



Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Deep learning-based automated terrain classification using high-resolution DEM data

Jiaqi Yang ^{a,b}, Jun Xu ^{a,*}, Yunshuo Lv ^{a,c}, Chenghu Zhou ^{a,b}, Yunqiang Zhu ^{a,d}, Weiming Cheng ^{a,b}

^a State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China

^b College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

^c School of Geography, Liaoning Normal University, Dalian 116029, China

^d Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China



ARTICLE INFO

Keywords:

Landform classification
AW3D30
Semantic segmentation
Fully convolutional network
Residual network

ABSTRACT

Landforms are a fundamental component of the natural environment, and digital terrain mapping on a large spatial scale is important when studying landforms. In this study, we adopted a semantic segmentation model in computer vision to classify elementary landform types using AW3D30 digital elevation model (DEM) data. We built a semantic segmentation model with an FCN-ResNet architecture that extracts features using a residual network (ResNet) and obtains pixel-level segmentation of the DEM using a fully convolutional network (FCN). A lightweight decoder based on skip connections was adopted to maintain detailed information at different scales. We used the 1:1,000,000 Chinese landform map as the label and tested different combinations of terrain factors. The experiments indicate that increasing the terrain factors has no significant influence on the model, and the semantic information can be learned using only DEM data. The model has strong feature extraction capability and is tolerant to noise and error. The results of landform category prediction confirm that deep learning methods have strong potential for landform classification and will have great application prospects in the field of geomorphological research.

1. Introduction

Landforms are one of the most fundamental elements of the natural environment and influence the spatial differentiation of aspects such as the ecological environment and natural resources. Landform mapping, particularly for large spatial scales, is an important geomorphological investigation method and is crucial to research in geoscientific disciplines (Evans, 2012). However, landform mapping for broad areas has been difficult due to a lack of systematic data. The rapid progress of remote sensing and geographical information system (GIS) technologies has provided abundant remote sensing images and digital elevation model (DEM) data as well as analytical tools that have made automatic or semiautomatic landform classification possible (Goodchild, 2005; MacMillan et al., 2004). For example, Cheng et al. (2011) compiled a new 1:1,000,000 digital geomorphologic atlas of China with visual interpretation from Landsat TM/ETM imagery and SRTM-DEM data;

Giano et al. (2020) realized semiautomatic landform classification tools using GIS software; and MacMillan et al. (2004) developed a toolkit for identifying spatial entities at different scales using DEMs. Using automatic terrain classification to improve the efficiency of digital landform mapping has drawn great attention (Iwahashi and Pike, 2007; Li and Hsu, 2020; Prima et al., 2006; Sinha and Mark, 2010; Theobald et al., 2015).

Landform mapping is usually based on the morphology, genesis, chronology and dynamic process of the topography (Minár and Evans, 2008). A DEM is a simulation of the Earth's surface using elevation data. A series of terrain factors can be derived from DEM data, such as aspect, slope, and curvature, through which landform information can be deduced (Wilson, 2012; Xiong et al., 2022). Therefore, DEM data are effective tools for terrain analysis and classification, and landform mapping using DEM is generally based on topographic morphology. Numerous studies addressing terrain classification with DEM data have

* Corresponding author at: State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China.

E-mail address: xujun@lreis.ac.cn (J. Xu).

emerged, which mainly use pixel-based and object-based approaches. Pixel-based approaches automatically cluster pixels by assigning each pixel to one or more landform classes according to threshold values of DEM and terrain factors (Graff and Usery, 1993; Iwashashi and Pike, 2007; Miliareis and Argialas, 1999; Prima et al., 2006). Pixel-based approaches use only pixel information and neglect geometric and contextual information, so the obtained results of class patches are fragmented. Moreover, the thresholds for the same landform in different regions could be different. Therefore, it is difficult to classify landforms using a set of fixed thresholds, especially in large regions. Object-based methods segment images into morphological units that have internal homogeneity and external heterogeneity (Blaschke, 2002; Drăguț and Blaschke, 2006; Drăguț and Eisank, 2012; Minár and Evans, 2008). The segmentation is either based on geometric characteristics, such as ridge and valley lines, convex and concave flexures that can obtain the units of geomorphological elements such as pit, peak, and slope (MacMillan et al., 2004; Minár and Evans, 2008), or based on characteristics of parameters, such as terrain factors and surface textures (Blaschke, 2002; Drăguț and Eisank, 2012; Iwashashi et al., 2018; Na et al., 2021). Object-based methods exploit the inherent characteristics of landforms and conform more to human cognition of terrain and are thus widely used in terrain classification. However, they are insensitive to continuously changing landforms, which are widely distributed on Earth (Li et al., 2020).

With the improvement in computing power and the increase in data volume, deep learning methods have developed rapidly. Deep learning methods can automatically extract deep features from data without domain experts (LeCun et al., 2015). Researchers have employed deep learning models in the field of geomorphology to identify certain types of terrain features (Li and Hsu, 2020; Li et al., 2017). Du et al. (2019) adopted a multimodal convolutional neural network (CNN) model and recognized different types of landforms using imagery and DEM data. Li et al. (2020) used the U-Net model to accomplish loess landform classification using imagery and DEM data. Yan et al. (2021) introduced an attention module based on the U-Net model and designed a method to extract glaciers using remote sensing images. This method performed better than previous traditional machine learning methods. However, most of these works focused on the identification of certain landform elements, such as craters and volcanoes (Li and Hsu, 2020; Li et al., 2017; Yan et al.; 2021). Few studies have been performed on classifying repeating landform types, which is considered more difficult than identifying component landform elements (MacMillan et al., 2004).

Recently, deep learning methods based on CNNs have become mainstream in the field of computer vision and have been widely used in tasks such as image classification, target detection and image segmentation (He et al., 2016; Krizhevsky et al., 2017; LeCun et al., 2015; LeCun et al., 1998; Simonyan and Zisserman, 2014). Image segmentation is an important component in vision understanding systems. In the field of image segmentation, semantic segmentation is formulated as a pixel classification problem with semantic labels. Semantic segmentation labels all image pixels at the pixel level and provides a set of object classes for all image pixels. Therefore, it is usually more difficult than image classification (Minaee et al., 2021). Using CNNs, Long et al. (2015) transformed fully connected layers into convolutional layers, realizing the first application of the CNN method in the field of semantic segmentation. Subsequently, U-Net was proposed and made a great breakthrough in medical image segmentation tasks (Ronneberger et al., 2015) and has since been widely used in the segmentation of remote sensing images with satisfactory results (Aghdami-Nia et al., 2022; Malik et al., 2021; Siddique et al., 2021; Zhang et al., 2021). The DeepLab series of methods has successfully introduced the strategy of atrous convolution to extend the perceptual domain and further improve the performance of many semantic segmentation tasks (Chen et al., 2014; Chen et al., 2017; Chen et al., 2018a; Chen et al., 2018b). Additionally, attentional mechanisms have improved the performance of semantic segmentation tasks as they continue to evolve (Oktay et al.,

2018; Vaswani et al., 2017). In recent years, an increasing number of semantic segmentation models have been employed in earth science, such as in forestry, terrain feature, and street scenes (Cheng et al., 2022; Guo et al., 2022; Zhang et al., 2022). Yuan et al. (2021) summarized the deep learning models and training strategies applied in the field of remote sensing image segmentation and compared the performance of different models on different remote sensing image datasets.

Traditional landform mapping requires extensive domain expert knowledge. Deep learning methods have been applied in identifying landform types due to their strong feature learning capacity. Nevertheless, current works using deep learning in geomorphology mostly have focused on identifying component landform elements, and there is a lack of work on mapping repeating landform types. Semantic segmentation models based on deep learning make it possible to classify repeating landform types with DEMs. However, semantic segmentation models were developed for natural images, such as animals, buildings, and plants. Unlike natural images, which have objects with distinct boundaries, DEMs are continuously changing land surfaces, and there are no distinct boundaries between different landform classes. Therefore, the ability of deep learning to segment landform types is unclear. In this study, we developed a semantic segmentation model based on deep learning to classify elementary landform classes by using 30 m-resolution DEM data at the pixel level. The purpose is to examine the applicability of deep learning in classifying repeating landform types. The main contributions of this study are as follows:

- i. A fully convolutional network-residual network (FCN-ResNet) model was constructed for both single-channel and multichannel segmentation, and a weighted cross-entropy loss function was adopted to mitigate unbalanced sampling.
- ii. Elementary landform types were obtained using the AW3D30 DEM data based on the FCN-ResNet model, and the results were analyzed and compared with those of other deep learning segmentation models.
- iii. The effects of using different terrain factors in the FCN-ResNet model were analyzed.
- iv. Three labeling strategies were used to investigate the effect of the labeling precision caused by scale differences on the prediction results of the model.

2. Data and methodology

2.1. Data description

The DEM used in this study was the AW3D30 data from the Japan Aerospace Exploration Agency.¹ The AW3D30 DEM has a spatial resolution of 30 m, which is recognized as a set of global DEM data products with high accuracy and is widely used in digital terrain analysis (Shumack et al., 2020; Zhao et al., 2022). Based on the DEM, some terrain factors, such as slope, hillshade, relief degree of land surface (RDLS), and curvature were calculated. The terrain factors were used with the DEM data as the input for the deep learning network. As manual labeling of the input data is time-consuming, we chose sampling and testing areas in China and adopted the first-level landform classes in China's 1:1,000,000 landform map (Cheng et al., 2011) to label the input data automatically.

China's 1:1,000,000 landform map is a digital geomorphology map based on remote sensing and GIS technology. Its classification system has five levels, which combine characteristics of morphology with genetic types (Cheng et al., 2011). The first level divides landforms into seven types based on morphology: plains, platforms, hills, low relief mountains, middle relief mountains, high relief mountains and highest relief mountains. The second level is divided based on altitude, which

¹ <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/data/index.html>.

can be obtained by setting different elevation thresholds. The combination of altitude types and relief types together forms 25 terrestrial geomorphologic types. The third and fourth levels divide the landforms according to the genesis and subgenesis of the landforms, and the fifth level divides the landforms further based on the subgenesis of the landforms.

