



Article

Geocomplexity Statistical Indicator to Enhance Multiclass Semantic Segmentation of Remotely Sensed Data with Less Sampling Bias

Wei He ^{1,2,†} , Lianfa Li ^{1,2,*,†} and Xilin Gao ^{1,2}

¹ State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; hew.20s@igsrr.ac.cn (W.H.); gaoxl@reis.ac.cn (X.G.)

² College of Resources and Environment, University of the Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: lilf@reis.ac.cn; Tel.: +86-10-6488-8362

† These authors contributed equally to this work.

Abstract: Challenges in enhancing the multiclass segmentation of remotely sensed data include expensive and scarce labeled samples, complex geo-surface scenes, and resulting biases. The intricate nature of geographical surfaces, comprising varying elements and features, introduces significant complexity to the task of segmentation. The limited label data used to train segmentation models may exhibit biases due to imbalances or the inadequate representation of certain surface types or features. For applications like land use/cover monitoring, the assumption of evenly distributed simple random sampling may be not satisfied due to spatial stratified heterogeneity, introducing biases that can adversely impact the model's ability to generalize effectively across diverse geographical areas. We introduced two statistical indicators to encode the complexity of geo-features under multiclass scenes and designed a corresponding optimal sampling scheme to select representative samples to reduce sampling bias during machine learning model training, especially that of deep learning models. The results of the complexity scores showed that the entropy-based and gray-based indicators effectively detected the complexity from geo-surface scenes: the entropy-based indicator was sensitive to the boundaries of different classes and the contours of geographical objects, while the Moran's I indicator had a better performance in identifying the spatial structure information of geographical objects in remote sensing images. According to the complexity scores, the optimal sampling methods appropriately adapted the distribution of the training samples to the geo-context and enhanced their representativeness relative to the population. The single-score optimal sampling method achieved the highest improvement in DeepLab-V3 (increasing pixel accuracy by 0.3% and MIoU by 5.5%), and the multi-score optimal sampling method achieved the highest improvement in SegFormer (increasing ACC by 0.2% and MIoU by 2.4%). These findings carry significant implications for quantifying the complexity of geo-surface scenes and hence can enhance the semantic segmentation of high-resolution remote sensing images with less sampling bias.

Keywords: sampling bias; optimal sampling; semantic segmentation; deep learning



Citation: He, W.; Li, L.; Gao, X. Geocomplexity Statistical Indicator to Enhance Multiclass Semantic Segmentation of Remotely Sensed Data with Less Sampling Bias. *Remote Sens.* **2024**, *16*, 1987. <https://doi.org/10.3390/rs16111987>

Academic Editor: Shuying Li

Received: 16 April 2024

Revised: 22 May 2024

Accepted: 23 May 2024

Published: 31 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Deep learning methods have achieved great success in semantic segmentation in traditional computer vision applications [1–3] and a superior performance in other domains [4–6]. Researchers are increasingly applying deep learning technology to remote sensing problems. The semantic segmentation of remote sensing data has been an important topic for decades and applied in many fields [7], such as environmental monitoring [8,9], crop cover and analysis [10–12], the detection of land cover and land use changes [13], the inventory and management of natural resources [14,15], etc. The complexity of the

geographical scene has considerably affected the accuracy of geographic feature classification [16–19], and the representativeness and quality of training samples have an important role in the performance of deep learning models for the semantic segmentation of remote sensing images [20–22].

Complexity, in contrast to simplicity, denotes a state of uncertainty, unpredictability, intricacy, or difficulty in terms of a description, explanation, or solution [23]. In the geographical sciences, the geocomplexity of the Earth's system comes from multiple interactions and feedback loops of the different spheres, including the geosphere, hydrosphere, lithosphere, atmosphere, cryosphere, and biosphere [24]. Various landforms, abundant types of surface objects, changeable meteorological conditions, human activities, and other factors [25,26] are important influential factors for the Earth's surface complexity that present challenges for the semantic segmentation of remote sensing images. There are three distinct characteristics of surface complexity, including scale dependence, non-linear driving, and a high level of uncertainty regarding evolutionary trends. For scale dependence, the spatial-temporal patterns of surface elements are diverse between low and high spatiotemporal scales, and the laws of different scales vary [27]. For non-linear driving, the surface elements are interrelated and constrained with complex nonlinearity interactions [28]. For a high level of uncertainty regarding the evolutionary trend, the overall evolutionary trend of the system could be changed by small changes in the surface system [29]. Practices have been conducted to investigate surface complexity in previous studies. Stand density and surface fraction, called geomorphic indicators, were used to quantify the spatial distribution to realize the mining of geospatial patterns of various land covers [30,31]. Shannon entropy was applied in urban research by quantifying information complexity [32]. The fractal dimension was employed in the ecological domain to understand spatial information [33]. Based on spatial neighbor dependence, a spatial local geocomplexity indicator was proposed to explain spatial errors between different spatial models [34]. It is a challenge to provide a geocomplexity indicator that can effectively improve the semantic segmentation of remote sensing. In addition, the impact of sampling bias, caused by simple random sampling, on model learning and the performance of the model is also a challenge.

