initial learning rate of the network was set to 0.01, and in the following training, the learning rate was reduced to 10% for every 30 generations of training. Under this training strategy, the network was trained for 100 generations.

3.3. Evaluation Metrics

To assess the performance of this approach quantitatively, the experiment employed four widely used metrics to evaluate the predictive performance of the model. These metrics include precision (P), recall (R), F1-score, and intersection over union (IoU). The definitions of these metrics are as follows:

$$P = \frac{TP}{TP + FP}, \quad (2)$$

$$R = \frac{TP}{TP + FN}, \quad (3)$$

$$F1 - Score = \frac{2PR}{P + R}, \quad (4)$$

$$IoU = \frac{TP}{TP + FP + FN} = \frac{PR}{P + R - PR} \quad (5)$$

where TP and FP represent the number of pixels correctly and incorrectly labeled as built-up areas, respectively, whereas FN refers to the number of pixels incorrectly labeled as non-built-up areas. The F1-Score is the harmonic mean of P and R, whereas IoU measures the ratio of the intersection of predicted and actual built-up area pixels to their union. F1-Score and IoU are comprehensive indicators that consider correctness (P) and completeness (R), providing a more holistic evaluation of the performance in built-up area detection.

3.4. Experimental Results and Analysis

3.4.1. Performance of the Proposed LMLFF-CNN

Based on the previously mentioned data sets and evaluation metrics, the proposed LMLFF-CNN model was quantitatively evaluated. Moreover, four state-of-the-art CNN models, including Inception [40], MobileNet [38], ShuffleNet [41], and EfficientNet [42] with lightweight designs, were tested for performance comparison. Table 1 presents the quantitative evaluation results of these models. On all test images, the proposed LMLFF-CNN obtains the highest F1-Score and IoU value among these models. Considering the precision and completeness of built-up area detection, the proposed network demonstrates notably superior recognition capabilities.

Table 1. Accuracy evaluation results of different CNN models.

Test Image	Model	P	R	F1-Score	IoU
1	InceptionV3	0.7461	0.9740	0.8450	0.7316
	MobileNet	0.8986	0.8747	0.8865	0.7962
	ShuffleNetV2	0.8854	0.8973	0.8913	0.8039
	EfficientNet-B0	0.8865	0.8957	0.8911	0.8036
	LMLFF-CNN	0.8930	0.8995	0.8962	0.8120
2	InceptionV3	0.7544	0.9686	0.8481	0.7364
	MobileNet	0.8952	0.9103	0.9027	0.8226
	ShuffleNetV2	0.9016	0.8831	0.8923	0.8056
	EfficientNet-B0	0.9097	0.8807	0.8949	0.8099
	LMLFF-CNN	0.9039	0.9106	0.9072	0.8302
3	InceptionV3	0.7946	0.9603	0.8696	0.7483
	MobileNet	0.8468	0.9153	0.8797	0.7852
	ShuffleNetV2	0.8417	0.9110	0.8750	0.7778
	EfficientNet-B0	0.8731	0.8848	0.8789	0.7840
	LMLFF-CNN	0.8657	0.9171	0.8906	0.8029
4	InceptionV3	0.7754	0.9708	0.8623	0.7565
	MobileNet	0.8863	0.8622	0.8741	0.7764
	ShuffleNetV2	0.8840	0.8556	0.8696	0.7692
	EfficientNet-B0	0.9030	0.8139	0.8562	0.7485
	LMLFF-CNN	0.8825	0.8714	0.8769	0.7808
5	InceptionV3	0.7295	0.9800	0.8364	0.7488
	MobileNet	0.9147	0.8710	0.8923	0.8056
	ShuffleNetV2	0.9023	0.8825	0.8923	0.8055
	EfficientNet-B0	0.9174	0.8668	0.8914	0.8041
	LMLFF-CNN	0.9048	0.8920	0.8984	0.8155

In addition to its accuracy advantage, the proposed LMLFF-CNN is a lightweight network. Table 2 compares these models in terms of parameter quantity and computational power. LMLFF-CNN only contains 0.18 G of parameters and 0.068 M of Flops, demonstrating substantially fewer parameters and higher computational efficiency compared with the four other networks. This result is attributed to the simple model structure of LMLFF-CNN, whose key components only include three FE modules and two FF modules. The use of a series of lightweight techniques such as a dual-branch structure, DSC, and global average pooling further reduces the number of parameters and computational burden.

Table 2. Comparison of model parameters and flops.

Model	Params (G)	Flops (M)
InceptionV3	3.51	0.102
MobileNet	2.24	0.583
ShuffleNetV2	5.015	0.957
EfficientNet-B0	5.288	0.705
LMLFF-Net	0.18	0.068

Furthermore, taking the Test 1 image (size: 8960×7840) as an example, the prediction time of these models are shown in Table 3. For the prediction of the built-up area in this image, the proposed LMLFF-CNN only takes 42.2 s, which is much less than the computational time of other models.

Table 3. Comparison of prediction time for different models.

