



## Multi-Resolution Feature Extraction and Fusion for Traditional Village Landscape Analysis in Remote Sensing Imagery



Qian Zhang<sup>1</sup>, Jing Zhang<sup>2</sup>, Shuang Lu<sup>1</sup>, Yi Liu<sup>1\*</sup>, Lei Liu<sup>1</sup>, Yingyi Wang<sup>1</sup>, Mingyu Cao<sup>1</sup>

<sup>1</sup> School of Art and Design, Zhengzhou University of Light Industry, Zhengzhou 450002, China

<sup>2</sup> Henan Provincial Civil Affairs School, Zhengzhou 450002, China

Corresponding Author Email: [2004009@email.zzuli.edu.cn](mailto:2004009@email.zzuli.edu.cn)

<https://doi.org/10.18280/ts.400344>

### ABSTRACT

**Received:** 2 February 2023

**Accepted:** 16 May 2023

#### Keywords:

*remote sensing images, traditional villages, landscape feature extraction*

The complexity and diversity of traditional village landscapes present significant challenges to remote sensing image analysis. Existing methods, such as pixel-level classification, object-based image analysis (OBIA), and deep learning techniques, are often computationally intensive and require powerful hardware support and optimization algorithms. To address these issues, a landscape feature analysis model based on multi-resolution feature extraction and fusion with attention pyramid decoding is proposed in this study. By employing multi-scale feature extraction and fusion, this model captures landscape features at various levels and scales, enabling more comprehensive and in-depth analysis of complex remote sensing images. Additionally, the attention pyramid decoding approach adaptively mines spatial and semantic information, enhancing the model's focus on pertinent features and consequently improving classification accuracy. Experimental results confirm the effectiveness of the proposed model for traditional village landscape analysis in remote sensing imagery.

## 1. INTRODUCTION

The rapid advancement of remote sensing technology has established remote sensing imagery as a crucial means for obtaining surface information [1-4]. Traditional village landscapes, as unique cultural heritages, embody rich historical, cultural, and natural values [5-8]. Effective feature extraction and analysis of these landscapes can offer vital technical support for their protection, planning, and utilization. However, the complexity and diversity of traditional village landscapes present significant challenges to remote sensing image analysis. This study investigates a method for extracting and applying features of traditional village landscapes based on remote sensing image analysis.

Existing remote sensing image analysis methods primarily include pixel-level classification, object-based image analysis (OBIA), and deep learning techniques. Pixel-level classification typically involves supervised or unsupervised classification of images based on their spectral information, but its performance in extracting complex traditional village landscape features is hindered by limitations and noises in spectral information [9-11]. While OBIA overcomes some of the pixel-level classification method's shortcomings by segmenting images into meaningful objects and classifying them, it still struggles with low extraction accuracy and high computational demands when dealing with intricate and diverse traditional village landscapes [12].

In recent years, deep learning techniques have demonstrated considerable success in remote sensing image analysis [13-18], with convolutional neural networks (CNNs) in particular exhibiting remarkable feature extraction and image classification capabilities. Nonetheless, deep learning methods exhibit certain limitations when analyzing traditional village

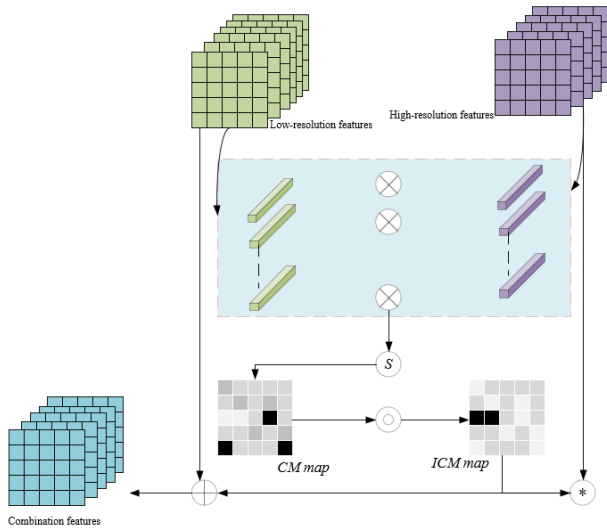
landscapes in remote sensing images [19-22]. These include insufficient ability to extract both high- and low-resolution features, resulting in limited identification accuracy, as well as high computational requirements and the need for powerful hardware support and optimization algorithms, which constrain their widespread practical application [23-25].

Considering the current research landscape and the limitations of existing remote sensing image analysis methods, this study focuses on the extraction and application of traditional village landscape features using remote sensing image analysis. Section 2 presents a multi-resolution feature balance strategy for extracting and fusing low- and high-resolution remote sensing image features in traditional village landscape feature extraction scenarios. In section 3, a landscape feature analysis model based on attention pyramid decoding is proposed. This model adaptively mines spatial and semantic information, enhancing the model's focus on pertinent features and, consequently, improving classification accuracy. Experimental validation demonstrates that the proposed model exhibits substantial advantages in feature extraction and classification accuracy compared to previous models.

## 2. MULTI-RESOLUTION FEATURE EXTRACTION AND FUSION

In traditional village landscape feature extraction scenes, high-resolution features have high spatial resolution, and provide richer detailed information for feature extraction, which helps identify and extract typical landscape features more accurately, such as architectural forms, spatial layout, and road network. Due to low spatial resolution, low-

resolution features cannot provide detailed information at the same level, but they provide a larger range of spatial information, which helps make a more global analysis of landscape features. When the two kinds of features are fused, the advantages of high- and low-resolution remote sensing images can be comprehensively utilized to improve the feature extraction accuracy (Figure 1).



