TEMCA-Net: A Texture-Enhanced Deep Learning Network for Automatic Solar Panel Extraction in High Groundwater Table Mining Areas

Min Tan, Weiqiang Luo, Jingjing Li, and Ming Hao

***Abstract*—Long-term coal mining has led to a series of ecological problems, constraining society's sustainable development. Ecological restoration is a crucial component of achieving sustainability, and with the continuous advancement of photovoltaic technology, the comprehensive utilization of photovoltaics has become one of the important restoration methods in mining areas. The area and location of solar panels, as key indicators for assessing the ecological restoration approach, require precise extraction and positioning. This article proposes a texture-enhanced multicontext attention network (TEMCA-Net). In the encoding part, the network utilizes residual connections in conjunction with the convolutional block attention module to preliminarily extract contextual information. Then, low-level features were input into the statistical texture learning (STL) texture enhancement module and high-level features into the horizontal atrous spatial pyramid pooling (H-ASPP) module. In the decoding part, the high-level features processed by the H-ASPP were combined module with the texture-enhanced features from the STL module. Experiments were conducted in the Peibei mining region located in Xuzhou City, Jiangsu Province. We established the solar panels of Peibei mining region (SPPMR) dataset. The experimental results on the SPPMR dataset demonstrate TEMCA-Net's outstanding performance in solar panel extraction, with precision at 90.24%, recall at 93.07%, an *F*1-score of 91.63%, and a mean intersection over union of 92.21%. It significantly outperforms three classic deep learning networks: Deeplabv3+, U-net, and PSPnet. In summary, this study provides an efficient and feasible solution for the extraction of solar panels in mining areas with high water tables.**

***Index Terms*—Semantic segmentation, solar panel extraction, texture enhancement.**

I. Introduction

COAL, as a crucial energy resource, plays an irreplaceable role in meeting global energy demands. However, with its long-term and excessive extraction, environmental issues arising from coal mining, such as surface subsidence and surface water accumulation, have become increasingly severe [1]. In the early stages, ecological restoration in coal mining subsidence areas primarily involved methods, such as farmland reclamation, afforestation, and the reclamation of wetlands for aquaculture [2]. In recent years, facing the dual challenges of climate change and energy transition, new restoration methods, such as agricultural and photovoltaic complementarity, photovoltaic wetlands, and aquaculture–photovoltaic integration, have been widely applied [3]. Photovoltaic power generation, as an efficient means of obtaining clean energy [4], when introduced into mining ecological restoration, not only provides a sustainable solution for energy supply but also opens up new prospects for environmental protection and ecological restoration, making it a forward-looking initiative worth exploring. When assessing the effectiveness of agricultural–photovoltaic complementarity, photovoltaic wetlands, and aquaculture–photovoltaic integration restoration models, information on the total area of efficiently obtained solar panels, geographical locations, and distribution is crucial for energy output evaluation and geographical spatial analysis. Satellite remote sensing images offer several advantages, including extensive coverage, rapid data acquisition, frequent updates, and independence from ground conditions [5]. Solar panels exhibit distinct features, such as texture, geometry, and spectral characteristics, in remote sensing images. Therefore, proposing a targeted method for extracting solar panels based on these features is of significant importance.

In recent years, researchers have used methods involving manual annotation and remote sensing visualization interpretation to create maps of photovoltaic power stations within regions [6]. However, manual visual interpretation is time-consuming, inefficient, and not suitable for large-scale classification tasks. Machine learning, as an empirical model for nonlinear system regression, has been widely applied to remote sensing classification tasks [7], [8], [9]. Traditional machine learning methods, such as random forest (RF) and support vector machines (SVMs), have been applied to various scales of remote sensing imagery to identify photovoltaic panels or solar power stations [10], [11], [12], [13], [14]. Automatic detection of solar photovoltaic arrays in high-resolution aerial imagery. Wang et al. [10] combined object-based image analysis with template matching techniques, achieving the precise extraction of photovoltaic panels from high-resolution aerial images and overcoming shape defects resulting from image segmentation.

Manuscript received 10 October 2023; revised 17 December 2023; accepted 23 December 2023. Date of publication 27 December 2023; date of current version 12 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42271368, Grant 52204190 and Grant U22A20569, and in part by the Fundamental Research Funds for the Central Universities under Grant 2021YCPY0113. *(Corresponding authors: Weiqiang Luo; Ming Hao.)*

Min Tan is with the School of Public Policy & Management (School of Emergency Management), China University of Mining and Technology, Xuzhou 221116, China (e-mail: db16160031b0@cumt.edu.cn).

Weiqiang Luo, Jingjing Li, and Ming Hao are with the School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China (e-mail: ts22160042a31@cumt.edu.cn; ts23160016a31tm@cumt.edu.cn; haomingcumt@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3347572

Muhammed et al. [11] utilized SVM and Bayesian machine learning algorithms to accurately detect solar panels on rooftops from satellite images, successfully estimating the photovoltaic potential in Egypt's Madinaty City. Zhang et al. [12], in their research, significantly improved the accuracy of the RF model in identifying photovoltaic power stations in Landsat-8 images by combining multiple input variables, including texture features calculated from gray-level co-occurrence matrix (GLCM), reflectance data, thermal spectral features, and image-normalized vegetation index. Chen et al. [13] selected Golmud, China, as their study area, establishing a photovoltaic power station recognition model based on original spectral features, photovoltaic extraction indicators, and terrain features. In their experiments, they compared the performance of different algorithms, such as XGBoost, RF, and SVM, confirming the effectiveness of combining original spectral features, photovoltaic extraction indicators, and terrain features for photovoltaic power station recognition. However, traditional machine learning algorithms involve relatively complex computation processes and lower model efficiency, making them less suitable for large-scale photovoltaic power station extraction tasks.

As an important branch of the machine learning field, deep learning has garnered widespread attention and application in recent years due to its efficiency, exceptional accuracy, and strong adaptability on large-scale data models [14], [15], [16], [17], [18], [19]. Classic deep learning models, such as U-Net, DeepLabv3+, PSPNet, SegNet, and others, have been extensively utilized for the task of extracting solar panels from remote sensing images. Yuan et al. [20] introduced the ConvNet method and successfully applied it to extract solar panels from aerial imagery covering 200 km^2 across two cities. Kouyama et al. [21] employed Landsat-8 imagery from the Kanto region for photovoltaic power station extraction, achieving favorable experimental results by combining convolutional neural networks (CNNs) with channel attention module (CAM) and demonstrating the effectiveness of this method. Arnaudo et al. [22] proposed a robust and efficient postprocessing polygon algorithm tailored for photovoltaic panels, converting coarse grid predictions into clearer and more accurate polygons. They tested this algorithm in the Piedmont region of Italy and compared it with methods, such as U-Net, showcasing its superior performance. Su et al. [23] introduced the filter-embedded network, embedding high-pass and low-pass filters as well as polarized self-attention into the high-resolution network to enhance its noise resistance and adaptive feature extraction capabilities, ultimately improving the extraction ability of photovoltaic power stations. da Costa et al. [24] utilized a semantic segmentation and large image classification stitching approach to identify photovoltaic power stations in Brazil. They compared the performance differences among four different architectures, U-Net, DeepLabv3+, pyramid scene parsing network, and feature pyramid network, in photovoltaic power station recognition. Ge et al. [25] proposed a hierarchical information extraction method, including localization information and shape information, for mapping the distribution of photovoltaic panels. In their experiments, they first employed the EfficientNet-B5 model for preliminary localization and then used U^2-net for precise segmentation. Through comparative experiments with DeepLabv3+, U-Net, SegNet, and FCN8s, the superiority of the hierarchical information extraction method in terms of accuracy and efficiency was verified. Wang et al. [26] introduced a novel semantic segmentation model called PVNet, consisting of a coarse prediction module and a fine optimization module. This research was tested in four different scenarios, providing a feasible solution for obtaining high-quality geographic spatial databases of large-scale photovoltaic systems. Castello et al. [27] employed a deep learning approach to extract rooftop solar panels. By adjusting data augmentation techniques and network parameters, they maximized model performance, contributing to the advancement of solar panel extraction efforts across Switzerland.

