

Identification of Built-up Urban Areas in High-Resolution, Remotely- Sensed Images Based on Deep Learning

Chujun Li^{1,2,a,*} and Md Gapar Md Johar^{3,b}

¹*School of Engineering, Xi'an International University, Xi'an 710077, Shaanxi, China*

²*School of Graduate Studies, Management and Science University, Shah Alam 40100, Malaysia*

³*Software Engineering and Digital Innovation Centre, Management and Science University, Shah Alam 40100, Malaysia*

The accuracy of traditional identification methods for built-up urban areas is often challenged when faced with complex urban environments comprising dense buildings and shadows, especially when identifying small targets and multi-scale building features. To address these problems, this paper constructs an built-up urban area identification model based on the Weighted Feature Fusion Enhanced UNet (WFFE-UNet). The data used in this paper is from the Massachusetts Buildings Dataset, which provides aerial images of the Boston metropolitan area with a resolution of 1 meter. Histogram equalization and principal component analysis denoising techniques are used to improve image quality and help the model capture key information in the image more accurately. Then, an identification model based on WFFE-UNet is constructed. Based on the UNet framework, this model uses an improved convolution module with multi-scale perception to enhance the ability to express multi-scale building structures in high-resolution images. The Convolutional Block Attention Module (CBAM) attention mechanism is integrated to improve the identification ability of small targets. By combining channel attention and spatial attention, the model's sensitivity to the key features of small targets is significantly enhanced. Finally, the strategy of joint optimization of cross-entropy loss and Dice loss is adopted to improve the model's convergence efficiency and identification precision. Experimental results show that the overall accuracy, precision, recall, and intersection over union (IoU) of the WFFE-UNet model are 97.34%, 96.76%, 95.42%, and 82.65%, respectively. In the cross-domain test, the overall identification accuracy of the model in different cities remains above 96%. For small targets B and P, the IoU of the model studied in this paper is 78.98% and 78.46%, respectively. The built-up urban area identification method based on WFFE-UNet and CBAM attention mechanism successfully solves the problems encountered by traditional methods when applied complex scenes by combining automatic feature learning, multi-scale information fusion, and an attention mechanism.

Keywords: deep learning, high-resolution remotely-sensed image, built-up urban area identification, Weighted Feature Fusion Enhanced UNet, Convolutional Block Attention Module

1. INTRODUCTION

With the continuous advancement of remote sensing technology, high-resolution remotely-sensed images can be

obtained, resulting in the adoption of this technology in many fields [1–3]. In particular, a precise identification of the features of built-up urban areas as well as the city center, is critical to understanding urban structure, optimizing resource allocation and improving governance capabilities. In complex urban environments such as high-density areas where there may be heavy shadow interference or scenes

*Address for correspondence: Chujun Li, School of Engineering, Xi'an International University, Xi'an 710077, Shaanxi, China, Email: lichujun2021@163.com, mdgapar@msu.edu.my

with diverse building forms, the precision and efficiency of traditional identification methods face challenges. With traditional methods, small targets and complex scenes cannot be adequately identified, making it difficult to meet the needs of smart cities where high-precision and high-efficiency identification are required.

In recent years, deep learning technology has made progress in the identification of built-up urban areas in high-resolution, remotely-sensed images [4–6]. In particular, convolutional neural networks (CNN) have performed well in computer vision tasks due to their excellent automatic feature learning and multi-level semantic understanding capabilities [7–9]. Applying it to built-up urban area identification in high-resolution, remotely-sensed images (such as images from the Gaofen-2 satellite with a resolution of up to 0.8 meters) not only automatically extracts complex spatial features; it also effectively reduces background noise, significantly improving the precision and efficiency of identification. This paper focuses on the identification of built-up urban areas in high-resolution remotely-sensed images and strives to achieve more efficient and accurate identification by means of an optimized deep learning model. The WFFE-UNet model is used to study the 1-meter-resolution aerial images of the Boston metropolitan area in the Massachusetts Buildings Dataset, focusing mainly on the problems of difficult target detection, large background noise, and difficulty in integrating multi-scale information. Through the CBAM mechanism, the model can dynamically allocate resources and focus more on small targets in complex backgrounds, thereby improving the identification precision of fine-grained targets. At the same time, the multi-scale feature fusion strategy also enhances the model's understanding and utilization of information at different scales and better copes with changes in target size.

Based on deep learning technology, a WFFE-UNet-based built-up urban area identification model is constructed. This model is based on the classic UNet framework and is improved to meet the special needs of remotely-sensed image identification. By adopting an improved multi-scale convolution module, the model can more effectively identify buildings and roads of different scales in remotely-sensed images. By using a weighted series integration mechanism and applying learnable fusion parameters, the importance of features at different levels in the identification process can be dynamically adjusted. CBAM is incorporated into WFFE-UNet, which significantly enhances the model's ability to handle few samples and small target identification tasks. By designing cross-domain identification experiments, the excellent adaptability and stability of the proposed model under various geographical, climatic, and urban structure conditions are verified, which supports the application of the model in practical engineering.

2. RELATED WORK

A remotely-sensed image built-up urban area identification refers to the use of high-resolution, remotely-sensed images and cutting-edge image processing technology to automatically identify information about developed areas in cities. It

can precisely depict the boundaries of built-up urban areas and provide scientific data for decision makers [10–12]. To make full use of the data attributes of remotely sensed images, Xie proposed a multimodal deep learning method with an attention mechanism and applied it to functional area identification in built-up urban areas [13]. Chen evaluated the built environment from the perspective of street view and found that not all areas of the built environment could be fully covered by street view data. Therefore, he proposed an intelligent evaluation technology for the urban built environment by using high-resolution, remotely-sensed image scene understanding [14]. Cheng took the Yuexiu District of Guangzhou as the research area, used support vector machine and random forest to derive street space quality information, and performed multi-scale segmentation based on object-based image analysis based on the pre-extraction results on the Gaofen-2 image [15]. Extracting built-up area boundaries using high-resolution satellite remotely-sensed images is of great significance for urban expansion monitoring and urban development planning. To obtain high-precision and high-spatial-resolution data for a built-up area, Liu constructed a dataset of remotely-sensed images of built-up urban areas by normalizing the building index and used four deep learning semantic segmentation networks to extract built-up areas from Sentinel-2 images [16]. Wang proposed a method for extracting and identifying urban river information based on frequency domain filtering. The high-resolution remotely-sensed image was subjected to a Fourier transform to obtain a spectrum diagram, and the radial and angular distribution diagrams were used to analyze the spectrum characteristics of urban rivers. This method can effectively identify urban rivers and extract relevant information [17].

At present, the identification of built-up urban areas relies mainly on manual feature extraction and empirical threshold setting. This method cannot cope with the complex and changeable land object structures and multi-scale changes in high-resolution, remotely-sensed images, and often leads to blurred boundaries and missed detection. Deep learning technology provides a new solution for urban identification in high-resolution, remotely-sensed images. By means of advanced technology, remotely-sensed images can be automatically analyzed and understood, thereby providing a more accurate identification of built-up areas in cities [18–20]. The extraction and identification of buildings from high-resolution remotely-sensed images is problematic due to the complexity of surface buildings. Zeng took the Gaofen-2 high-resolution remote sensing satellite as the research object and proposed a correction neural network model, namely the boundary adjustment network, which extracts the features of remotely-sensed images through convolution, pooling, and classification [21]. Song proposed a new technology for identifying water bodies based on remotely-sensed data by integrating region proposal generation and target detection based on CNN, which improved the detection of image targets [22]. Traditional image segmentation methods cannot fully represent the information in high-spatial resolution images and often produce unsatisfactory segmentation precision. Zheng used U-Net to perform multi-classification and binary classification on Gaofen-2 high spatial resolution remotely-sensed image data [23]. Freudenberg



Figure 1 Examples of dataset images.

proposed a deep learning-based framework for identifying individual tree crowns in satellite and aerial data. The method improves the precision of identification by generating irregular polygons instead of bounding boxes [24]. Wang proposed a building extraction technology based on CNN, which extracts multi-scale and multi-resolution features from remotely-sensed photos by incorporating multi-scale cyclic residual convolution into the generative network [25].