Model	Time (s)
InceptionV3	96.7
MobileNet	208.3
ShuffleNetV2	159.6
EfficientNet-B0	241
LMLFF-Net	42.2

3.4.2. Integrated Prediction through Majority Voting

Although grid blocks, as basic processing units, are beneficial for the feature representation and discrimination of built-up areas in complex scenes, traditional single-grid-based methods do not consider the contextual information of the target when splitting images. This limitation not only increases the probability of erroneous discrimination but also leads to evident jagged boundaries, especially when the size of the grid cells is large. This study achieved multiple partitions of the test image by offsetting the grid horizontally and vertically in multiple steps. Using the proposed LMLFF-CNN, the image blocks generated by each grid partition are classified, and a series of preliminary prediction results are obtained. Subsequently, through the majority voting method, the final refined extraction result is generated. Figure 7 presents the accuracy evaluation results of each test image using this strategy. Compared with prediction methods based on a single grid partition, voting methods based on multiple grids achieve varying degrees of performance improvement in terms of evaluation metrics. For example, for test image 1, the F1-score increases from 0.8962 to 0.9170, and the IoU value increases from 0.8120 to 0.8468, representing increases of 2.08% and 3.48%, respectively. Similarly, consistent performance improvements can be observed on other test images. This outcome is primarily attributed to the efficacy of this integrated decision method in leveraging label information from contextual image blocks, thereby augmenting the robustness of discrimination.

The proposed method can achieve superior performance in the extraction of built-up areas, primarily due to two key techniques: the LMLFF-CNN model and majority voting. To further understand their performances in the extraction of built-up areas, Figure 8 shows a percentage stacked bar chart in terms of F1 score and IoU, showing how much they contribute to accuracy. In the proposed method, the LMLFF-CNN model plays a crucial role, whereas majority voting contributes less than 5% to accuracy. Nevertheless, majority voting remains important in the proposed methods.

To illustrate the advantages of this integrated discrimination method more intuitively, Figure 9 takes the Test 2 image as an example to show the extraction results using the single-grid method and the multigrid-based voting method. Compared with the commonly used single-grid division method, the majority voting method based on multigrid division can remarkably improve the detection results of built-up areas, resulting in better shape integrity and boundary smoothness.