**Figure 1.** Schematic diagram of a multi-resolution feature balance strategy

Remote sensing images of traditional village landscapes in practical applications may be influenced by various factors, such as atmospheric conditions, illumination variation, and sensor capability, which may lead to decreased image quality and affect the feature extraction accuracy. By balancing low- and high-resolution features, the impact of these unfavorable factors on feature extraction results was reduced, thus improving the model’s generalization ability. When the balanced fusion features were sent to the classification network, it effectively improved the network’s performance.

The remote sensing image feature correlation  $V_{uk}$  of traditional village landscapes was quantified using the sigmoid function of pixel-level cosine similarity, as shown in the following equation. Let  $D_{me}^{u,k} \in E^{1 \times 1 \times V}$  and  $D_{ae}^{u,k} \in E^{1 \times 1 \times V}$  be the feature vectors of low- and high-resolution remote sensing image features at  $(u,k)$ . The higher the value of  $V_{e_{u,k}}$ , it was be considered as a stronger feature among the high-resolution remote sensing image features of traditional village landscapes.

$$V_{e_{u,k}} = \text{Sigmoid} \left( \frac{D_{me}^{u,k} \cdot D_{ae}^{u,k}}{\|D_{me}^{u,k}\| \|D_{ae}^{u,k}\|} \right) \quad (1)$$

$Cr$  was mapped based on inverse correlation metric, multiplied by  $D_{ae} \in E^{Q \times G \times V}$ , and differentially integrated with  $D_{me}$ . Let  $W, H$  and  $C$  be the width, height, and number of channels of corresponding features, respectively;  $Cr \in R^{W \times H}$  be the correlation metric graph,  $D_v$  be the combination features, and  $*$  be the pixel-level multiplication, then there was an equation as follows:

$$D_v = D_{me} + (1 - Cr) * D_{ae} \quad (2)$$

There were few weak features in the high-resolution

features  $D_{ae}$  at this time, and the weak features in  $D_{ae}$  were fully utilized by multiplying  $(1 - Cr)$  with  $D_{ae}$ . To verify the effectiveness of balancing low- and high-resolution features, the following comparison scheme of centralized feature integration was used for verification. Let  $\oplus$  be the pixel-level addition operation. Combination features  $D_v$  were obtained by adding  $D_{me}$  and  $D_{ae}$ , then there was the following equation:

$$D_v = D_{me} \oplus D_{ae} \quad (3)$$

Let  $\cup$  be the join operation of feature channel, and  $Conv$  be the convolution operation. If  $D_v$  was obtained by joining  $D_{me}$  with  $D_{ae}$ , there was the following equation:

$$D_v = \text{Conv}(D_{me} \cup D_{ae}) \quad (4)$$

The third scheme was to monitor low-resolution features through relationship transfer loss, i.e., to improve low-resolution features by transferring context relationship of high-resolution features. Let  $A_{ae}$  and  $A_{me}$  be the high- and low-resolution feature similarity matrices, respectively;  $D'$  be a channel feature,  $u, k$  be the pixel position index, and  $Q$  and  $G$  be the length and width of the feature matrix. The adopted relationship transfer loss was calculated by the following equations:

$$M_{EY} = \frac{1}{Q^2 G^2} \sum_{u=1}^{QG} \sum_{k=1}^{QG} \|A_{ae}(u, k) - A_{me}(u, k)\|_1 \quad (5)$$

$$A(u, k) = \left( \frac{D'(u)}{\|D'(u)\|_2} \right)^y \left( \frac{D'(k)}{\|D'(k)\|_2} \right) \quad (6)$$

If the combination features  $D_v$  were obtained through attention feature weighting operation of  $D_{me}$  and  $D_{ae}$ , the attention weighting operation was as follows:

$$D_v = D_{me} \oplus (D_{ae} * \text{Sigmoid}(CN(D_{ae} \cup D_{me}))) \quad (7)$$

where,  $*$  is the pixel-level multiplication operation.

L1 loss was used to supervise LR features. The features were improved by transferring the knowledge of SR features, and L1 loss was used in the process:

$$M_1 = \|\|VN(D_{ae}) - VN(D_{me})\|_1 \quad (8)$$

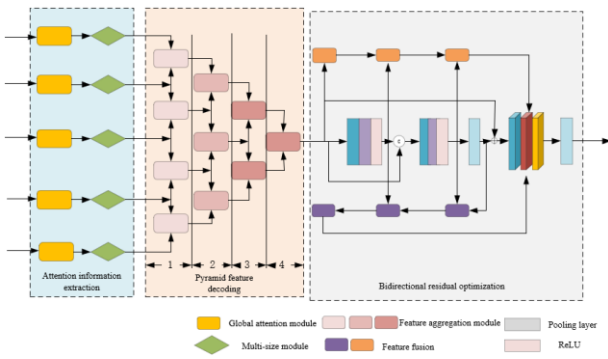
where,  $\|\cdot\|_1$  is the first normal form.

