scientific reports

Check for updates

OPEN

River extraction from high-resolution remote sensing images based on non-uniform sampling and semi-supervised learning

Kun Wang[1,2,4], Lin Han[2,3] & Liangzhi Li[2,3]✉

Accurate river extraction is crucial for agricultural irrigation, water conservancy planning, and flood warning. To mitigate the issues of excessive detail loss and scarcity of labeled data in existing encoder-decoder networks, we propose a non-uniform sampling method combined with graph-based semi-supervised learning to leverage unlabeled data effectively. The method samples more points in high-frequency regions (e.g., river edges) and fewer in low-frequency regions, followed by bilinear interpolation for feature fusion. Experimental results on the Gaofen-2 dataset demonstrate that our method improves Unet, Linknet, and DeeplabV3 by 0.9, 1.5, and 1.6% in accuracy, and by 1.7, 2.9, and 1.9% in IoU, respectively. With semi-supervised learning, using all unlabeled data boosts pixel accuracy by 5.0% and IoU by 9.3%. Additionally, evaluations on the OpenEarthMap dataset and comparisons with state-of-the-art SSL methods further confirm the robustness and generalization capability of our framework.

Keywords River extraction, Non-uniform sampling, Semi-supervised learning, Remote sensing images, Semantic segmentation

Advances in remote sensing satellite technology have facilitated the acquisition of high-resolution remote sensing images, and river extraction of remote sensing images has become one of the important applications of high-resolution remote sensing images[1,2]. As an important part of the ecosystem, rivers play a vital role in people's production and life, and river extraction with high-resolution remote sensing images is of great significance in agricultural irrigation, water conservancy planning and ecological monitoring[3,4].

In recent years, significant progress has been made in deep learning, and various networks represented by fully convolutional neural networks and encoder-decoder network have achieved some results in semantic segmentation of remote sensing images[5–8]. OLAF R et al.[9] use deconvolution as an upsampling structure, connect the features of the encoder to the decoder, and fuse the low-dimensional information and high-dimensional information to construct the U-net network, which restores more spatial information, which is of great significance for fine-grained segmentation. Jianmin Su et al.[10] realized end-to-end pixel-level semantic segmentation based on U-net's improved deep convolutional neural network. A binary classification model is trained for each type of feature target, and then the prediction subgraphs are combined to generate the final semantic segmentation image. CHAURASIA A et al.[11] proposed that the Linknet network adopts a U-shape structure to add richer connections to transmit more shallow information to the deeper layers of the network and avoid excessive parameter increases. WEI X et al.[12] use the encoder network to extract the high-level semantic features of ultra-high-resolution images, and use the decoder network to map the low-resolution encoder features to the full-input resolution feature mapping to achieve pixel-level marking. In order to solve the contradiction between feature map resolution and receptive fields, the Deeplab series introduces extended convolution and pyramid pooling structure, and uses multi-scale information to further enhance the segmentation effect[13–15]. The existing codec structure semantic segmentation network is to evenly divide the image, and each pixel is a sampling point, which is classified by the segmentation algorithm. For low-frequency areas of the image, relatively smooth, small color changes, without sampling too many points. For the high-frequency region, the change is obvious, mostly the edge of the object, if the sampling point is too sparse, it will eventually lead to the

[1]School of Computer Science and Technology, Weinan Normal University, Wei Nan, Shaanxi, China. [2]Xi'an Key Laboratory of Territorial Spatial Information, Xi'an, Shaanxi, China. [3]College of Land Engineering, Chang'an University, Xi'an, China. [4]Weinan Key Laboratory of Digital Technology Innovation and Application, LTD., Wei Nan, Shaanxi, China. ✉email: 2017027010@chd.edu.cn

segmented object boundary is too smooth. Therefore, the segmented network has oversampling of smooth areas and undersampling of boundaries[16,17].

Due to the difficulty of making labeled data, the semi-supervised learning method of using unlabeled data for network training has gradually attracted attention[18–21]. Liu Kun et al.[22,23] used generative adversarial networks in semi-supervised learning, replaced the final output layer with softmax, and combined with automatic classification diagnosis for experiments. Geng Yanlei et al.[24] use ensemble prediction to train end-to-end semantic segmentation networks by optimizing the standard supervised classification loss on labeled samples and the unsupervised consistency loss on unlabeled data. Zhang Guimei et al.[25,26] use adaptive learning rate to adjust the weights of adversarial loss and cross-entropy loss, perform supervised learning at different feature layers, and use SegNet structured network discriminator to use maximum pooling for nonlinear upsampling. These methods merely incorporate unlabeled data in training without fully leveraging its potential.

Recent semi-supervised learning (SSL) methods, such as Mean Teacher and Cross Pseudo Supervision, still struggle to preserve fine edge details in river extraction due to uniform sampling strategies. Our work introduces an adaptive non-uniform sampling strategy integrated into a graph-based SSL framework, explicitly designed to address edge detail loss and label scarcity… To the best of our knowledge, this combined approach represents a novel contribution, as it has not been previously explored for river extraction.

In view of the difficulty of extracting details of river edges and the limited nature of labeled data, this paper applies non-uniform sampling to the semantic segmentation network of encoded and decoded structural images, and uses semi-supervised learning in the training to effectively use unlabeled data, and the effectiveness of the proposed method is proved by experiments, and it has strong generalization.

Materials and methods

Overall framework

Both labeled and unlabeled data are fed into the encoder-decoder semantic segmentation network of the codec structure to obtain coarse-grained and fine-grained semantic maps. By non-uniform sampling of the point sampling strategy of multi-sampling in high-frequency regions and low-frequency regions with less sampling, only those points that are most likely to be different from the surrounding pixels are calculated, and the calculation formula for high-resolution images can greatly reduce the amount of calculation. The selected key points are then bilinearly interpolated to fuse coarse-grained semantic features with fine-grained ones. By varying the ratio of labeled to unlabeled data during training, we perform semi-supervised learning to achieve optimal river extraction. The overall scheme diagram is shown in Fig. 1.

Non-uniform sampling

Fine-grained features can enhance detail, but not information that contains well-defined areas, and boundary points will have the same fine-grained features. Therefore, when the image is segmented, predicting different labels at the same point requires additional area feature information. At the same time, some fine-grained feature maps only contain some low-level information, and feature maps with more context information and semantic information can provide global information.