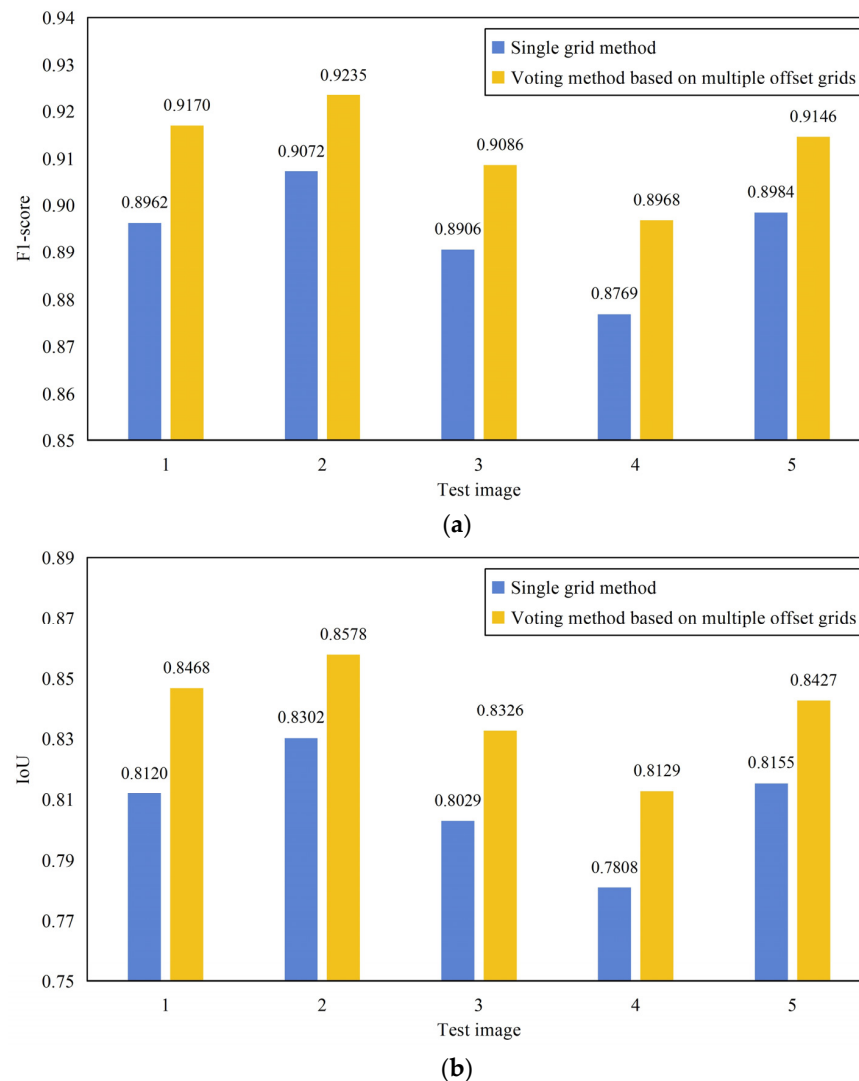


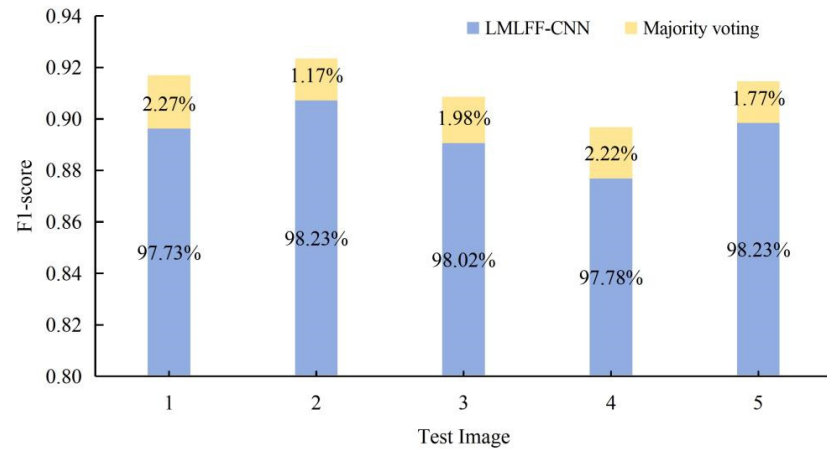
Figure 7. F1-scores (a) and IoU (b) values using different grid partitioning methods.

3.4.3. Comparison with State-of-the-Art Built-Up Area Detection Methods

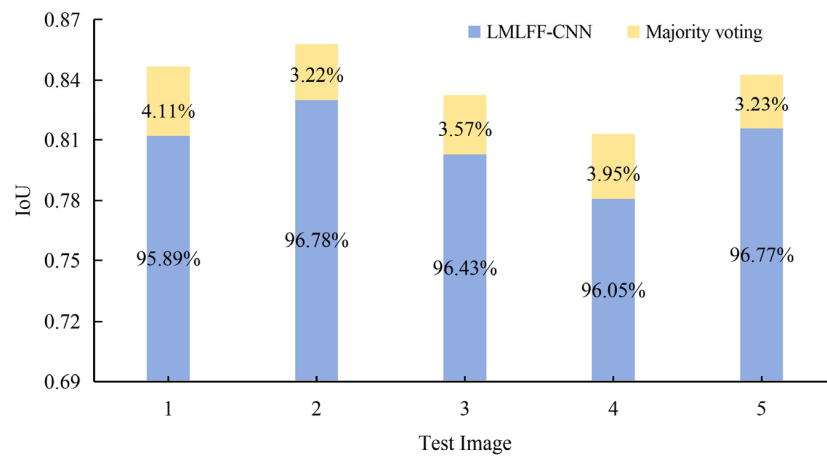
To evaluate the overall performance of the proposed method in the task of extracting built-up areas, several representative built-up area extraction methods were also tested on the same dataset to compare their performance differences. The compared methods include MSTSD [7], GLCM + SVM, LMB-CNN [41], and GHS-S2Net [37]. MSTSD is a pixel-based unsupervised method that utilizes multiscale textures and spatial dependencies to construct saliency maps of built-up areas; then, it segments built-up areas through Ostu thresholding. GLCM+SVM adopts a block-based supervised method, which uses the grayscale co-occurrence matrix to measure the texture features of built-up areas in image blocks and then inputs them into the SVM classifier to distinguish between built-up and non-built-up areas. Both methods are traditional methods based on manually designed features. LMB-CNN and GHS-S2Net are deep-learning-based methods. In the LMB-CNN method, a regular grid is used to partition images to generate image blocks, and then LMB-CNN is used for the block-level feature representation and discrimination of built-up areas. GHS-S2Net is also based on CNN, and it uses a sliding-window-based method for each pixel to predict the category of the center pixel.

Figure 10 shows the extraction results obtained using these methods separately. The results obtained using the proposed method are generally closer to the actual situation of the built-up area and maintain better shape integrity. The results of the MSTSD method have a large number of missed detections, which makes the extracted built-up areas

appear extremely incomplete, whereas the GLCM+SVM method has a more evident false detection phenomenon, resulting in a large number of redundant detections. Compared with these two classic methods, the detection performance of the LMB-CNN and GHS-S2Net methods is greatly improved, but the resulting map contains a large amount of salt-and-pepper noise.

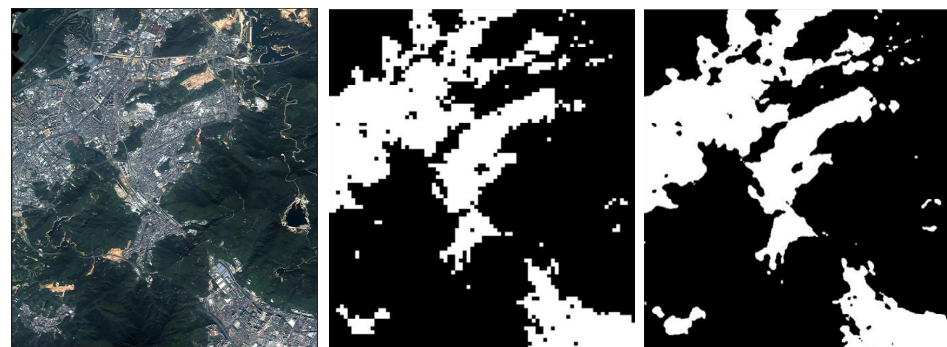


(a)



(b)

Figure 8. Percentage stacked bar charts for F1 score (a) and IoU (b), showing how much LMLFF-CNN and majority voting contribute to accuracy.