At present, researchers have made various attempts in the field of solar panel extraction, and different extraction methods have shown feasibility. However, traditional machine learning approaches are not suitable for large-scale solar panel extraction. Furthermore, in deep learning, there is still room for improvement in utilizing texture features from shallow-level characteristics, and there is a lack of specificity in network design. To address these issues, this article proposes a deep learning model called texture-enhanced multicontext attention network (TEMCA-Net) based on an encoder–decoder framework. Given the distinct texture features of photovoltaic power stations in images, the TEMCA-Net incorporates a statistical texture learning (STL) module in the network to enhance texture details within shallow-level features. In addition, to improve the network's horizontal perception, modifications are made to the atrous spatial pyramid pooling (ASPP) module to adapt it to the horizontal orientation of solar panels in images. Furthermore, residual networks (RNs) and convolutional block attention module (CBAM) modules are introduced into the network to achieve fine-grained extraction of solar panels. In summary, this article's primary contributions are given as follows.

1) Based on the texture and horizontal features of solar panels in remote sensing images, we designed a deep learning network, TEMCA-Net, to achieve finer extraction of solar panels. By introducing modules, such as horizontal atrous spatial pyramid pooling (H-ASPP) and STL, the performance of the TEMCA-Net network has been significantly improved.
2) Utilizing remote sensing images of solar panels in the Peibei mining area, we constructed a deep learning dataset, solar panels of Peibei mining region (SPPMR), providing support for the research and application of deep learning techniques in the monitoring of solar panels in high groundwater-level mining areas.
3) The trained TEMCA-Net model was applied to extract solar panels in the southern part of Peibei mining area. The extraction results were used to create a distribution map of solar panels in the southern part of Peibei mining area.

II. Materials

A. Study Area

This experiment selected Peibei mining region as the study area, as shown in Fig. 1(a). Peibei mining region is a typical high water table mining area located in the northwest part of Jiangsu

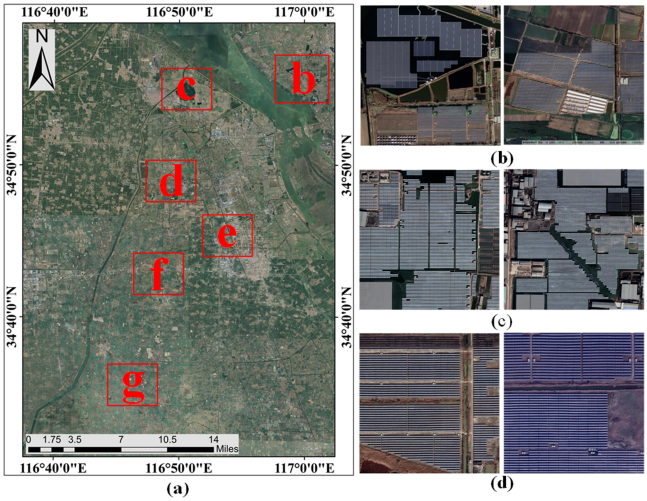

Fig. 1. Illustrates the geographical location and environment of the study area, with (a) indicating the position of Pei County. Regions b–g represent the primary distribution areas of solar panels in Pei County. Subfigures (b), (c), and (d) showcase some of the solar panels in these areas.

Province, approximately 90 km from the center of Xuzhou City. The terrain in this area is flat, with elevations ranging from 31.5 to 41 m. Peibei mining region hosts the main industrial and coal mining activities of Pei County, including five administrative towns: Longgu, Yangtun, Anguo, Lulou, and Zhuzhai, as well as four subdistricts: Datun, Peicheng, Hanyuan, and Hanxing, covering a total area of 744.3476 km^2. Prolonged mining activities have led to damage to the topography, landscape, and land resources in this region, resulting in ecological degradation. Geological hazards induced by mining activities in Peibei mining region mainly include ground subsidence, collapses, and ground fissures. The coal mining subsidence area is mainly located in the northern part of Pei County (116°41′–117°09′E, 34°28′–34°59′N), extending east to the west dike of Weishan Lake, west to the border with Fengxian County, south to Fengpei Road (extending east to the west dike of Weishan Lake), and north to the border with Longgu Town and Shandong Yutai County. To address the challenges posed by different subsidence depths, various restoration methods have been implemented locally. Specifically, for areas with subsidence depths exceeding 3 m, Pei County has utilized solar and aquatic resources, constructing photovoltaic power stations on subsided water surfaces, thereby achieving the "agricultural–photovoltaic complementarity" and "aquaculture–photovoltaic integration" power generation modes. This experiment aims to extract information regarding the area and location of solar panels constructed under this power generation model.

B. Dataset

1) The experiment utilized QuickBird satellite imagery data, sourced from Google Earth. QuickBird is a high-resolution commercial satellite operated by DigitalGlobe. Since its launch in 2001, it has been in a sun-synchronous orbit, circling the Earth multiple times daily, thereby offering comprehensive global coverage. QuickBird is renowned for its 0.65-m Panchromatic band and 2.62-m multispectral resolution, making it a standout choice across various domains, including urban planning, land management, agriculture, forestry, geological exploration, and environmental monitoring. The satellite's swift data acquisition capabilities are particularly invaluable in situations requiring urgent responses. Within the high-resolution remote sensing domain, QuickBird satellites play a pivotal role by providing researchers and decision makers with invaluable observational data. The QuickBird data parameters are listed in Table I.