The non-uniform sampling diagram is shown in Fig. 2, and the coarse-grained semantic map output by the encoder and the fine-grained semantic map output by the decoder are non-uniformly sampled, and the high-frequency region is sampled more and the low-frequency region is less sampled to obtain the sampling map. The underlying features are obtained by using bilinear interpolation on the coarse-grained semantic map, and then the high-level features are obtained by 2-fold bilinear interpolation sampling on the fine-grained semantic

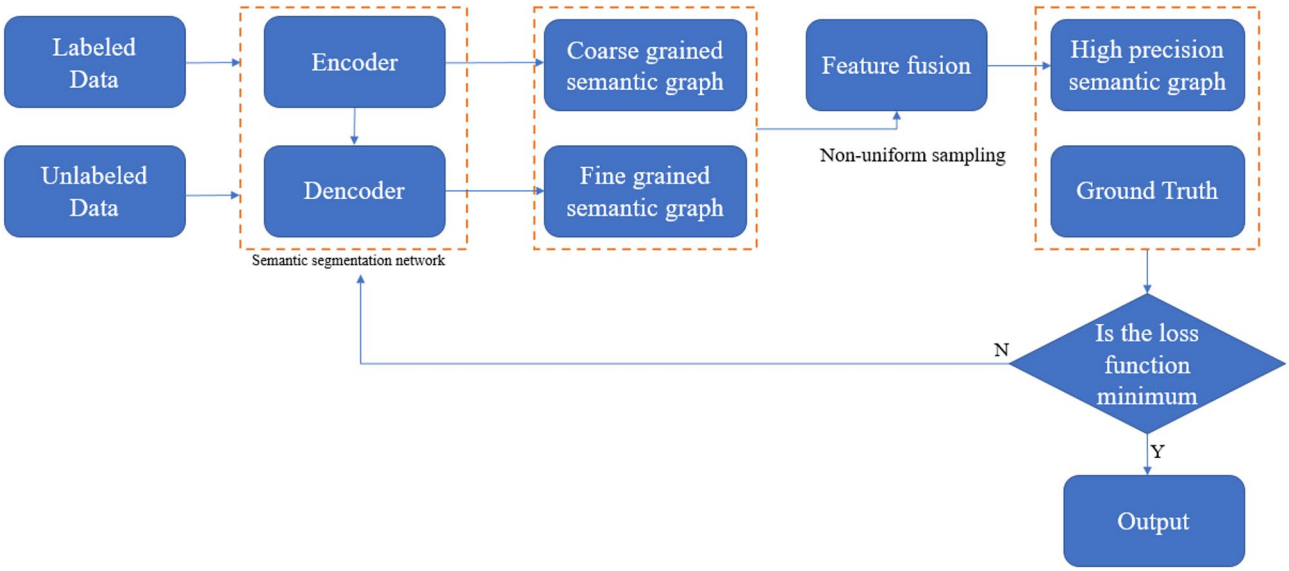

Fig. 1. Overall scheme.

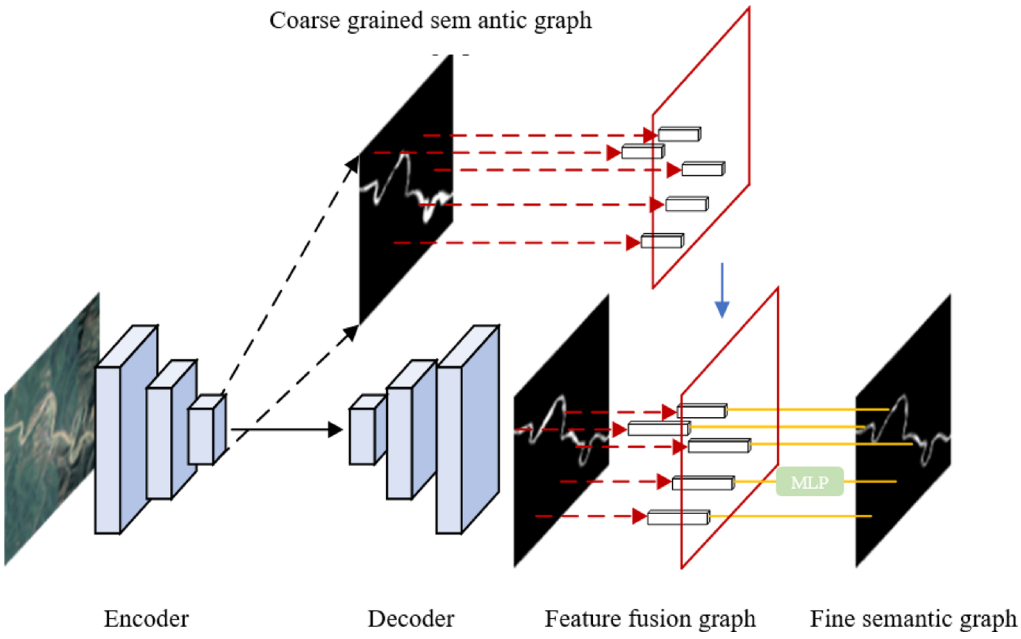

Fig. 2. Non-uniform sampling.

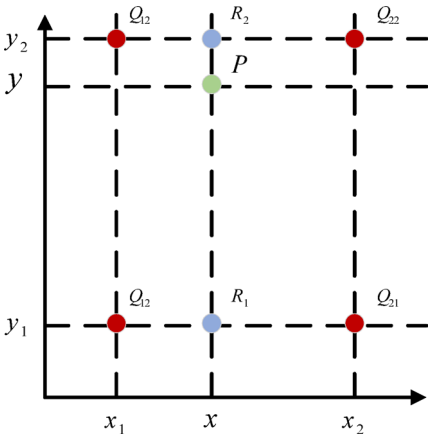

Fig. 3. Bilinear interpolation.

map, and the feature vectors are obtained by combining the underlying features and the high-level features. The concise multilayer perceptron is used to predict the feature vector, and the segmentation result of each sample point is calculated, and the multilayer perceptron shares the weight at all points and maps it to the semantic map of the same size. This process is then repeated until you have a semantic map of the desired resolution.