In this study, our goal is to distinguish the fundamental landform types by morphology, so we only address the first-level landform types. There are seven first-level landform classes, four of which are mountains with different reliefs. Platforms, referred to terraces in many studies, are similar to plains in that they both have flat surfaces, but platforms are different from plains in that they have raised horizontal surfaces. Cheng et al. (2011) characterized platforms as having an altitude relief of more than 30 m. Usually, platforms are on transition zones between plains and hills or mountains, so they are difficult to accurately divide from plains or hills. Compared to mountains, hills have gentler slopes and rounder peaks. As the four types of mountains have similar morphological characteristics of sharp peaks and steep slopes that can be distinguished by RDLS, we merged these classes into a single mountain class. Therefore, only four classes were labeled: plain, platform, hill, and mountain. The sampling and testing areas were selected from different geographic subdivisions, with a rich variety of landform types in each area. Fig. 1 shows China's 1:1,000,000 landform map, as well as seven sampling and six testing areas selected for the experiment.

The data in the sampling areas were used for model training and accuracy assessment. They were cropped into 1024×1024 pixel images to ensure that a single image contained sufficient spatial information to help the model perform semantic segmentation. There were no overlapping regions between any two images in case of information leakage between the training and validation sets. We obtained 2935 images from

the sampling areas and randomly selected 2364 images as the training set and the remaining 571 as the validation set. The data in the testing areas were used to test the performance and generalization ability of the model. The cropping method was the same as that used for the sampling areas. A total of 483 images were obtained from the testing area.

Table 1 shows the statistics for the sampled landform types in the sampling areas. The distribution is unbalanced, i.e., “mountain” and “plain” types occupy a large proportion, while “platform” only comprises a very small proportion. The influence of sample imbalance must be considered during model training.

2.2. Segmentation model

An FCN-ResNet architecture was used in the experiments. ResNet (He et al., 2016) was used to extract features from terrain factors, and a classifier with an FCN structure (Long et al., 2015) was implemented to segment the region based on the extracted features.

While training a neural network, as the depth of the network and the number of parameters of the model increase, the model is expected to have a stronger fitting ability. However, owing to issues such as vanishing/exploding gradients, directly increasing the depth of the model

Table 1
Proportion of each landform type in the sampling areas.

Landform type	Label	Proportion (%)
No data	0	1.60
Plain	1	28.48
Platform	2	10.52
Hill	3	20.14
Mountain	4	39.26

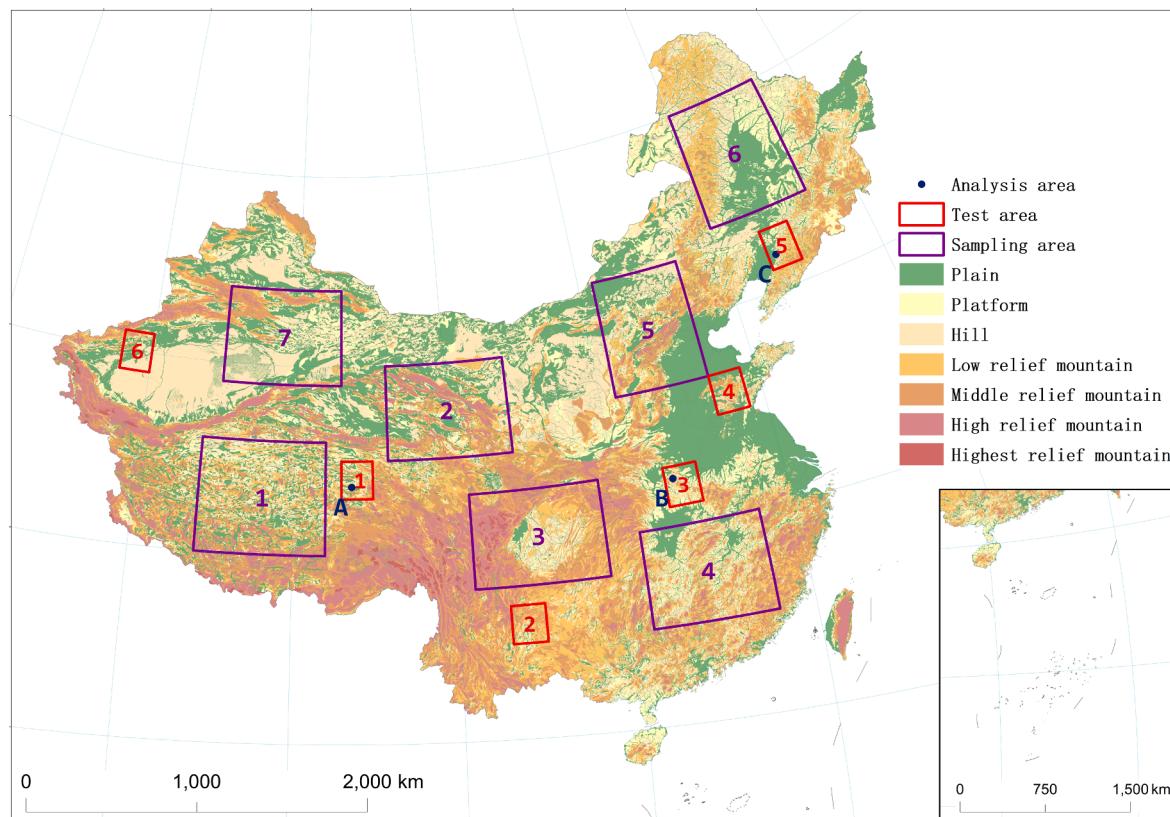


Fig. 1. First-level landform classes in China's 1:1,000,000 landform map (Cheng et al., 2011). The areas within the purple boxes are sampling areas used for training and validation sets in the experiments, and the areas within the red boxes are testing areas used for testing the model. The black dots mark the three places where the predicted results were analyzed, as discussed in Section 3.3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by stacking convolutional layers makes training more difficult, and the model may degenerate. Deep ResNets use a 3-layer bottleneck structure in each residual block, which introduces shortcut connections (He et al., 2016), as shown in Fig. 2, making the network easier to optimize and solving the training problem of deep networks. In addition, this connection does not require any additional parameters. Thus, ResNet can be used to extract image features and has a wide range of applications in computer vision tasks.

We used the ResNet-50 (50-layer ResNet) model pretrained on the ImageNet dataset (Deng et al., 2009) to extract image features. Fig. 3 shows the model for the single-channel segmentation. The ResNet-50 structure has four “Layers” of different depths, with each “Layer” consisting of a number of bottlenecks that extract features at different scales. The features extracted from the last three “Layers” are passed into the classifier for classification to refine the semantic and spatial accuracy by incorporating detailed information at different scales.

The model structure for multichannel tasks was basically the same as that for single-channel tasks. However, different to the 3-channel natural images, the multichannel data used in this study are multi-modal data, such as slope, aspect, hillshade, RDLS. To fully extract different features from multi-modal data, we used a ResNet module for each channel to extract its features instead of simply stacking channels, as Du et al. (2019) did in their work. The features of the different channels were then concatenated and passed into the classifier to complete the classification. There is still only one classifier, and its structure is the same as that of the classifier in Fig. 3, except there are more channels for the features at each depth.

The FCN was the first network to achieve pixel-level segmentation of images (Long et al., 2015). It is mainly based on CNN, and all its fully connected layers are transformed into convolutional layers, so that the model does not depend on the size of the input image. To avoid losing accuracy in the prediction process, the FCN model uses skip connections that incorporate features of different depths during upsampling to ensure the integrity of detailed information. Based on the FCN-8s model (Long et al., 2015), we designed a relatively lightweight classifier that contained three 3×3 convolution layers and one 1×1 convolution layer at the end to adjust the number of channels (Fig. 3). The entire model is trained in an end-to-end manner.

After the last convolution layer of the classifier, a result was obtained containing five channels of the same size as the input images, representing the model’s predicted values for five categories: no data, plain, platform, hill, and mountain. The higher the value is, the higher the probability that the pixel belongs to this category. The final map of landform types was obtained by performing the *argmax* operation on

these five channels. Except for the final results, we also analyzed the heatmaps of the predicted values for each landform type, which we believe provide more valuable information than the final result.

2.3. Loss function and evaluation metrics

In this experiment, weighted cross entropy (WCE) was used as the loss function. It is a common loss function in semantic segmentation and is expressed as follows (Chong and Zak, 2013):

$$Loss = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C w_c * \log \frac{\exp(x_{n,c})}{\sum_{i=1}^C \exp(x_{n,i})} y_{n,c} \quad (1)$$

where N represents the size of the batch, C represents the number of landform types, x and y represent the predicted result and the label, respectively, and w_c represents the weight for each labeled type. In this experiment, the weight w for the “no data” type was set to 1.0, and the weights for the other types were set as 1.0 plus the inverse of their percentages in the sampling dataset. Thus, the influence of the sample imbalance was weakened, and the relative proportional difference between types was maintained. A lower $Loss$ value indicated a better fit for the training set.

The evaluation metrics used in this experiment were Pixel Accuracy (PA) and Mean Intersection over Union (MIoU). PA and MIoU are common metrics for semantic segmentation, and their computations are based on confusion matrices. Assuming k types, we define p_{ii} as true positives (TP), p_{ij} as false positives (FP), and p_{ji} as false negatives (FN). The formulae for the PA and MIoU are as follows:

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (2)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (3)$$

3. Results

Terrain factors, including slope, aspect, hillshade, RDLS, surface curvature, profile curvature and plane curvature were first calculated, and correlation analysis was performed on these factors as the reference for choosing input channels. The model was then trained with each terrain factor and their different combinations to determine the best terrain factors for landform classification. Finally, an appropriate factor combination was used to obtain segmentation results for a detailed analysis.