Sampling bias can have undesired effects on the performance of deep learning remote sensing applications [35–39]. Deep learning has been utilized to realize the recognition and classification of remote sensing images in recent years [40–43] because of its efficient learning and powerful prediction abilities. While altering the network structure and increasing the number of layers were proven to be effective strategies for enhancing information extraction in remote sensing images, the selection of training samples played a crucial role in this improvement [44]. A common assumption in machine learning methods is that the training sample consists of identically distributed examples which are simply and randomly drawn from a population [45,46]. However, in remote sensing applications such as land use/cover analyses, this simplistic assumption may not hold true due to the presence of diverse and varied geographical features with imbalanced spatial and/or temporal distributions [44,47–49]. In geosciences, geo-surfaces, including land use/cover, are often characterized by spatially stratified heterogeneity (SSH), making it challenging to obtain representative training samples solely through the random sampling of a population [50,51]. If the study regions are spatially stratified and heterogenous and the sample size is not large enough with few or no samples in some strata, the problem of prediction bias will be exasperated [51]. This issue is commonly referred to as “sampling bias” or “sample selection bias” in the context of remote sensing applications. Quality training samples should effectively represent the population while minimizing sampling bias [50]. The common sampling methods [52], such as simple random sampling, systematic sampling, and strata sampling, have been applied by collecting labeled samples in remote sensing and selecting training samples of deep learning methods with a potentially unbalanced spatial distribution of samples and an unbalanced proportion of classes in samples [44]. Inadequate and non-representative training samples may be the primary causes of seman-

tic segmentation errors in remote sensing images [44,53,54]. Especially in the multiclass semantic segmentation of remote sensing images, category diversity, boundary ambiguity, and complex spatial distribution require higher-quality training samples than those required for binary class segmentation [55,56]. To achieve good performance in the multiclass semantic segmentation of remote sensing images, it is important to acquire training samples representing the overall characteristics of the study area.

Within a geospatial context, our paper aims to explore the effectiveness of utilizing complexity-related indicators for optimal sampling, with the goal of reducing sampling bias in remote sensing. For this purpose, we derived two indicators based on information entropy and Moran's I to quantify the multiclass complexity in remote sensing images. Based on the complexity scores, two optimal sampling strategies were proposed to reduce bias in the training samples representative of the population. To evaluate the effectiveness and extensibility of the optimal sampling method, several representative deep learning models, such as UNet, SegNet, Global CNN, DeepLab V3, FCN-ResNet, UperNet, and SegFormer, were selected for model comparison. The deep learning models selected above, with different architectures, are mainstream models commonly employed in remote sensing semantic segmentation applications. Additionally, we also did a sensitivity test using our optimal sampling methods for other classical machine learning methods such as Random Forest and XGBoost. By testing our proposed optimal sampling method against these representative models, we can effectively demonstrate the generalizability and extensibility of our approach. In summary, this paper made the following three contributions:

- (1) The entropy-based and gray-based multiclass indicators could effectively recognize the multiclass complexity features of remote sensing images and quantify geocomplexity information, providing a quantitative basis for selecting training samples.
- (2) An optimal sampling strategy was proposed to obtain training samples that were preferably representative of the population. Compared to the simple random sampling method, the optimal sampling method could improve the performance of the multiclass semantic segmentation and adeptly select samples with rich feature information while simultaneously reducing the sampling bias.
- (3) The optimal sampling method was effective and applicable to representative machine learning algorithms, particularly for those involving deep learning.

2. Methods

We proposed two geocomplexity indicators to quantify the complexity of multiclass remote sensing images, and based on the result of the complexity quantification, we designed the multiclass optimal sampling method. To verify the effectiveness and extensibility of the multiclass optimal sampling method, we carried out several control experiments for semantic segmentation of remote sensing images. The flow chart of the research is shown in Figure 1. There were three parts: multiclass complexity quantification, multiclass optimal sampling, and model evaluation. In the multiclass complexity quantification, the entropy-based indicator and Moran's I indicator were used to quantify the complexity of the ground truth and grayscale images, respectively, and the complexity scores were obtained to be the stratification factor and sampling weight in the optimal sampling stage. In the multiclass optimal sampling, the feature of the samples was described by the complexity score, and two optimal sampling methods were designed to select optimal training samples. In the model evaluation of multiclass semantic segmentation, the training samples were selected by two optimal sampling strategies (single-score vs. multi-score) and compared with the simple random training samples. The results of complexity quantification and semantic segmentation were analyzed to explain the usefulness of geocomplexity indicators for enhancement of multiclass segmentation.