Bilinear interpolation is a linear interpolation extension of an interpolation function with two variables, and its core idea is to perform a linear interpolation in each direction. This is shown in Fig. 3:

If the value of the unknown function at the point is*f* obtained, assume that the known function $P = (x, y)$ is in *f*, $Q_{11} = (x_1, y_1)$ $Q_{12} = (x_1, y_2)$and the$Q_{21} = (x_2, y_1)$ $Q_{22} = (x_2, y_2)$values of the four points. It is the *f*pixel value of a pixel, which is first *x*linearly interpolated in the direction and obtained.

$$f(R_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \tag{1}$$

$$f(R_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \tag{2}$$

where, $R_1 = (x, y_1)$ $R_2 = (x, y_2)$

Then *y* linearly interpolate in the direction and get

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \tag{3}$$

Synthesis:

$$f(x,y) = \frac{f(Q_{11})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y_2 - y) + \frac{f(Q_{21})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y_2 - y)$$
$$+ \frac{f(Q_{12})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y - y_1) + \frac{f(Q_{22})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y - y_1) \tag{4}$$

Point selection strategy

Predict segmentation labels by flexibly and adaptively selecting points on the image plane, which are located near denser high-frequency regions. High-resolution imagery is effectively segmented by calculating only locations where there is a high probability that values are significantly different from their neighborhoods, and values are obtained from the difference calculated output values (obtained from the coarse prediction plot) for other locations.

For each area, iteratively output a prediction graph in a coarse-to-fine fashion, making the coarsest prediction of points on the rule network. In each iteration, bilinear interpolation is used to up-sample its previously predicted segmentation plot, and then N points with a 50% probability are selected on this denser network, and the feature representation of these N points is computed, and so on until the desired resolution is reached. Figure 4 shows the process of refining a semantic map with resolution 4×4 to 8×8, using bilinear interpolation to ×4 The predicted value on the network is increased by a factor of 2, and then the prediction is made for N uncertain points, restoring details on a finer mesh.

It is expected to output M×M pixels, the starting resolution M 0×M0, and the number of prediction points is required to not exceed to $N\log_2 \frac{M}{M_0}$ reduce the amount of computation in training.

In the exercise, a non-iterative strategy based on random sampling is adopted, N points are selected on the feature map, and uncertain regions are selected, while maintaining a certain degree of uniform coverage, following the following rules:

(1) Randomly sample kN points as candidate points (k > 1) in a uniform distribution;
(2) We prioritize uncertain points among the kN candidates after interpolation and estimate the uncertainty, and select the αN most uncertain points in kN.$\alpha \in [0,1]$
(3) The remaining (1-α) N points are randomly sampled from a uniform distribution.

When the non-iterative strategy of random sampling is trained, only the difference between the predicted value and the true value of N sampling points is calculated, which makes the calculation more efficient.

Semi-supervised learning

The dataset is mapped as a graph, each sample corresponds to a node, and the graph operation is treated as a matrix operation for the derivation of semi-supervised learning algorithms.

Given$M_t = \{(x_1, y_1), (x_2, y_2), \cdots, (x_t, y_t)\}$ and$M_n = \{x_{t+1}, x_{t+2}, \cdots, x_{t+n}\}$,$t \ll n$,$t + n = m$. Build a graph$G = (V, S)$ =$M_t \cup M_n$where the set of nodes$V = \{x_1, \cdots x_t, x_{t+1}, \cdots, x_{t+n}\}$and the similarity set *s* are an affinity matrix, based on the Gaussian function defined as:

$$S\_\{ij\} = \begin{cases} exp\left(\frac{-?x_i - x_j?_2^2}{2\sigma^2}\right) & i \neq j \\ 0 & else \end{cases} \tag{5}$$

where $i, j \in \{1, 2, \cdots, m\}$, $\sigma > 0$is the Gaussian function bandwidth parameter.

The semantic segmentation network $G = (V, S)$learns a real-valued function from the graph $\mathrm{Re}\,call = \frac{TP}{TP+FN}$with the corresponding segmentation rule :, $ACC = \frac{TP}{TP+FP}.y_i \in \{0, 1\}$ Define the loss function about and the $IOU = \frac{TP}{TP+FN+FP}$loss function $E(f)$is the smallest to obtain the optimal segmentation result of the network.

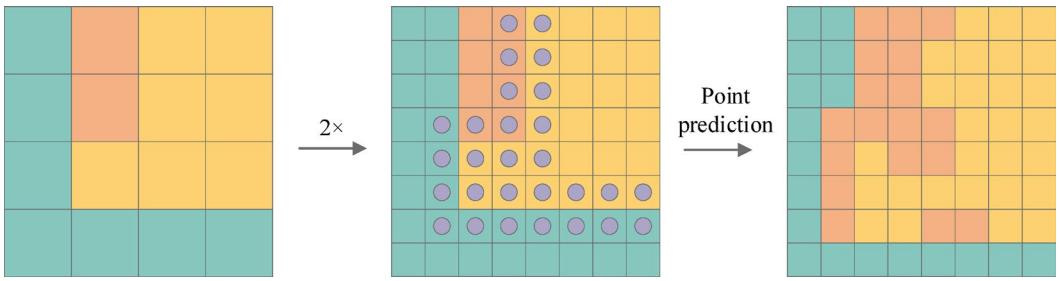

Fig. 4. Resolution 4×4 refined to resolution 8×8.

$$E(f)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}S_{ij}(f(x_i)-f(x_j))^2$$

$$=\frac{1}{2}\left(\sum_{i=1}^{m}m_i f^2(x_i)+\sum_{j=1}^{m}m_i f^2(x_j)-2\sum_{i=1}^{m}\sum_{j=1}^{m}S_{ij}f(x_i)f(x_j)\right)$$

$$=\sum_{i=1}^{m}m_i f^2(x_i)-\sum_{i=1}^{m}\sum_{j=1}^{m}S_{ij}f(x_i)f(x_j)=f^T(M-S) \tag{6}$$

where$f=\left(f_t^T f_n^T\right)^T$ $f_t=(f(x_1);f(x_2);\cdots;f(x_t))$, $f_n=(f(x_{t+1});f(x_{t+2});\cdots;f(x_{t+n}))$is f a diagonal matrix whose diagonal elements are the sum of the elements $M=diag(m_1,m_2,\cdots,m_{t+n})$of the $m_i=\sum_{j=1}^{t+n}S_{ij}$matrix's row of elements, respectively, on labeled S and unlabeled data i.