The code in this study was written in PyTorch and run on a server with an NVIDIA GeForce RTX 3090. During training, the Adam optimization method (Kingma and Ba, 2014) was used. The initial learning rate was set to 1e-5, the batch size was set to 2 for multichannel tasks and to 4 for single-channel tasks, and the training was performed through 50 iterations. The initial learning rate of 1e-5 was maintained for the first 20 iterations, and it gradually decayed in the latter 30 iterations until it finally reached 0.

3.1. Terrain factor selection

We first performed a correlation analysis between the DEM and all seven terrain factors (Fig. 4). There was a strong correlation between DEM, slope and RDLS. There was a weak correlation between hillshade and aspect. The three curvatures exhibited strong correlations.

Considering the strong correlations among the three curvatures, only the surface curvature was maintained for further exploration. Compared with hillshade, aspect can only reflect the orientation of the slope and is limited in expressing the form of the Earth’s surface. Therefore, aspect was not considered in the experiment. As slope and RDLS are terrain factors widely used in terrain analysis, we retained them in further

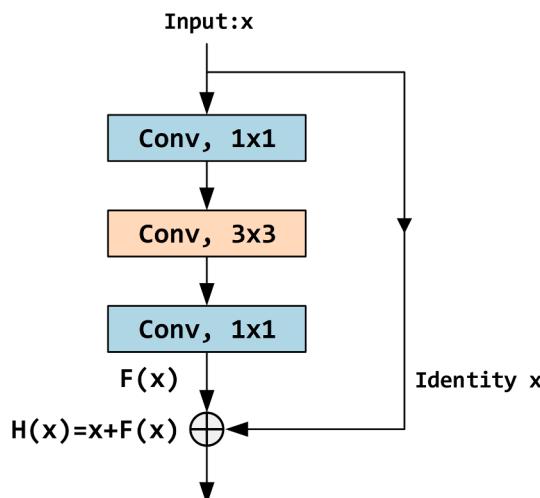


Fig. 2. The 3-layer “bottleneck” building block for deep ResNets (He et al., 2016).

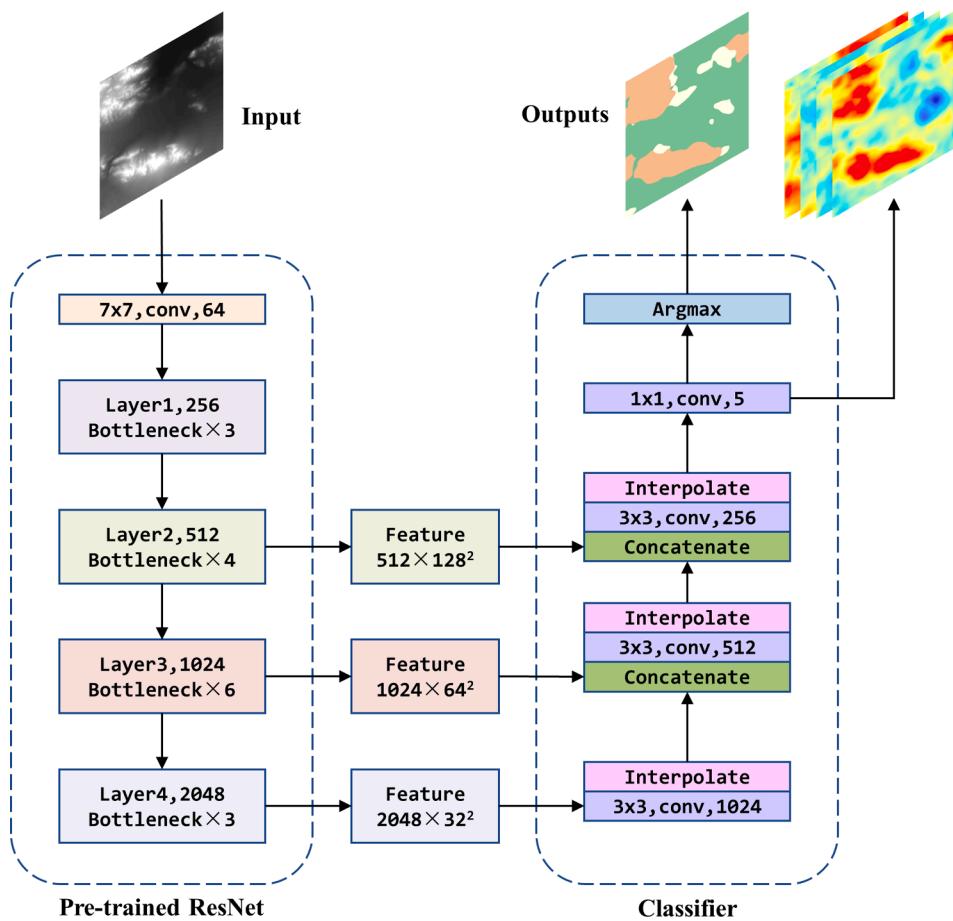


Fig. 3. The single-channel FCN-ResNet deep learning model structure. On the left is a ResNet-50 model pretrained on the ImageNet dataset, and on the right is a classifier whose design is based on the FCN.

analysis. Finally, five terrain factors, namely, DEM, slope, RDLS, hill-shade, and surface curvature, were selected for landform classification using the FCN-ResNet model.

3.2. Model training results

DEM data, terrain factors, and their combinations were used to train the FCN-ResNet model, and the results were compared.

Fig. 5 shows the training procedure using only DEM as the input, and the three solid lines (task A) represent the results of using ResNet pretrained on the ImageNet dataset. It can be seen that with the continuous decrease in *Loss*, the values of PA and MIoU continue to increase, but they tend to stabilize after a certain number of iterations. PA is stable at approximately 80%, and MIoU is stable at approximately 68%. In the first 10 iterations, the PA and MIoU were close to the highest level, and the *Loss* continued to decrease after 10 iterations. The three dashed lines (task B) in Fig. 5 present the results of ResNet with no pretraining. The *Loss* values decreased very slowly, whereas the PA and MIoU values increased very slowly. After 50 iterations, PA and MIoU were still very low and could not reach the values of PA and MIoU of the pretrained ResNet at the 10th iteration. This indicates that the use of pretrained weights is helpful for improving the convergence rate and model performance. Therefore, ResNet pretrained on the ImageNet dataset was used in subsequent experiments.

The results for the different factors are listed in Table 2. The last two columns, “PA” and “MIoU”, show the performance of the model on the validation set. Except for the curvature, the training results of the DEM and other terrain factors are not markedly different, all the *Loss* values are below 0.05, and all the PA values are close to 80%. The surface

curvature performed the worst regardless of which matrix was used for the judgment. For the multifactor training, we can see in Table 2 that they all performed slightly better than the single-factor task on the validation set: both PA and MIoU have less than a 2% increase compared to the results of single-factor tasks, except for curvature.

Fig. 6 shows the confusion matrices of several results for the validation set. The confusion matrices for the different factors are very similar. The classification of plains and mountains was basically accurate. The accuracy of plains was above 83% and that of mountains was above 90%, which are the two categories with the highest proportion in the sample. The results for the hill type were fair, with an accuracy rate of higher than 68%. However, the results for the platform type were relatively poor, with an accuracy rate of 50%. The confusion rate with the plain type was more than 25%, and that with the hill type was approximately 18%. The reason for this phenomenon is that on the one hand, the proportion of platform types in the entire sampling area is relatively low, and on the other hand, the features of platform types are not distinguishable from the other three types. For example, river terraces and premountain flood terraces can easily be confused with plains.

Owing to the large number of parameters in the ResNet network, each additional input channel will drastically increase the parameters. The number of parameters for the single-channel model is 54.18 M, while the total number of parameters for the 3-channel model is 332.42 M. Increasing the number of parameters considerably reduces the training speed substantially and intensively increases the computational overhead. However, the performance did not markedly improve. We believe that DEM alone can fulfill the task of terrain classification. Therefore, we chose a model trained with DEM data for terrain classification.