(a)

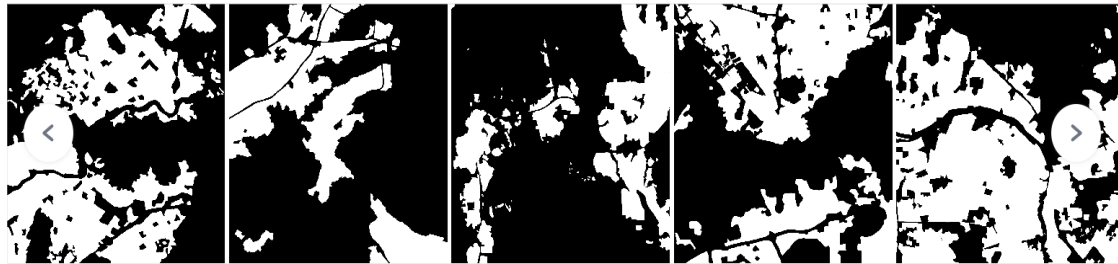
(b)

(c)

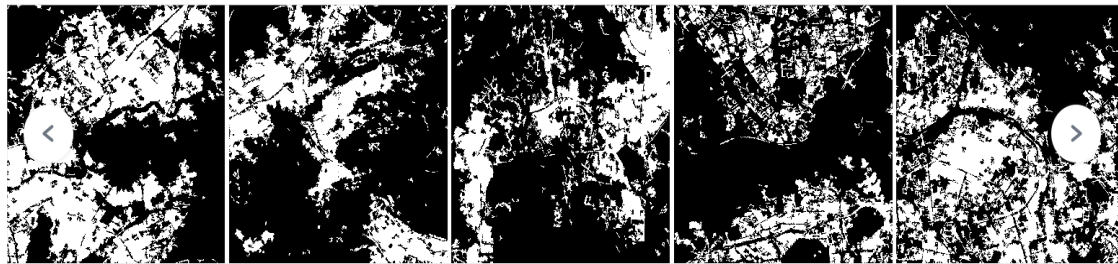
Figure 9. Detection results using different grid partitioning methods. (a) Test image; (b) single-grid method; (c) multigrid-based majority voting method.