Although existing deep learning models such as UNet and CNN have made achievements in terms of identifying features in built-up urban areas, they still face obvious challenges when processing high-resolution, remotely-sensed images. Although the CNN model has excellent feature extraction capabilities, its fixed size of the convolution kernel limits the capture of multi-scale building features, resulting in unstable performance when processing targets of different scales. Hence, this paper constructs the WFFE-UNet model that effectively solves the problems of the missed detection of small targets and insufficient multi-scale feature fusion through weighted feature fusion and a multi-scale perception improved convolution module, demonstrating the innovation and practical value of the model.

3. BUILT-UP URBAN AREA IDENTIFICATION METHOD BASED ON DEEP LEARNING

3.1 Preprocessing and Enhancement of High-Resolution, Remotely-Sensed Images

(1) Image data collection

The Massachusetts Buildings Dataset (download link: <https://www.cs.toronto.edu/~vmnih/data/>) is a classic dataset used for built-up urban area research. This dataset uses 1-meter-resolution aerial photography images to show the coherent urban landscape of Boston from the bustling commercial center to the quiet suburban residential areas. Using

1,510 images of $1,500 \times 1,500$ pixels, it comprehensively records the diverse urban landscape. The experimental sample set includes training and test samples. Figure 1 displays several examples of the collected dataset.

(2) Image preprocessing

When extracting urban-related targets from remotely-sensed images, it is essential to distinguish between pixels in built-up urban areas and those in non-built-up urban, which requires detailed pixel-level annotation of high-resolution, remotely-sensed images of the selected area. In this paper, the image annotation tool Labelme is used, and the data is carefully annotated according to the standard format of the dataset. Figure 2 shows the results.

Figure 2 shows the original image that was processed using Labelme software. After opening the image, the edges of the city are marked carefully, and labels are assigned to these marks. After completing the marking of all the buildings in the image of the built-up urban area, it is saved.

When identifying built-up urban areas in high-resolution, remotely-sensed images, image quality has a significant impact on the feature extraction and identification performance of subsequent deep learning models [26–27]. However, due to the variability of shooting conditions and the noise that may be captured by the remote sensor, problems such as uneven brightness, low contrast, and high-frequency noise often occur in the image. These interference factors can significantly reduce the model's ability to identify core information such as textures and building area boundaries. To solve this problem, two image enhancement strategies are used when preprocessing the image: histogram equalization and principal component analysis (PCA) noise reduction. These two methods improve the quality of original remotely-sensed images from the perspectives of illumination compensation and noise suppression. Histogram equalization is used to improve the overall contrast of the image. Stretching the grayscale value distribution of the original image makes it more evenly distributed in the entire grayscale space, thereby improving the visibility of details in the darker parts of the image [28–30]. The expression is:

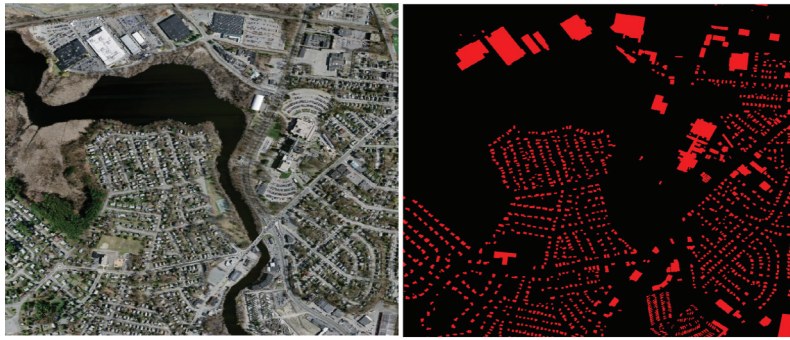


Figure 2 Image annotation.



Figure 3 Effects of image quality enhancement.

$$L_{eq}(x, y) = C(I(x, y)) * (L - 1) \quad (1)$$

where, $I(x, y)$ represents the gray value of the original image pixel; C is the cumulative distribution function of the image gray value; L is the gray level of the image; $L_{eq}(x, y)$ is the output image after equalization.

In this study, PCA is used to reduce image noise by compressing redundant channels in multispectral or hyperspectral remotely-sensed images and separating noise features [31]. PCA is used to perform feature decomposition on the covariance matrix between image channels, which can extract the components that carry the main information and filter out those components dominated by noise and with low variance. The equation used for PCA noise reduction is:

$$X_{PCA} = W^T(X - \gamma) \quad (2)$$

where, X represents the pixel matrix of the original image; γ is the mean vector of each dimension; W is the eigenvector matrix obtained after eigenvalue decomposition; X_{PCA} is the image representation with the main components retained after dimensionality reduction. Figure 3 shows the image enhancement effect.

Figure 3 shows the effects of using histogram equalization and the PCA technique to improve image quality. Figure A is the original image, and Figure B is the enhanced outcome. To improve the contrast of the image, the image data of each band undergoes histogram equalization. The processed three-dimensional image data is converted into a two-dimensional matrix for PCA processing, thereby reducing the amount of noise and redundant information.

3.2 WFFE-UNet Model Architecture Design

This paper constructs a WFFE-UNet model to solve the problem of sample scarcity and image complexity in high-resolution, remotely-sensed image identification of built-up urban areas. This model is based on the UNet framework and improves the convolution module and feature integration based on the characteristics of remotely-sensed images. In the convolution module, multi-scale perception is applied to better capture multi-scale structures such as buildings and roads. A weighted series integration method is designed for feature integration. By adopting learnable fusion parameters, the contribution of different levels of identified features is dynamically adjusted, solving the problem of insufficient utilization of traditional UNet features. In addition, to improve identification precision, a fusion-enhanced UNet (FE-UNet) is applied. By fusing and weighting the feature maps of each layer of the encoder, the network can more comprehensively absorb information at different semantic levels and reduce semantic conflicts caused by differences between layers.

The feature integration structure involves feature fusion, which integrates feature layers according to the channel dimension so as to reuse features. At each feature fusion node, a lightweight weight prediction subnetwork is designed. This subnetwork takes the feature group to be fused as input, captures the global statistical information of the feature map through global average pooling, and then passes through two layers of fully-connected layers and a

nonlinear activation, and finally outputs the fusion weight of each feature map. Unlike the traditional series connection method, before concatenating in the channel dimension, each feature map to be fused is multiplied by its corresponding weight coefficient, which not only retains the information of all feature maps but also achieves the adaptive screening of features through weights.

Multi-level feature integration can be mathematically expressed as:

$$\tilde{Y}_I = \text{concat}(Y_1, \dots, Y_L) \quad (3)$$

where, $\tilde{Y}_I$ is the feature layer after feature integration; $\text{concat}()$ is the feature layer integration function; Y_i is the feature layer that needs to be integrated ($i \in [1, L]$).

In CNN, due to multiple convolution operations and activation function processing, there are obvious semantic differences between feature layers at all levels. Shallow features focus more on image details such as color and shape, and deep features are rich in semantic information. Directly integrating them may introduce invalid or interfering information. To address this problem, this paper uses the weighted feature integration (WFI) structure, which can dynamically adjust the contribution of different levels of features in the fusion through learnable weights. This not only retains the effective information of features at each level, but also reduces the interference of useless information through the weight mechanism, thereby improving the efficiency and accuracy of feature integration. The innovation of WFFE-UNet lies in the way that feature integration has been transformed from traditional passive reception to active regulation. Before the feature map is concatenated, an adaptive weight adjustment process is added. For feature map F_i from the encoder i layer and feature map D_i of the corresponding decoder layer, dynamic weight coefficients are generated through a lightweight weight prediction module. These coefficients change according to the content of the input image. First, the feature dimension is compressed through the first layer of full connection to improve efficiency, and then restored to the original number of channels through the second layer of full connection. The weight is limited to 0-1 by the Sigmoid function. The weighted feature maps are concatenated in the channel dimension to form new fusion features.