### 3. CONSTRUCTING A LANDSCAPE FEATURE ANALYSIS MODEL BASED ON ATTENTION PYRAMID DECODING

Compared with the conventional CNN pooling layer, a pyramid decoding structure was designed in this study, which reduced feature information loss caused by feature scale reduction, meaning that more original information was retained to improve the feature extraction quality when features were reconstructed in the decoder. The pyramid decoding structure achieved better multi-level feature expression by gradually integrating encoding features at

different levels, which helped fully analyze complex features in remote sensing images and improve the accuracy and robustness of feature extraction. Meanwhile, a bidirectional residual optimization module was introduced in this study, which enhanced the boundary integrity of the salient target object in initial detection results, helping refine the contour of salient target and improve the accuracy of remote sensing image analysis.

The model consisted of four parts, including the SR and LR feature fusion module based on feature balance, attention extraction, feature pyramid decoding, and bidirectional residual optimization, which effectively addressed the challenges in remote sensing image analysis based on the characteristics of remote sensing image feature extraction of traditional village landscapes. ResNet was taken as the backbone model, which helped provide rich feature information for subsequent modules, such as feature fusion, attention extraction, and pyramid decoding, because it had strong feature extraction ability as a deep CNN. The SR and LR feature fusion module based on feature balance improved the accuracy and robustness of feature extraction by integrating the advantages of high- and low-resolution remote sensing images, which was of great significance for extracting complex traditional village landscape features. Contact information between different types of targets was extracted by utilizing the global attention mechanism, which helped the model better capture salient feature targets during the feature extraction stage, thus improving the accuracy of remote sensing image analysis. The feature pyramid decoding structure fully integrated feature information at different levels and improved the resolution of remote sensing images during the decoding process, which helped more accurately extract the salient features of traditional village landscapes and effectively segmented foreground and background. The bidirectional residual optimization module optimized the initial salient prediction graph, and enhanced the boundary integrity of the salient target object, thus improving the accuracy of remote sensing image analysis (Figure 2).



**Figure 2.** Structure diagram of the constructed model

The global attention mechanism captured the feature dependencies of the entire dataset, leading to better understanding the connection between different types of targets, which helped improve the model's ability to distinguish between different scene features and various types of salient targets during the feature extraction stage, thus improving the accuracy of remote sensing image analysis. Dimension of the input fusion feature  $d^1 \in E^{V1 \times G1 \times Q1}$  was first converted to  $R^{V1 \times B1}$ , with  $B1 = G1 \times Q1$ . Then the feature connection between  $d^1$  and two external storage units ( $L_g \in E^{B1 \times j}$

and  $L_v \in E^{j \times B1}$ ) was calculated separately. Finally feature relationship graphs  $S_a \in E^{V1 \times j}$  and  $S_v \in E^{V1 \times B1}$  were output:

$$S_a = d^1 L_g \quad (9)$$

$$S_v = S_a L_v \quad (10)$$

Then the dimension of  $S_v$  was converted to  $E^{V1 \times G1 \times Q1}$  again. Let  $\sigma$  be the learnable hyper-parameter, then the calculation equation of the output global attention features was as follows:

$$E = d^1 + \sigma \cdot S_v \quad (11)$$

The pyramid decoding structure fused feature information at different levels, including shallow and deep features. The former contained more detailed information, while the latter contained more semantic information. This integration of multi-scale features enabled the model to better extract salient targets from remote sensing images. Compared with the single branch CNN method, the pyramid decoding structure more accurately detected salient targets by fusing features at different levels, which helped improve the accuracy of remote sensing image analysis, thus providing more reliable technical support for protecting, planning, and utilizing traditional village landscapes.

The key component of the pyramid decoding structure was the feature aggregation module. Let  $conv(\cdot)$  be the convolutional layer,  $CA(\cdot)$  be the channel concatenation operation,  $UP(\cdot)$  be the upsampling operation, and  $u \in \{0, 1, 2, 3\}$  be the depth of the feature aggregation module from top to bottom in the pyramid encoding structure. Calculation process of the feature aggregation module in the first stage from left to right was given by the following equation:

$$L_1^u = Conv(CA(P^{a+1}, UP(P^a))) \quad (12)$$

Let  $v \in \{1, 2, 3\}$  be the remaining stages. Calculation process of the feature aggregation module in the remaining stages was given by the following equation:

$$L_v^u = Conv(CA(L_{a-1}^{u-1}, UP(L_{a-1}^u))) \quad (13)$$

The bidirectional residual optimization module captured the bidirectional connection between multiple layers, which enhanced the structural integrity of the target object in the initial salient prediction graph, helping improve the accuracy and reliability of the model in extracting traditional village landscape features. The bidirectional residual optimization module not only captured information transmission from shallow to deep layers, but also transmitted information reversely from deep to shallow layers, which enabled the model to better preserve the boundary information of the target object, thus improving the accuracy of object segmentation and identification.