The function with the smallest loss f is satisfied on labeled data $f(x_i)=y_i\,(i=1,2,\cdots,t)$and satisfied on unlabeled data$\frac{\partial E(f)}{\partial f_n}=0$. Delimited by t row and column, expressed t as:, using a tile matrix as, $S=\begin{bmatrix}S_{tt}&S_{tn}\\S_{nt}&S_{nn}\end{bmatrix}M=\begin{bmatrix}M_{tt}&0_{tn}\\0_{nt}&M_{nn}\end{bmatrix}$then $E(f)$it can be expressed as:

$$E(f)=\left(f_t^T f_n^T\right)\left(\begin{bmatrix}M_{tt}&0_{tn}\\0_{nt}&M_{nn}\end{bmatrix}-\begin{bmatrix}S_{tt}&S_{tn}\\S_{nt}&S_{nn}\end{bmatrix}\right)\begin{bmatrix}f_t\\f_n\end{bmatrix}$$

$$=f_t^T(M_n-S_n)f_t-2f_n^T S_{nt}f_t+f_n^T(M_{nn}-S_{nn})f_n \tag{7}$$

$\frac{\partial E(f)}{\partial f_n}=0$Available:

$$f_n=(M_{nn}-S_{nn})^{-1}S_{nt}f_t \tag{8}$$

Cause:

$$W=M^{-1}S=\begin{bmatrix}M_{tt}^{-1}&0_{tn}\\0_{nt}&M_{nn}^{-1}\end{bmatrix}\begin{bmatrix}S_{tt}&S_{tn}\\S_{nt}&S_{nn}\end{bmatrix}=\begin{bmatrix}M_{tt}^{-1}S_{tt}&M_{tt}^{-1}S_{tn}\\M_{nn}^{-1}S_{nt}&M_{nn}^{-1}S_{nn}\end{bmatrix} \tag{9}$$

Namely: $W_{nn}=M_{nn}^{-1}S_{nn}, W_{nt}=M_{nn}^{-1}S_{nt}$, then .

$$f_n=\left(M_{nn}\left(I-M_{nn}^{-1}S_{nn}\right)\right)^{-1}S_{nt}f_t=\left(I-M_{nn}^{-1}S_{nn}\right)^{-1}M_{nn}^{-1}S_{nt}f_t=(I-W_{nn})^{-1}W_{nt}f_t \tag{10}$$

The $M_t$label information on is used as $f_t=(f(y_1);f(y_2);\cdots;f(y_t))$a substitute (10), that is, the f_n prediction of unlabeled data is completed.

Justification of Non-uniform sampling for river extraction

Non-uniform sampling prioritizes high-frequency regions where pixel values change rapidly, such as river boundaries. This reduces computational cost while preserving details, unlike uniform sampling which over-samples smooth regions. For river extraction, this is critical since edges are often blurred in standard encoder–decoder outputs. Our method ensures that more sampling points are allocated to boundary regions, thereby enhancing edge accuracy without increasing overall computation.

Results

Sources and processing of data

In this paper, the Gaofen Image Dataset (GID) (https://opendatalab.org.cn/OpenDataLab/GID), a large-scale dataset for land use and land cover classification, is selected as the labeled experimental dataset. It contains 150 high-quality Gaofen II images from more than 60 different cities in China, covering an area of more than 50,000 km² and an image size of 6908×7300Pixel. Twenty scenes were selected as the test and validation set, and the rest as the training set. At the same time, 20 unlabeled images of the same size were selected, including rivers, lakes, mountains, buildings, farmland, roads and railways. This study is grounded in deep learning methodologies, and the remote sensing images are selected as a combination of bands that can characterize the real features, i.e., multispectral 3, 2, 1 band combination.

In order to improve the training speed of the network, the training image is split into 512×512 size, and the corresponding label image is segmented in the same way, and the data enhancement operation is performed at the same time: rotate 90°, 180°, to the right270°; Flip left and right; Flip up and down. The image and the label map need to be synchronized for data enhancement, but the unlabeled image is not affected, and the data enhancement diagram is shown in Fig. 5. After data segmentation enhancement, 115,000 sets of labeled data and 15,000 sheets of unlabeled data were displayed.

In order to accelerate the convergence of the network, the data is normalized. Using maximum and minimum normalization, normalize the data to [0,1], and the calculation formula is:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{11}$$

Formula: x ——Original image
x'——Normalized image.
min(x)——Image pixel minimum.
max(x)——Maximum image pixels.

Evaluation indicators

The test platform is CPU I7-9700, GPU RTX2080Ti, memory 128GB, Ubuntu16 operating system, and the program running environment is TensorFlow. Two sets of experiments were set up to verify the generalization and effectiveness of non-uniform sampling and semi-supervised learning, respectively. The initial learning rate of all experimental networks is set to 2.5e-4 and the network input image batch size is set to 16. The number of iterations is 400, and the optimizer selects the Adam optimizer.

Recall, Accuracy (ACC), and Intersection Over Union (IoU) were used as evaluation indicators. Take the river pixel 1 and the background pixel as 0.

The formula for Recall is:

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{12}$$

where TP – the number of correctly predicted river pixels.

FN – The number of river pixels incorrectly predicted as background.

The pixel accuracy calculation formula is:

$$ACC = \frac{TP}{TP + FP} \tag{13}$$

where FP – the number of background pixels incorrectly predicted as the river.

The intersection-union ratio *(IoU)* is calculated as:

$$IoU = \frac{TP}{TP + FN + FP} \tag{14}$$

Extraction analysis

In order to verify the impact of non-uniform sampling on network performance and generalization, and to verify it in combination with a variety of semantic segmentation networks, Unet, Linknet and Deeplabv3 networks are selected for the network. Network parameters are initialized using the uniform variance scaling