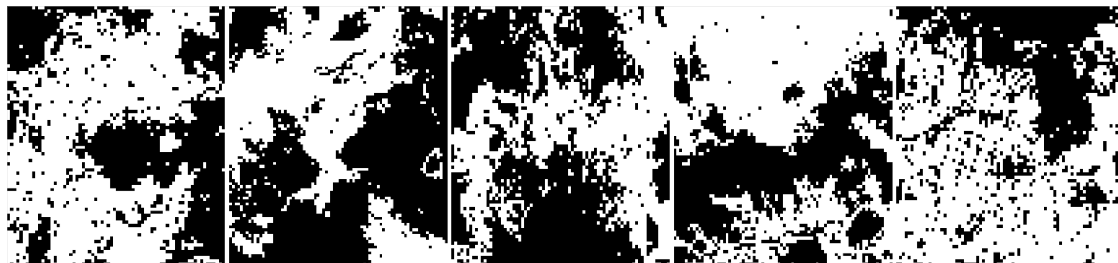
(a) Original image



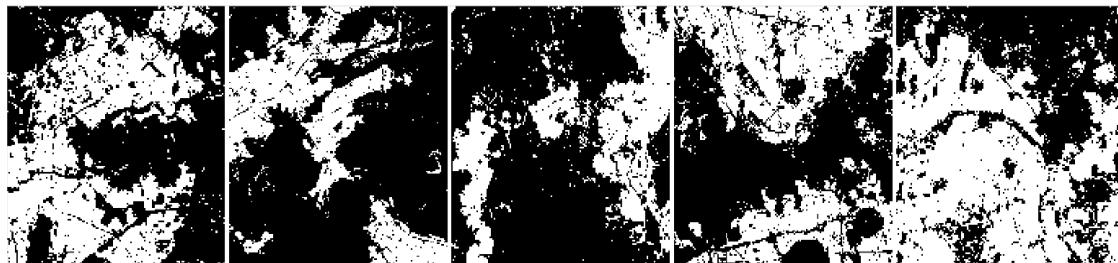
(b) Ground truth



(c) MSTSD

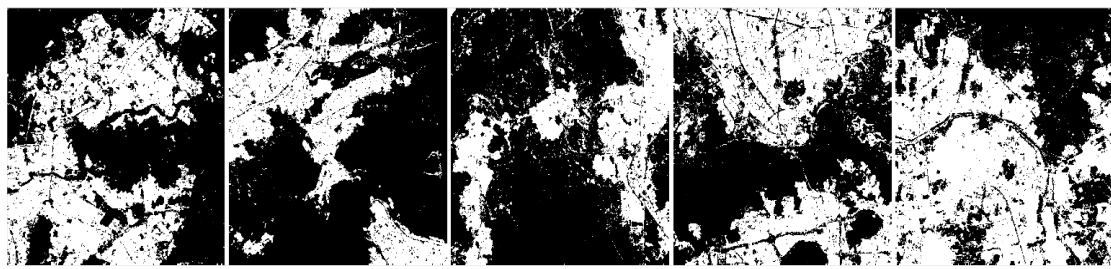


(d) GLCM + SVM

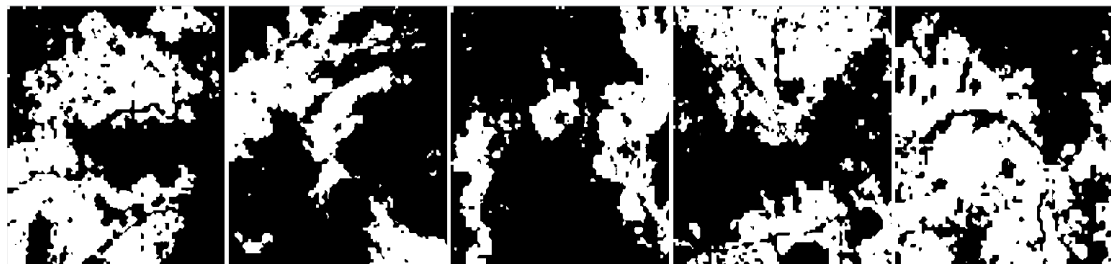


(e) LMB-CNN

Figure 10. Cont.



(f) GHS-S2Net



(g) Proposed (LMLFF CNN + Majority Voting)

Figure 10. Detection results of built-up areas using different methods.

The accuracy statistics of the quantitative evaluation of these methods are shown in Table 4. According to the evaluation indicators, deep-learning-based methods obtain higher F1-score and IoU values than traditional manually designed feature-based methods. This outcome indicates a better balance between accuracy and completeness in extracting built-up areas and a higher consistency between the extraction results and the actual situation. The proposed method achieves the highest F1-score and IoU values for all test images, with only slight fluctuations in different test areas. Especially in Region 3, the scene is extremely complex, but our method still achieves stable performance with an F1-score of 0.9086 and an IoU value of 0.8326.

Table 4. Comparison of detection accuracies of built-up areas using different methods.

Test Image	Method	P	R	F1-Score	IoU
1	MSTSD	0.8938	0.8171	0.8537	0.7448
	GLCM + SVM	0.7180	0.9296	0.8102	0.6810
	LMB-CNN	0.8974	0.8922	0.8948	0.8096
	GHS-S2Net	0.9182	0.8747	0.8959	0.8115
	Proposed (LMLFF CNN + Majority Voting)	0.9101	0.9241	0.9170	0.8468
2	MSTSD	0.8401	0.8033	0.8213	0.6968
	GLCM + SVM	0.6576	0.9814	0.7875	0.6495
	LMB-CNN	0.8988	0.8880	0.8933	0.8072
	GHS-S2Net	0.8933	0.9039	0.8986	0.8158
	Proposed (LMLFF CNN + Majority Voting)	0.9169	0.9301	0.9235	0.8578
3	MSTSD	0.6940	0.7848	0.7366	0.5830
	GLCM + SVM	0.5195	0.9287	0.6663	0.4996
	LMB-CNN	0.8110	0.9244	0.8640	0.7606
	GHS-S2Net	0.7838	0.9327	0.8518	0.7419
	Proposed (LMLFF CNN + Majority Voting)	0.8844	0.9342	0.9086	0.8326
4	MSTSD	0.8977	0.5232	0.6611	0.4938
	GLCM + SVM	0.6471	0.9043	0.7544	0.6056
	LMB-CNN	0.8522	0.8693	0.8607	0.7555
	GHS-S2Net	0.7991	0.8970	0.8452	0.7320
	Proposed (LMLFF CNN + Majority Voting)	0.8985	0.8951	0.8968	0.8129

Table 4. Cont.

Test Image	Method	P	R	F1-Score	IoU
5	MSTSD	0.9130	0.6545	0.7625	0.6161
	GLCM + SVM	0.7527	0.8530	0.7997	0.6663
	LMB-CNN	0.8650	0.9278	0.8953	0.8105
	GHS-S2Net	0.8812	0.9038	0.8923	0.8056
	Proposed (LMLFF CNN + Majority Voting)	0.9182	0.9111	0.9146	0.8427

4. Discussion

4.1. Ablation Study on FF Module

In the proposed LMLFF-CNN, the FF module can integrate different levels of features for built-up area discrimination. To evaluate the effectiveness of the module, the network prediction performance without the FF module was further tested. In Table 5, the results of the ablation experiment indicate that the addition of the FF module substantially improves the predictive performance of the model for each test area. Especially in test image 4, the performance improvement is most remarkable, with F1 scores increasing from 0.7976 to 0.8769 and IoU increasing from 0.6634 to 0.7808. This result indicates the effectiveness of the FF module. Furthermore, the FF module substantially improved the P of built-up area discrimination, achieving a balance between P and R. This result is attributed to the module's ability to effectively integrate the multilevel convolutional features for a more accurate discrimination between built-up and non-built-up areas.

Table 5. Results of ablation experiment on FF module.