TABLE I
QUICKBIRD SPECTRAL PARAMETERS

Data Type	Bandwidth Range(μm)	Resolution
Multi-Spectral	Band1: 0.45–0.52	2.44
	Band2: 0.52–0.60	2.44
	Band3: 0.63–0.69	2.44
	Band4: 0.76–0.90	2.44
Full Spectrum	0.45–0.90	0.61

2) Process the acquired data through image processing. First, perform radiometric correction and geometric correction on the images, where radiometric correction involves radiometric calibration and atmospheric correction. Next, conduct orthorectification on the images. Finally, mosaic and crop the processed images to obtain a remote sensing image of the entire study area.
   a) *Atmospheric correction of the images*—Remote sensing images are influenced by factors, such as atmospheric molecules and aerosols during the acquisition process. To ensure the spectral information of the land features during the change detection process, atmospheric correction is necessary for remote sensing images. In this study, the ENVI atmospheric correction module, specifically the fast line-of-sight atmospheric analysis of spectral hypercubes was utilized for rapid atmospheric correction.
   b) *Orthorectification of the images*—Orthorectification effectively addresses radiometric distortions caused by terrain factors, ensuring the accuracy of the geometric position of the images.
   c) *Image mosaicking and cropping*—Mosaic the remote sensing images to cover the entire study area, and then utilize vector data of the study area to crop the mosaicked remote sensing images, obtaining remote sensing images within the study area.
3) The training set in SPPMR is constructed from three areas at locations b, c, and d, as shown in Fig. 1. The distribution of solar panels in these areas is illustrated in Fig. 2, where in the *b* region, bar-shaped solar panels are densely distributed on the ground. In both b and c regions, there are both bar-shaped solar panels on the ground and panel-shaped solar panels on the water surface.

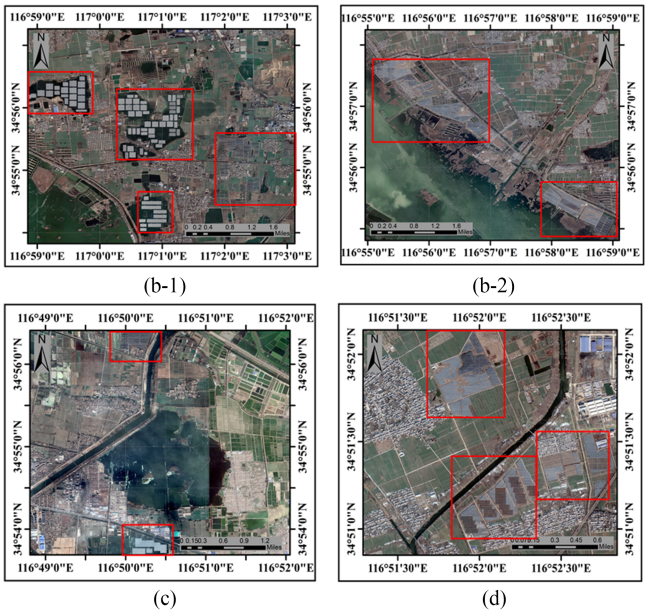

Fig. 2. (b), (c), and (d)—Schematic distribution of solar panels at three locations.

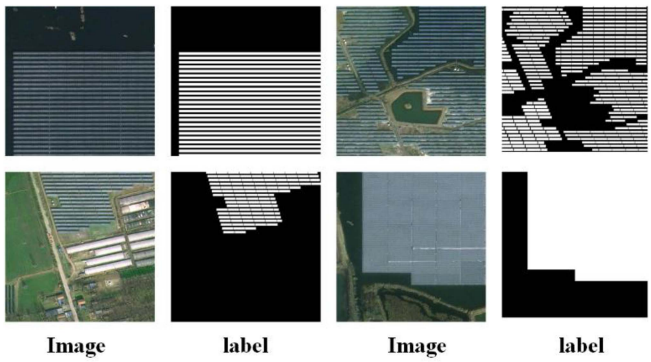

Fig. 3. Partial original images and labels.

For the testing set in SPPMR, three areas at locations e, f, and g in Fig. 1 are selected for construction. The types of solar panels in these areas are also diverse, facilitating the reduction of differentiation between the training and testing sets.

The selected region images are pixelwise annotated, with the original images segmented into solar panel class and background class. Background pixels are represented as (0, 0, 0), while solar panel pixels are represented as (255, 255, 255). Using a sliding window with a stride of 300 pixels, the original images and annotated images are segmented into 512×512-sized images, as depicted in Fig. 3.

The entire training set consists of images from regions b, c, and d, comprising a total of 13 246 samples. The testing set is composed of images from regions e, f, and g, totaling 5326 samples. Overall, the SPPMR dataset includes 18 572 samples, with a training-to-testing set ratio of approximately 7:3.

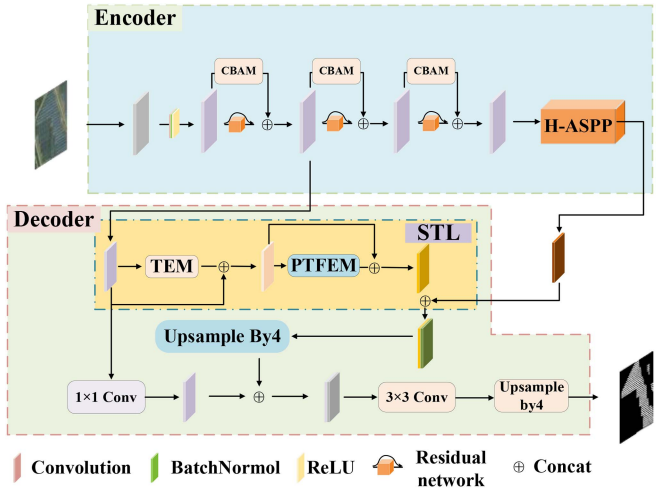

Fig. 4. Presents the overall framework of the TEMCA-Net network.

III. Methods

This study proposes the TEMCA-Net network based on an encoder–decoder framework with the aim of achieving high-precision semantic segmentation of solar panels. The complete structure of the network is depicted in Fig. 4. In the encoding section, the residual edge connection technique [28] is employed along with the collaborative CBAM mixed attention mechanism module to extract features of solar panels from shallow to deep layers. The initially generated shallow features are input into the decoder, and simultaneously, high semantic features passing through multiple residual blocks are fed into the H-ASPP module for horizontal enhancement.

In the decoding section, the shallow features from the encoder are input into the STL [29] texture enhancement module (TEM) to improve the network's perception of the texture of solar panels. The texture features processed by STL are combined with the high-level features from the H-ASPP module, and the final solar panel extraction results are generated through usampling.

A. Encoder Featuring RNs, CBAM Attention, and H-ASPP

1) *RNs*: Given the dense distribution of solar panels in the Pei Bei mining area, they are susceptible to strong noise interference from factors, such as sunlight and shadows. Therefore, it is imperative to employ deep CNNs for extracting multiscale and high-level features of solar panels to facilitate their recognition. In the realm of deep learning, the quality of feature extraction is intricately linked to the network's depth, with deeper network layers capable of generating more complex and abstract features compared with shallower layers. However, as network depth increases, there is a risk of encountering the problem of network degradation, which can degrade the performance of deep networks. To mitigate the challenges associated with increased network depth, we have adopted an RN [30] approach, merging the outputs of shallow network layers

Fig. 5. Schematic diagram of CBAM.

with those of deeper layers. This strategy ensures that the performance of the deep network is, at the very least, on par with that of the shallow network, and possibly even superior. In consideration of the aforementioned factors, we have incorporated RNs into TEMCA-Net to extract various contextual information.