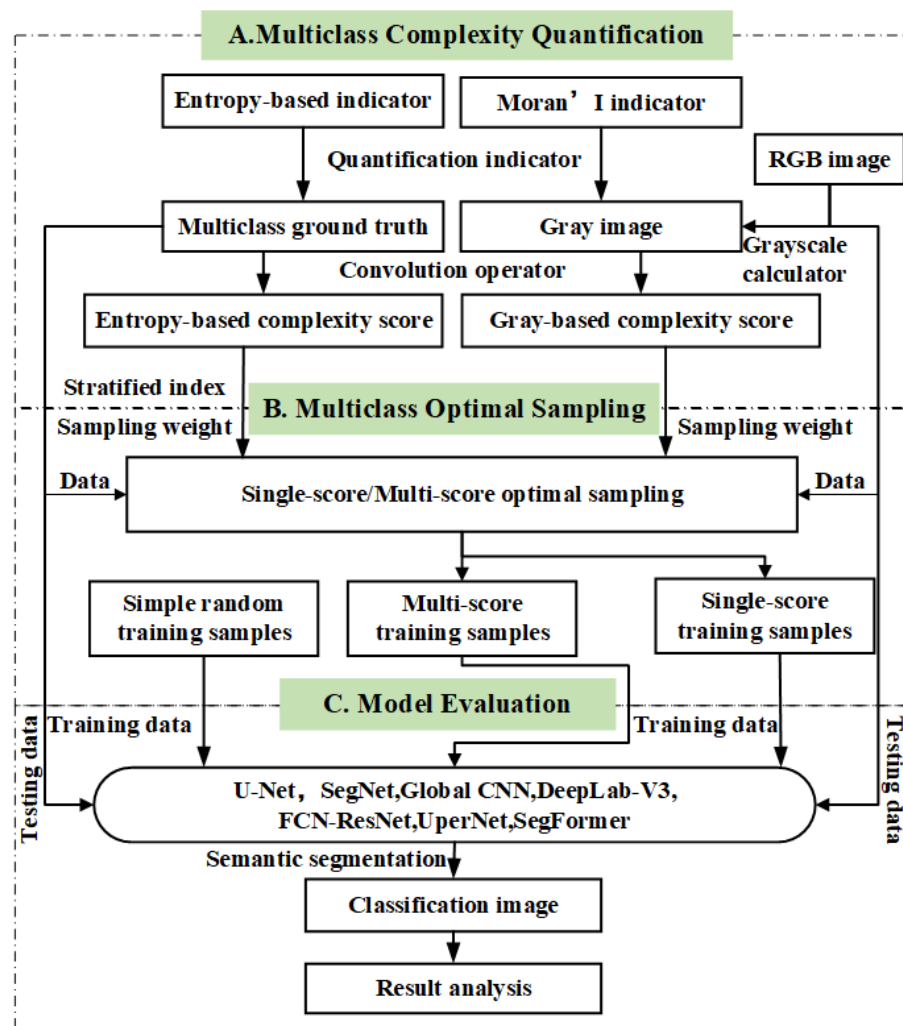


Figure 1. The flow chart of the research.

2.1. Definition of Geocomplexity Statistical Indicators and Complexity Quantification

The choice of statistical indicators to measure geocomplexity was based on geoscience context of our study. For fine-scale remote sensing semantic segmentation, we defined complexity of geo-scenes as a statistical indicator of the context surrounding a target pixel or spatial location of interest. Our study focuses on understanding geocomplexity from the perspective of spatial variogram/dependence [34] and spatial heterogeneity [51]. Spatial variation or uncertainty can be intuitively measured by information entropy based on information theory, and its higher value indicates higher levels of uncertainty and randomness (and thus, a higher complexity), showing multiple classes located or clustered in a location of interest. Moran's I can be used to measure spatial dependence or distribution of spatial locations, and its value at a location reflects the spatial distribution pattern of its context or surrounding neighbors. Similar complexity measures based on information entropy and spatial statistics have also been proposed and led to important improvements in other remote sensing applications [32,34,57,58].

(a) Entropy-based indicator and complexity quantification

Information entropy, proposed by Shannon in 1948 [59], is described as the probability of the occurrence of discrete random events and used to quantify the complexity of information. Based on the definition of information entropy, we designed an entropy-based indicator to quantify the multiclass non-linear complexity from remote sensing images and

used a convolution operator to achieve the extraction of the multiclass complexity features. The formula of the entropy-based indicator is defined as follows:

$$C = E \left[-\ln \left(P \left(x^{(d_k)} \right) \right) \right] = - \sum_{i=1}^c p \left(x_i^{(d_k)} \right) \ln \left(p \left(x_i^{(d_k)} \right) \right) \quad (1)$$

where x denoted the ground truth of the multiclass, d_k was the kernel size of the convolution operator, i was the class index within context, x_i was the class i for classification, c was the number of all classes, and $p \left(x_i^{(d_k)} \right)$ was the probability for x_i within the kernel size d_k . C was the entropy-based complexity score, and its value range was $[0, 1]$. This meant that as the complexity decreased, the entropy-based complexity score neared 0, and vice versa.