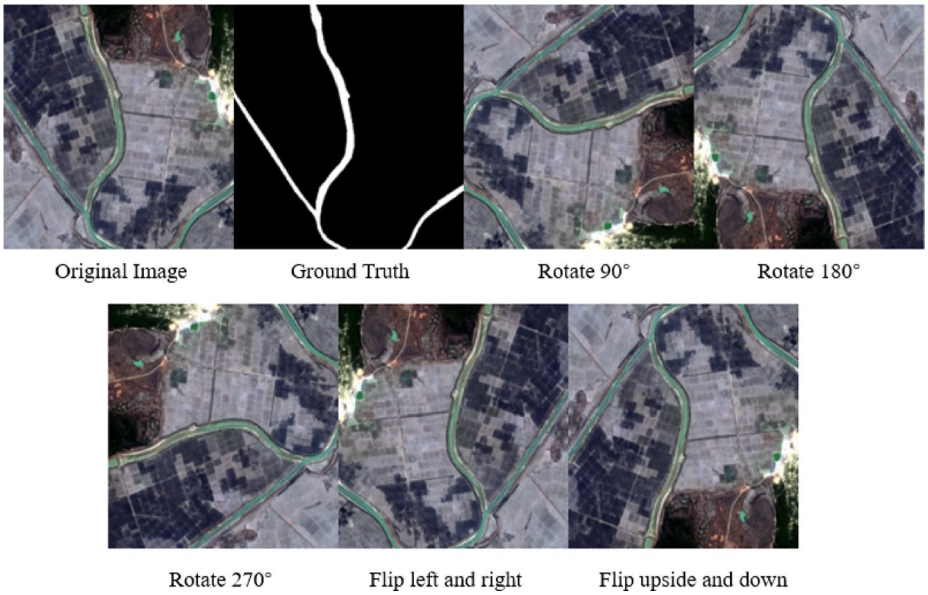

Fig. 5. Data enhancement diagram. Background/processed images: Derived from the GID under CC BY 4.0 license. Figure composition and annotation: Created by the authors using ArcGIS Desktop (version 10.8).

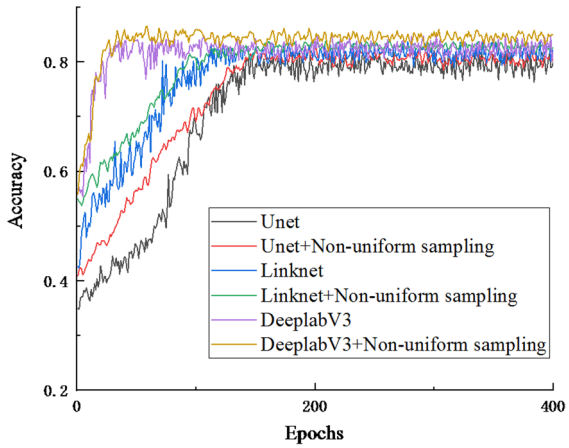

Fig. 6. Curves of pixel accuracy during each network test.

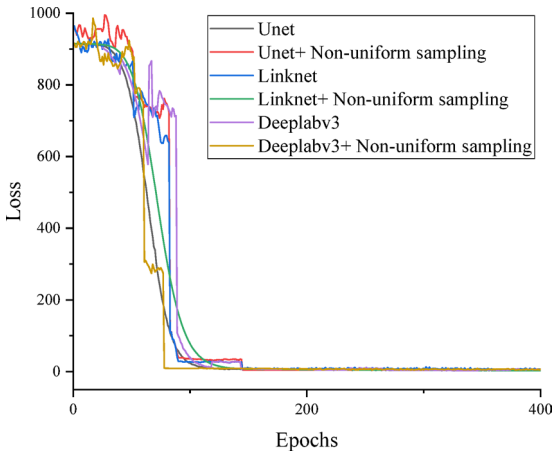

Fig. 7. Loss function curve.

method[24]. A sample is drawn from a uniform distribution in [-limit, limit], where limit is and $\sqrt{6/\text{fan_in}}$fan_in is the number of input units in the weight tensor. The loss function of all networks uses the Dice coefficient loss function. Maximize the overlap between the predicted true class and the underlying truth class (that is, maximize the dice coefficient). Therefore, the target is usually minimized and the equation is:

$$DiceLoss = 1 - \frac{2|T \cap P|}{|T| + |P|} \tag{15}$$

where $|T|$– the number of true semantic pixels.

$|P|$- Predict the number of semantic pixels.

The pixel accuracy curve in the network test is shown in Fig. 6, the loss function curves for the different network models are shown in Fig. 7, and the evaluation index is shown in Table 1. As shown in Fig. 6; Table 1 that the non-uniform sampling strategy proposed in this paper can make the original network converge quickly during the training process, and the accuracy is higher than that of the original Unet, Linknet, and the Deeplab V3 network increased by 0.9%, 1.5% and 1.6%, respectively, and the cross-merger ratio increased by 1.7% and 2.9%, respectively and 1.9%. Figure 7 presents visual segmentation results on the test data. It can be seen from the figure that the rivers extracted by the original Unet network have a high false lifting rate for roads and buildings, and certain details of the extracted rivers are lost. From the figure, it can be seen that the speed of convergence is accelerated, and the reason for this phenomenon may be due to the fact that due to the semi-supervised network, fewer features need to be learned, and therefore the speed of network training is increased, and the speed of convergence is also incremented. The rivers extracted by the original Linknet network have a certain improvement in the false extraction of buildings and roads, but the extracted rivers are lost in the edge details, and the false lifting rate of the roads adjacent to the river is high. The river extracted by the original Deeplab V3 network is the best of the three networks, but there is still some misextraction and loss of detail. After introducing non-uniform sampling, all three networks reduced the mis-extraction of mountain shadow,