2) *CBAM*: The perception of features by a network is largely dependent on the size of its receptive field. However, the feature maps obtained through RN are insufficient to comprehensively capture rich global features and contextual information. Therefore, we introduce the lightweight attention module CBAM [31] from the features extracted by DRN. As shown in Fig. 5, the CBAM module comprises two components: the spatial attention module and the CAM.

   The computation process of the CBAM module can generally be divided into two stages. First, the intermediate feature map F is taken as input and subjected to global max-pooling and average-pooling operations. The one-dimensional (1-D) vectors generated by these two pooling operations are passed through a multilayer perceptron (MLP) for computation, and their results are added together. After applying an activation function to the input, a 1-D channel attention weight M_c is generated and multiplied with the input elements to adjust channel attention, resulting in the feature map F'.

   Second, a global max-pooling and average-pooling operation is applied to F', and the two 2-D vectors generated by pooling are concatenated. Through a convolution operation, a 2-D spatial attention weight M_s is generated. This obtained M_s is then multiplied with the feature map F' to obtain a feature map enhanced by mixed attention. The introduction of the CBAM module primarily addresses the limitation of feature representation within the receptive field of a CNN. When embedded into RN, it further enhances the model's perceptual capability of features during the encoding phase, enabling the model to capture richer semantic information.

3) Improved H-ASPP: The orientation of solar panels is influenced by the solar altitude angle. In the Peibei mining area, the majority of solar panels are positioned horizontally in the imagery. In the context of semantic segmentation tasks, the salient features of solar panels predominantly lie in the horizontal direction. Therefore, this study introduces an enhancement to dilated convolutions, where dilation convolution is performed exclusively in the horizontal direction. This means that dilated convolutions are applied along the width dimension of the feature layer, enabling the more concentrated capture of contextual information in the horizontal direction. Simultaneously, this modification significantly reduces the computational workload in this stage.

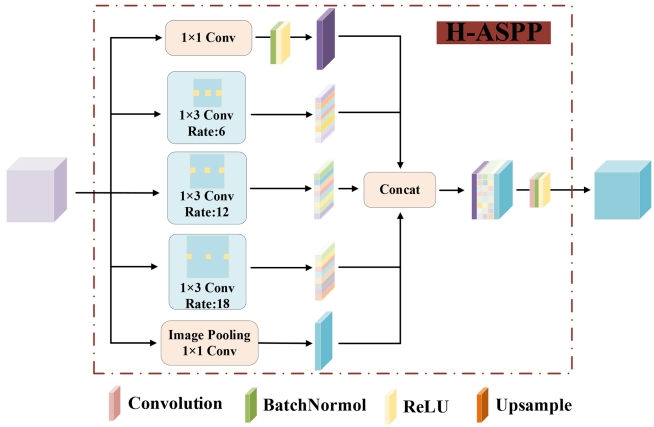

Fig. 6. Improved H-ASPP.

The improved ASPP [32] maps feature into five parallel branches for processing. These branches consist of a 1×1 convolution branch, three 1×3 horizontal dilated convolution layers with dilation rates of 12, 24, and 36 in the horizontal direction, and a final branch comprising pooling layers and a 1×1 convolution layer. While the 1×3 convolution operation enhances the network's receptive field, it diminishes the network's ability to capture context information across the entire image. To ensure the network's performance, we introduce global average pooling to assist in capturing global information. Simultaneously, we incorporate batch normalization layers and ReLU activation functions into each branch following convolution, effectively mitigating gradient vanishing issues and accelerating model convergence. These measures not only maintain network stability but also enhance training efficiency. Upon completion of the H-ASPP module, we connect and resample the feature maps generated at multiple scales, fully leveraging global contextual information. This empowers TEMCA-Net to achieve highly accurate and efficient classification across a variety of spatial scales. The improved spatial pyramid H-ASPP structure is shown in Fig. 6.

B. Decoder Enhanced by STL Texture

Texture features are a highly significant aspect of visual characteristics used for describing the texture and structure within an image [33]. These features aid in the network's understanding of homogeneity in the image, specifically the presence of similar textures. By analyzing factors, such as the distribution of pixel values, texture frequencies, and orientations, we can extract both local and global information about the image. In remote sensing imagery, solar panel textures are distinctive and can be readily differentiated from other features. Consequently, in this study, we incorporate the STL enhancement module into the decoding section. This module takes low-level features obtained from the RN and applies texture enhancement to improve the network's perception of texture.

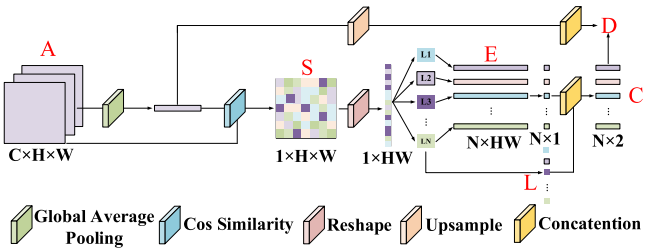

Fig. 7. Structure of 1D-QCO.

1) *STL module:* The STL module introduces statistical texture information for semantic segmentation to fully utilize texture features. Traditional CNNs are primarily designed for extracting local texture features and local shape features, such as edges, smoothness, and roughness. Although convolutional kernels excel as feature extraction tools in image processing, shallow-level features inherently contain local texture information. However, there is no explicit mechanism for extracting and utilizing texture information for semantic segmentation [34]. Relying solely on convolutional operations for information and feature extraction is insufficient. On the other hand, statistical information within texture features, such as grayscale histograms, holds significant value and finds extensive application in traditional image processing algorithms. Low-level features are crucial for enhancing semantic segmentation performance. On the other hand, statistical information within texture features, such as grayscale histograms, holds significant value and finds extensive application in traditional image processing algorithms [35], [36]. Low-level features are crucial for enhancing semantic segmentation performance.
To effectively capture statistical texture within deep neural networks, the STL module introduces a novel feature encoding approach known as the quantization and counting operator (QCO). The QCO operator resembles convolution kernels in CNNs and comprises three critical components: quantization, counting, and average feature encoding. Furthermore, there are two types of QCO, namely 1D-QCO and 2D-QCO, utilized in constructing the TEM and the pyramid texture feature extraction module (PTFEM) texture feature extraction module. These innovative designs facilitate a better understanding and utilization of statistical texture information within deep neural networks, ultimately enhancing the performance and efficiency of semantic segmentation tasks.
2) *1D-QCO module:* The structure of 1D-QCO is illustrated in Fig. 7. First, the input feature map A, with dimensions $C\times H\times W$, undergoes quantization. The feature map A is first processed through global average pooling to obtain the global mean feature g, which has dimensions $C\times 1\times 1$. Then, the cosine similarity between each spatial position $A_{ij}(i\in[1,W],j\in[1,H])$ in the feature map A and g is calculated, resulting in cosine similarity features S, with dimensions $1\times H\times W$. Subsequently, the feature map is transformed into a 1-D vector S with dimensions $H\times W$, from which L_x is calculated by dividing the difference between the maximum and minimum values in the 1-D vector S by N. Starting from the minimum value min (S) in S, each layer of information L_x is cumulatively accumulated to obtain L_n. Finally, E_i is computed from S_i,L_n using the formula provided in (1). This quantization encoding approach ensures smoother quantization and mitigates the issue of gradient vanishing during the backpropagation process. In the second step, statistics are performed on the encoding matrix obtained through the quantization process, leading to the computation of C. Here, the first dimension represents each quantization level, and the second dimension represents the corresponding normalized count value. The second step involves conducting further statistical analysis on the quantized results obtained from the encoding matrix E, resulting in the generation of a count matrix C. In this matrix, the first dimension represents different quantization levels, while the second dimension represents their corresponding normalized count values