U-Net has a simple structure and limited use of feature layers, integrating only those feature layers that are symmetrical from beginning to end [32–34]. This paper uses FE-UNet to make full use of the feature layers extracted by the encoder. FE-UNet deeply mines the information of features at all levels through multi-level feature fusion, thereby improving the precision of small target identification and boundary identification. For the integration, bilinear interpolation is used to scale the layers based on the layer size in the decoder, thereby making full use of the feature layers extracted by the encoder while adding a small number of convolution parameters. FE-UNet effectively narrows the gap between feature layers at different levels by means of a weighted series integration structure, further improving the performance of the model. The specific expression is as follows:

$$F_{fused}^{(k)} = \sum_{i=1}^L \text{Resize}(F_i^{(e)}, g_d) \quad (4)$$

where, $F_{fused}^{(k)}$ represents the fused feature map on the k -th decoder level; $F_i^{(e)}$ is the feature map extracted by the i -th encoder level; $\text{Resize}(F_i^{(e)}, g_d)$ represents the feature map adjusted to the size g_d of the d -th decoding layer using bilinear interpolation; L is the total number of encoder layers.

FE-UNet uses a weighted series connection mechanism to control the weight of the integrated features:

$$F_{output}^{(k)} = \vartheta\left(\sum_{j=1}^M w_j * F_{fused,j}^{(k)}\right) \quad (5)$$

where, $F_{fused,j}^{(k)}$ represents the j -th feature map after fusion at the k -th decoding layer; $w_j \in [0, 1]$ is the fusion weight obtained through the learnable mechanism; $\vartheta(\cdot)$ represents the nonlinear activation function; $F_{output}^{(k)}$ is the final feature output after fusion of this layer.

3.3 CBAM Attention Mechanism and Multi-Scale Optimization

When identifying features in high-resolution, remotely-sensed images of built-up urban areas, the accurate capture of small targets, such as narrow roads, small buildings, and street facilities, is important for the precision of the overall identification [35]. However, deep learning models find it difficult to accurately identify small targets in complex backgrounds due to their small size and weak expression of information. Especially in built-up areas, small targets are dense, and they have unclear edges and various shapes, which increases the difficulty of identification. This paper incorporates CBAM into WFFE-UNet to improve the identification precision of deep learning models for small targets. CBAM contains both channel attention and spatial attention modules. It can enhance the model's sensitivity to the key features of small targets and can also flexibly allocate resources in feature maps, helping the model to better focus on small targets in complex or low-contrast backgrounds, thereby solving the problem of insufficient expression of shallow features.

3.3.1 CBAM Attention Mechanism

The channel attention submodule generates two matrix vectors of dimension $C \times 1 \times 1$ by applying maximum pooling and average pooling to the fusion feature F [36–38]. These two vectors are then processed by the shared weight WFFE-UNet, and then two vectors measuring $3C \times 1 \times 1$ are output. M_c represents the channel attention map obtained by adding these two matrix vectors element by element. The expression is:

$$M_c = \text{MLP}(\text{Max Pool}(F)) + \text{MLP}(\text{Avg Pool}(F)) \quad (6)$$

$$F_c = a * F_1 + b * F_2 + c * F_3 \quad (7)$$

where, a , b , and c represent the attention weights obtained by the original feature maps F_1 , F_2 , and F_3 after applying the Softmax layer, reflecting the contribution of different channels to the edge detail features of the building. The original feature maps F_1 , F_2 , and F_3 are element-wise multiplied with the attention weights in different channels to obtain feature map F_c .

Firstly, the input feature map is compressed in the channel dimension to obtain two feature maps measuring $H \times W \times 1$. These two feature maps have their own characteristics: one captures the most significant response in the entire channel, and the other reflects the average response. By concatenating these two in the channel dimension, a more comprehensive $H \times W \times 2$ spatial feature representation is formed. This concatenation result is input into a convolutional layer equipped with a 7×7 convolution kernel to deeply mine local spatial information and generate a spatial attention map. After processing, it is activated by the Sigmoid function so that the value range is between $[0,1]$, indicating the importance of each spatial position.

3.3.2 Multi-scale Optimization

This paper adopts a multi-scale optimization strategy to enhance the WFFE-UNet model's ability to identify multi-scale building features in remotely-sensed images. Since traditional convolution operations are limited by the size of fixed convolution kernels, it is difficult to effectively capture target features of different scales in images. Therefore, by incorporating a multi-scale optimization strategy, the model can more flexibly adapt to and identify building features of various scales, thereby improving the performance of the model. This paper constructs a cross-level feature interaction network (CFIN) to solve the problem of multi-scale feature optimization. CFIN cleverly integrates shallow high-resolution features with deep semantic features by means of a learnable attention mechanism, thereby greatly improving the model's ability to capture multi-scale building features. In the decoding stage, CFIN innovatively adds two modules, feature alignment and feature enhancement, to address the problem of spatial information distortion in multi-scale feature fusion. The feature alignment module uses deformable convolution to intelligently adjust the spatial relationship of feature maps, ensuring that spatial information can be accurately aligned during multiple sampling processes, significantly reducing information distortion. The feature enhancement module effectively filters noise and strengthens key features by combining residual connections with 1×1 convolutions, which not only improves the model's sensitivity to multi-scale features but also improves its stability in complex backgrounds. Through these innovative designs, CFIN successfully integrates multi-scale optimization strategies into the built-up urban area identification model, significantly improving the identification precision and efficiency of the model and providing a new solution for high-precision identification of features in built-up urban areas.

In each upsampling stage, the feature enhancement module uses residual connections and 1×1 convolutions to suppress noise and enhance expression capabilities. Its update expression is as follows:

$$F_{enhanced} = Conv_{1 \times 1}(F_{input}) + F_{input} \quad (8)$$

F_{input} represents the feature map after upsampling in the current stage; $Conv_{1 \times 1}$ denotes a 1×1 convolution operation; $F_{enhanced}$ is the enhanced feature map.

3.4 Loss Function and Cross-Domain Training Strategy

3.4.1 Loss Function

The spatial information and ground details contained in high-resolution images are extremely complex, which can make model training difficult. This is because a single loss function often fails to strike a balance between overall semantic understanding and pixel-level precision. In this paper, the cross-entropy (CE) loss and the Dice loss (DL) are jointly optimized during the training of the deep learning network. This not only improves the convergence speed of the model, but also significantly improves the precision of identification of built-up areas. Given a set of training images and corresponding true values, the core goal is to optimize the objective function of the precise change detection network. Since the change detection task can be regarded as a pixel-level binary classification problem, the cross-entropy loss function is applied during the training phase to effectively measure the difference between the predicted value and the true value, ensuring that the generated change map is as close to the actual situation as possible. Its equation is:

$$L_{CE} = -\frac{1}{m} \sum_{k=1}^m (a_k \log \hat{a}_k + (1 - a_k) \log (1 - \hat{a}_k)) \quad (9)$$

where, m represents the number of samples; a_k represents the true value of a given sample; $\hat{a}_k$ is the predicted probability value.