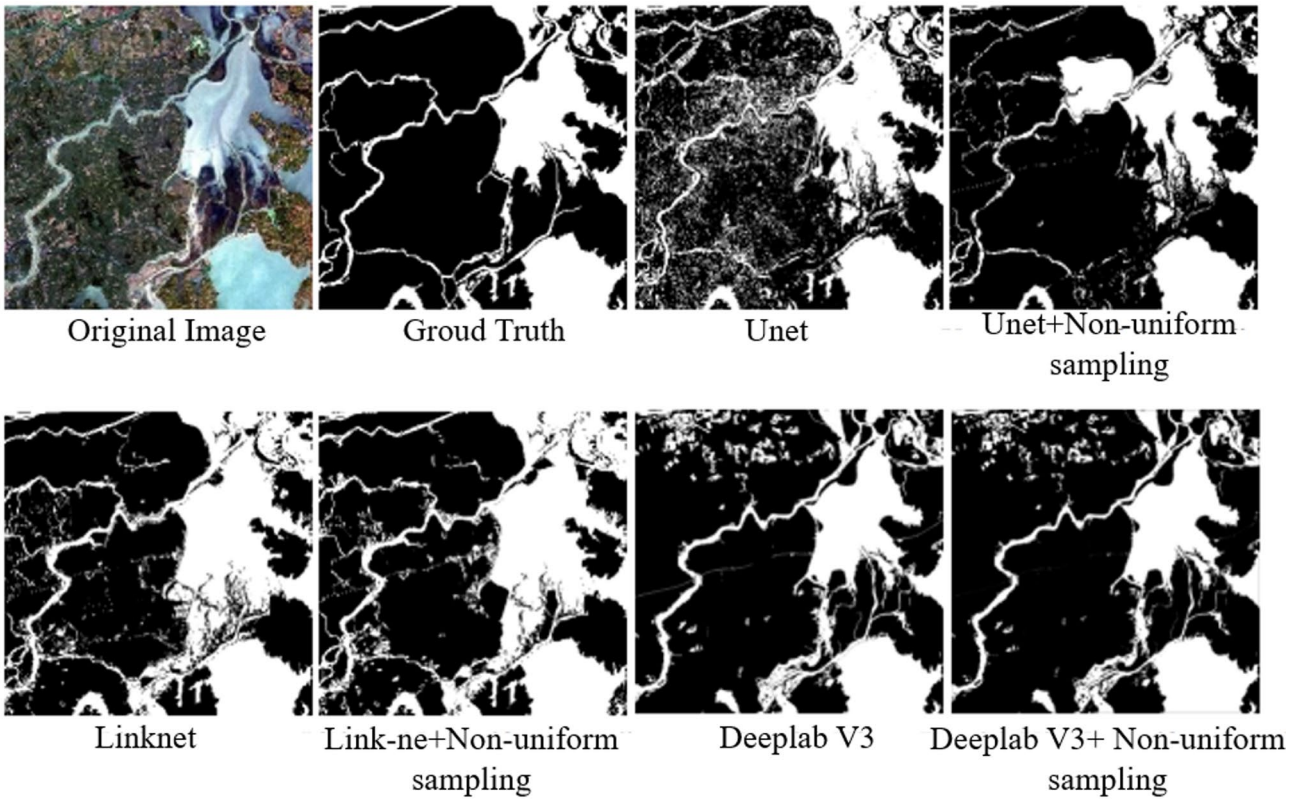

Fig. 8. Comparison of river extraction results across different networks. Background images: GID, CC BY 4.0. Result overlays and visualizations: Generated by the authors using Python (Matplotlib v3.5.1, OpenCV v4.5.5).

Network	Recall	ACC	IoU
Unet	0.796	0.794	0.538
Unet + Non-uniform sampling	0.817	0.813	0.558
Linknet	0.816	0.821	0.612
Linknet + Non-uniform sampling	0.827	0.836	0.641
Deeplabv3	0.829	0.833	0.637
Deeplabv3 + Non-uniform sampling	0.838	0.849	0.656

Table 1. Evaluation indicators of each network.

roads, vegetation and other features, and avoided the interruption of small rivers and the extraction of river edge details, there is a big improvement.

The impact of semi-supervised learning on network performance

In order to verify the effect of semi-supervised learning on water extraction from high-resolution remote sensing images, the non-uniformly sampled Deeplab V3 network was used as the benchmark network, and the unlabeled data were 1/8, 1/4, 1/2 and 1, respectively the proportion of network training was compared with labeled data. In 1/8, 1/4, 1/2 and all unlabeled data combined with labeled data, the pixel accuracy of the benchmark network is improved by 0.9% ,2.0%, 3.8% and 5.0% respectively, the cross-merge ratio increased by 0.8%, 3.1%, 5.6% and 9.3% respectively, accuracy and intersection ratio contrast line chart are shown in Fig. 9. The change in accuracy is shown in Fig. 10, and the evaluation indicators are shown in Table 2.

Figure 10; Table 2 indicate that the introduction of unlabeled data can not only effectively improve the accuracy of river extraction, but also does not affect the convergence of the entire network, which proves the effectiveness of semi-supervision. Figure 11 shows the effect of river extraction on the dataset with different proportions of unlabeled data, with the increase of unlabeled data, the false extraction of buildings and roads gradually decreases, the continuity of river extraction is improved, small rivers are more coherent, and edge details extraction is more accurate, which can effectively optimize the details of the edge of the water body and provide accurate data for subsequent data analysis.

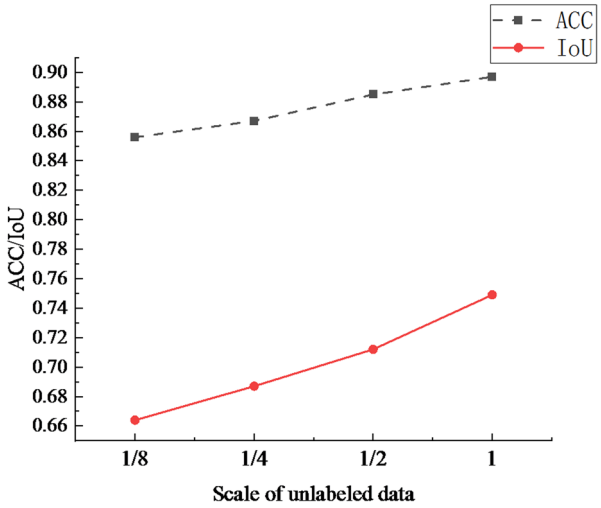

Fig. 9. Accuracy and contrast line chart.

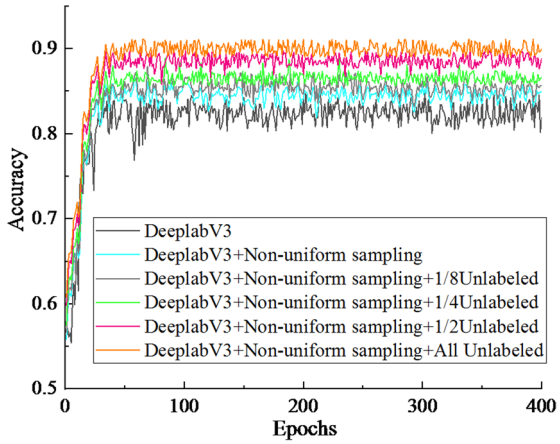

Fig. 10. Network pixel accuracy change.