$$E_{i,n}=\begin{cases}1-|L_n-S_i| & \text{if } -\frac{0.5}{N}\le L_n-S_i<\frac{0.5}{N}\\ O & \text{else.}\end{cases} \tag{1}$$

3) *2D-QCO module:* 1D-QCO output reflects the feature distribution at various spatial positions. However, in 1D-QCO, it does not include information about the spatial relationships between pixels, and spatial relationships play a crucial role in describing texture features.
2D-QCO aims to compute the spatial relationships between pixels within the feature map, extending upon the foundation of 1D-QCO. The specific process is as follows: the input feature A undergoes 1D-QCO, resulting in quantization encoding E and quantization level L. Then, E is transformed into an $R\times N\times H\times W$ tensor, and the product of every pair of adjacent pixels E_{ij} is calculated to obtain $\hat{E}$, using $\hat{E}$ to represent the quantized values of adjacent pixel relationships.
$\hat{E}$ is statistically analyzed to generate a 3-D mapping C, where the first two dimensions represent each possible quantized relationship, and the third dimension signifies the corresponding normalized count value. Finally, mean feature encoding is performed, and the mean feature within the processing region is denoted as g. It is upsampled, resulting in the input for 2D-QCO.
4) *TEM module:* In situations where image contrast is low, the low-level features extracted from the backbone network may suffer from reduced quality. The extracted low-level features might struggle to capture the fine texture details of the image clearly, posing challenges for subsequent tasks [37]. Hence, as depicted in Fig. 8, we introduce the TEM specifically for enhancing the fine texture details within the low-level features. The primary objective of TEM is to enhance the quality of these low-level features, making them more adept at capturing crucial information related to image textures.

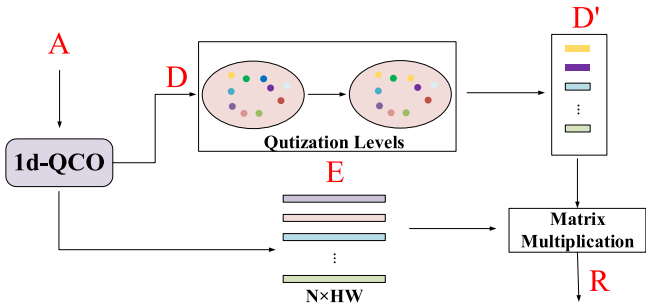

Fig. 8. Structure of TEM.

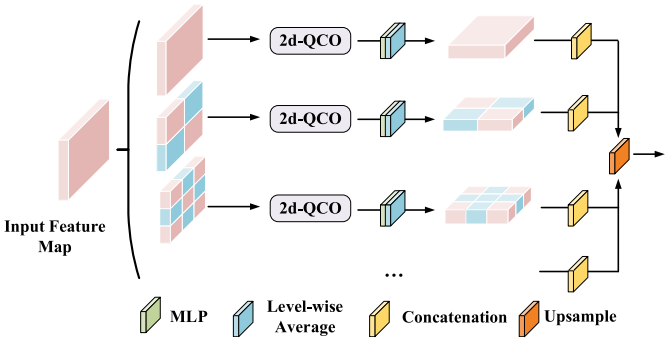

Fig. 9. Structure of PTFEM.

The inspiration for the texture enhancement method is derived from the classical image quality enhancement technique called histogram equalization, where the horizontal and vertical axes of a histogram represent each grayscale level and its count, respectively. TEM primarily utilizes the quantization encoding E and statistical features D obtained through 1D-QCO calculations. It defines a statistical feature for each quantization level, employing an improved approach over the traditional histogram equalization algorithm by extending it into a learning matrix. The softmax function is used to normalize the calculation along the first dimension of the statistical feature D, resulting in the reconstructed quantization levels D'. Finally, the reconstructed quantization levels D' are multiplied with the quantization encoding E to obtain the ultimate output R.

5) *PTFEM module:* As shown in Fig. 9, the PTFEM aims to extract texture-related information from multiscale feature maps. Since texture features are highly correlated with the statistical information of spatial relationships between pixels, the method for extracting texture information in PTFEM draws inspiration from GLCM. In GLCM, a co-occurrence matrix is first generated, and then statistical descriptors, such as contrast and uniformity, which are manually designed, are used to represent the texture information of the region. In 2D-QCO, a similar principle to GLCM is employed to extract co-occurrence statistical features. However, unlike the manually designed statistical descriptors used in GLCM, 2D-QCO automatically learns effective statistical representations from samples through deep learning. Subsequently, an MLP is employed to further extract texture features.

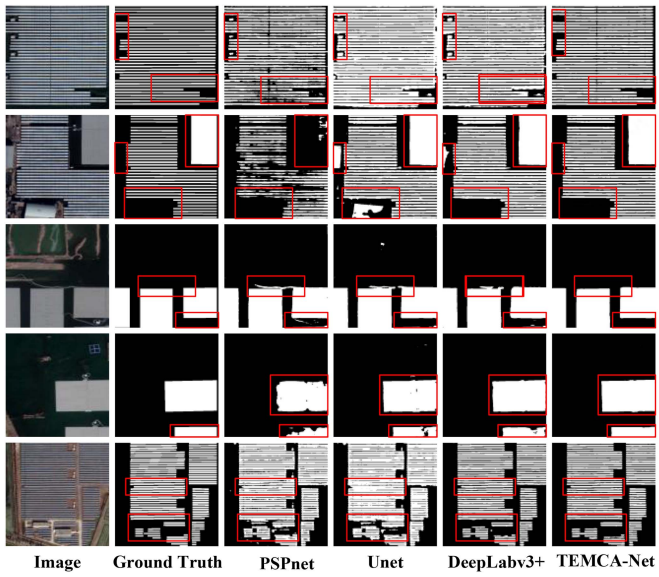

Fig. 10. Solar panel extraction results of different methods on the SPPMR dataset. From left to right: Original image, ground truth labels, PSPNet [38], U-Net [39], DeepLabv3+ [40], and TEMCA-Net.