In the identification of built-up urban areas in remotely-sensed images, CE loss is used to measure the difference between the predicted and the true values. However, when faced with extremely unbalanced categories, such as when the area of built-up areas and non-built-up areas is very different, CE loss may cause small category areas to be ignored. Therefore, it is supplemented by Dice loss, which is originally used to measure the degree of set overlap and can intuitively reflect the spatial overlap between the prediction and the true label. By combining these two loss functions, the problem of category imbalance can be handled more effectively, and the identification accuracy of a few target areas can be improved, thereby optimizing the training effect and performance of the model. Dice loss is expressed with:

$$L_{Dice} = 1 - \frac{1}{n} \sum_{i=1}^n \frac{2 \sum_{j=1}^M \hat{a}_{i,j} a_{i,j}}{\sum_{j=1}^M \hat{a}_{i,j} + \sum_{j=1}^M a_{i,j}} \quad (10)$$

where, $\hat{a}_{i,j}$ and $a_{i,j}$ represent the predicted image and the true label, respectively. The two loss functions are mixed as the optimization objective of the network, and the expression is as follows:

$$L = L_{CE} + L_{Dice} \quad (11)$$

Table 1 Experimental results of learning rate comparison.

Learning rate	Training epochs	Training set accuracy	Test set accuracy	Training time (minutes)
0.01	6	0.9887	0.9742	20.41
0.001	6	0.9903	0.9836	21.58
0.0001	94	0.9997	0.9924	158.62
0.00001	52	0.9795	0.9806	131.28

Table 2 Experimental results of batch size comparison.

Batch size	Training epochs	Training set accuracy	Test set accuracy	Training time (minutes)
2	18	0.9987	0.9941	116.34
4	46	0.9993	0.9945	294.14
6	73	0.9994	0.9955	359.72
8	99	0.9997	0.9991	451.63

During the optimization process, the Dice loss function pays special attention to the quality of the overall prediction area and performs well in dealing with imbalanced categories. It can significantly improve the precision of identifying small targets or boundary areas. By jointly using cross-entropy and Dice loss functions, the model can not only learn the differences between different semantic categories from a global perspective, but can also more precisely depict the boundaries and contours of built-up areas in local areas, thereby improving the accuracy of the final identification results. During the model training process, key hyperparameters are fine-tuned to improve training efficiency and model stability. Learning rate, as a key factor in controlling the parameter update step size, has a decisive influence on the convergence speed and performance of the model. Through the warm-up strategy and the method of gradual decay of the learning rate, rapid exploration in the early stage of training and stable optimization in the later stage are achieved. At the same time, the batch size is also adjusted to balance the amount of information and computational efficiency in gradient updates to further improve the performance of the model. Tables 1 and 2 list the learning rates and batch size adjustments.

The data in Table 1 shows the impact of different learning rates on model training and performance. When the learning rate is 0.01, although the training time is the shortest, the test set accuracy is not high, indicating that the model does not fully learn the data features. When the learning rate is reduced to 0.001, the test set accuracy is significantly improved, showing a better performance balance, indicating that the model's learning ability and generalization performance are the strongest under this setting. When the learning rate is reduced to 0.0001, although the test set accuracy is the highest, the training time increases significantly, and the performance is not balanced. When the learning rate is too low, the model's convergence speed is significantly reduced, and the accuracy also decreases, indicating that too low a learning rate is not conducive to model optimization. Considering the training efficiency and stability, the configuration of a learning rate of 0.001 achieves good model optimization

while ensuring high precision, which is the best choice for this paper.

In Table 2, the batch size is very important in the model training process and has a significant impact on training efficiency and model performance. When the batch size is set to 2, the model quickly reaches a training set accuracy of 0.9987 and a test set accuracy of 0.9941 after only 18 training epochs. The entire training process takes 116.34 minutes. This result highlights the advantage of better training efficiency for small batch settings. When the batch size is increased to 8, although the training epochs and time increase to 99 epochs and 451.63 minutes, respectively, the accuracy of the model on the test set is significantly improved to 0.9991, which fully demonstrates the role of the maximum batch size in optimizing model performance. However, after comprehensively considering the adjustment strategy of the learning rate, it is found that although the accuracy of large batches is relatively high, in the long term, the configuration of batch size 2 is a better choice for obtaining more stable training and better final performance.

3.4.2 Cross-Domain Training Strategy

This paper designs and implements a cross-domain training strategy to further verify the WFFE-UNet model's generalization ability and robustness under diverse conditions. Boston is selected as the training dataset, and detailed data cleaning and annotation work are carried out. The model is fully trained using the rich urban geomorphological information in the Boston dataset, enabling it to deeply learn the general features and identification patterns of built-up urban areas. Subsequently, the trained model is applied to remotely-sensed images of cities such as Beijing, Chengdu, and New York. The significant differences in geographical features, climatic conditions, and urban planning layouts of these cities provide a very challenging testing environment for the model. During cross-domain testing, the consistency of the model parameters is maintained, and only necessary preprocessing operations are performed on the input images so as to adapt to the differences between different datasets.

Table 3 Sample distribution of training set and test set.

Identified target	Training set	Test set	Total
B	125	32	157
P	98	25	123
G	142	36	178
I	117	29	146
R	135	34	169
W	108	27	135
S	126	31	157
Total	851	214	1065

Table 4 Experimental environment and hardware configuration.

Serial number	Name	Parameter
1	Learning rate	0.001
2	Weight decay	0.0001
3	Non-maximum suppression	0.7
4	Operating system	Ubuntu18.04
5	Compiled language	Python3.6
6	Deep learning framework	Pytorch1.5
7	CPU (Central Processing Unit)	AMD Ryzen 7 4700U with Radeon Graphics 2.00 GHz

4. EXPERIMENTAL ANALYSIS OF BUILT-UP URBAN AREA IDENTIFICATION MODEL

4.1 High-resolution, Remote Sensing Experimental Dataset and Experimental Environment

4.1.1 Dataset

Since the observation altitude of satellites can exceed 30,000 meters when taking high-resolution remotely-sensed images, medium and large targets commonly seen in daily life, such as single buildings (B) and small squares (P), are defined as small targets in remotely-sensed images because they occupy very few pixels in the image. Green space and square land (G), industrial land (I), residential land (R), water areas (W), and road networks (S) have more obvious spatial scales in remotely-sensed images and are considered medium and large targets. This scale shift places higher demands on the task of identifying features in built-up urban areas. This paper selects seven built-up urban areas from the dataset studied as research objects. These objects cover typical urban functional areas, which are helpful in more finely depicting the urban spatial structure and enabling the automatic identification of built-up areas. Table 3 lists detailed results.

In Table 3, the training set contains 851 samples, and the test set contains 214 samples, making a total of 1,065. In high-resolution remotely-sensed images, although the two types of targets, B and P, occupy a small proportion of pixels and are defined as small targets, they are important in identifying urban functional areas. Therefore, these two types of targets still account for a considerable proportion of the number of samples. Medium and large targets with more obvious spatial characteristics, such as G, I, R, W,

and S, are more widely distributed in the samples, which undoubtedly helps the model better learn and understand the spatial structural characteristics of these targets, thereby further improving the overall identification precision of built-up urban areas.

4.1.2 Experimental Environment

This experiment uses a deep learning model. Table 4 gives details about the experimental hardware environment and deep learning environment.

Table 4 lists the deep learning environment and hardware configuration used in the experiment, including key training parameters and system settings. In terms of software settings, the learning rate is set to 0.001, and the weight decay is set to 0.0001. This configuration helps to improve the model's convergence speed and generalization performance. At the same time, the non-maximum suppression threshold is set to 0.7 to further optimize the detection results.