The bidirectional residual optimization module in this study consisted of two convolutional layers, a batch standardization layer, and a ReLU layer. After the two residual modules generated features  $J^e_1$  and  $J^e_2$  respectively, let  $X_n$  be the backward output features, and  $X_d$  be the forward output features, then there were:

$$X_n = SN(J_0^e, SN(J_0^e, J_1^e)) \quad (14)$$

$$X_d = SN(J_0^e, SN(J_1^e, J_0^e)) \quad (15)$$

Let  $conv(\cdot)$  be the convolutional layer,  $CA(\cdot)$  be the channel concatenation operation, and  $SN(\cdot)$  be the feature aggregation operation, then there were:

$$SN = Conv(CA(J_u^e, J_k^e, CA(J_u^e, J_k^e))) \quad (16)$$

#### 4. EXPERIMENTAL RESULTS AND ANALYSIS

Two different types of sample sets were set up in this study for the extraction and application study of traditional village landscape features based on remote sensing image analysis. The first was the regional classification sample set, which divided traditional village landscape features into different categories based on geographical location and regional characteristics. Specifically, the traditional villages in

Northeast China included their local landscape features, such as various traditional buildings, courtyards, and streets, etc. The traditional villages in South China included their local landscape features, such as Lingnan architectural style, and watery place features, etc. The second was the architectural style classification sample set, which divided traditional village landscape features into different categories based on architectural style and historical and cultural background. For example, the residential building sample set included remote sensing images of quadrangle dwellings, earth buildings of Hakka, cave dwellings and other residential building styles. The religious architecture sample set included remote sensing images of various religious architectural styles, such as temples, Taoist temples, and churches. By setting these two different types of sample sets, in-depth research was conducted on the traditional village landscape features in remote sensing images from different perspectives, thus better addressing various challenges faced by protecting, planning, and utilizing traditional villages. At the same time, this also helped improve the generalization ability of the model based on remote sensing image analysis in practical applications.

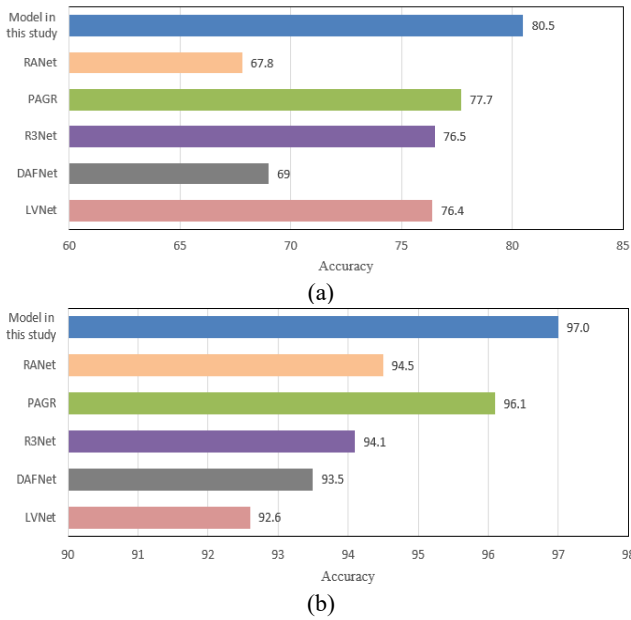
**Table 1.** TOP=1 accuracy (%) results of different feature balance methods

Feature categories	DA	WL	GANs	Method in this study
01. Architectural forms	100.0	100.0	100.0	100.0
02. Roads and alleys	87.3	72.1	73.2	70.3
03. Village boundaries	53.9	95.3	85.4	85.3
04. Waters or water systems	92.8	94.3	89.4	92.6
05. Farmlands	83.4	89.5	92.5	83.5
06. Green spaces and vegetation	85.2	48.3	85.4	89.3
07. Topography and landform	64.9	72.4	78.4	74.3
08. Sites and historical and cultural heritages	87.3	68.4	83.5	85.3
09. Residential areas	100.0	75.8	74.5	73.1
10. Public facilities and services	79.7	85.6	84.3	85.0
11. Transportation and communication facilities	86.7	49.3	82.5	93.1
12. Land use and cover	74.7	92.6	93.7	91.3
Overall accuracy	84.7	83.5	82.0	99.3

According to the data in Table 1, the TOP-1 accuracy (%) results of different feature balance methods, namely, data augmentation (DA), WL, generative adversarial networks (GANs), and the method in this study, can be compared and analyzed in the traditional village landscape feature extraction scenes based on remote sensing image analysis. As shown in the Table 1, DA shows a high accuracy in most feature categories, and especially reaches an accuracy of over 80% in architectural forms, waters or water systems, farmlands, green spaces and vegetation. However, the accuracy of DA is relatively low in roads and alleys, topography and landform. WL shows a high accuracy in village boundaries, waters or water systems, farmlands, public facilities and services, land use and cover, etc., while its accuracy is relatively low in roads and alleys, green spaces and vegetation, transportation and communication facilities. GANs exhibit a high accuracy in farmlands, green spaces and vegetation, topography and landform, sites and historical and cultural heritages, transportation and communication facilities, land use and cover. However, its accuracy is relatively low in roads and alleys, village boundaries, and residential areas, etc. The method in this study shows a high accuracy in all feature categories, and especially reaches an accuracy of more than 70% in several aspects, such as architectural forms, roads and alleys, village boundaries, waters or water systems, farmlands, green spaces and vegetation, topography and landform, sites and

historical and cultural heritages, residential areas, public facilities and services, transportation and communication facilities, land use and cover, etc. The overall accuracy reaches 99.3%, which is significantly better than that of the other three methods.