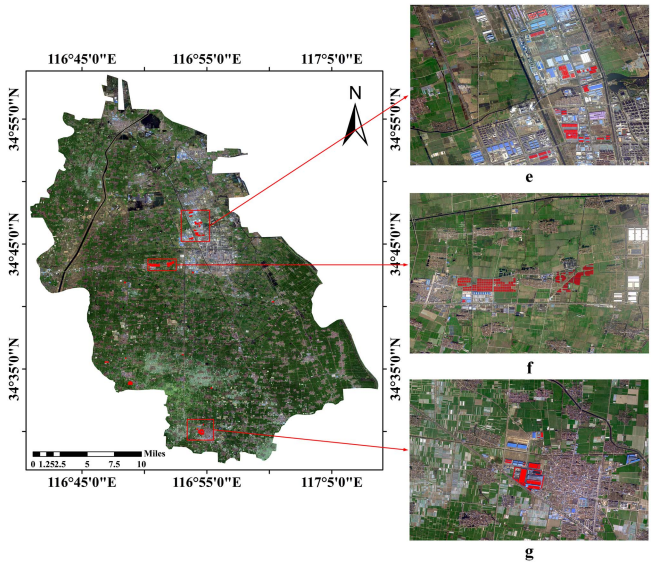

Fig. 11. Distribution of solar panels in three locations within the Peibei mining area.

Furthermore, to comprehensively consider texture information at different scales, PTFEM adopts a pyramid structure. It operates on multiple scale levels, performing the same operations to generate feature maps at multiple scales, and then combines them for subsequent processing. This multiscale processing contributes to improving the performance of semantic segmentation models, allowing them to better capture object details at different scales, thereby enhancing the model's adaptability to various scenes.

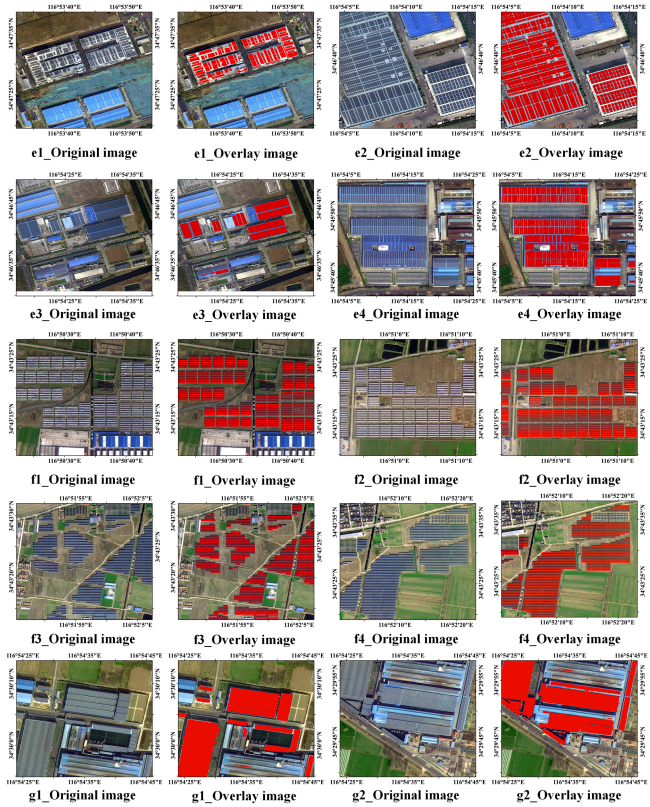

Fig. 12. Extraction results of TEMCA-Net in scenes e, f, and g. e1, e2, …, g2, etc., represent different regions within areas e, f, and g.

IV. Experiment

A. Experimental Environment and Evaluation Measures

1) In this experiment, we employed the PyTorch 1.11.7 framework to construct the TEMCA-Net network. All tasks were executed on a workstation equipped with a CPU (13th Gen Intel(R) Core (TM) i7-13700KF) and GPU (NVIDIA GeForce RTX 4090). The training epochs were set to 100, with a batch size of 10 and an initial learning rate of 7e-3. The loss function used was cross-entropy loss. For optimizing the network parameters, we chose the stochastic gradient descent optimizer with a momentum parameter of 0.8, alongside the application of weight decay at a rate of 1e-5. To ensure the rigor of this experiment, we maintained consistency across all methods by using the same hyperparameters for training, including learning rate, batch size, and regularization parameters.
2) For the evaluation of solar panel extraction results, metrics, such as precision (Pr), recall (Re), $F1$-score (F1), and mean intersection over union (mIoU), are employed for a fair assessment. These metrics are defined as follows:

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{2}$$

$$\mathrm{Re} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{3}$$

TABLE II
Confusion Matrix

Ground Truth	Predicted results	
	Solar panel	Non-solar panel
Solar panel	TP	FN
Non-solar panel	FP	TN

TABLE III
On the SPPMR Dataset, We Compare the Semantic Segmentation Accuracy and Time Between TEMCA-Net and Other Methods

Method	Pr	Re	F1	mIoU	Training time(h)
PSPNet	66.24	78.65	71.91	77.83	3.6
U-Net	91.5	65.67	76.61	80.89	7.2
DeepLabV3+	**90.49**	80.88	85.41	87.16	5.1
TEMCA-Net	90.24	**93.07**	**91.63**	**92.21**	5.4

Bold numbers in the table represent the highest accuracy.

TABLE IV
FLOPS and Parameter Quantities Are Computed for Different Models, and the Results Are Summarized in the Table

Method	Input Data	FLOPs	Parameters
PSPNet	(1,3,512,512)	9896.40M	2.38M
U-Net		225836.01M	24.89M
DeepLabV3+		26433.37M	5.81M
TEMCA-Net		28332.38M	6.53M

$$\mathrm{F1} = \frac{(2 \times \mathrm{Re} \times \mathrm{Pr})}{\mathrm{Re}+\mathrm{Pr}} \tag{4}$$

$$\mathrm{MIoU} = \frac{1}{2}\left(\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} + \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}\right). \tag{5}$$

In which, TP, FP, FN, and TN represent the true positives, false positives, false negatives, and true negatives, as illustrated in the confusion matrix in Table II.

We introduced the $F1$-score, which is the harmonic mean of precision and recall, as it provides a comprehensive evaluation of model performance. When a model achieves high precision and high recall simultaneously, the $F1$-score value is higher, indicating a more balanced performance in the task. In addition, we used mIoU to evaluate the similarity between ground truth solar panel pixels and predicted solar panel pixels. A higher mIoU value signifies a closer resemblance between the model's segmentation results and the ground truth.