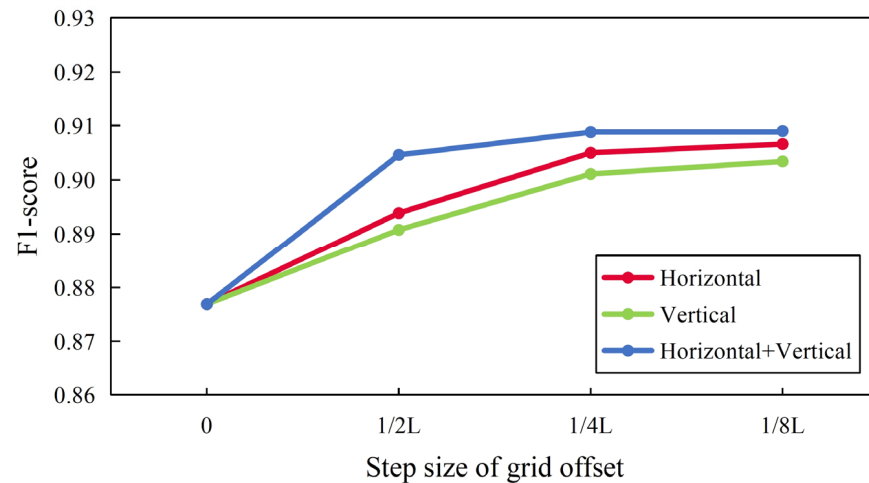
Test Image	Models	P	R	F1-Score	IoU
1	With FF	0.8930	0.8995	0.8962	0.8120
	Without FF	0.8007	0.9812	0.8818	0.7886
2	With FF	0.9039	0.9106	0.9072	0.8302
	Without FF	0.8179	0.9671	0.8862	0.7958
3	With FF	0.8657	0.9171	0.8906	0.8029
	Without FF	0.7853	0.9647	0.8658	0.7634
4	With FF	0.8825	0.8714	0.8769	0.7808
	Without FF	0.6733	0.9782	0.7976	0.6634
5	With FF	0.9048	0.8920	0.8984	0.8155
	Without FF	0.7943	0.9701	0.8734	0.7754

4.2. Effect of Grid Offset Parameters on Extraction Results

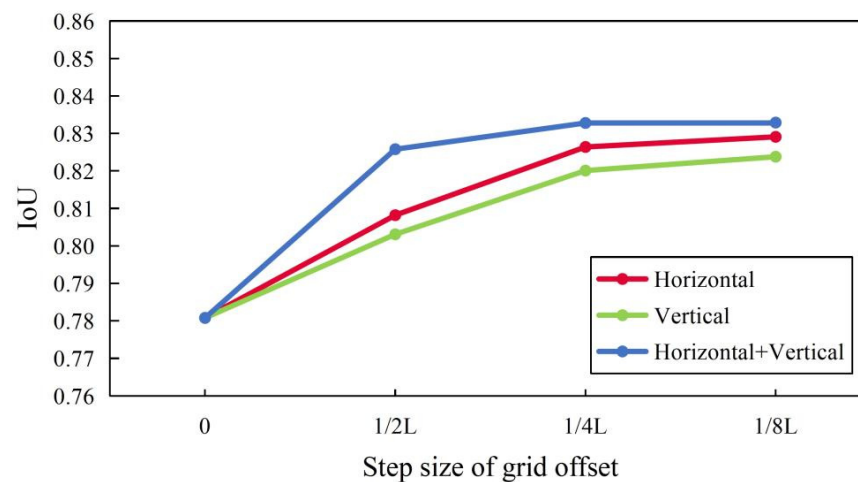
This study used a grid-offset-based method to generate image blocks with a certain overlap rate, which provides rich contextual information for the final discrimination of built-up areas. It not only improved the accuracy of built-up area recognition but also refined the severely jagged boundaries of the extracted results. Furthermore, taking a test image as an example, experiments were conducted on different grid offset strategies to evaluate the effect of their two parameters, namely, direction and step size, on the final extraction results. The offset directions included horizontal (moving to the right) and vertical (moving down). In each direction, the offset steps were $1/2 L$ (56 pixels), $1/4 L$ (28 pixels), and $1/8 L$ (14 pixels), where $L = 112$ pixels is the length of the original grid cell; correspondingly, the number of offsets required were 2, 4, and 8, respectively.

Figure 11 shows the overall performance of the built-up area detection using offset grids with different directions and step sizes on the test data. As the offset step size decreases, the F1 score and IoU value show a similar trend of first increasing and then stabilizing. Furthermore, adopting grids that integrate the horizontal and vertical directions achieves superior performance compared with using a single direction because multidirectional, multistep integration methods can more effectively utilize the contextual

information of objects. In addition, although smaller offset step sizes can result in a higher resolution of the extraction results, the use of more offset grids means greater computational consumption. Considering accuracy and computation time, $1/4 L$ (28 pixels) or $1/8 L$ (14 pixels) was taken as the offset step size for this study. For image data of other resolutions, this parameter can be adjusted accordingly.



(a)



(b)

Figure 11. F1-scores (a) and IoU values (b) of built-up area detection using offset grids with different directions and step sizes. The directions of grid offset include horizontal, vertical, and their integration. The offset steps include 0, $1/2 L$, $1/4 L$, and $1/8 L$, where $L = 112$ is the length of the original grid.