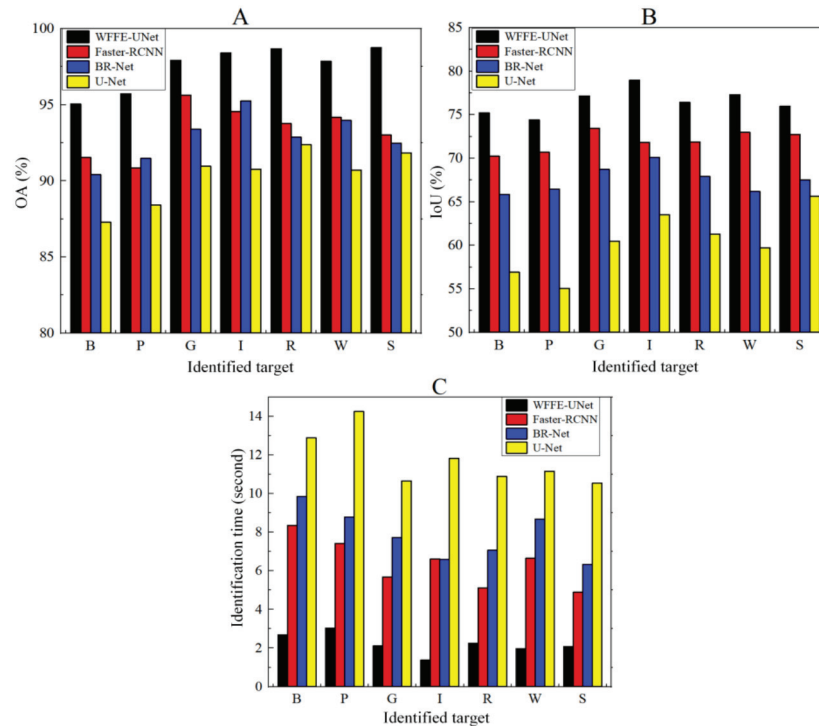
4.2 Identification Performance Indicators

4.2.1 Comparative Experiment

The performance of the WFFE-UNet model studied in this paper is comprehensively evaluated in terms of its efficiency when identifying features in built-up urban areas. Several current mainstream identification models are selected for comparative experiments: Faster Region-based Convolutional Neural Network (Faster-RCNN), Boundary Refinement Network (BR-Net), Fully Convolutional Early Fusion (FC-EF), Dual-Task Constrained Deep Siamese Convolutional Network (DTC-DSCN), Fusion Enhanced U-Net (FE-U-Net), etc. These models are trained and tested on the same remotely-sensed image dataset under unified training

Table 5 Performance comparison of different models in the task of built-up urban area identification.

Algorithm model	OA (%)	Precision (%)	Recall (%)	F1 (%)	IoU (%)
WFFE-UNet	97.34	96.76	95.42	96.08	82.65
Faster-RCNN	93.12	92.94	92.17	92.55	75.89
BR-Net	90.53	90.91	90.46	90.68	66.42
FC-EF	91.04	90.27	90.09	90.18	68.33
DTC-DSCN	93.86	92.48	91.57	92.02	76.25
FE-U-Net	92.07	90.13	89.32	89.72	72.14
UNet	88.50	85.81	84.93	85.37	58.88

**Figure 4** Comparison of the identification effects of different identification models on selected targets. (A) OA, (B) IoU, (C) Identification time.

conditions to ensure the fairness of the comparison. In the evaluation process, indicators such as overall accuracy (OA), precision, recall, IoU, and F1-score are used to conduct research to comprehensively measure the performance of each model. Table 5 shows the specific research results.

Table 5 shows that the WFFE-UNet model has apparent advantages in all evaluation indicators. The OA of the WFFE-UNet model is 97.34%, far exceeding other compared models, proving the extremely high accuracy of the model in the task of built-up urban area identification. The precision of WFFE-UNet is 96.76%, indicating that the actual correct proportion of pixels predicted as built-up urban areas is very high, effectively reducing misjudgments. The recall of WFFE-UNet is 95.42%, which is higher than that of other models, indicating that it can more comprehensively identify built-up urban area pixels in the image and reduce the risk of missed detection. In terms of the two comprehensive performance indicators, IoU and F1-score, the values of WFFE-UNet are 82.65% and 96.08%, respectively, further highlighting its excellent identification ability. The outstanding performance of the WFFE-UNet model is due to its ability to more effectively extract and utilize key

information from remotely-sensed images by combining feature weighting and frequency enhancement mechanisms, especially having a higher identification sensitivity to the subtle differences between built-up urban areas and the surrounding environment. At the same time, the improved U-Net architecture enhances the ability to fuse features and capture contextual information, further improving the identification precision and generalization performance of the model.

This paper selects seven typical targets from the dataset for experiments to more comprehensively evaluate the identification effect of the WFFE-UNet model in different types of urban functional area targets. By comparing and analyzing the performance of the model on OA, IoU, and identification time, the WFFE-UNet model's excellent performance and practicality in various scenes are further verified. Figure 4 presents the specific research results.

In Figure 4, WFFE-UNet is compared with other models, and it is found that it has obvious advantages in terms of OA, IoU, and identification time when identifying the seven selected research objects. In Figure 4A, the OA values of WFFE-UNet for various types of targets are higher than those

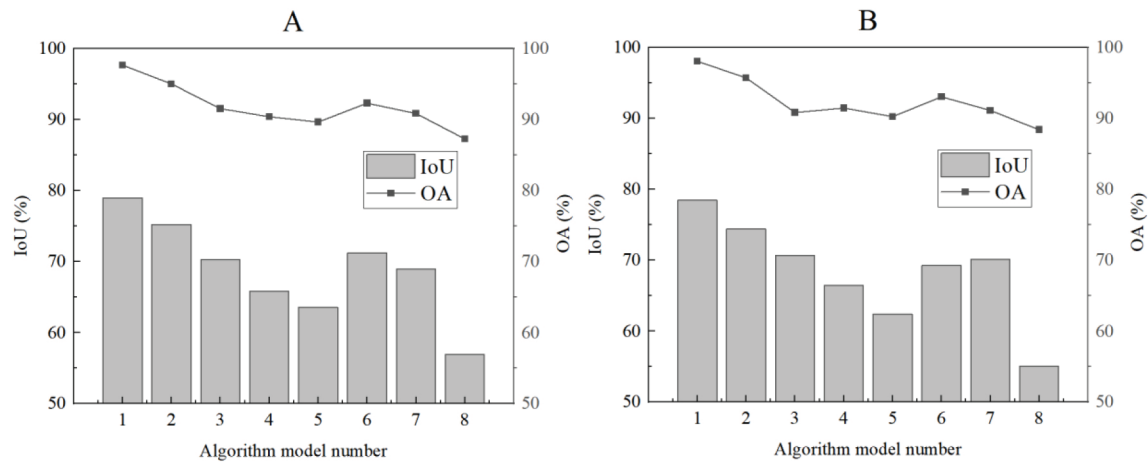


Figure 5 Performance comparison of different models on small target identification. (A) Single buildings (B), (B) Small squares (P).

of other models, especially in the identification of medium and large targets such as industrial land (I) and residential land (R). WFFE-UNet shows excellent performance, with OA reaching 98.41% and 98.68%, respectively. This shows that the WFFE-UNet model has excellent feature expression capabilities for land objects with obvious spatial structures. In Figure 4B, WFFE-UNet performs well in the IoU values of various functional areas, especially in residential areas, road networks, and water areas, where its IoU values far exceed those of other models. This shows the high reliability and accuracy of WFFE-UNet in practical applications. Compared with other models, WFFE-UNet shows stronger generalization ability and adaptability when dealing with complex and changeable targets in built-up urban areas. In Figure 4C, WFFE-UNet also has significant advantages in regard to identification time, and its overall identification speed is stable within 3.1 seconds, showing high real-time performance and practicality. By adding a weighted feature fusion mechanism, the WFFE-UNet model can better integrate multi-scale information so that it can effectively capture key details even when facing small targets. The edge enhancement module is used to optimize the identification precision of the built-up area boundary, thereby improving the IoU. In summary, WFFE-UNet achieves an excellent balance of precision, stability, and efficiency and is very suitable for the automatic identification and fine characterization of complex built-up urban areas in high-resolution, remotely-sensed images.