According to Figure 3, the TOP-1 accuracy (%) results of different models, namely, latent variable network (LVNet), dense attention fluid network (DAFNet), recurrent residual refinement network (R3Net), PAGR, resolution adaptive network (RANet), and the model in this study, in the regional classification sample set 1 can be compared and analyzed in the traditional village landscape feature extraction scenes based on remote sensing image analysis. As shown in the Figure 3, the accuracy of the LVNet model is 76.4%, which is acceptable, but there is still room for improvement. The DAFNet model has a relatively low accuracy of 69%, and optimization or parameter adjustment is needed. The accuracy of the R3Net model is 76.5%, which is comparable to the performance of LVNet, but there is still room for improvement. The accuracy of the PAGR model is 77.7%, which performs better compared with other models, but there is still some room for improvement. The RANet model has an accuracy of 67.8%, which performs the worst among these models, requiring further optimization or adjustment. The accuracy of the model in this study is 80.5%, which performs the best among all models with a high accuracy.



**Figure 3.** TOP=1 accuracy (%) comparison of different models (a) sample set 1 (b) sample set 2

**Table 2.** Comparative experimental results of different models

Methods	Sample set 1			Sample set 2		
	MAE↓	F-measure↑	S-measure↑	MAE↓	F-measure↑	S-measure↑
Model in this study	0.0123	0.938	0.947	0.0069	0.846	0.940
RANet	0.0132	0.974	0.912	0.0604	0.884	0.956
PAGR	0.0583	0.764	0.701	0.0493	0.784	0.710
R3Net	0.0382	0.762	0.803	0.0195	0.782	0.839
DAFNet	0.0382	0.772	0.889	0.0284	0.792	0.894
LVNet	0.0285	0.863	0.873	0.0198	0.802	0.874

According to the data in Table 2, the comparative experimental results, including mean absolute error (MAE), F-measure, and S-measure, of different models (e.g., the model in this study, RANet, PAGR, R3Net, DAFNet, and LVNet) in sample sets 1 and 2 can be compared and analyzed in the traditional village landscape feature extraction scenes based on remote sensing image analysis. In the two sample sets, the model in this study performs the best in MAE, and also performs well in F-measure and S-measure, indicating that the model has advantages in terms of accuracy and stability. RANet has the best performance in F-measure in sample set 1, and a relatively poor performance in MAE and S-measure, meaning that RANet has advantages in certain landscape categories, with room for improvement in its overall performance. In both sample sets, PAGR generally performs poorly in all indexes and needs further optimization and improvement. R3Net performs relatively evenly in both sample sets, but does not have optimal performance in all indexes, indicating that its overall performance needs to be improved. DAFNet performs well in S-measure in sample set 1, but performs relatively average in other indexes, meaning that DAFNet has certain advantages in structural similarity, with room for improvement in accuracy and stability. LVNet performs relatively well in both sample sets, but does not have optimal performance in all indexes, indicating that there is still room for improvement in its overall performance. It can be seen from the comparative analysis that the model in this study has the most prominent performance in sample sets 1 and 2, with lower MAE and higher F-measure and S-measure, in the

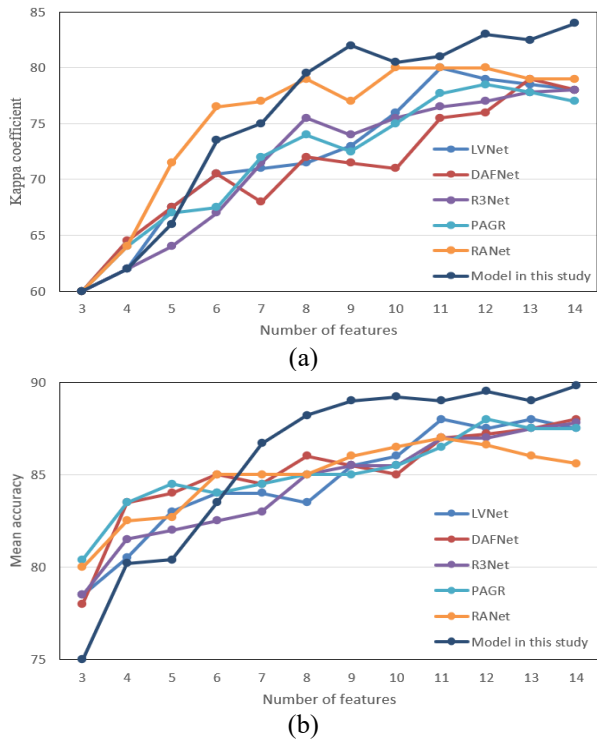
In architectural style classification sample set 2, the accuracy of the LVNet model is 92.6%, which is acceptable, but there is a certain gap compared with other models. The accuracy of the DAFNet model is 93.5%, which performs slightly better than LVNet, but there is still room for improvement. The accuracy of the R3Net model is 94.1%, which performs well among these models, but there is still room for improvement. The accuracy of the PAGR model is 96.1%, which performs better compared with other models with a high accuracy. The accuracy of the RANet model is 94.5%, which performs well among these models, but is still inferior to PAGR and the model in this study. The accuracy of the model in this study is 97.0%, which performs the best among all models with the highest accuracy.

It can be seen that the TOP-1 accuracy of the model in this study is significantly better than the other five models (e.g. LVNet, DAFNet, R3Net, PAGR, and RANet) in sample sets 1 and 2 in the traditional village landscape feature extraction scenes based on remote sensing image analysis, indicating that the model in this study has strong generalization ability and accuracy, and can be used as an effective method for traditional village landscape feature extraction tasks in remote sensing image analysis.

traditional village landscape feature extraction scenes based on remote sensing image analysis, which verifies that the model has strong generalization ability and advantages in terms of accuracy, stability, and structural similarity.