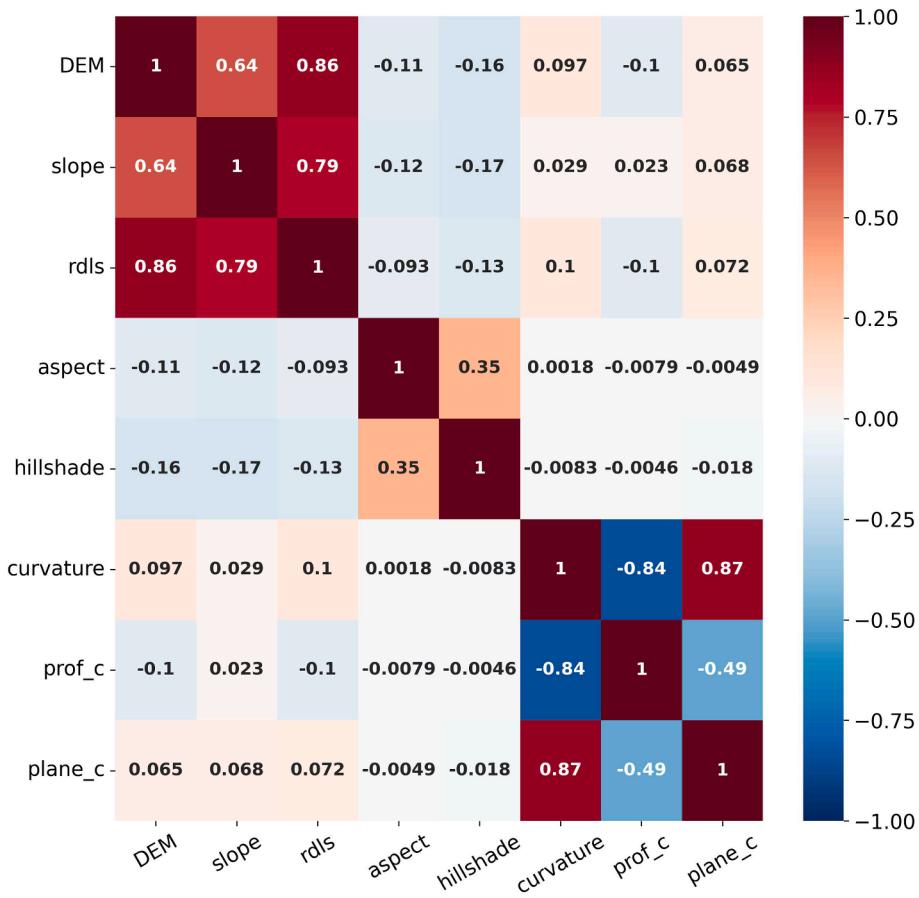
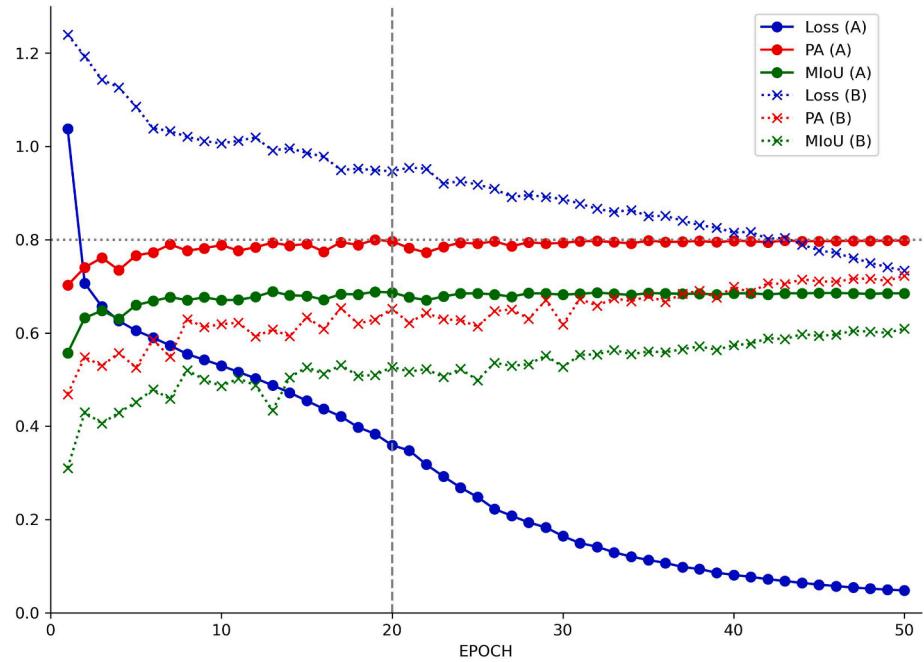
**Fig. 4.** Correlations between different terrain factors.**Fig. 5.** The training procedure of DEM data with the single-channel FCN-ResNet model. The solid lines (A) show the results of the ResNet network pretrained with the ImageNet dataset, and the dashed lines (B) show the results of the ResNet network without pretraining. Both PA and MIoU in the figure are the results on the validation set.

Table 2
Results for different terrain factors.

Terrain factors	Loss	PA	MIoU
DEM	0.0482	79.80	68.49
Slope	0.0374	79.23	67.53
Hillshade	0.0337	79.52	68.00
RDLS	0.0388	79.96	68.64
Curvature	0.0756	75.83	63.69
DEM + Slope + Hillshade	0.0244	80.91	69.76
DEM + Hillshade + Curvature	0.0281	80.52	69.24
DEM + RDLS + Hillshade	0.0257	81.18	70.14

3.3. Classification results

The FCN-ResNet model trained with DEM was used to predict the four elementary landforms in the test areas. Six test areas were chosen in six different regions of China (Fig. 1). The PA value was 75.24 and the MIoU was 63.15, which were slightly lower than the results on the validation sets. Fig. 7 is the confusion matrix of the result, which shows a similar distribution to Fig. 6, in which plains and mountains are well classified and hills and platforms are more confusing.

To examine the model performance in different regions, we compared the results in different test areas. Fig. 8 shows the confusion matrices of the models in six different test regions. The model performed differently on different landform classes in different regions.

Test area 1 is located in the central part of the Tibetan Plateau, and the landforms in the region are mainly formed by glacial action. In the classification results, the accuracy rate of plains and mountains is

acceptable (PA 69.69, MIoU 54.44), but the platforms produce great confusion with plains, and hills are also confused with mountains. The main reason for this is related to the topographical characteristics of the high-altitude areas of the Tibetan Plateau, and the morphological characteristics of the glacially acted platforms are not obvious. The relevant reasons will be elaborated in the detailed analysis of area A.

Test area 2 is on the Yunnan-Guizhou Plateau, where more than 70% of the landforms are mountains, with a small proportion of plains and platforms. The mountains were well classified with an accuracy of 96.95%, but the accuracies for other types were low, resulting in high PA (82.03) and low MIoU (49.52) in this area.

Test areas 3, 4 and 5 are all located in eastern China, and the model performance is acceptable overall. Test area 3 is located in eastern China along the Dabie Mountains. The distribution of the four types in this region is relatively balanced, and therefore, the performance was relatively good (PA 73.26, MIoU 66.31). Test area 4 is located in the Shandong Hills, where platforms only occupy 2.0% of the area. Therefore, the model performed well except for platforms (PA 72.12, MIoU 56.03). Test area 5 is located in the southern part of the Northeastern Plain, and the model performance is quite good for all the four landform types, resulting in high PA (81.95) and high MIoU (71.81).

Test area 6 is located in the Taklamakan Desert in northwestern China. The landforms in this region are mainly aeolian sand dunes that belong to the hill type, mountains, and a small number of alluvial plains. The model performed well on plains and mountains but did not perform well on platforms and hills. The accuracy of platforms was poor because platforms only take 4% of the area. The model classified many hills as plains. Considering that the hills in this region are mostly sand dunes

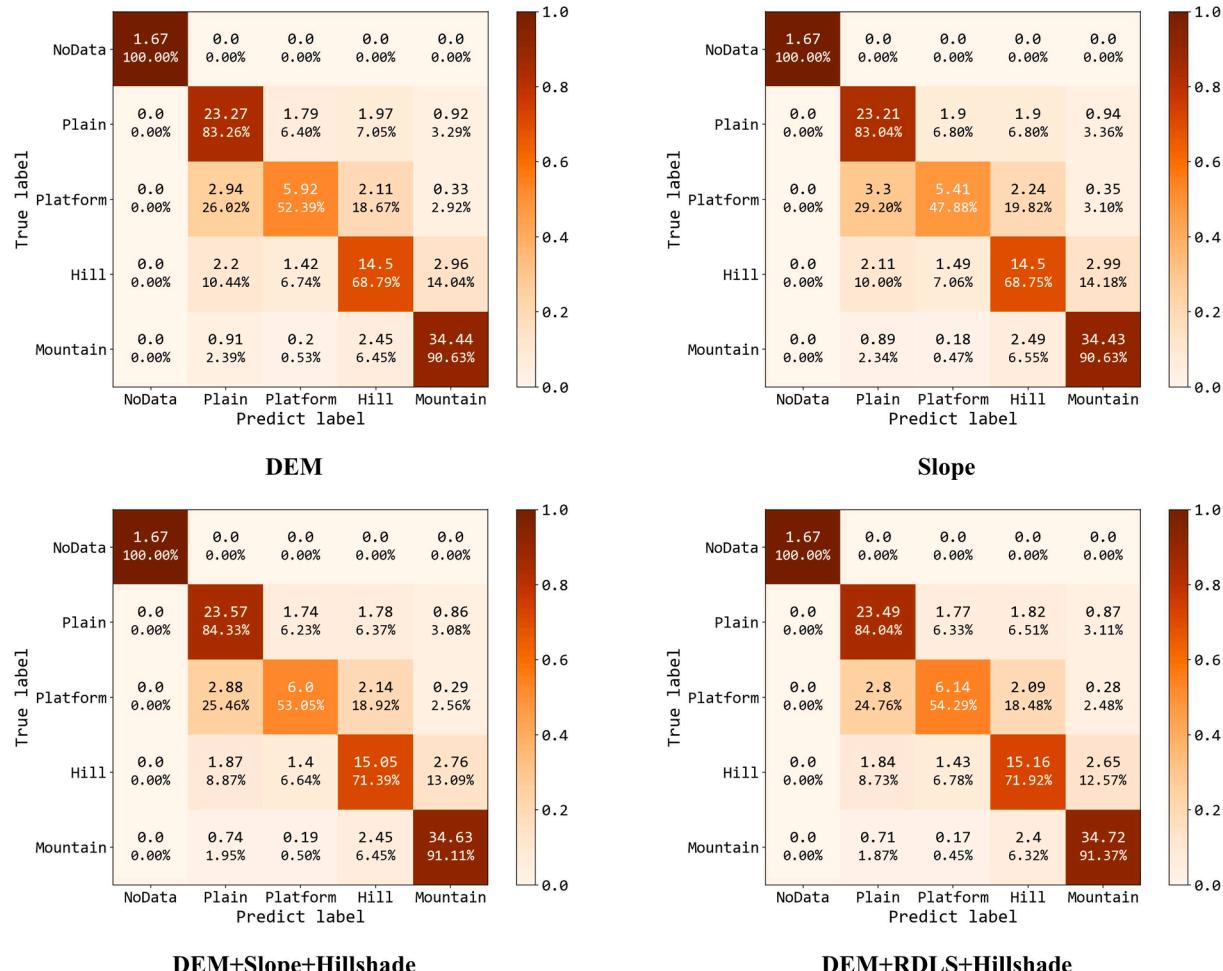


Fig. 6. Confusion matrices on the validation set for models trained with different terrain factors.