4.3. Generation of Urban-Scale Built-Up Area Maps with a Resolution of 1 m

The distribution map of built-up areas at the urban scale is valuable foundational geographic data for urban planning, management, construction, and research. To demonstrate the feasibility of the proposed method in large-scale, HR mapping of built-up areas, it was further applied to the extraction of built-up areas from Gaofen-2 satellite images covering the entire Shenzhen City. The image data used included $90,698 \times 46,444$ pixels, occupying 31.3 G of memory. Considering that our computer could directly process such large image data, the original data were first split into five sub images, which were then fed into our model for prediction. Finally, the extracted results from these sub images were combined to obtain a distribution map of the entire city's built-up areas. Figure 12 shows the urban built-up area map of Shenzhen, with a spatial resolution of 1 m. Compared with commonly

used mapping products based on Landsat or Sentinel data, it has a higher spatial resolution and contains richer details. According to the quantitative evaluation results of the five test areas described in Section 3, their average F1 score and IoU value are 0.9121 and 0.8386, respectively. Compared with previous similar studies [12], this result also demonstrates superior performance in the accuracy and shape integrity of urban built-up area detection.



Figure 12. Distribution map of urban built-up areas in Shenzhen with a resolution of 1 m.

5. Conclusions

Urban built-up areas are large-scale composite object classes that exhibit remarkable spatial heterogeneity and scene complexity in their characteristics. Currently, the mapping of built-up areas using HR satellite imagery still faces considerable challenges. This study adopted a block-based image processing strategy and constructed an LMLFF-CNN model for the feature representation and discrimination of built-up areas in HR images. Furthermore, to improve the problem of incorrect discrimination and severe jagged boundaries caused by block-based processing, a majority voting method based on a grid offset was adopted to achieve a refined extraction of built-up areas. The Gaofen-2 satellite images covering Shenzhen, China, were used to evaluate the performance of the proposed method experimentally. The main findings are as follows: (1) the proposed LMLFF-CNN model has fewer parameters and computational power than the classical CNN model but achieves a higher discrimination accuracy. (2) The integrated discrimination method based on grid offset can effectively utilize block-level spatial contextual information and considerably improve the accuracy of built-up area detection and the smoothness of target boundaries. (3) The proposed built-up area detection method achieved good experimental results in five selected test areas, with F1 scores of 0.9170, 0.9235, 0.9086, 0.8968, and 0.9146, and IoU of 0.8468, 0.8578, 0.8326, 0.8129, and 0.8427, respectively. Compared with the current representative built-up area extraction algorithm, it demonstrated a higher recognition accuracy and maintains better shape integrity in the extraction results. (4) The proposed method was used to generate a 1 m resolution distribution map of built-up areas throughout Shenzhen, demonstrating its feasibility in large-scale HR mapping of built-up areas.

In the future, image data from more regions and sensors will be utilized to evaluate the performance of the proposed method. In addition, how to achieve large-scale, multi-resolution urban built-up area mapping through the transfer learning of cross-regional or sensor data will be explored. In addition to satellite imagery, the processing and mapping of underwater images [43,44] are currently a promising research field. Considering the unique features of underwater images, including challenges related to lighting conditions,

environmental factors, and unique features of underwater scenes, whether the proposed method can be applied to underwater images for object detection and mapping will be another research focus.

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

Funding: This research was funded by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (KF-2021-06-090), and by the University-Industry Collaborative Education Program (202102245033).

Data Availability Statement: The datasets used in this study are available at <https://github.com/3539386390/Urban-Built-Up-Area-Extraction> (accessed on 17 January 2024).

Acknowledgments: The authors thank the editors and reviewers for their valuable comments and suggestions.

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

References

- Liu, C.; Huang, X.; Zhu, Z.; Chen, H.; Tang, X.; Gong, J. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: Analysis of 45 global cities. *Remote Sens. Environ.* **2019**, *226*, 51–73. [CrossRef]
- Wang, H.; Gong, X.; Wang, B.; Deng, C.; Cao, Q. Urban development analysis using built-up area maps based on multiple high-resolution satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102500. [CrossRef]
- Verma, A.; Bhattacharya, A.; Dey, S.; López-Martínez, C.; Gamba, P. Built-up area mapping using Sentinel-1 SAR data. *ISPRS J. Photogramm. Remote Sens.* **2023**, *203*, 55–70. [CrossRef]
- Hu, Z.; Li, Q.; Zhang, Q.; Wu, G. Representation of block-based image features in a multi-scale framework for built-up area detection. *Remote Sens.* **2016**, *8*, 155. [CrossRef]
- Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [CrossRef]
- Shao, Z.; Tian, Y.; Shen, X. BASI: A new index to extract built-up areas from high-resolution remote sensing images by visual attention model. *Remote Sens. Lett.* **2014**, *5*, 305–314. [CrossRef]
- Tao, C.; Tan, Y.; Zou, Z.-r.; Tian, J. Unsupervised detection of built-up areas from multiple high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1300–1304. [CrossRef]
- Chen, Y.; Lv, Z.; Huang, B.; Jia, Y. Delineation of built-up areas from very high-resolution satellite imagery using multi-scale textures and spatial dependence. *Remote Sens.* **2018**, *10*, 1596. [CrossRef]
- Chen, Y.; Lv, Z.; Huang, B.; Zhang, P.; Zhang, Y. Automatic extraction of built-up areas from very high-resolution satellite imagery using patch-level spatial features and gestalt laws of perceptual grouping. *Remote Sens.* **2019**, *11*, 3022. [CrossRef]
- Ali, A.; Nayyar, Z.A. A Modified Built-up Index (MBI) for automatic urban area extraction from Landsat 8 Imagery. *Infrared Phys. Technol.* **2021**, *116*, 103769. [CrossRef]
- Misra, M.; Kumar, D.; Shekhar, S. Assessing machine learning based supervised classifiers for built-up impervious surface area extraction from sentinel-2 images. *Urban For. Urban Green.* **2020**, *53*, 126714. [CrossRef]
- Chen, Y.; Yao, S.; Hu, Z.; Huang, B.; Miao, L.; Zhang, J. Built-up Area Extraction Combing Densely Connected Dual-Attention Network and Multi-Scale Context. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5128–5143. [CrossRef]
- Tan, Y.; Xiong, S.; Li, Y. Automatic extraction of built-up areas from panchromatic and multispectral remote sensing images using double-stream deep convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3988–4004. [CrossRef]
- Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
- Wahbi, M.; El Bakali, I.; Ez-zahouani, B.; Azmi, R.; Moujahid, A.; Zouiten, M.; Alaoui, O.Y.; Boulaassal, H.; Maatouk, M.; El Kharki, O. A deep learning classification approach using high spatial satellite images for detection of built-up areas in rural zones: Case study of Souss-Massa region-Morocco. *Remote Sens. Appl. Soc. Environ.* **2023**, *29*, 100898. [CrossRef]
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

19. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
20. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [[CrossRef](#)]
21. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
22. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Proceedings, Part III 18, Munich, Germany, 5–9 October 2015; pp. 234–241.
23. Persello, C.; Stein, A. Deep fully convolutional networks for the detection of informal settlements in VHR images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2325–2329. [[CrossRef](#)]
24. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [[CrossRef](#)]
25. Wu, F.; Wang, C.; Zhang, H.; Li, J.; Li, L.; Chen, W.; Zhang, B. Built-up area mapping in China from GF-3 SAR imagery based on the framework of deep learning. *Remote Sens. Environ.* **2021**, *262*, 112515. [[CrossRef](#)]
26. Li, T.; Wang, C.; Wu, F.; Zhang, H.; Tian, S.; Fu, Q.; Xu, L. Built-Up area extraction from GF-3 SAR data based on a dual-attention transformer model. *Remote Sens.* **2022**, *14*, 4182. [[CrossRef](#)]
27. Lv, Z.; Zhang, P.; Sun, W.; Benediktsson, J.A.; Lei, T. Novel Land-Cover Classification Approach with Nonparametric Sample Augmentation for Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4407613. [[CrossRef](#)]
28. Lv, Z.; Zhang, P.; Sun, W.; Lei, T.; Benediktsson, J.A.; Li, P. Sample Iterative Enhancement Approach for Improving Classification Performance of Hyperspectral Imagery. *IEEE Geosci. Remote Sens. Lett.* **2023**. [[CrossRef](#)]
29. Cao, Y.; Huang, X.; Weng, Q. A multi-scale weakly supervised learning method with adaptive online noise correction for high-resolution change detection of built-up areas. *Remote Sens. Environ.* **2023**, *297*, 113779. [[CrossRef](#)]
30. Hafner, S.; Ban, Y.; Nascetti, A. Unsupervised domain adaptation for global urban extraction using sentinel-1 SAR and sentinel-2 MSI data. *Remote Sens. Environ.* **2022**, *280*, 113192. [[CrossRef](#)]
31. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [[CrossRef](#)]
32. Mboga, N.; Persello, C.; Bergado, J.R.; Stein, A. Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sens.* **2017**, *9*, 1106. [[CrossRef](#)]
33. Corbane, C.; Syrris, V.; Sabo, F.; Politis, P.; Melchiorri, M.; Pesaresi, M.; Soille, P.; Kemper, T. Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Comput. Appl.* **2021**, *33*, 6697–6720. [[CrossRef](#)]
34. Huang, F.; Yu, Y.; Feng, T. Automatic extraction of impervious surfaces from high resolution remote sensing images based on deep learning. *J. Vis. Commun. Image Represent.* **2019**, *58*, 453–461. [[CrossRef](#)]
35. Li, Y.; Huang, X.; Liu, H. Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 567–579. [[CrossRef](#)]
36. Tan, Y.; Xiong, S.; Li, Z.; Tian, J.; Li, Y. Accurate detection of built-up areas from high-resolution remote sensing imagery using a fully convolutional network. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 737–752. [[CrossRef](#)]
37. Tan, Y.; Xiong, S.; Yan, P. Multi-branch convolutional neural network for built-up area extraction from remote sensing image. *Neurocomputing* **2020**, *396*, 358–374. [[CrossRef](#)]
38. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861 2017.
39. Lv, Z.; Liu, J.; Sun, W.; Lei, T.; Benediktsson, J.A.; Jia, X. Hierarchical Attention Feature Fusion-Based Network for Land Cover Change Detection With Homogeneous and Heterogeneous Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [[CrossRef](#)]
40. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
41. 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.
42. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
43. Zhang, X.; Wu, H.; Sun, H.; Ying, W. Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10835–10853. [[CrossRef](#)]
44. Yang, P. An imaging algorithm for high-resolution imaging sonar system. *Multimed. Tools Appl.* **2023**, 1–17. [[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.