To correctly evaluate the complexity on the edge of the image, we expanded the input patch of complexity convolution with half the kernel size to avoid the edge effects of convolution computations [60] (Figure 2). The score of the target point or pixel reflected the entropy-based complexity of the surrounding context within the kernel of convolution. Thus, the kernel size was an important factor in complexity quantification [61].

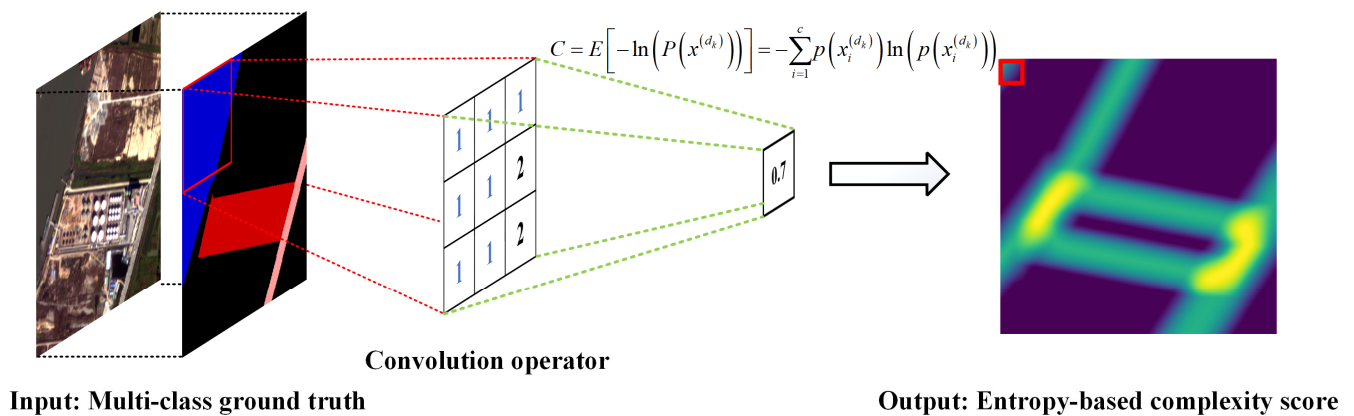


Figure 2. Flow chart for entropy-based complexity quantification. (The number in convolution operator represent different classes within receptive fields, and the red box indicates the complexity of target pixel.)

(b) Gray-based indicator and complexity quantification

Moran's I, developed by Patrick Alfred Pierce Moran [62], is a statistical measure of spatial autocorrelation and reflects the spatial pattern and spatial structure of data. Based on the features of Moran's I, we designed Moran's I indicator to identify the complexity of spatial patterns from remote sensing images and used the convolution operator to calculate multiclass complexity score. The formula of the gray-based indicator was as follows:

$$L = \frac{N \sum_{i \in B} \sum_{j \in B} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i \in B} (x_i - \bar{x})^2} \quad (2)$$

where N denotes the number of pixels within the sliding window. i was the pixel i , and w_{ij} was the spatial weight of pixel i and pixel j (w_{ij} was 1 if pixel i and pixel j were adjacent; otherwise, it was 0). x_i was the gray value for pixel i , while \bar{x} meant the average gray value of pixels within the sliding window. B was the collection of pixels, and W was the sum of spatial weight w . L was the gray-based complexity score, and its value range was $[0, 1]$. This meant that as the spatial complexity decreased, the gray-based complexity score neared 0, and vice versa.

To effectively calculate the spatial complexity on the edge of the image, we also expanded the input patch of complexity convolution with half the kernel size to avoid the edge effects of convolution computations (Figure 3). The gray-based complexity score of

the target pixel reflected the spatial complexity corresponding to the sliding window region. The kernel size of convolution was used to decide the size of the region, and we used the same kernel size to quantify the gray-based complexity in this paper corresponding to entropy-based complexity.