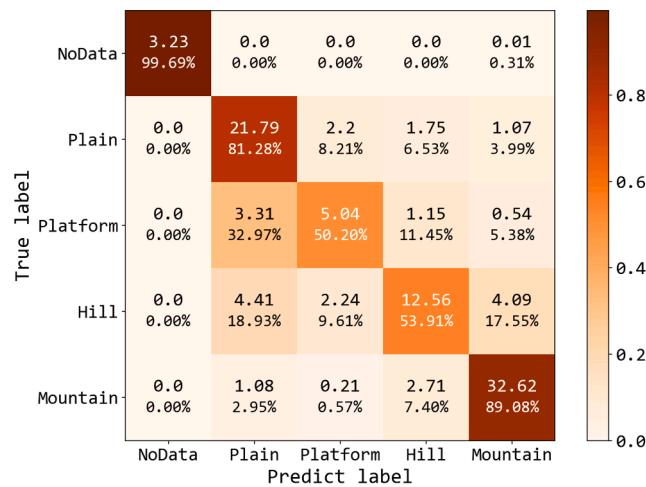


Fig. 7. Confusion matrix of the model on the test set.

spread in the desert that are small and unstable, it is not easy to identify them correctly. Since platforms only take 4% of the area, the overall performance is acceptable (PA 75.24, MIoU 63.15).

Three small areas in different regions (A, B, and C in Fig. 1) were chosen to show the detailed results. As shown in Fig. 3, before the final classification, the model obtains a predicted value for each category, which represents the confidence in assigning a certain pixel to this category, and the image is labeled with the category of the value with the highest probability. In addition to the final predicted result, we acquired the predicted values for each category for analysis. Figs. 9–11 show the maps of the true and predicted categories, as well as the predicted values of different categories displayed as heatmaps.

Area A in Fig. 9 is located in the middle of the Qinghai-Tibet Plateau at an altitude of 4460–5081 m (Fig. 9a), where the main forces are fluviation and periglacial processes. The predicted categories (Fig. 9c) are consistent with the labeled categories (Fig. 9b), except that some platforms at the left boundary were not well identified and some small areas at the edge of the mountains were classified as platforms. Additionally, the platform in the middle was disconnected. As shown by the DEM (Fig. 9a), the small areas at the edge of the mountains could be piedmont terraces, and it was reasonable to classify them as platforms. We can also observe from the DEM that the platform in the middle is not disconnected. As we labeled the data with the categories in the 1:1,000,000 landform map that was delineated by humans, people can generalize when they identify the classes. Thus, the results classified by our model were more accurate. There is also a small area at the end of the mountain that was identified as a hill. From the DEM, we found that this low-altitude mountain is discrete from the main body of the mountain. It morphologically looks like a hill but belongs to a mountain. In this case, the manually classified landform map was correct. The error on the left boundary was due to insufficient information caused by image cropping at the boundary. This can be solved using the image mosaic. Fig. 9d–g show the heatmaps of the predicted values for plains, platforms, hills and mountains. The predicted value is high in the middle of each landform class and low near the boundary, which means that the certainty of the pixels near the boundary belonging to the class is low. However, most pixels on the boundary were well classified. For pixels with predicted categories inconsistent with the labeled categories, the predicted values are low for all categories, indicating that the model is uncertain about the results.

Area B in Fig. 10 is located in the western part of the Dabie Mountains, Henan Province, at an altitude of 101–1018 m (Fig. 10a). It mainly contains alluvial plains, platforms, and eroded and denuded hilly mountains. The difference between the labeled and predicted results in the middle of this area is obvious (Fig. 10b–c). Although the category

was labeled as a continuous platform, the model predicted it as a plain interlaced with platforms. The high predicted values in Fig. 10d and 10e indicate that the model did not confuse the results. The DEM (Fig. 10a) also shows the rolling terrain of this area. The predicted results for the hills and mountains were accurate, but the predicted hilly area on the right was slightly larger than the labeled area. The heatmap of the hill (Fig. 10f) shows that the predicted value of the extra hill area is high, which means that the model is certain of the result. From the DEM (Fig. 10a), it can be inferred that it could be a hill in front of the mountain. Therefore, the predicted results of the model in area B were reasonable.

Area C in Fig. 11 is located in the southern part of the Northeast China Plain at an altitude of –10–320 m (Fig. 11). The area mainly contains a large plain and some low mountains and hills. A large city, Shenyang, is located on the left corner of this area and causes noise in the DEM data. Fig. 11c and Fig. 11d show that the plain is well identified with high confidence, and noise does not have a significant influence on the result. The hill was also well identified with relatively high confidence, except for a small peak that was identified as a mountain but with low confidence, indicating that the model can differentiate between mountain and hill landforms. The platform in the middle area has a slightly different contour than the labeled one, as the outlines delineated by humans are more generalized, and the model-predicted outlines are more accurate.

Summarizing the results, the model performed much better on plains and mountains than on platforms and hills. Platforms are mostly mixed with plains since they all have flat surfaces, and hills are mixed with both plains and mountains since they all have rolling surfaces. However, the model performance varied in space. The overall accuracy of the results is higher when the four landform types are evenly distributed. When one kind of landform dominates in a region, then the accuracy for this kind of landform is extremely high, while the overall accuracy is not very high. The deep learning model proposed in this paper classified elementary landform types when only information was obtained from DEM data, and the results are consistent with human perception regarding land surface morphology. The detailed presentation of the result in the three typical areas indicated that the model performs well in many easily confused regions and in some regions containing noise. To demonstrate the detailed results, this paper only displays the results of small test areas on a large scale. Since the model has learned sufficient features to classify these elementary landform types, it is no problem to classify repeating landform types in a broad region. These results demonstrate the great potential of deep-learning methods in landform mapping.

4. Discussion

The experiment proved that the FCN-ResNet model can be applied to DEM data to segment terrain classes. However, some issues remain to be addressed, such as the effectiveness of the proposed model, the accuracy of the results, the scale, and the effects of label precision and pretraining on the results.

To test the effectiveness of the proposed model, we compared the performance of the models in this study with other common semantic segmentation models, including FCN-8s, U_Net (Ronneberger et al., 2015), U_Net-attention (Oktay et al., 2018), and DeepLabv3 (Chen et al., 2017). The training results for the different models are listed in Table 3. The ResNet-101 model in the last two rows of the table was also pre-trained on the ImageNet dataset. Compared with the FCN-ResNet model used in this study, the U_Net and FCN-8s models are simpler and have fewer parameters. Therefore, they can be easily trained. However, because of the lack of pretraining, these lightweight models have a large deficiency in fitting ability, which is indicated by a higher Loss and slightly lower MIoU and PA. With an increasing number of parameters in the ResNet-50 and ResNet-101 models, the Loss value dropped significantly. However, MIoU and PA increased marginally.

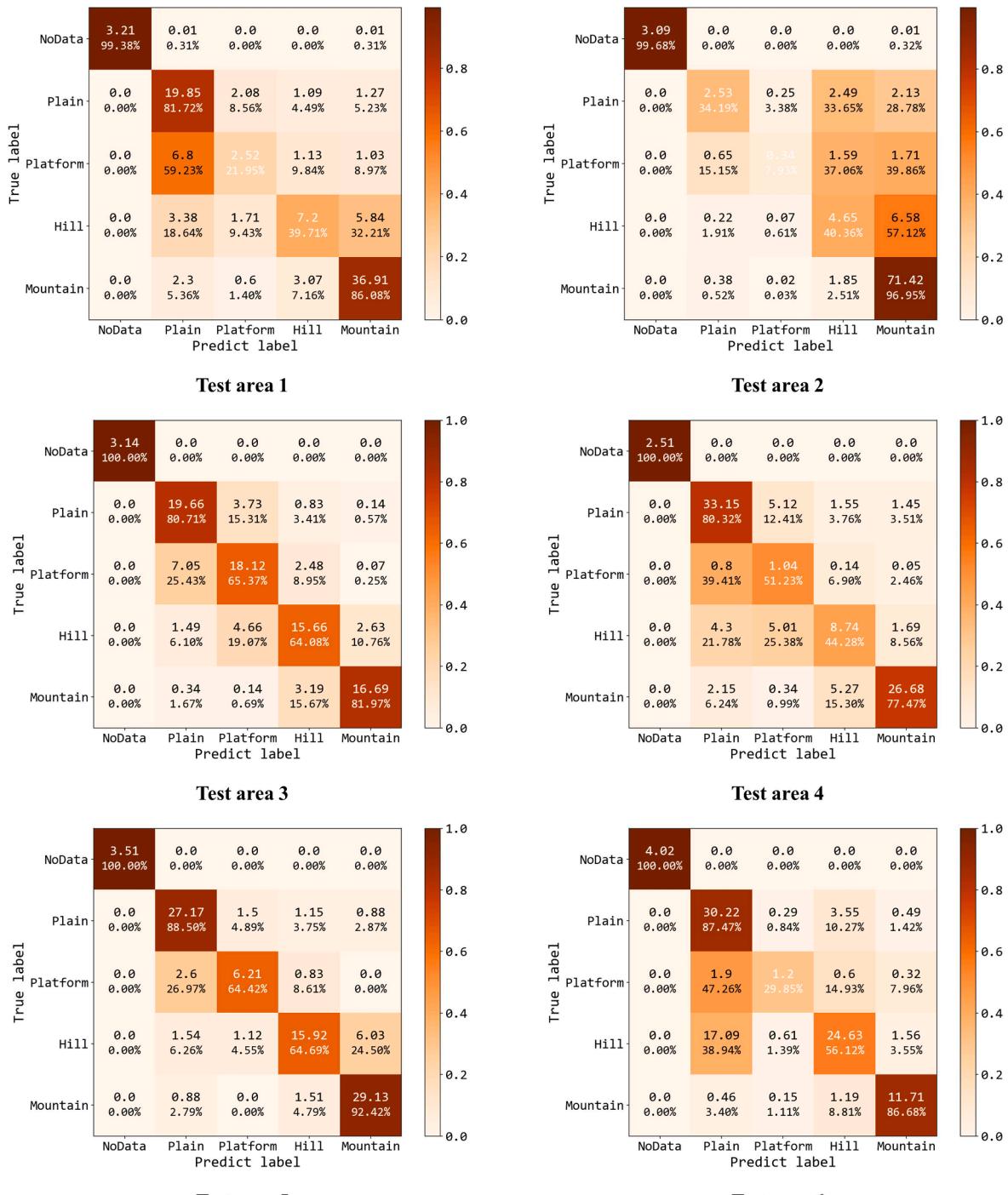


Fig. 8. Confusion matrices of the model on different test areas.