According to the data in Figure 4, the relationship between the classification accuracy of sample sets 1 and 2 and the number of features ranging from 3 to 14, extracted by different models (e.g., LVNet, DAFNet, R3Net, PAGR, RANet, and the model in this study), can be compared and analyzed in traditional village landscape feature extraction scenes based on remote sensing image analysis. In sample set 1, the overall classification accuracy of LVNet shows an upward trend as the number of features increases. The classification accuracy reaches 80 especially when the number of features is 11, and then fluctuates slightly, with a good overall performance. The classification accuracy of DAFNet is relatively low when the number of features is 4 and 6, and then shows a good upward trend as the number of features increases. The classification accuracy reaches 79 when the number of features is 13. The classification accuracy of R3Net shows an upward trend on the whole as the number of features increases, and reaches 78 when the number of features is 14. The classification accuracy of PAGR is relatively low when the number of features is small, and improves as the number of features increases. However, the classification accuracy fluctuates when the number of features is 9, and reaches 78.5 when the number of features is 12. The classification accuracy of RANet is relatively low when the number of features is small, and shows an upward trend on the whole as the number of features

increases. The classification accuracy reaches 80 when the number of features is 12, and then fluctuates slightly, with a good overall performance. The classification accuracy of the model in this study is relatively low when the number of features is small, and shows a clear upward trend as the number of features increases. The classification accuracy reaches the highest of 84 when the number of features is 14.

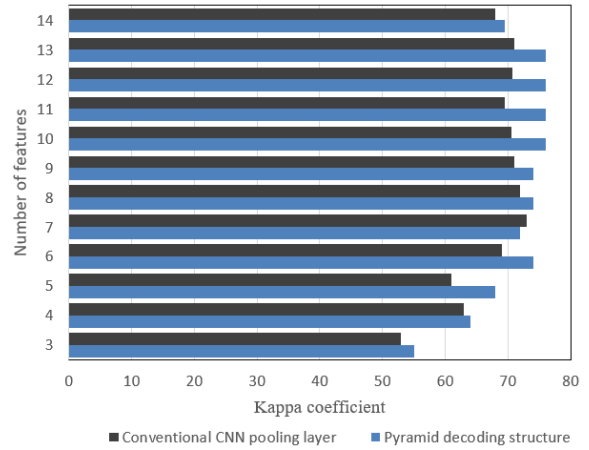


**Figure 4.** Relationship between classification accuracy and the number of features extracted by different models (a) sample set 1 (b) sample set 2

In sample set 2, the classification accuracy of LVNet shows an upward trend on the whole as the number of features increases. The classification accuracy reaches 85.5 when the number of features is 9, and then fluctuates slightly, with a good overall performance. The classification accuracy of DAFNet is relatively low when the number of features is small, and then shows a good upward trend when the number of features increases. The classification accuracy reaches 88 when the number of features is 14. The classification accuracy of R3Net shows an upward trend on the whole as the number of features increases, and reaches 87.8 when the number of features is 14. The classification accuracy of PAGR is relatively high when the number of features is small, and improves as the number of features increases, with a relatively small overall fluctuation. The classification accuracy reaches 88 when the number of features is 12. The classification accuracy of RANet is relatively low when the number of features is small, and shows an upward trend on the whole as the number of features increases. The classification accuracy reaches 86.5 when the number of features is 10, and then fluctuates slightly, with a good overall performance. The classification accuracy of the model in this study is relatively low when the number of features is small, and shows a clear upward trend as the number of features increases. The classification accuracy reaches the highest of 89.8 when the number of features is 14.

In the traditional village landscape feature extraction scenes

based on remote sensing image analysis, it can be seen from the comparative analysis that the overall classification accuracy of each model shows an upward trend as the number of features increases. The model in this study has the most prominent performance with different numbers of features, and has a high classification accuracy in both sample sets.



**Figure 5.** Kappa coefficient in different model structures

According to Figure 5, the Kappa coefficient of different model structures (e.g., conventional CNN pooling layer and pyramid decoding structure) with different numbers of features ranging from 3 to 14 can be compared and analyzed in traditional village landscape feature extraction scenes based on remote sensing image analysis. The Kappa coefficient of the conventional CNN pooling layer fluctuates significantly when the number of features increases from 3 to 14, and is relatively high ranging from 74 to 76 when the number of features is 6, 8, 9, 10, 11, and 12. However, the Kappa coefficient decreases to 69.5 with a poor performance when the number of features is 14. The Kappa coefficient of the pyramid decoding structure has a small fluctuation when the number of features increases from 3 to 14, and is high ranging from 68 to 73 when the number of features is 5, 6, 8, 11, 12, 13, and 14. Compared with the conventional CNN pooling layer, the Kappa coefficient of the pyramid decoding structure performs well and reaches 68 when the number of features is 14. It can be seen from the comparative analysis that the Kappa coefficient of the pyramid decoding structure has a relatively small fluctuation with different numbers of features, and has a relatively stable overall performance. However, the Kappa coefficient of the conventional CNN pooling layer is relatively high with certain numbers of features, but fluctuates greatly. The performance is poor especially when the number of features is 14. Therefore, when dealing with remote sensing image analysis tasks, the pyramid decoding structure proposed in this study is a more stable and reliable choice.