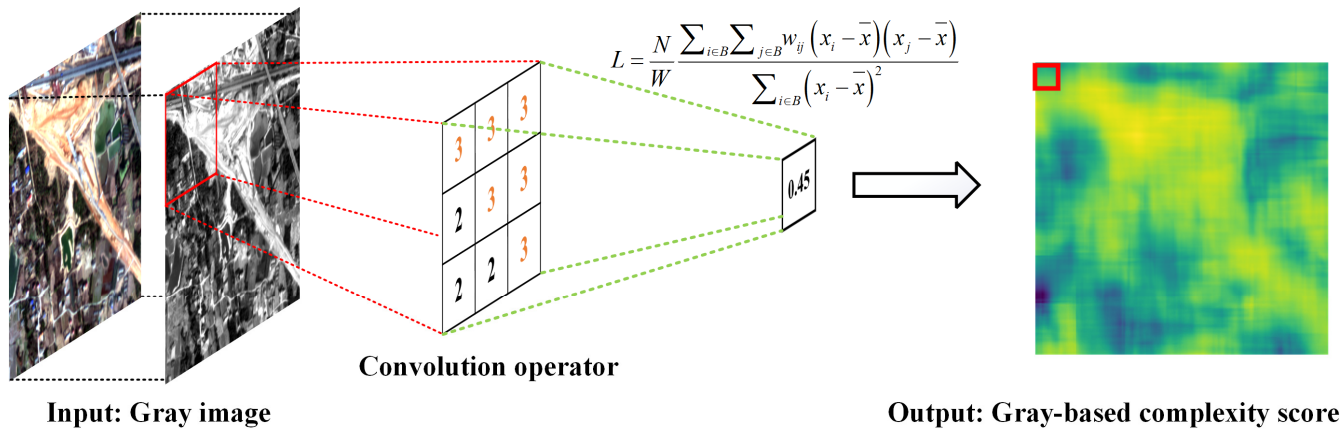


Figure 3. Flow chart for gray-based complexity quantification. (The number in convolution operator represent grayscale values within receptive fields, and the red box indicates the complexity of target pixel.)

2.2. Multiclass Optimal Sampling Method

It is evident that nature is lawful and structured rather than purely random, and there is spatial stratified heterogeneity (SSH) in geographical data [51]. And the presence of SSH is a source of bias in the simple random selection of training samples. Our proposed geocomplexity indicators could detect spatial correlation and spatial structure information from geo-surface, thereby indirectly reflecting SSH within context. Based on quantitative results, we designed a multiclass optimal sampling method to increase the representativeness of the selected samples, thus decreasing sampling bias during remote sensing information extraction.

There are two strategies for optimal sampling, single-score optimal sampling and multi-score optimal sampling, which differ in stratification factors and sampling weights (Figure 4). Single-score optimal sampling uses entropy-based complexity score as the stratifying and weighing factors. On the other hand, for the multi-score optimal sampling strategy, entropy-based complexity score is the stratifying factor, and the gray-based complexity score is used as the sampling weight. For optimal sampling, complexity scores of the stratification factor were firstly summarized, and the samples were divided into a certain number of stratum, which was determined according to the distribution of the average of these scores. Then, the same proportion was used in the sampling within each stratum, and the samples drawn from every stratum were combined into training samples. Complexity scores of the sampling weight were probability of sampling within each stratum so that the patches with higher complexity scores were more likely to be selected, because regions with high levels of complexity were more difficult to learn than those with low levels of complexity, and we put higher sampling weight on their samples. Therefore, the distribution of complexity in training samples was closer to the population distribution, while sampling bias was decreased.

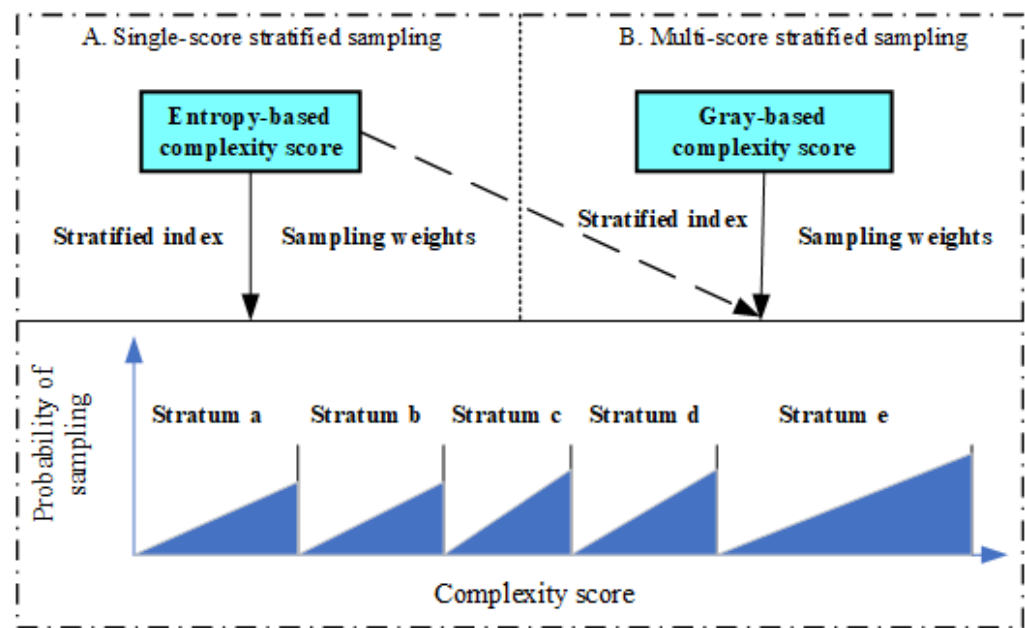


Figure 4. Flow chart for multiclass optimal sampling.