The advantage of pretraining is worth mentioning. ImageNet is a dataset of natural images in RGB format, which is different from the data used in our experiments and has a different density of semantic information. However, the results show that the ImageNet-based pretraining results can be well applied to this experiment, as shown in Fig. 5. Marmanis et al. (2015) transferred the results learned from ImageNet tasks to remote sensing tasks and explained that remote sensing tasks benefit from ImageNet because of the structure of the convolutional neural network itself. The network can easily decompose images after pretraining and detect similarities at different abstraction levels, and the pretrained weights can be easily adjusted. Our experiments also proved that pretraining on ImageNet can improve the performance of deep

learning models on DEM data and that the landform classification task relies on the spatial feature extraction capability of convolutional networks.

In traditional semantic segmentation tasks, such as segmentation on the VOC2012 (Everingham et al., 2010) and Cityscapes (Cordts et al., 2016) datasets, the value of MIoU changes significantly for different models. As shown by Minaee et al. (2021), the MIoU value on the Cityscapes dataset was 81.3 when using DeepLabv3_ResNet101, and only 65.3 when using FCN-8s. In our experiments, the simple U-Net model was able to achieve an MIoU of 67.01, but the highest MIoU that could be achieved by the best-performing model was only 68.94 in Table 3, which was not a significant improvement. This is because in

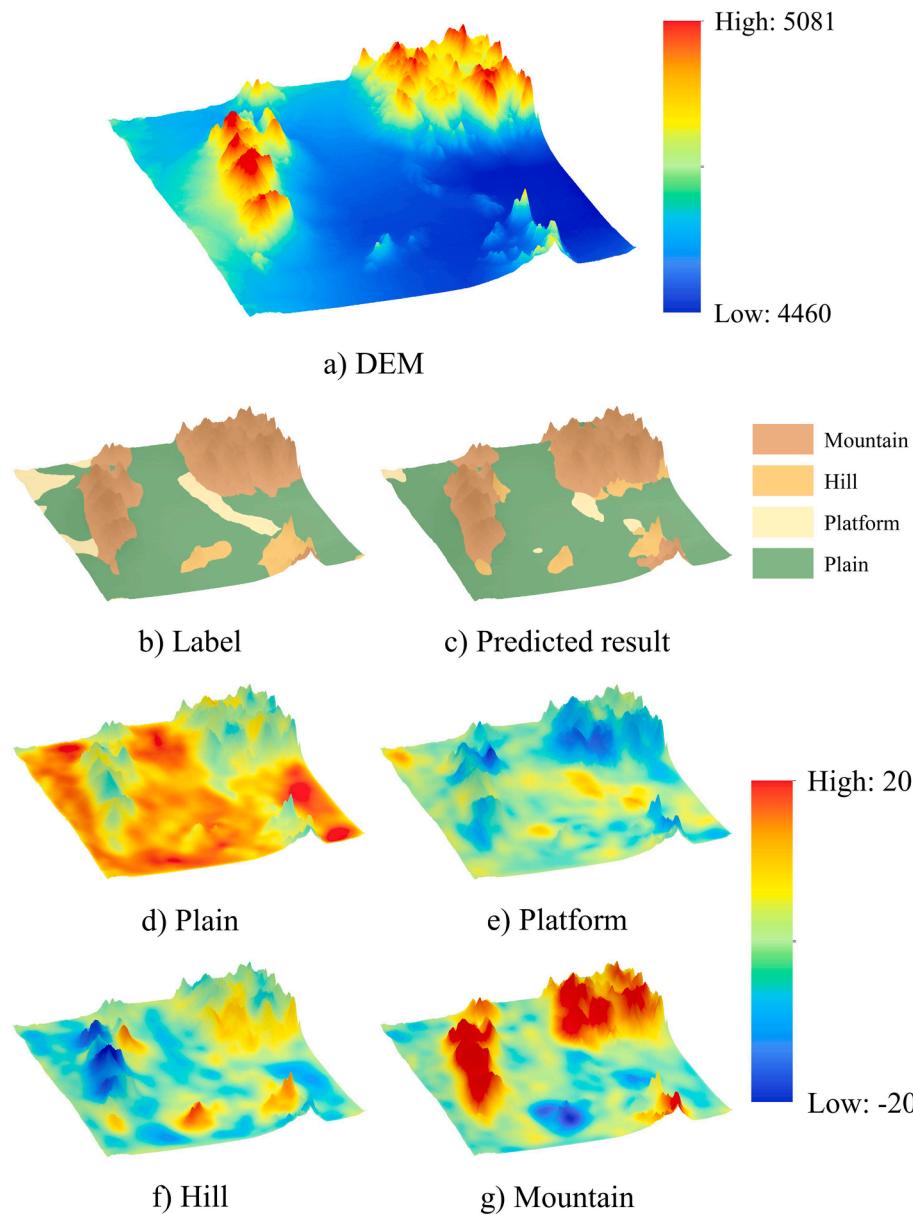


Fig. 9. Prediction result of area A in Fig. 1.

datasets such as VOC2012 and Cityscapes, the objects have clear boundaries that can reduce the error to a very low level on the boundary. However, unlike natural images and entity types, there is usually a transition region between different landform types in nature, where the boundaries of landform types are inherently difficult to determine. Therefore, it is difficult to improve the overall accuracy of landform classification tasks using more complex networks.

The scale difference is another issue that should be discussed. We labeled the types and evaluated the results using the first-level landform classes in the 1:1,000,000 landform map of China, which has a different resolution than the 30 m DEM data. This may have caused two problems. One is the label precision. Labeling the types with the 1:1,000,000 landform map may result in several pixel mismatches on the boundaries, which will reduce the accuracy of training. To demonstrate the effect of label error on the boundary, we tested three labeling strategies on the boundaries. The first one directly used the boundaries of a 1:1,000,000 landform map; the second created a 5-pixel width buffer on boundaries, discarded the pixels in the buffer areas, and used only accurately labeled pixels in the model training processes; and the third randomly assigned

labels in buffer areas to blur the boundaries and incorporated all inaccurately labeled pixels in model training (Fig. 12).

Table 4 indicates that the results of the three strategies are not significantly different between the validation and testing sets. It can be concluded that labeling precision does not have a significant influence on the prediction of the model. As deep learning methods are based on large amounts of training data, they have strong error tolerance. The accurate labeling of boundary pixels does not have a significant influence on the predicted results as long as other pixels are correctly labeled. Therefore, it is feasible to label the data with the 1:1,000,000 landform map, although the 30 m resolution DEM data can make larger scale maps.

Another problem caused by scale difference arose in the result evaluation. The landform classes in the 1:1,000,000 landform map of China were identified by humans. Because of the vagueness of landform boundaries, there is an inevitable displacement of the boundaries delineated by humans. It is not appropriate to evaluate the result with the landform map, and there is no absolute criterion to evaluate the result. As indicated in Section 3.3, in some places where the results are

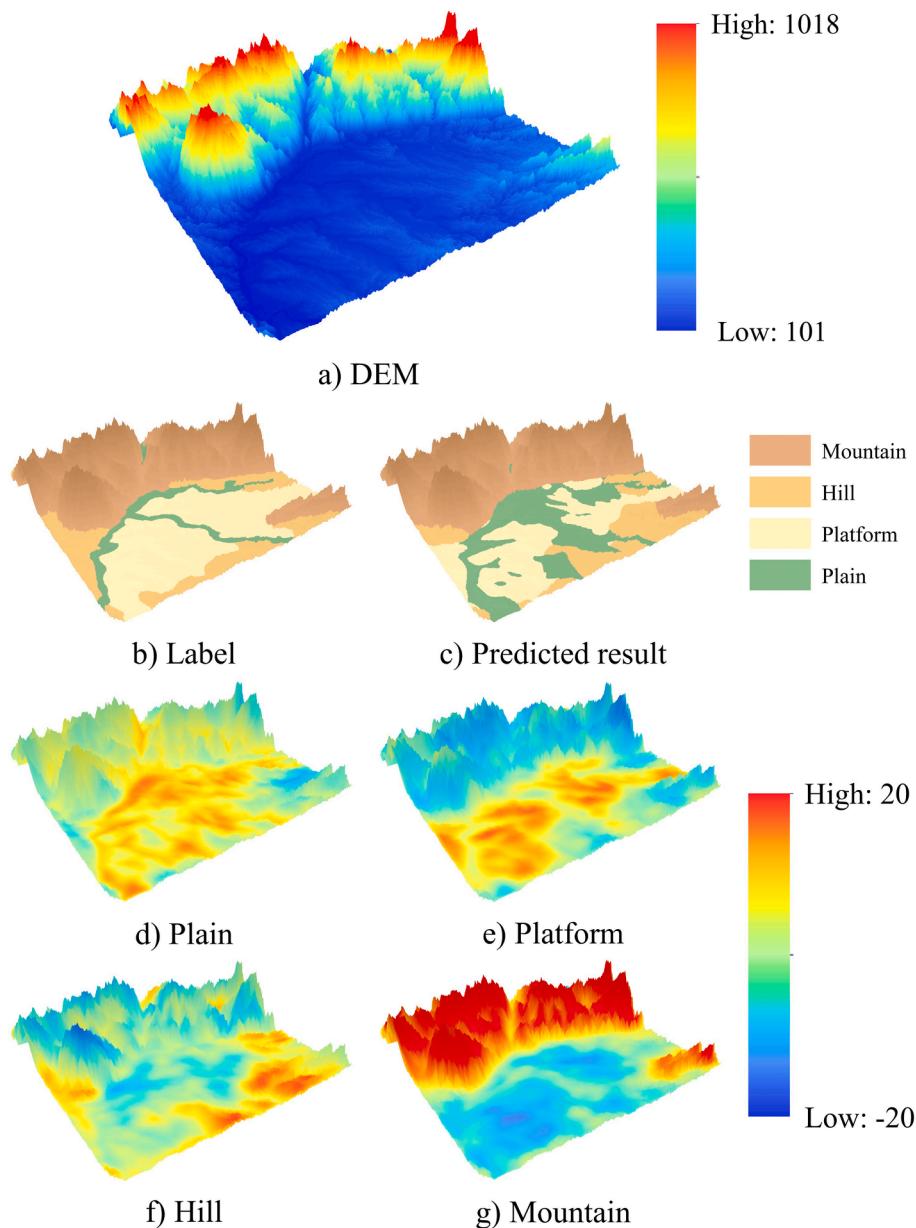


Fig. 10. Prediction results of area B in Fig. 1.

different from the landform map, the model prediction is more accurate than the landform map, and the model is very certain about the results. In some places where we are unsure about the landform type, the model also has low confidence, indicated by low prediction values. Therefore, we believe that the model prediction result is plausible, and the actual accuracy of the result should be higher than that disclosed by the values of PA and MIoU.