This paper integrates the CBAM module into the WFFE-UNet network, which significantly improves the accuracy of small target identification. To verify the performance of this method, single buildings (B) and small squares (P) with an area of less than 32×32 pixels are selected as research objects. The performance of different methods in small-scale built-up area object identification is quantified by the two evaluation indicators of IoU and OA. The experimental results are compared with those of WFFE-UNet, Faster-RCNN, BR-Net, FC-EF, DTC-DSCN, FE-U-Net, and classic U-Net. The above algorithms are numbered from 1 to 8. Figure 5 shows the experimental results.

In Figure 5, the method studied in this paper achieves good results in both single-building and small square target

identification tasks. In Figure 5A, in the single-building identification task, the IoU and OA of the method in this paper reach 78.98% and 97.68%, respectively, which are 3.75% and 2.64% higher than the second-best WFFE-UNet. The IoU of U-Net is only 56.91%, showing its obvious inadequacy in capturing small target features. In Figure 5B, for the target P with a more discrete spatial distribution, the proposed method is also superior, with IoU and OA of 78.46% and 98.07%, respectively. In terms of OA, it is 5.01% higher than DTC-DSCN, which shows that the proposed method has a strong identification ability for irregular square contours. Although WFFE-UNet outperforms other traditional models in building identification, it is still inferior to the proposed method, further confirming that the model is effectively improved for small target feature extraction and expression after applying the CBAM module. The method in this paper applies the CBAM attention mechanism to the WFFE-UNet model and highlights the key features in remotely-sensed images by weighted enhancement of channel and spatial dimension information, thereby improving the identification of small targets. By deeply fusing CBAM, the method in this paper provides assistance for fine-grained target identification in high-resolution, remotely-sensed images.

An original remotely-sensed image containing rich urban landform information is selected as an evaluation benchmark to study the performance of different deep learning models in the task of identifying built-up urban areas in high-resolution, remotely-sensed images. The image covers various complex landform types such as buildings, roads, and green spaces, providing a real test environment for the evaluation of model performance. Four deep learning models, WFFE-UNet, Faster-RCNN, BR-Net, and U-Net, are used to process the image, and the identification results for each model are recorded in detail. Figure 6 shows the specific research results.

In Figure 6, Figure A is the original remotely-sensed image, and Figures B-E show the results of built-up urban area identification using four different models: WFFE-UNet, Faster-RCNN, BR-Net, and U-Net. After comparative analysis, WFFE-UNet (Figure B) has the best effect, as it precisely captures the boundaries and shapes of buildings. It is highly consistent with the original image, with very few

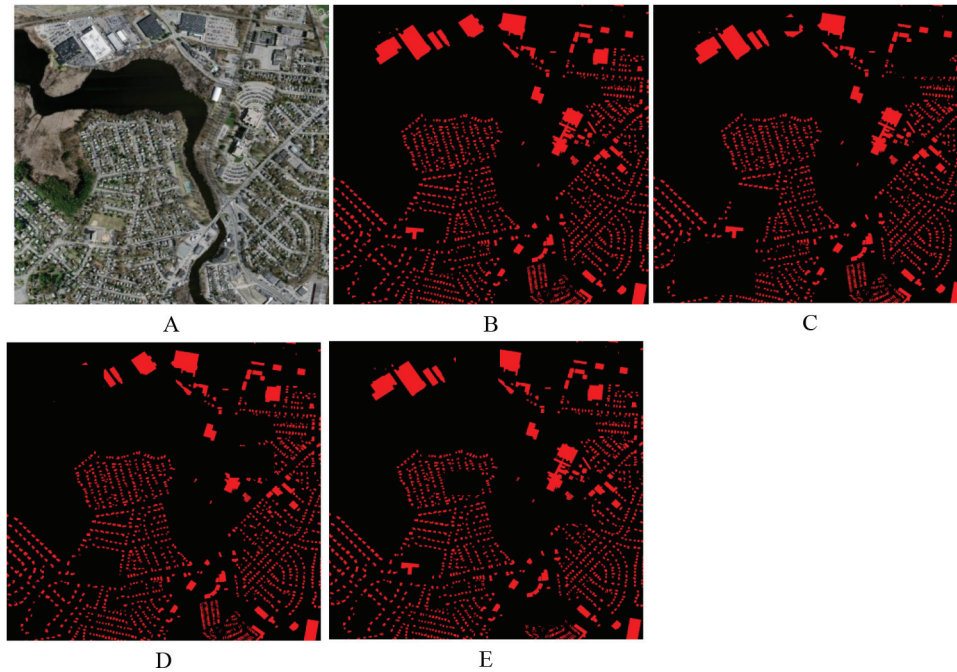


Figure 6 Comparison of identification performance of different models in the task of built-up urban area identification.

missed detections and false detections, indicating that WFFE-UNet has obvious advantages in regard to feature extraction and target positioning. Although Faster-RCNN (Figure C) can identify most buildings, it lacks detail, has blurred edges, and a small amount of mislabeling. For small or irregular buildings, the identification stability is not as good as that of WFFE-UNet. BR-Net (Figure D) performs even worse, with more false detections and missed detections, which may be related to the limitations of its feature fusion and multi-scale processing capabilities. U-Net (Figure E) performs the worst, with a large amount of noise and false targets, blurred boundaries, and much lower precision than WFFE-UNet. In summary, WFFE-UNet, with its powerful features and positioning capabilities, has shown significant advantages in the identification of built-up urban areas, providing reliable support for subsequent urban planning and management.

4.2.2 Cross-Domain Generalization Test

A cross-domain identification experiment is designed to evaluate the WFFE-UNet model's generalization ability and robustness when identifying built-up areas in remotely-sensed images. In this experiment, several cities with significant differences in terms of geography, climate, and urban structure are selected as test areas, including Beijing, Chengdu, and New York. First, the WFFE-UNet model is trained using the Boston-related remotely-sensed images in the dataset, and then tested on the data of the other cities to verify its generalization performance. Figure 7 shows the specific results.

In Figure 7, WFFE-UNet demonstrates excellent identification performance and cross-domain adaptability in tests in multiple cities. In Figure 7A, on the Boston training set, the OA and F1 values of the WFFE-UNet model are 97.34% and 96.08%, respectively, far exceeding those of other

models, highlighting its excellent performance in standard environments. In Figure 7B, in the Beijing test, despite the challenges of seasonal changes and building density, WFFE-UNet maintains high performance close to the training set, demonstrating strong urban adaptability, with OA, F1-score, and IoU of 96.86%, 95.93%, and 80.64%, respectively. In Figure 7C, for the foggy Chengdu, the model also performs well, with an OA of 98.42%, an F1-score of 98.03%, and an IoU of 83.11%, significantly ahead, proving the robustness of WFFE-UNet to low visibility and atmospheric interference. In Figure 7D, in the test for New York, the WFFE-UNet model is also stable, showing good generalization ability across geography and urban forms. The performance of other models generally declines in cross-city tests, especially in cities with large environmental differences, showing the limitations of their feature extraction capabilities. The advantage of WFFE-UNet lies in its multi-layer feature fusion strategy, which enhances the model's semantic expression ability and enables it to maintain high discrimination in cross-domain data. In general, WFFE-UNet not only performs well in a single city, but also demonstrates strong generalization ability and environmental adaptability in cross-domain, remote-sensing tasks, providing strong support for the automatic identification of built-up urban areas.