**Figure 6.** Schematic diagram of feature extraction (a) previous moment (b) next moment

Figure 6 shows the remote sensing images of a traditional village landscape in 2005 and 2017. Some significant changes can be effectively captured by combining with the processing of the model in this study. First, the scale and building density in the village have changed, which may be related to population growth and economic development. Second, architectural style and materials have changed, which reflects the development of local culture and technology. Third, topography and vegetation cover have changed, which may be related to environmental protection and urbanization process. These further verified that the model proposed in this study was highly effective in remote sensing image analysis, providing strong support for extracting, monitoring, and planning traditional village landscape features.

## 5. CONCLUSION

A landscape feature analysis model of multi-resolution feature extraction and fusion based on attention pyramid decoding was proposed in this study, which captured landscape features at different levels and scales through multi-scale feature extraction and fusion, thus analyzing complex remote sensing images more comprehensively and deeply. Meanwhile, the method based on attention pyramid decoding adaptively mined spatial and semantic information, and improved the model's attention to features, thus improving the classification accuracy. Compared with previous models, the model proposed in this study exhibited strong advantages in feature extraction and classification accuracy.

Combined with experimental results, the following conclusions were drawn:

1. Remote sensing image analysis of traditional village landscapes was of great significance for understanding regional cultural characteristics, protecting and inheriting cultural heritages, and guiding urbanization process and sustainable development.

2. The landscape feature analysis model of multi-resolution feature extraction and fusion based on attention pyramid decoding proposed in this study exhibited strong advantages in feature extraction and classification accuracy of remote sensing images. Compared with other models, this model had higher generalization ability and effect.

3. As the number of features increased, the overall classification accuracy of each model showed an upward trend. The model proposed in this study exhibited the most prominent performance with different numbers of features and had a high classification accuracy.

4. By comparing the remote sensing images of a traditional village landscape in 2005 and 2017, it was found that the model proposed in this study was highly effective in remote sensing image analysis, which helped analyze and study the variation trend of village landscapes.

In summary, the proposed landscape feature analysis model of multi-resolution feature extraction and fusion based on attention pyramid decoding had strong advantages in remote sensing image analysis and provided strong support for extracting, monitoring, and planning traditional village landscape features. Although selecting the optimal model structure and the number of features still should be determined based on specific tasks and data sets in practical applications, the model proposed in this study is undoubtedly an effective method and is worth further studying and applying in the remote sensing image analysis field.

## ACKNOWLEDGMENT

2023 Henan Provincial Science and Technology Development Plan Soft Science Research Project, Research on the Construction Strategy of Immersive Experience of Central Plains Red Tourism Classic Scenic Spot Based on Cultural and Tourism Integration, Grant No.: 232400411118; 2023 Henan Provincial Department of Education Humanities and Social Science Research Project, Landscape Narrative Research of Central Plains Red Culture Memorial Park Based on Tourists' Perception, Grant No.: 2023-ZDJH-645; 2023 Henan Provincial Department of Education Humanities and Social Science Research Project, Research on Regional Landscape Color Protection and Planning of Mountainous Traditional Villages, Grant No.: 2023-ZDJH-348; 2022 Henan Revitalization Cultural Engineering Cultural Research Project, Research on the Inheritance Strategy of "Yellow River Culture" in the Protection and Development of Traditional Villages in Henan, Grant No.: 2022XWH127; 2022 Henan Provincial Science and Technology Department Key R&D and Promotion Special Project (Science and Technology Research) Project "Research on Selection of Characteristic Resources and Green Development Path in the Construction of Beautiful Countryside in Henan", Grant No.: 222102320323; 2021 Philosophy and Social Science planning project of Henan Province, Research on the strategy of screening, protection and Utilization of rural red cultural resources in Central Plains, Grant No.: 2021BYS051; 2021 Philosophy and Social Science project of Henan Province, Research on the protection of traditional Village landscape features in Henan province, Grant No.: 2021BYS048.

## REFERENCES

- [1] Nanal, W., Hajiarbabi, M. (2023). Captioning Remote Sensing Images Using Transformer Architecture. In 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Bali, Indonesia, pp. 413-418. <https://doi.org/10.1109/ICAIIIC57133.2023.10067039>
- [2] Xiao, R., Zhong, C., Zeng, W., Cheng, M., Wang, C. (2023). Novel convolutions for semantic segmentation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61(3): 1538-1551. <https://doi.org/10.1109/TGRS.2023.3265752>
- [3] Najim, S.A., Ahmed, B.Y. (2022). Insightful visualization of remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 20(1): 1-5. <https://doi.org/10.1109/LGRS.2022.3228874>
- [4] Chen, G., Lu, H., Di, D., Li, L., Emam, M., Jing, W. (2022). StfMLP: Spatiotemporal Fusion Multi-Layer Perceptron for Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 20(3): 431-435. <https://doi.org/10.1109/LGRS.2022.3230720>
- [5] Yin, S. (2022). Study on the Protection and Development of Farming-Type Traditional Village Landscape Under the Concept of Sustainability—A Case Study of Western Henan. In *Proceedings of the 2022 International Conference on Green Building, Civil Engineering and Smart City*, Guilin, China, pp. 299-310. [https://doi.org/10.1007/978-981-19-5217-3\\_29](https://doi.org/10.1007/978-981-19-5217-3_29)
- [6] Liu, J., Wu, X., Zhang, Y., Wang, L. (2023). Visualization system of Hlai ethnic village landscape