It is worth mentioning that this study only used the first-level landform types in China's 1:1,000,000 landform map, which divided landforms according to morphology. More subtypes based on genesis can be obtained from the landform map. DEM data, a morphological representation of the Earth's surface, can recognize morphological landform types well. However, whether they can distinguish genetic landform types is still a question. Du et al. (2019) successfully identified different types of genetic landform elements by using multichannel data, including DEM, slope, flow accumulation, curvature, topographic wetness index, and relief. Their work indicated that with auxiliary data, it is possible to use DEM data to classify genetic landform types.

5. Conclusions

We proposed an FCN-ResNet semantic segmentation model and applied it to elementary landform classification using AW3D30 DEM data. Based on correlation analysis, DEM, slope, RDLS, hillshade and surface curvature were selected as inputs in single-channel tasks, and combinations of factors were used as inputs in multichannel tasks to train the model. The data were labeled and the results were evaluated with first-level landform classes with a 1:1,000,000 landform map. To investigate the effect of labeling precision caused by scale differences on the prediction results, we tested three labeling strategies. The experiments demonstrated the effectiveness of using a semantic segmentation model for landform classification based on DEM data. The conclusions are as follows:

- i. A semantic segmentation model based on a deep learning segmentation model can classify repeating landform types and thus can be used in landform mapping. The prediction results of the FCN-ResNet model indicate that deep learning segmentation

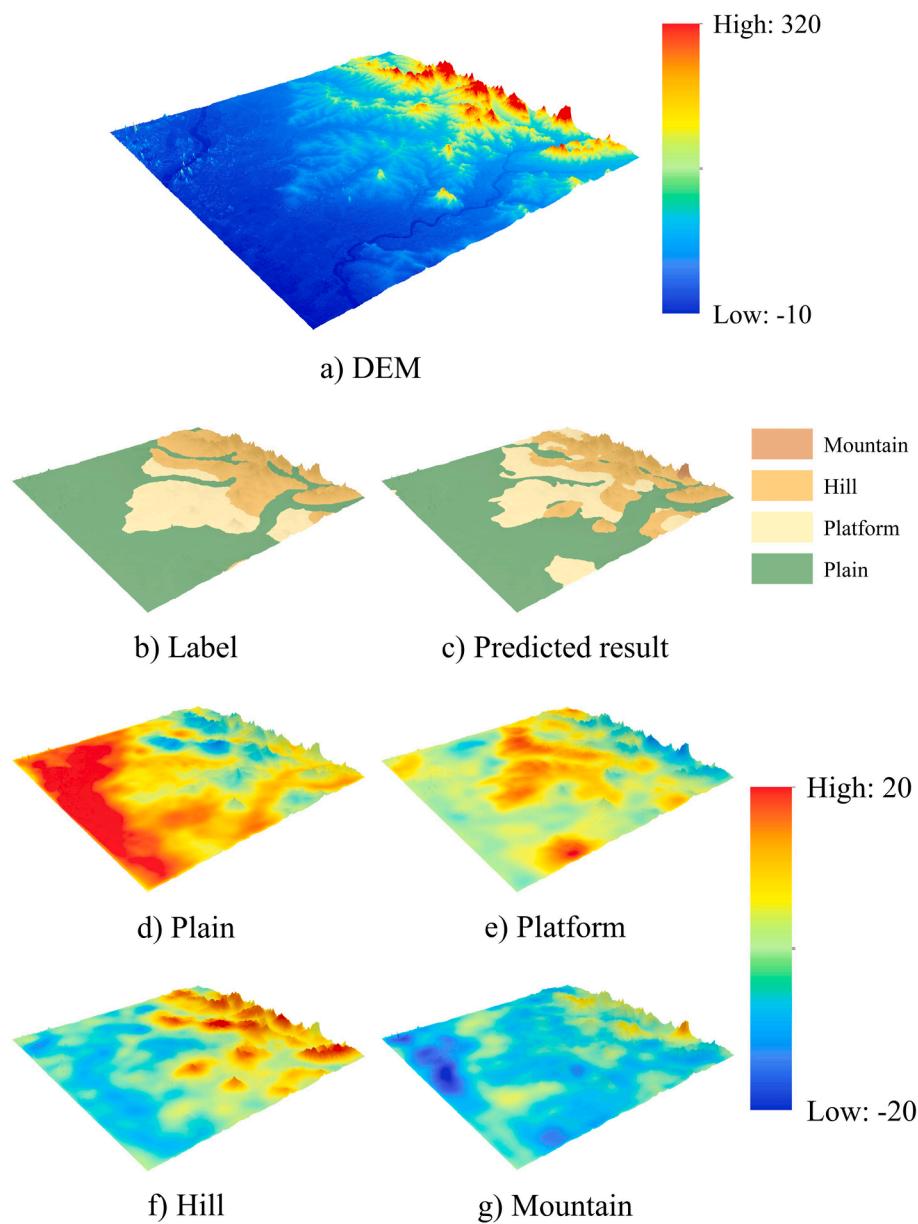


Fig. 11. Prediction results of area C in Fig. 1.

Table 3
Results of different models*.

Method	Parameters	Loss	PA	MIoU
U_Net	34.53 M	0.5494	77.91	67.01
U_Net-attention	34.88 M	0.5244	78.48	67.64
FCN-8s	15.12 M	0.5437	78.14	67.25
FCN_ResNet50	54.18 M	0.0482	79.80	68.49
DeepLabv3_ResNet101	60.99 M	0.0669	79.94	68.65
FCN_ResNet101	73.18 M	0.0452	80.24	68.94

*All models were trained using DEM data only.

methods have great potential for classifying elementary landform types, even in some easily confused regions, and high tolerance to noise.

ii. The terrain factors only play the role of data augmentation and do not introduce additional information into model training. As the terrain factors are all derived from the DEM, they are different forms of the same semantic information. In contrast to other

traditional rule-based landform classification methods, deep learning methods have a strong feature extraction ability and can extract enough features using only DEM data.

iii. Deep learning methods have a strong error tolerance. Therefore, labeling precision has no significant influence on the accuracy of the prediction result.

This study illustrated the applicability of classifying morphological landform types using semantic segmentation models based on deep learning. Future work could focus on examining the feasibility of deep learning models in classifying genetic landform types, probably with auxiliary data, such as remote sensing images and temperature and precipitation data.

CRediT authorship contribution statement

Jiaqi Yang: Methodology, Software, Writing – original draft. **Jun Xu:** Supervision, Writing – review & editing. **Yunshuo Lv:** Data curation, Formal analysis. **Chenghu Zhou:** Conceptualization. **Yunqiang**

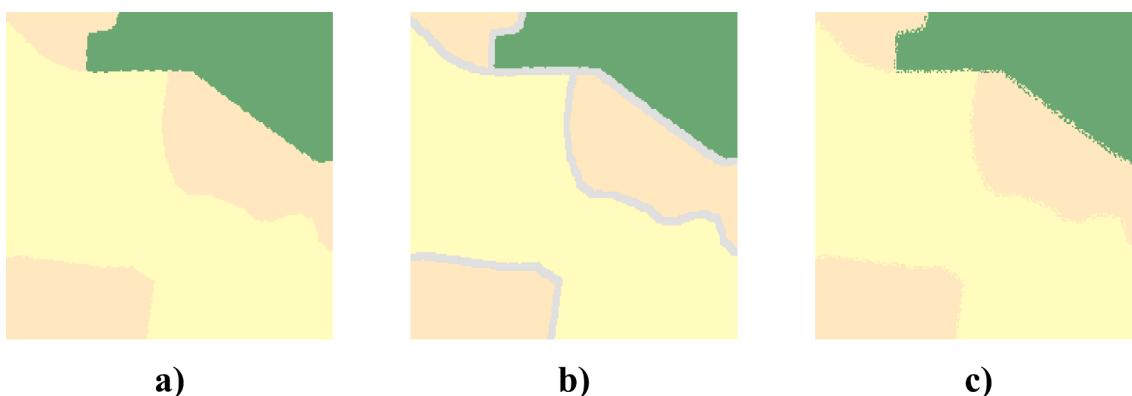


Fig. 12. Three labeling strategies: a) original boundary from China 1:1,000,000 landform map; b) discarding samples in buffer area of boundary; and c) blurring the boundary by randomly assigning labels in buffer area.

Table 4

The results of different labeling strategies*. (V) denotes the results on the validation set and (T) denotes the results on the testing set.

Labeling strategy	Loss	PA (V)	MIoU (V)	PA (T)	MIoU (T)
Label with origin boundary	0.0482	79.80	68.49	75.24	63.15
Remove pixels in boundary buffers	0.0231	79.68	68.33	75.25	63.27
Randomly assign labels in boundary buffers	0.0521	79.77	68.37	75.44	63.52

* All were trained using only DEM data.