Network	Unlabeled data	Recall	ACC	IOU
DeeplabV3	0	0.829	0.831	0.637
DeeplabV3 + Non-uniform sampling	0	0.838	0.849	0.656
DeeplabV3 + Non-uniform sampling	1/8	0.841	0.856	0.664
DeeplabV3 + Non-uniform sampling	1/4	0.848	0.867	0.687
DeeplabV3 + Non-uniform sampling	1/2	0.854	0.885	0.712
DeeplabV3 + Non-uniform sampling	1	0.862	0.897	0.749

Table 2. Network evaluation indicators.

Comparison with semi-supervised learning methods

We compare our approach with two established SSL methods: Mean Teacher[27] and Cross Pseudo Supervision (CPS)[28]. Results in Table 3 show that our method outperforms both in terms of IoU and F1-score, especially when limited labeled data are available. To demonstrate generalization, we evaluated our method on the OpenEarthMap dataset.

To demonstrate generalization, we evaluated our method on the OpenEarthMap dataset[29], a global-scale land cover mapping dataset. Our approach achieved an IoU of 77.8% and an F1-score of 85.2% for river extraction, confirming its robustness across different geographical regions. The results suggest that our method is not limited to specific sensor characteristics and can generalize to diverse remote sensing data sources.

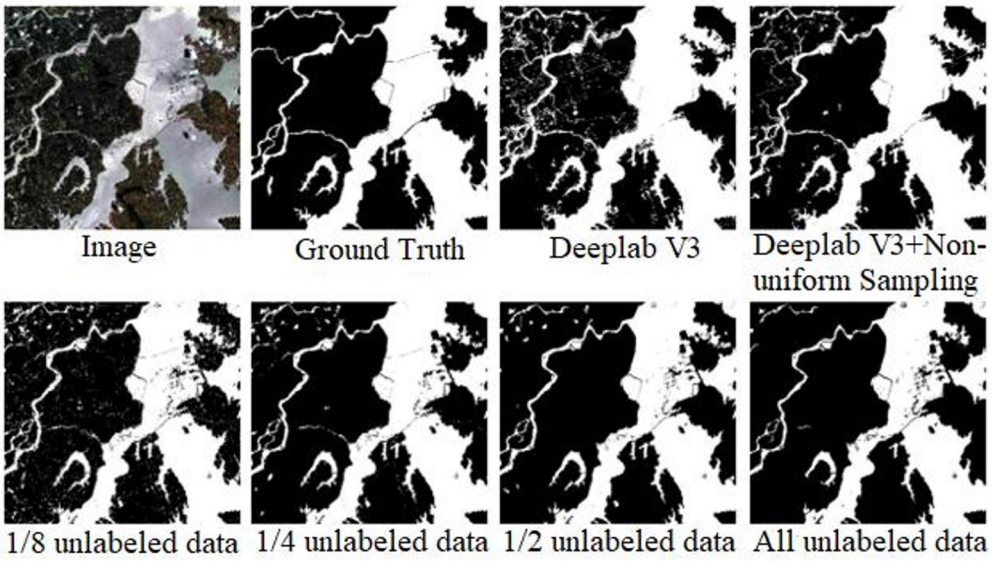

Fig. 11. Semi-supervised learning network extracts comparison results. Background images: GID, CC BY 4.0. Result overlays and visualizations: Generated by the authors using Python (Matplotlib v3.5.1, OpenCV v4.5.5).

Method	IoU (%)	F1-score (%)
Mean teacher	64.3	77.8
CPS	65.1	78.5
Ours	77.8	85.2

Table 3. Performance comparison with semi-supervised learning methods.

Model	FLOPs (G)	Inference Time (ms)	GPU Memory
Unet	45.2	32	3.2
DeeplabV3	52.1	41	3.8
Ours	48.7	35	3.5

Table 4. Computational cost Comparison.

Generalization test on openearthmap dataset

We report the computational cost of our method compared to baseline networks in Table 4. The non-uniform sampling strategy reduces the number of points processed, leading to lower FLOPs and inference time compared to DeeplabV3, while maintaining higher accuracy. All experiments were conducted on an NVIDIA RTX2080Ti GPU (Table 4).

Discussion

This study focuses on river extraction from remote sensing images, employing deep learning as the foundational approach, and proposes a non-uniform sampling method for the existing coded structure extraction network, which has the problems of too much detail loss and the sample labels are not easy to obtain. Semi-supervised learning is introduced to make full use of unlabeled data to solve the problems of insufficient detailed feature extraction during river extraction and low accuracy in the case of fewer sample labels. By experimental comparison, this paper's method improves 0.9%, 1.5%, and 1.6% over Unet, Linknet, and Deeplab V3 networks, and 1.7%, 2.9%, and 1.9% over intersection and merger ratios, respectively. With the introduction of semi-supervised learning, it is found experimentally that the benchmark network improves the pixel accuracy by 0.9%, 2.0%, 3.8%, 5.0%, and the intersection and merger ratios by 0.8%, 3.1%, 5.6%, and 9.3% for the cases of 1/8, 1/4, 1/2, and all unlabeled data in combination with labeled data, respectively. Our experimental results demonstrate that the introduction of unlabeled data can not only effectively improve the river extraction accuracy, but also did not affect the convergence of the whole network, proving the effectiveness of introducing semi-supervised learning. The experimental results and output maps demonstrate that our method mainly improves on the river edge features, the distinguishing features from other regular land classes such as roads and buildings, and the recognition of the shadow image features present in the images.

Although this paper is based on remote sensing images of rivers as the extraction object, the proposed modeling method has a certain degree of generalization, which is also applicable to other typical land classes such as roads and buildings, but this requires further experiments and further modification of the model based on the data, which is also part of the future research content.

Conclusions

Aiming at the problems of excessive detail loss and less labeled data in the existing river extraction network of codec structure, a non-uniform sampling method is proposed to solve the problem of detail loss, and the coarse-grained semantic map output by the encoder and the fine-grained semantic map output by the decoder are interpolated bilinearly. By sampling more in the high-frequency region and less sampling at low frequencies, it is avoided to count all pixels excessively. The selected key points are interpolated bilinearly to complete the fusion of coarse-grained semantic features and fine-grained semantic features. Through semi-supervised learning training network, network training using easily available unlabeled data is used to reduce costs and improve network accuracy. The experimental results validate the strong generalization capability and effectiveness of the proposed method.

Data availability

Please contact the first author Dr. Kun Wang if someone wants to request the data from this study.