2.3. Model Evaluation

To verify the effectiveness and extensibility of the optimal sampling method, we used seven representative deep learning models and two machine learning algorithms to verify the improvement of multiclass segmentation in remote sensing images by optimal sampling strategy. These deep learning methods included U-Net, SegNet, Global CNN, DeepLab V3, FCN-ResNet, UperNet, and SegFormer. UNet is known for its use of skip connections, containing a context path to learn context information and a spatial path to preserve spatial information [63]. SegNet, like UNet, is a classic encoder–decoder architecture network using the combined index and nonlinear upsampling to minimize the number of parameters during model learning and improve the recognition effect of the segmentation task [3]. Compared to small filters (1×1 or 3×3) used in network architecture, Global CNN adopt symmetric, separable large filters and boundary refinement block to reduce the parameters and model the boundary alignment as a residual structure [64]. DeepLab V3 was another typical convolutional neural network architecture incorporating atrous spatial pyramid pooling (ASPP) modules enable the network to capture multi-scale context in images efficiently [1]. FCN-ResNet integrated the fully convolutional network [2] with residual connections from ResNet [65]. By utilizing skip connections, it effectively merged low-level and high-level features to improve segmentation performance. UperNet leveraged a combination of the pyramid pooling module (PPM) [66] and the feature pyramid network [67] to unify parsing of visual attributes across multiple levels. This approach exploited multi-level feature representations in an inherent and pyramidal hierarchy, enabling the incorporation of global prior representations [68]. SegFormer is an efficient and powerful semantic segmentation framework that unifies Transformers [69] with lightweight multilayer perceptron decoders, avoiding the interpolation of positional codes and complex decoders [70]. As for the other two machine learning methods selected, random forest [71] is a combination of tree predictors, which are independent with same distribution. It is robust with respect to noise. XGBoost [72] is a scalable and end-to-end tree boosting system, with a sparsity-aware algorithm for sparse data and weighted quantile sketch. Three-group experiment for each model was designed to compare results with those of simple random sampling, single-score optimal sampling, and multi-score optimal sampling. Pixel accuracy (ACC) and mean intersection over union (MIoU) were used to measure the performance of the learned models in semantic segmentation. For semantic segmentation of remote sensing images, MIoU can better measure the classification performance than pixel accuracy [73].

2.4. Evaluation and Prediction

(a) Pixel Accuracy

The pixel accuracy was defined as the ratio of the number of correctly classified pixels to the number of total pixels, and its formula was as follows:

$$ACC = \frac{\sum_{i=1}^K x_i}{N} \quad (3)$$

where K denoted the number of classes, x_i was the number of pixels correctly classified as class i , and N was the total number of pixels.

(b) Mean Intersection over Union

The mean intersection over union was defined as the size of the intersection for the correctly classified region and ground truth masks divided by the size of the union for the prediction region and ground truth masks, which was used to measure the degree of overlap between the intersection region and the union region. Its formula was as follows:

$$MIoU = \frac{1}{K} \frac{|\hat{y} \cap y|}{|\hat{y} \cup y|} = \frac{1}{K} \frac{\sum_{k=1}^K x_{kk}}{\sum_{i=1}^K x_{ik} + \sum_{j=1}^K x_{kj} - x_{kk}} \quad (4)$$

where K was the number of classes, \hat{y} was the number of pixels of prediction, and y was the ground truth. x_{kk} was the number of predicted pixels correctly classified as target class k , x_{ik} was the number of pixels classified as class k whose ground truth was class i , and x_{kj} was the number of pixels classified as class j whose ground truth was class k .

3. Experiment and Result

3.1. Experiment

3.1.1. Dataset

The Five-Billion-Pixels dataset [74], derived from GF-2 satellite images, was used to evaluate our proposed method. It consisted of 150 images with a size of 6800×7200 pixels and a ground resolution of 4 m, including a spectral range of blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.63–0.69 μm), and near-infrared (0.77–0.89 μm). There were 24 categories in this dataset, and the images were collected in various regions of China, containing rich and diverse information on complex multiclass geographic scenarios. Experiments based on this dataset can effectively verify the applicability of geocomplexity statistical indicators in multiclass classification. We only used the RGB band of the images in this paper to fairly compare it with other works. The Five-Billion-Pixels dataset is available online (<https://x-ytong.github.io/project/Five-Billion-Pixels.html> (accessed on 24 October 2023)).