Zhu: Resources. **Weiming Cheng:** Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the national key R&D program [Grant Number: 2022YFB3904204]; the NSFC Program [grant numbers 42,071,376 and 42130110]; the Deep-time Digital Earth (DDE) Big Science Program; and the Key Project of Innovation LREIS [Grant Number: KPI009].

References

- Aghdam-Nia, M., Shah-Hosseini, R., Rostami, A., Homayouni, S., 2022. Automatic coastline extraction through enhanced sea-land segmentation by modifying Standard U-Net. *Int. J. Appl. Earth Obs. Geoinf.* 109, 102785. <https://doi.org/10.1016/j.jag.2022.102785>.
- Blaschke, T., 2002. A multiscalar GIS / image processing approach for landscape monitoring of mountainous areas. In: Bottarin, R., Trappeiner, U. (Eds.), *Interdisciplinary Mountain Research*. Blackwell Science, Southampton, pp. 12–25.
- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv preprint arXiv:1412.7062*. <https://doi.org/10.48550/arXiv.1412.7062>.
- Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*. <https://doi.org/10.48550/arXiv.1706.05587>.
- Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818.
- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018b. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>.
- Cheng, X., Doosthosseini, A., Kunkel, J., 2022. Improve the deep learning models in forestry based on explanations and expertise. *Front. Plant Sci.* 13, 902105. <https://doi.org/10.3389/fpls.2022.902105>.
- Cheng, W., Zhou, C., Li, B., Shen, Y., Zhang, B., 2011. Structure and contents of layered classification system of digital geomorphology for China. *J. Geogr. Sci.* 21, 771–790. <https://doi.org/10.1007/s11442-011-0879-9>.
- Chong, E.K., Zak, S.H., 2013. *An Introduction to Optimization*, fourth ed. John Wiley & Sons.
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3213–3223. <https://doi.org/10.1109/CVPR.2016.350>.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.
- Drăguț, L., Blaschke, T., 2006. Automated classification of landform elements using object-based image analysis. *Geomorphology* 81, 330–344. <https://doi.org/10.1016/j.geomorph.2006.04.013>.
- Drăguț, L., Eisank, C., 2012. Automated object-based classification of topography from SRTM data. *Geomorphology* 141–142, 21–33. <https://doi.org/10.1016/j.geomorph.2011.12.001>.
- Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. *ISPRS J. Photogramm. Remote Sens.* 158, 63–75. <https://doi.org/10.1016/j.isprsjprs.2019.09.018>.
- Evans, I.S., 2012. Geomorphometry and landform mapping: What is a landform? *Geomorphology* 137, 94–106. <https://doi.org/10.1016/j.geomorph.2010.09.029>.
- Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* 88, 303–338. <https://doi.org/10.1007/s11263-009-0275-4>.
- Giano, S.I., Danese, M., Gioia, D., Pescatore, E., Siervo, V., Bentivenga, M., 2020. Tools for semi-automated landform classification: A comparison in the basilicata region (Southern Italy), in: Gervasi, O., Murgante, B., Misra, S., Garau, C., Bleićić, I., Taniar, D., Apduhan, B.O., Rocha, A.M.A.C., Tarantino, E., Torre, C.M., Karaca, Y. (Eds.), *Computational Science and Its Applications – ICCSA 2020*. Springer International Publishing, Cham, pp. 709–722. https://doi.org/10.1007/978-3-030-58802-1_51.
- Goodchild, M.F., 2005. GIS, spatial analysis, and modelling overview. In: Maguire, D.J., Batty, M., Goodchild, M.F. (Eds.), *GIS*. ESRI Press, Spatial Analysis and Modeling, pp. 1–18.
- Graff, L., Usery, E.L., 1993. Automated classification of terrain features in digital elevation models. *Photogramm. Eng. Remote Sens.* 59, 1409–1417.
- Guo, D., Ge, S., Zhang, S., Gao, S., Tao, R., Wang, Y., 2022. DeepSSN: A deep convolutional neural network to assess spatial scene similarity. *Trans. GIS.* 26, 1914–1938. <https://doi.org/10.1111/tgis.12915>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Iwahashi, J., Kamiya, I., Matsuoka, M., Yamazaki, D., 2018. Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification. *Prog. Earth Planet. Sci.* 5, 1–31. <https://doi.org/10.1186/s40645-017-0157-2>.
- Iwahashi, J., Pike, R.J., 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* 86, 409–440. <https://doi.org/10.1016/j.geomorph.2006.09.012>.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://arxiv.org/abs/1412.6980>.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. <https://doi.org/10.1145/3065386>.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. <https://doi.org/10.1109/5.726791>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539>.
- Li, W., Hsu, C.-Y., 2020. Automated terrain feature identification from remote sensing imagery: a deep learning approach. *Int. J. Geogr. Inf. Sci.* 34, 637–660. <https://doi.org/10.1080/13658816.2018.1542697>.

- Li, S., Xiong, L., Tang, G., Strobl, J., 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology* 354, 107045. <https://doi.org/10.1016/j.geomorph.2020.107045>.
- Li, W., Zhou, B., Hsu, C.-Y., Li, Y., Ren, F., 2017. Recognizing terrain features on terrestrial surface using a deep learning model – An example with crater detection. First ACM SIGSPATIAL Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery, Los Angeles, CA, USA. <https://doi.org/10.1145/3149808.3149814>.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>.
- MacMillan, R.A., Jones, R.K., McNabb, D.H., 2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). *Comput. Environ. Urban Syst.* 28, 175–200. [https://doi.org/10.1016/S0198-9715\(03\)00019-X](https://doi.org/10.1016/S0198-9715(03)00019-X).
- Malik, K., Robertson, C., Braun, D., Greig, C., 2021. U-Net convolutional neural network models for detecting and quantifying placer mining disturbances at watershed scales. *Int. J. Appl. Earth Obs. Geoinf.* 104, 102510 <https://doi.org/10.1016/j.jag.2021.102510>.
- Marmanis, D., Datcu, M., Esch, T., Still, U., 2015. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geosci. Remote Sens. Lett.* 13, 105–109. <https://doi.org/10.1109/LGRS.2015.2499239>.
- Miliaresis, G.C., Argialas, D.P., 1999. Segmentation of physiographic features from the global digital elevation model/GTOPO30. *Comput. Geosci.* 25 [https://doi.org/10.1016/S0098-3004\(99\)00025-4](https://doi.org/10.1016/S0098-3004(99)00025-4).
- Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* <https://doi.org/10.1109/TPAMI.2021.3059968>.
- Minár, J., Evans, I.S., 2008. Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping. *Geomorphology* 95, 236–259. <https://doi.org/10.1016/j.geomorph.2007.06.003>.
- Na, J., Ding, H., Zhao, W., Liu, K., Tang, G., Pfeifer, N., 2021. Object-based large-scale terrain classification combined with segmentation optimization and terrain features: A case study in China. *Trans. GIS.* 25, 2939–2962. <https://doi.org/10.1111/tgis.12795>.
- Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. <https://doi.org/10.48550/arXiv.1804.03999>.
- Prima, O.D.A., Echigo, A., Yokoyama, R., Yoshida, T., 2006. Supervised landform classification of northeast honshu from DEM-derived thematic maps. *Geomorphology* 78, 373–386. <https://doi.org/10.1016/j.geomorph.2006.02.005>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, vol 9351. Springer, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Shumack, S., Hesse, P., Farebrother, W., 2020. Deep learning for dune pattern mapping with the AW3D30 global surface model. *Earth Surf. Process. Landf.* 45, 2417–2431. <https://doi.org/10.1002/esp.4888>.
- Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* 9, 82031–88205. <https://doi.org/10.1109/ACCESS.2021.3086020>.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. <https://doi.org/10.48550/arXiv.1409.1556>.
- Sinha, G., Mark, D.M., 2010. Cognition-based extraction and modelling of topographic eminences. *Cartographica* 45, 105–112. <https://doi.org/10.3138/cart.45.2.105>.
- Theobald, D.M., Harrison-Atlas, D., Monahan, W.B., Albano, C.M., 2015. Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PLoS ONE* 10, e0143619.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. arXiv preprint arXiv: 1706.03762. <https://doi.org/10.48550/arXiv.1706.03762>.
- Wilson, J.P., 2012. Digital terrain modeling. *Geomorphology* 137, 107–121. <https://doi.org/10.1016/j.geomorph.2011.03.012>.
- Xiong, L., Li, S., Tang, G., Strobl, J., 2022. Geomorphometry and terrain analysis: Data, methods, platforms and applications. *Earth-Sci. Rev.* 233, 104191 <https://doi.org/10.1016/j.earscirev.2022.104191>.
- Yan, S., Xu, L., Yu, G., Yang, L., Yun, W., Zhu, D., Ye, S., Yao, X., 2021. Glacier classification from Sentinel-2 imagery using spatial-spectral attention convolutional model. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102445 <https://doi.org/10.1016/j.jag.2021.102445>.
- Yuan, X., Shi, J., Gu, L., 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* 169, 114417 <https://doi.org/10.1016/j.eswa.2020.114417>.
- Zhang, H., Liu, M., Wang, Y., Shang, J., Liu, X., Li, B., Song, A., Li, Q., 2021. Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102557 <https://doi.org/10.1016/j.jag.2021.102557>.
- Zhang, Y., Yu, W., Zhu, D., 2022. Terrain feature-aware deep learning network for digital elevation model superresolution. *ISPRS J. Photogramm. Remote Sens.* 189, 143–162. <https://doi.org/10.1016/j.isprsjprs.2022.04.028>.
- Zhao, S., Liu, J., Cheng, W., Zhou, C., 2022. Fusion Scheme and Implementation Based on SRTM1, ASTER GDEM V3, and AW3D30. *ISPRS J. Photogramm. Remote Sens.* 11, 207. <https://doi.org/10.3390/ijgi11030207>.