5. CONCLUSION

This paper uses deep learning technology to construct an efficient and accurate identification method for the problem of built-up urban area identification in high-resolution, remotely-sensed images. Through experimental verification, it is found that the combination of WFFE-UNet and CBAM has significant advantages in dealing with complex urban environments and small target identification. WFFE-UNet

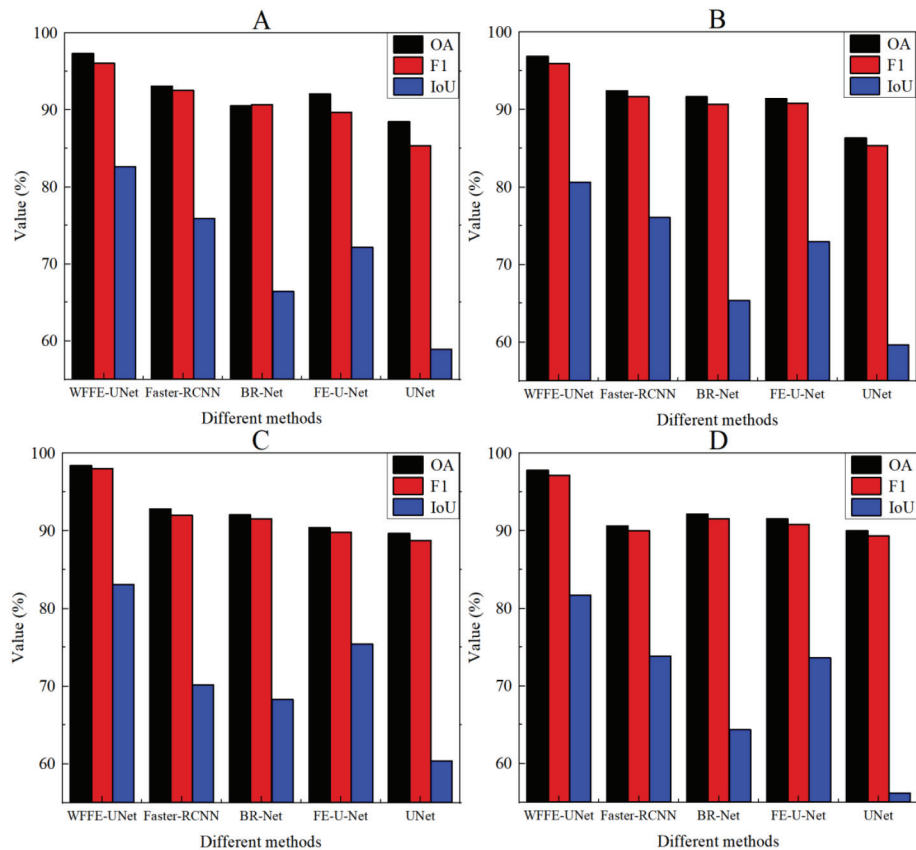


Figure 7 Comparison of the generalization capabilities of different identification models in cross-domain identification tasks. (A) Boston, (B) Beijing, (C) Chengdu, (D) New York.

improves the model's efficiency in utilizing multi-scale features and features of different levels through multi-scale convolution modules and weighted feature integration, thereby significantly improving the precision of identifying features in built-up urban areas. In built-up areas, this method shows higher robustness than traditional methods due to its automatic feature learning and background noise suppression capabilities. After incorporating the CBAM attention mechanism, the model can focus more on small targets in complex backgrounds, significantly improving the accuracy of identifying small targets. Cross-domain identification experiments show that the proposed model can maintain stable performance under different conditions, showing good generalization ability and environmental adaptability. In summary, this study not only successfully addresses the limitations of traditional methods in the identification of built-up urban areas in high-resolution, remotely-sensed images: it also provides technical support for the construction of urban dynamic monitoring through an innovative model design.

FUNDING

This research was supported by the 14th Five-Year Plan for Educational Science in Shaanxi Province in 2023 (No. SGH23Y2905). The project is titled "Research on the Construction of Curriculum System for the Major of Computer Science and Technology in Application-oriented Undergraduate Universities Facing New Engineering Education".

REFERENCES

- Han, Wei, Chen, J., Wang, L., Feng, R., Li, F., Wu, L. et al. "Methods for small, weak object detection in optical high-resolution remote sensing images: A survey of advances and challenges". *IEEE Geoscience and Remote Sensing Magazine*, 9.4 (2021): 8–34. DOI: 10.1109/MGRS.2020.3041450.
- Chen, Bingyu, Xia, M., Qian, M., & Huang, J. "MANet: A multi-level aggregation network for semantic segmentation of high-resolution remote sensing images". *International Journal of Remote Sensing*, 43.15–16 (2022): 5874–5894. <https://doi.org/10.1080/01431161.2022.2073795>.
- Zhang, Hongyan, Lin, M., Yang, G., & Zhang, L. "ESNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images". *IEEE Transactions on Neural Networks and Learning Systems*, 34.1 (2021): 28–42. DOI: 10.1109/TNNLS.2021.3089332.
- Adegun, Adeganmi Adeyinka, Serestina Viriri, and Jules-Raymond Tapamo. "Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis". *Journal of Big Data*, 10.1 (2023): 93. <https://doi.org/10.1186/s40537-023-00772-x>.
- Kislov, Dmitry E., Korznikov, K. A., Altman, J., Vozmishcheva, A. S., & Krestov, P. V. "Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images". *Remote Sensing in Ecology and Conservation*, 7.3 (2021): 355–368. <https://doi.org/10.1002/rse2.194>.
- Zhang, Shichao, Wang, C., Li, J., & Sui, Y. "MF-Dfnet: A deep learning method for pixel-wise classification of very high-resolution remote sensing images". *Interna-*