1) We employed data augmentation to increase both the quantity and quality of training samples in each iteration. The implementation of three different types of data augmentation is given as follows:
   a) randomly flipping the original image along both the horizontal and vertical axes;
   b) performing random cropping on the original image, followed by resizing to match the original image's dimensions;

TABLE V
PRESENTS THE ABLATION EXPERIMENTS OF THE TEMCA-NET NETWORK ON SPPMR

Method	CNN	RN	CBAM	ASPP	H-ASPP	STL	Pr	Re	F1	mIoU
TEMCA-Net	✓						0.786	0.733	0.758	0.724
	✓	✓					0.780	0.745	0.762	0.726
	✓	✓					0.791	0.764	0.773	0.782
	✓	✓	✓				0.857	0.805	0.873	0.836
	✓	✓	✓				0.846	0.783	0.812	0.823
	✓	✓	✓	✓			0.891	0.871	0.884	0.832
	✓	✓	✓		✓		0.889	0.876	0.897	0.849
	✓	✓	✓		✓		0.873	0.867	0.882	0.862
	✓	✓	✓		✓	✓	0.902	0.930	0.91	0.922

c) randomly varying the brightness and contrast of the image, with the variation intensity being around 30% of the original values.

B. Results of the Experiment

To demonstrate the performance of TEMCA-Net, we conducted comparative experiments in a selected testing area, evaluating TEMCA-Net alongside three commonly used semantic segmentation models: U-Net, DeepLabv3+, and PSPNet. The comparative experimental results are presented in Fig. 10.

From the comparative experimental results, as shown in Fig. 10, the following observations can be made regarding solar panel extraction.

1) When it comes to solar panel extraction, U-Net and PSPnet exhibit higher rates of both false negatives (missed detections) and false positives (incorrect detections). In addition, their edge quality in the extracted solar panels is poor, making them susceptible to interference from other objects. In the case of strip-shaped solar panel extraction, the low segmentation accuracy often results in overlapping panels
2) In contrast, DeepLabv3+ shows a significant reduction in false negatives and false positives compared with U-Net and PSPnet. However, the quality of the extracted solar panel edges remains somewhat rough, and some instances of overlap are still observed.
3) TEMCA-Net, on the other hand, exhibits lower rates of false positives, particularly when the TEM is incorporated. This leads to significantly improved accuracy in extracting rectangular solar panels and superior performance in segmenting the edges of strip-shaped solar panels compared with the other three methods. Furthermore, TEMCA-Net demonstrates less susceptibility to panel overlap.

In Table III, the evaluation metrics on the SPPMR dataset are presented. TEMCA-Net achieves accuracy levels close to the DeepLabv3+ model but boasts a recall rate of 93.07%, an $F1$-score of 91.63%, and an mIoU of 92.21%, all of which outperform the other methods by a considerable margin. Simultaneously, we compared the training times of each model. From the statistical perspective, our proposed method demonstrates relatively higher efficiency while ensuring accuracy. Finally, the statistical analysis of model parameters in Table IV indicates that the proposed method in this article has a relatively low model complexity.

C. Solar Panel Extraction Results in the Peibei Mining Area

To assess the performance of TEMCA-Net in real-world large-scale scenarios, we selected three locations within the Peibei mining area, as shown in Fig. 11. These three locations are denoted as e, f, and g, constituting the test set. These scenes are representative and showcase different texture characteristics. In locations e and f, the solar panels are distributed in a planar fashion.

Fig. 12 displays partial extraction results from these three scenes, indicating that TEMCA-Net can adapt to a variety of different scenarios. It can accurately extract solar panels in different environments, demonstrating a low false-negative rate and precise delineation of solar panel edges.

D. Ablation Study

When evaluating the effectiveness of TEMCA-Net with the incorporation of RN, CBAM, H-ASPP, and STL modules, we can draw some key observations from the results in Table IV. First, in the first row, we used a basic CNN to extract features and then directly performed upsampling operations to generate the feature extraction results. Starting from this baseline, we gradually introduced the RN, CBAM, H-ASPP, and STL modules. To ensure the fairness of the ablation experiments, when conducting ablation experiments on each module, we replaced them with convolutional layers with approximately equal parameter sizes. In the results, as shown in Table V, when a module is unchecked, it indicates that it has been replaced with a convolutional layer of roughly the same parameters.

In observing the performance improvements after the introduction of different modules, although precision and recall metrics may show slight fluctuations in some cases, the overall trend indicates a gradual enhancement of the $F1$ and mIoU metrics with the introduction of each module. In the ablation experiments, we replaced modules with convolutional layers having approximately equal parameters. Through the comparison of various metric data, it is evident that merely increasing the parameter count does not enhance the network's performance. In the ablation experiments involving the STL module, increasing the parameter count actually led to a decline in network performance. This indicates that the introduction of these modules is highly effective for improving the performance of image segmentation tasks.

Specifically, the last row displays the metrics for MFPA-Net, and compared with the first row, all metrics show significant improvements. Precision increased by 11%, recall increased by 19.7%, $F1$ improved by 15.2%, and mIoU increased by 19.8%. These numbers reflect that by introducing the RN, CBAM, H-ASPP, and STL modules, the model's performance in image segmentation tasks has been successfully enhanced, particularly in terms of precision and recall. The combined effects of these modules enable our model to more accurately identify objects and perform pixel-level segmentation, thereby improving the overall performance level.

V. Conclusion

To accurately extract and locate solar panels in the Peibei mining area, this study proposes an efficient deep learning network called TEMCA-Net based on an encoder–decoder structure. The network incorporates RN, CBAM, H-ASPP, and STL modules, enabling automatic and precise extraction and localization of solar panels in the images of the Peibei mining area. In order to effectively extract solar panels, this article constructs the SPPMR dataset using QuickBird imagery. The experimental results on SPPMR using various methods show that adding the STL module to the TEMCA-Net significantly reduces false positives and false negatives. In addition, the inclusion of the horizontal H-ASPP enhances the network's ability to perceive details, effectively eliminating the issue of overlapping patterns of solar panels due to insufficient segmentation accuracy. Finally, ablation experiments demonstrate that the use of RN, CBAM, H-ASPP, and STL modules can significantly improve the performance of TEMCA-Net.

This research enhances the efficiency and accuracy of solar panel extraction, providing key information on the location and area of solar panels. This information is crucial for evaluating ecological restoration models, such as agricultural–photovoltaic complementarity, photovoltaic wetlands, and fishery–photovoltaic complementarity. It holds significant practical value for ecological model assessment.

In future studies, we intend to extract solar panels in high groundwater areas of Jiangsu Province using the method proposed in this article. By analyzing the trend of solar panel installation areas in recent years, we aim to investigate the impact of solar panels on the agricultural–photovoltaic complementary restoration mode in these areas.

References

[1] W. Xiao, Y. Fu, T. Wang, and X. Lv, "Effects of land use transitions due to underground coal mining on ecosystem services in high groundwater table areas: A case study in the Yanzhou coalfield," *Land Use Policy*, vol. 71, pp. 213–221, 2018.

[2] Z. Li, J. Chang, C. Li, and S. Gu, "Ecological restoration and protection of National Land Space in coal resource-based cities from the perspective of ecological security pattern: A case study in Huaibei City, China," *Land*, vol. 12, no. 2, 2023, Art. no. 442.

[3] I. Kougias et al., "The potential of water infrastructure to accommodate solar PV systems in Mediterranean islands," *Sol. Energy*, vol. 136, pp. 174–182, 2016.