3.1.2. Experimental Detail

We performed the calculation of the complexity-based statistical indicators and compared seven representative deep learning models and two other representative machine learning algorithms to investigate the enhancement of optimal sampling strategies for multiclass semantic segmentation with PyTorch. The experiments were conducted on a Linux server with a single Nvidia GeForce RTX 3090 GPU. The RGB image and the ground truth were cut into image patches with the size of 296×296 pixels without overlapping, and the size of the image patches included the target output size (256×256 pixels) along with a border size of 20 pixels to eliminate boundary effects in the complexity calculations. The gray image patches were generated from the RGB image patches. Three sampling methods were used to obtain the training dataset for multiclass segmentation, which was composed of 5000 image patches with a size of 256×256 pixels. The testing dataset, obtained by simple random sampling, included 2000 image patches with a size of 256×256 pixels. After performing a sensitivity test, we found that a kernel size of 41 was the optimal choice for the convolution operator when computing the complexity in a geo-surface scene from

the GID dataset, as employed in our experiment. The model was trained for 80 epochs with the Adam optimizer, and the learning rate, initially set to 0.001, was adjusted during the process of training.

3.1.3. Loss Function

The dice coefficient loss [75], designed to solve the class imbalance problem, was effective in learning better boundary representations. The dice coefficient loss L_{dice} can be written as follows:

$$L_{dice} = 1 - \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (5)$$

where the sums run over the N voxels, the predicted segmentation volume is $p_i \in P$, and the ground truth volume is $g_i \in G$.

The cross-entropy loss [76] can measure the differences of information between two probability distributions by cross-entropy and minimize it to judge a model. There are many variations of this loss function that apply to different situations [77]. The binary cross-entropy loss (BCE loss) was chosen to train the model in our research, which was used as a stable loss function for segmentation. We made a slight change to the BCE loss to make it suitable for multiclass segmentation, and L_{BCE} can be written as follows:

$$L_{BCE} = \sum_K -\frac{1}{N} \sum_i [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \quad (6)$$

where K was the number of classes, N was the number of total pixels, g_i was the probability of the ground truth mask in pixel i belonging to class k whose value was 0 or 1, and p_i was the probability of the prediction in pixel i belonging to class k .

The BCE loss gauged the pixel-wise similarity between the predicted and target masks, whereas the dice coefficient loss emphasized the spatial overlap and boundary localization. The combination of the dice coefficient loss and BCE loss, which was selected to judge the model in our experiments, could be beneficial in achieving a balanced optimization that captured both the pixel-wise classification and spatial accuracy of the semantic segmentation. Its formula was as follows:

$$L_{total} = \frac{1}{2} L_{dice} + \frac{1}{2} L_{BCE} \quad (7)$$

3.2. Result

3.2.1. Quantification of Geocomplexity

The geocomplexity was effectively detected by the entropy-based indicator and gray-based indicator. The geocomplexity quantification results of six different scenes (an irrigated field, urban residential area, road, rural residential area, arbor forest, and fish pond) show the different characteristics of the two indicators in the complexity extraction (Figure 5). The entropy-based indicator was sensitive to the boundaries of different classes and the contours of geographical objects. The target pixel exhibited a higher entropy-based complexity score, indicating the intricate presence of a greater number of different categories (classes) in its mixed vicinity. The entropy-based complexity score decreased when the environment around the target pixel was homogeneous. The gray-based (Moran's I) indicator had a better performance in extracting the spatial structure information of geographical objects in the remote sensing images. A lower gray-based (Moran's I) score indicated a bigger difference in gray values, as well as a higher level of randomness of spatial distributions in the pixels. The gray-based (Moran's I) score increased when the environment around the target pixel tended to be homogeneous, indicating a strong spatial autocorrelation based on the statistical implication of Moran's I [78].