- tional Journal of Remote Sensing, 43.1 (2022): 330–348. <https://doi.org/10.1080/01431161.2021.2018147>.
7. Peng, Daifeng, Bruzzone, L., Zhang, Y., Guan, H., Ding, H., & Huang, X. “SemiCDNet: A semi-supervised convolutional neural network for change detection in high resolution remote-sensing images”. *IEEE Transactions on Geoscience and Remote Sensing*, 59.7 (2020): 5891–5906. DOI: 10.1109/TGRS.2020.3011913.
 8. Wang, Guojie, Wu, M., Wei, X., & Song, H. “Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks”. *Remote Sensing*, 12.5 (2020): 795. <https://doi.org/10.3390/rs12050795>.
 9. Lv, Xianwei, Shao, Z., Ming, D., Diao, C., Zhou, K., & Tong, C. “Improved object-based convolutional neural network (IOCNN) to classify very high-resolution remote sensing images”. *International Journal of Remote Sensing*, 42.21 (2021): 8318–8344. <https://doi.org/10.1080/01431161.2021.1951879>.
 10. Ghodieh, Ahmed. “Built-up urban area estimation and change detection of the occupied West Bank, Palestine, using multi-temporal aerial photographs and satellite images”. *Journal of the Indian Society of Remote Sensing*, 48.2 (2020): 235–247. <https://doi.org/10.1007/s12524-019-01073-8>.
 11. Li, Jiayi, Huang, X., Tu, L., Zhang, T., & Wang, L. “A review of building detection from very high resolution optical remote sensing images”. *GI Science & Remote Sensing*, 59.1 (2022): 1199–1225. <https://doi.org/10.1080/15481603.2022.2101727>.
 12. Guo, Wenkai, Yang, G., Li, G., Ruan, L., Liu, K., & Li, Q. “Remote sensing identification of green plastic cover in built-up urban areas”. *Environmental Science and Pollution Research*, 30.13 (2023): 37055–37075. <https://doi.org/10.1007/s11356-022-24911-8>.
 13. Xie, Lijian, Feng, X., Zhang, C., Dong, Y., Huang, J., & Liu, K. “Identification of urban functional areas based on the multimodal deep learning fusion of high-resolution remote sensing images and Social Perception Data”. *Buildings*, 12.5 (2022): 556. <https://doi.org/10.3390/buildings12050556>.
 14. Chen, Jie, Dai, X., Guo, Y., Zhu, J., Mei, X., Deng, M., et al. “Urban built environment assessment based on scene understanding of high-resolution remote sensing imagery”. *Remote Sensing*, 15.5 (2023): 1436. <https://doi.org/10.3390/rs15051436>.
 15. Cheng, C. U. I., Lu, Z. H. A. O., Hongyan, R. E. N., Weili, L. U., & Yaohuan, H. U. A. N. G. “Integrating high-resolution remote sensing and street view images to identify urban villages: A case study in Yuexiu District, Guangzhou City”. *National Remote Sensing Bulletin*, 26.9 (2023): 1802–1813. <https://dx.doi.org/10.11834/jrs.20210202>.
 16. Liu Zhao, Liao Feifan, and Zhao Tong. “Remote sensing image built-up urban area extraction and optimization method based on PSPNet”. *Remote Sensing for Land & Resources*, 32.4 (2020): 84–89. DOI: 10.6046/gtzyyg.2020.04.12.
 17. Wang Ke, Xiao Pengfeng, Feng Xuezhai, Wu Guiping, & Li Hui. “Extraction of urban rivers from high spatial resolution remotely sensed imagery based on filtering in the frequency domain”. *NATIONAL REMOTE SENSING BULLETIN*, 17.2 (2021): 277–285. <https://dx.doi.org/10.11834/jrs.20131331>.
 18. Barkur, Rahasya, Suresh, D., Lal, S., Reddy, C. S., & Diwakar, P. G. “RSCDNet: A robust deep learning architecture for change detection from bi-temporal high resolution remote sensing images”. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7.2 (2022): 537–551. DOI: 10.1109/TETCI.2022.3230941.
 19. Nurkarim, Wahidya, and Arie Wahyu Wijayanto. “Building footprint extraction and counting on very high-resolution satellite imagery using object detection deep learning framework”. *Earth Science Informatics*, 16.1 (2023): 515–532. <https://doi.org/10.1007/s12145-022-00895-4>.
 20. Zhang, Ju, Hu, Q., Li, J., & Ai, M. “Learning from GPS trajectories of floating car for CNN-based urban road extraction with high-resolution satellite imagery”. *IEEE Transactions on Geoscience and Remote Sensing*, 59.3 (2020): 1836–1847. DOI: 10.1109/TGRS.2020.3003425.
 21. Zeng, Yifu, Yi Guo, and Jiayi Li. “Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning”. *Neural Computing and Applications*, 34.4 (2022): 2691–2706. <https://doi.org/10.1007/s00521-021-06027-1>.
 22. Song, Shiran, Liu, J., Liu, Y., Feng, G., Han, H., Yao, Y., et al. “Intelligent object recognition of urban water bodies based on deep learning for multi-source and multi-temporal high spatial resolution remote sensing imagery”. *Sensors*, 20.2 (2020): 397. <https://doi.org/10.3390/s20020397>.
 23. Zheng, Xiaoxiong, and Tao Chen. “High spatial resolution remote sensing image segmentation based on the multiclassification model and the binary classification model”. *Neural Computing and Applications*, 35.5 (2023): 3597–3604. <https://doi.org/10.1007/s00521-020-05561-8>.
 24. Freudenberg, Maximilian, Paul Magdon, and Nils Nolke. “Individual tree crown delineation in high-resolution remote sensing images based on U-Net”. *Neural Computing and Applications*, 34.24 (2022): 22197–22207. <https://doi.org/10.1007/s00521-022-07640-4>.
 25. Wang, Zhen, Xu, N., Wang, B., Liu, Y., & Zhang, S. “Urban building extraction from high-resolution remote sensing imagery based on multi-scale recurrent conditional generative adversarial network”. *GIScience & Remote Sensing*, 59.1 (2022): 861–884. <https://doi.org/10.1080/15481603.2022.2076382>.
 26. Ding, Wei, Li Zhang, and Guangliang Yang. “Building detection algorithm in multi-scale remote sensing images based on attention mechanism”. *Evolutionary Intelligence*, 16.5 (2023): 1717–1728. <https://doi.org/10.1007/s12065-023-00849-9>.
 27. Zhu, Panpan, Tan, Y., Zhang, L., Wang, Y., Mei, J., Liu, H. et al. “Deep learning for multilabel remote sensing image annotation with dual-level semantic concepts”. *IEEE Transactions on Geoscience and Remote Sensing*, 58.6 (2020): 4047–4060. DOI: 10.1109/TGRS.2019.2960466.
 28. Agrawal, Sanjay, Panda, R., Mishro, P. K., & Abraham, A. “A novel joint histogram equalization based image contrast enhancement”. *Journal of King Saud University-Computer and Information Sciences*, 34.4 (2022): 1172–1182. <https://doi.org/10.1016/j.jksuci.2019.05.010>.
 29. Vijayalakshmi, Dhurairajan, and Malaya Kumar Nath. “A novel contrast enhancement technique using gradient-based joint histogram equalization”. *Circuits, Systems, and Signal Processing*, 40.8 (2021): 3929–3967. <https://doi.org/10.1007/s00034-021-01655-3>.
 30. Mondal, Kalimuddin, Rinku Rabidas, and Rajdeep Dasgupta. “Single image haze removal using contrast limited adaptive histogram equalization based multiscale fusion technique”. *Multimedia Tools and Applications*, 83.5 (2024): 15413–15438. <https://doi.org/10.1007/s11042-021-11890-0>.
 31. Hasan, Basna Mohammed Salih, and Adnan Mohsin Abdulazeez. “A review of principal component analysis algorithm for dimensionality reduction”. *Journal of Soft Computing and Data Mining*, 2.1 (2021): 20–30. <https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/8032>.
 32. Su, Zhongbin, Li, W., Ma, Z., & Gao, R. “An improved U-Net method for the semantic segmentation of remote sens-

- ing images". *Applied Intelligence*, 52.3 (2022): 3276–3288. <https://doi.org/10.1007/s10489-021-02542-9>.
33. Yin, MeiJie, Wang, P., Hao, W., & Ni, C. "Cloud detection of high-resolution remote sensing image based on improved U-Net". *Multimedia Tools and Applications*, 82.16 (2023): 25271–25288. <https://doi.org/10.1007/s11042-023-14655-z>.
 34. Zhang, Guanjin, Roslan, S. N. A. B., Wang, C., & Quan, L. "Research on land cover classification of multi-source remote sensing data based on improved U-net network". *Scientific Reports*, 13.1 (2023): 16275. <https://doi.org/10.1038/s41598-023-43317-1>.
 35. Wang, Yetong, Hang, Guoz, Xing, Kongduo, Alfred, Kongduo. "Automatic classification and recognition of spatiotemporal high-resolution image data based on deep neural networks". *Engineering Intelligent Systems*, 30.4(2024): 309–317.
 36. Yu, Wenling, Liu, B., Liu, H., & Gou, G. "Recurrent residual deformable conv unit and multi-head with channel self-attention based on U-Net for building extraction from remote sensing images". *Remote Sensing*, 15.20 (2023): 5048. <https://doi.org/10.3390/rs15205048>.
 37. Liu, Jingyi, and Xiaomin Yang. "Remote sensing image super-resolution based on cross residual compensation mechanism and attention mechanism". *Multimedia Tools and Applications* 83.18 (2024): 55641–55657. <https://doi.org/10.1007/s11042-023-17645-3>.
 38. Shi, Aiye, Sen Wang, and Xin Wang. "Self-supervised remote sensing image change detection based on high frequency feature and gate attention-guided optimization unit". *Journal of Applied Remote Sensing*, 17.2 (2023): 024518–024518. <https://doi.org/10.1117/1.JRS.17.024518>.