- design based on machine learning. *Soft Computing*, 27(14): 10001-10011. <https://doi.org/10.1007/s00500-023-08196-8>
- [7] Jia, X., Liu, R., Qiao, Z. (2023). Optimization design of mountain and water landscape of traditional mountain village based on new robot visual technology. *Journal of Robotics*, 2023, Article ID: 4155090. <https://doi.org/10.1155/2023/4155090>
- [8] Cai, Y., Ding, R., Li, Q. (2023). An improved geographic modeling method to the vulnerability of the traditional village architectural landscape. In *International Conference on Geographic Information and Remote Sensing Technology (GIRST 2022)*, 12552: 171-177. <https://doi.org/10.1117/12.2667412>
- [9] Yan, L., Fan, B., Liu, H., Huo, C., Xiang, S., Pan, C. (2019). Triplet adversarial domain adaptation for pixel-level classification of VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5): 3558-3573. <https://doi.org/10.1109/TGRS.2019.2958123>
- [10] Ming, D., Du, J., Zhang, X., Liu, T. (2013). Modified average local variance for pixel-level scale selection of multiband remote sensing images and its scale effect on image classification accuracy. *Journal of Applied Remote Sensing*, 7(1): 073565-073565. <https://doi.org/10.1117/1.JRS.7.073565>
- [11] Hossain, M.D., Chen, D. (2019). Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150: 115-134. <https://doi.org/10.1016/j.isprsjprs.2019.02.009>
- [12] Zheng, C., Hu, C., Chen, Y., Li, J. (2023). A Self-Learning-Update CNN model for semantic segmentation of remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 20(9): 1593-1597. <https://doi.org/10.1109/LGRS.2023.3261402>
- [13] Zhao, L., Jiao, J., Yang, L., Pan, W., Zeng, F., Li, X., Chen, F. (2023). A CNN-based layer-adaptive GCPs extraction method for TIR remote sensing images. *Remote Sensing*, 15(10): 2628. <https://doi.org/10.3390/rs15102628>
- [14] Arun, P.V., Buddhiraju, K.M., Porwal, A., Chanussot, J. (2020). CNN based spectral super-resolution of remote sensing images. *Signal Processing*, 169: 107394. <https://doi.org/10.1016/j.sigpro.2019.107394>
- [15] Konstantinidis, D., Argyriou, V., Stathaki, T., Grammalidis, N. (2020). A modular CNN-based building detector for remote sensing images. *Computer networks*, 168: 107034. <https://doi.org/10.1016/j.comnet.2019.107034>
- [16] Song, G., Wang, Z., Bai, L., Zhang, J., Chen, L. (2020). Detection of oil wells based on faster R-CNN in optical satellite remote sensing images. In *Image and Signal Processing for Remote Sensing XXVI*, 11533: 114-121. <https://doi.org/10.1117/12.2572996>
- [17] Han, C., Li, G., Ding, Y., Yan, F., Bai, L. (2020). Chimney detection based on faster R-CNN and spatial analysis methods in high resolution remote sensing images. *Sensors*, 20(16): 4353. <https://doi.org/10.3390/s20164353>
- [18] Yang, J., Zhi, J., Zhang, Y., Wu, J., Zhou, Y., Zuo, B. (2020). Small aircraft target detection using cascade FP-CNN in remote sensing images. In *IET International Radar Conference (IET IRC 2020)*, 2020: 609-614. <https://doi.org/10.1049/icp.2021.0797>
- [19] Wei, Y., Zhang, K., Ji, S. (2020). Simultaneous road surface and centerline extraction from large-scale remote sensing images using CNN-based segmentation and tracing. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12): 8919-8931. <https://doi.org/10.1109/TGRS.2020.2991733>
- [20] Gan, Y., You, S., Luo, Z., Liu, K., Zhang, T., Du, L. (2020). Object detection in remote sensing images with mask R-CNN. In *Journal of Physics: Conference Series*, 1673: 012040. <https://doi.org/10.1088/1742-6596/1673/1/012040>
- [21] Nie, X., Duan, M., Ding, H., Hu, B., Wong, E. K. (2020). Attention mask R-CNN for ship detection and segmentation from remote sensing images. *IEEE Access*, 8: 9325-9334. <https://doi.org/10.1109/ACCESS.2020.2964540>
- [22] Wu, W., Gao, X., Fan, J., Xia, L., Luo, J., Zhou, Y. (2020). Improved Mask R-CNN-Based Cloud Masking Method for Remote Sensing Images. *International Journal of Remote Sensing*, 41(23): 8908-8931. <https://doi.org/10.1109/ACCESS.2020.2964540>
- [23] Lei, L., She, Y., Feng, X., Xiong, R., Liu, S. (2020). Aircraft detection of remote sensing images based on faster R-CNN and Yolov3. In *2020 International Conference on Culture-oriented Science & Technology (ICCST)*, Beijing, China, pp. 166-170. <https://doi.org/10.1109/ICCST50977.2020.00038>
- [24] Uss, M., Vozel, B., Lukin, V., Chehdi, K. (2020). Efficient discrimination and localization of multimodal remote sensing images using CNN-based prediction of localization uncertainty. *Remote Sensing*, 12(4): 703. <https://doi.org/10.3390/rs12040703>
- [25] Dong, Z., Lin, B. (2020). Learning a robust CNN-based rotation insensitive model for ship detection in VHR remote sensing images. *International Journal of Remote Sensing*, 41(9): 3614-3626. <https://doi.org/10.1080/01431161.2019.1706781>