[4] G. K. Singh, "Solar power generation by PV (photovoltaic) technology: A review," *Energy*, vol. 53, pp. 1–13, 2013.

[5] H. Lu, Q. Liu, X. Liu, and Y. Zhang, "A survey of semantic construction and application of satellite remote sensing images and data," *J. Org. End User Comput.*, vol. 33, no. 6, pp. 1–20, 2021.

[6] L. Duarte, A. C. Teodoro, D. Maia, and D. Barbosa, "Radio astronomy demonstrator: Assessment of the appropriate sites through a GIS open source application," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 11, 2016, Art. no. 209.

[7] B. Fu et al., "Quantifying scattering characteristics of mangrove species from Optuna-based optimal machine learning classification using multi-scale feature selection and SAR image time series," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 122, 2023, Art. no. 103446.

[8] T. Hou, W. Sun, C. Chen, G. Yang, X. Meng, and J. Peng, "Marine floating raft aquaculture extraction of hyperspectral remote sensing images based decision tree algorithm," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 111, 2022, Art. no. 102846.

[9] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Front.*, vol. 7, no. 1, pp. 3–10, 2016.

[10] M. Wang, Q. Cui, Y. Sun, and Q. Wang, "Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 100–111, 2018.

[11] E. Muhammed, A. El-Shazly, and S. J. S. Morsy, "Building rooftop extraction using machine learning algorithms for solar photovoltaic potential estimation," *Sustainability*, vol. 15, no. 14, 2023, Art. no. 11004.

[12] X. Zhang, M. Zeraatpisheh, M. M. Rahman, S. Wang, and M. Xu, "Texture is important in improving the accuracy of mapping photovoltaic power plants: A case study of Ningxia autonomous region, China," *Remote Sens.*, vol. 13, no. 19, 2021, Art. no. 3909.

[13] Z. Chen, Y. Kang, Z. Sun, F. Wu, and Q. Zhang, "Extraction of photovoltaic plants using machine learning methods: A case study of the pilot energy city of Golmud, China," *Remote Sens.*, vol. 14, no. 11, 2022, Art. no. 2697.

[14] B. Fu et al., "Combination of super-resolution reconstruction and SGA-net for marsh vegetation mapping using multi-resolution multispectral and hyperspectral images," *Int. J. Digit. Earth*, vol. 16, no. 1, pp. 2724–2761, 2023.

[15] D. Guo and W. Shi, "A task decoupled framework for enhancing the deep learning-based spatiotemporal fusion method," *Int. J. Remote Sens.*, vol. 44, no. 13, pp. 4163–4189, 2023.

[16] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Oct. 2023, Art. no. 5527812.

[17] D. Hong et al., "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, 2023, Art. no. 113856.

[18] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5513412.

[19] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jun. 2023, Art. no. 5514415.

[20] J. Yuan, H.-H. L. Yang, O. A. Omitaomu, and B. L. Bhaduri, "Large-scale solar panel mapping from aerial images using deep convolutional networks," in *Proc. IEEE Int. Conf. Big Data*, 2016, pp. 2703–2708.

[21] T. Kouyama, N. Imamoglu, M. Imai, and R. Nakamura, "Verifying rapid increasing of mega-solar PV power plants in Japan by applying a CNN-based classification method to satellite images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 4104–4107.

[22] E. Arnaudo et al., "A comparative evaluation of deep learning techniques for photovoltaic panel detection from aerial images," *IEEE Access*, vol. 11, pp. 47579–47594, 2023.

[23] B. Su et al., "FEPVNet: A network with adaptive strategies for cross-scale mapping of photovoltaic panels from multi-source images," *Remote Sens.*, vol. 15, no. 9, 2023, Art. no. 2469.

[24] M. V. C. V. da Costa et al., "Remote sensing for monitoring photovoltaic solar plants in Brazil using deep semantic segmentation," *Energies*, vol. 14, no. 10, 2021, Art. no. 2960.

[25] F. Ge, G. Wang, G. He, D. Zhou, R. Yin, and L. Tong, "A hierarchical information extraction method for large-scale centralized photovoltaic power plants based on multi-source remote sensing images," *Remote Sens.*, vol. 14, no. 17, 2022, Art. no. 4211.

[26] J. Wang, X. Chen, W. Jiang, L. Hua, J. Liu, and H. Sui, "PVNet: A novel semantic segmentation model for extracting high-quality photovoltaic panels in large-scale systems from high-resolution remote sensing imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 119, 2023, Art. no. 103309.
[27] R. Castello, S. Roquette, M. Esguerra, A. Guerra, and J.-L. Scartezzini, "Deep learning in the built environment: Automatic detection of rooftop solar panels using convolutional neural networks," *J. Phys., Conf. Ser.*, vol. 1343, no. 1, 2019, Art. no. 012034.
[28] F. He, T. Liu, and D. Tao, "Why resnet works? Residuals generalize," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5349–5362, Dec. 2020.
[29] Q. Shen, H. Deng, X. Wen, Z. Chen, and H. Xu, "Statistical texture learning method for monitoring abandoned suburban cropland based on high-resolution remote sensing and deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3060–3069, Mar. 2023.
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[31] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[33] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[34] L. Zhu, D. Ji, S. Zhu, W. Gan, W. Wu, and J. Yan, "Learning statistical texture for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 12532–12541.
[35] S. Arivazhagan and L. Ganesan, "Texture segmentation using wavelet transform," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3197–3203, 2003.
[36] A. Ramola, A. K. Shakya, and D. van Pham, "Study of statistical methods for texture analysis and their modern evolutions," *Eng. Rep.*, vol. 2, no. 4, 2020, Art. no. e12149.
[37] H. Wang et al., "SFNet-N: An improved SFNet algorithm for semantic segmentation of low-light autonomous driving road scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21405–21417, Nov. 2022.
[38] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239.
[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
[40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 833–851.

Min Tan received the Ph.D. degree in land resources management from the China University of Mining and Technology, Xuzhou, China, in 2021.

She is a Postdoctoral Fellow with the China University of Mining and Technology. Her research focuses on ecological restoration and greenhouse gas emissions reduction in mining areas.

Weiqiang Luo received the B.S. degree in surveying engineering from the Xi'an University of Science and Technology, Xi'an, China, in 2021. He is currently working toward the M.S. degree in photogrammetry and remote sensing with the China University of Mining and Technology, Xuzhou, China.

Jingjing Li received the B.S. degree in surveying and mapping engineering in 2023 from the China University of Mining and Technology, Xuzhou, China, where she is currently working toward the M.S. degree in photogrammetry and remote sensing.

Ming Hao received the Ph.D. degree in surveying and mapping from the China University of Mining and Technology, Xuzhou, China, in 2015.

He is currently an Associate Professor with the School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, China. He has authored or coauthored more than 40 peer-reviewed articles in international journals, such as IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING, *Remote Sensing*, *International Journal of Remote Sensing*, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. His current research interests include change detection, image processing, and image fusion for remote sensing images.