Received: 27 August 2025; Accepted: 29 January 2026
Published online: 31 January 2026

References

1. Rekha, B., Desai, V., Ajawan, .P, et al. Remote sensing technology and applications in agriculture. *ICCTEMS*, 193–197 (2018).
2. Zhe, L. I. U. et al. Review on crop type fine identification and automatic mapping using remote sensing. *Trans. Chin. Soc. Agricultural Mach.* **49** (12), 1–12 (2018).
3. Sun, F. et al. Projecting meteorological, hydrological and agricultural droughts for the Yangtze river basin. *Sci. Total Environ.* **696**, 134076 (2019).
4. Yuan, X. U. E. et al. Automatic extraction of small mountain river information and width based on China-made GF-1 satellites remote sensing images. *Bull. Surveying Mapp.* **03**, 12–16 (2020).
5. Ding, H. & Jiang, X. Semantic segmentation with context encoding and Multi-Path Decoding. *IEEE Trans. Image Process.* **29**, 3520–3533 (2020).
6. Fu, X. Qu H. Research on semantic segmentation of high-resolution remote sensing image based on full convolutional neural network. *ISAPE* 978–984. (2018).
7. Tong, Z. H. A. N. G. et al. Remote sensing image scene classification based on deep Multi-branch feature fusion network. *Acta Photonica Sinica*. **49** (5), 166–177 (2020).
8. Hai-quan, F. A. N. G. et al. River extraction from high-resolution satellite images combining deep learning and multiple chessboard segmentation. *Acta Scientiarum Naturalium Universitatis Pekinensis*. **55** (4), 692–698 (2019).
9. Olaf, R. Philipp, F. & Thomas, B. U-Net: convolutional networks for biomedical image segmentation. *MICCA*, 234–241. (2015).
10. Jian-min, S. U., Lan-xin, Y. A. N. G. & Wei-peng, J. I. N. G. U-Net based semantic segmentation method for high resolution remote sensing image. *Comput. Eng. Appl.* **55** (7), 207–213 (2019).
11. Chaurasia, A. & Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. *VCIP* 462–466. (2017).
12. Wei., X, Guo, Y. J, Gao, X. A new semantic segmentation model for remote sensing images. *IGARSS* 1776–1779. (2017).
13. Chen, L., Papanderou, G. & Kokkinos, I. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40** (4), 834–848 (2016).
14. Developing an intelligent. Cloud attention network to support global urban green spaces mapping[J]. *ISPRS J. Photogrammetry Remote Sens.* **198** (04), 197–209. https://doi.org/10.1016/j.isprsjprs.2023.03.005 (2023).
15. Papandreou, G. et al. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. *ICCV* 1742–1750. (2015).
16. Chen, L., Papandreou, G. & Schroff, F. Rethinking atrous Convolution for semantic image segmentation[J/OL]. https://arxiv.org/abs/1706.05587 (2017).
17. A Novel Spectral Indices-Driven Spectral-. Spatial-context attention network for automatic cloud detection. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.*, **16**(04): 3092–3103. https://doi.org/10.1109/JSTARS.2023.3260203. (2023).
18. Chen, H., Song, J., Wu, C., Du, B. & Yokoya, N. *ISPRS J. Photogrammetry Remote Sens.* **206**, 87–105. https://doi.org/10.1016/j.isprsjprs.2023.11.004 (2023).
19. Lan, D. U. et al. SAR target detection network via Semi-supervised learning. *J. Electron. Inform. Technol.* **42** (01), 154–163 (2020).
20. Isikdogan, F., Bovik, A. C. & Passalacqua, P. Surface water mapping by deep learning. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **10** (11), 4909–4918. https://doi.org/10.1109/JSTARS.2017.2735443 (2017).
21. Han Yan-ling, Z. H. A. O. et al. Cooperative active learning and semi-supervised method for sea ice image classification. *Haiyang Xuebao*. **42** (01), 123–135 (2020).
22. Chen, H., Song, J., Han, C., Xia, J. & Yokoya, N. ChangeMamba: Remote sensing change detection with spatiotemporal state space model. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–20, https://doi.org/10.1109/TGRS.2024.3417253 (2024).
23. Kun, L. I. U., Dian, W. A. N. G. & Meng-xue, R. O. N. G. X-Ray image classification algorithm based on Semi-Supervised generative adversarial networks. *Acta Optica Sinica*. **39** (08), 117–125 (2019).
24. Yan-lei, G. E. N. G. et al. High-resolution remote sensing image semantic segmentation based on semi-supervised full Convolution network method. *Acta Geodaeticaet Cartogr. Sinica*. **49** (04), 499–508 (2020).
25. Chen, H. et al. ObjFormer: Learning Land-Cover Changes From Paired OSM Data and Optical High-Resolution Imagery via Object-Guided Transformer. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–22, https://doi.org/10.1109/TGRS.2024.3410389 (2024).
26. Wang, K., Liangzhi, Li & Ling Han & A decoupled search deep network framework for high-resolution remote sensing image classification. *Remote Sens. Lett.* **14** (3), 243–253. https://doi.org/10.1080/2150704X.2023.2185110 (2023).
27. Tarvainen, A. & Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural. Inf. Process. Syst.* ;**30**. (2017).
28. Chen, X., Yuan, Y., Zeng, G. & Wang, J. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio.*. pp. 2613–2622. (2021).

29. OpenEarthMap A benchmark dataset for global high-resolution land cover mapping. *Int. J. Appl. Earth Obs. Geoinf.* **117**, 103201 (2023).

Acknowledgements
We are grateful to those involved in data processing and manuscript writing revision.

Author contributions
Kun Wang authored the primary manuscript, conducting analysis and implementation of its core methodologies. Liangzhi Li provided conceptual guidance and methodological discussions for the manuscript. Lin Han edited and processed the images and tables. All authors participated in the manuscript's review and verification process.

Funding
This work is supported by Xi'an Key Laboratory of Territorial Spatial Information (300102355505). Natural Science Basic Research Program of Shaanxi Province (No. 2025JC-YBQN-407). National Science Foundation of China (No. 211035210511). Shaanxi Province Youth Science and Technology Rising Star Project (NO. 2025ZC-KJXX-148), Natural Science Basic Research Program of Shaanxi (No. 2025JC-YBMS-333, 2025JC-YBMS-334). The Fundamental Research Funds for the Central Universities, CHD (300102355201).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.