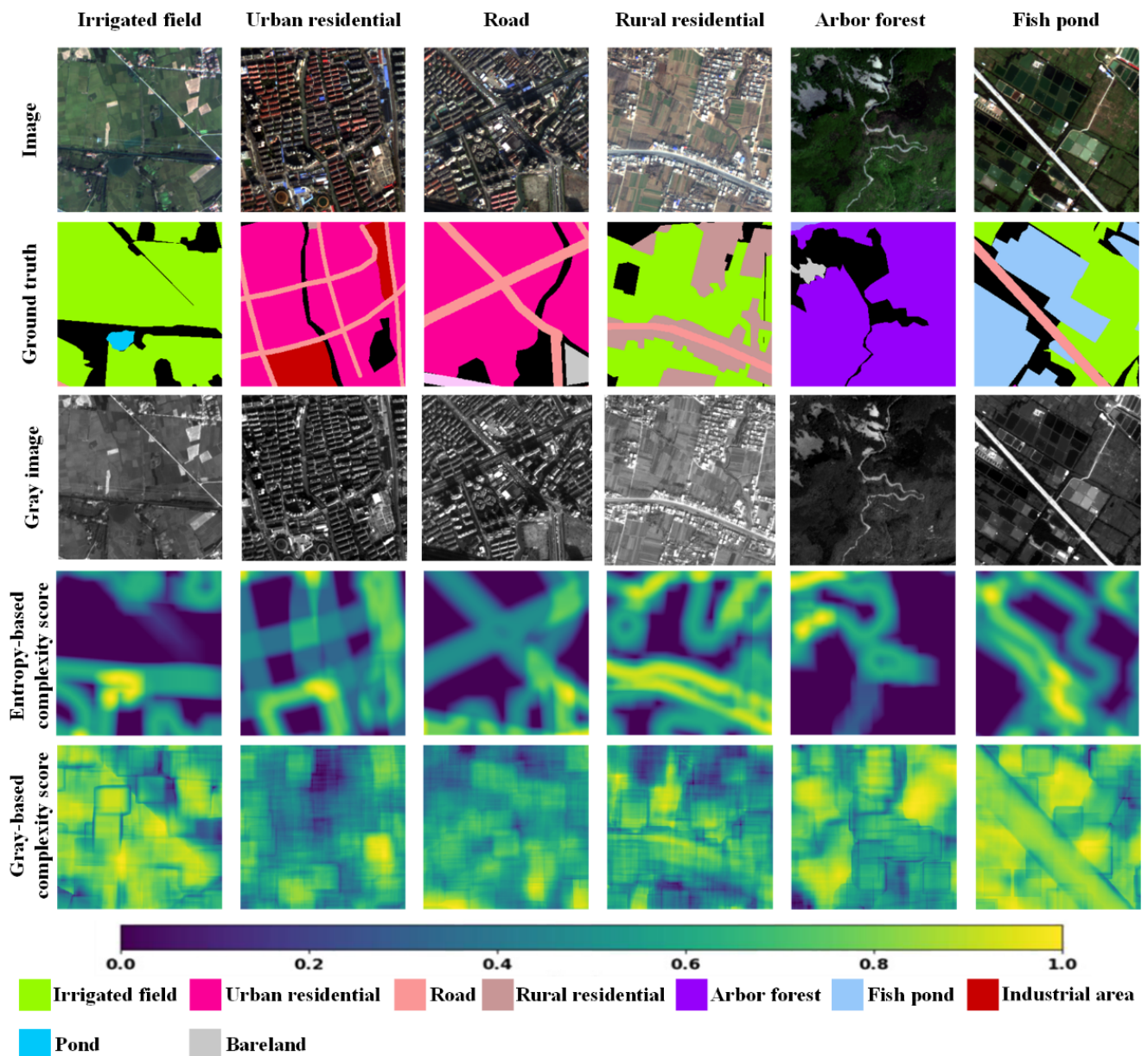


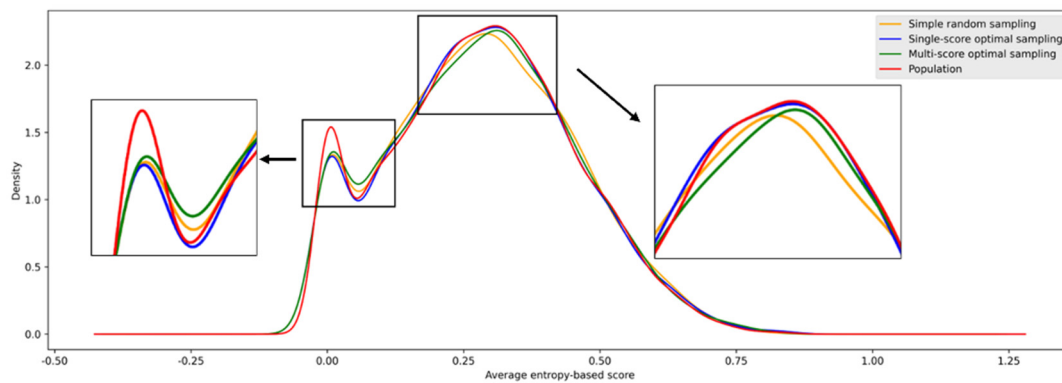
Figure 5. Quantification of geocomplexity in different scenes.

3.2.2. Complexity Score Distribution of Training Samples

The optimal sampling method was capable of selecting training samples with more abundant feature information compared to the simple random sampling method. As illustrated in Figure 6, the complexity score distributions of the training samples obtained through optimal sampling better aligned with the population distribution than the distribution resulting from simple random sampling. The average complexity score distribution was analyzed to discover the differences between training samples obtained by simple random sampling, single-score optimal sampling, and multi-score optimal sampling (Figure 6). The kernel density curve was drawn to observe the complexity distribution of different samples. There were two peaks in the curve of the kernel density for the entropy (entropy-based) complexity score, which indicated that the entropy complexity of the training sample was concentrated within the region close to 0 and the region of [0.20, 0.50]. Compared to the training samples of the simple random sampling, the training samples of the single-score optimal sampling had more patches with an entropy complexity score within the range of [0.20, 0.50], and the training samples of the multi-score optimal sampling had more

patches with an entropy complexity score within the range of [0.30, 0.40]. This suggested that the two optimal training samples contained rich feature information with moderate levels of entropy complexity compared to the simple random training samples. Regarding the distribution of the Moran's I (gray-based) complexity, the score distributions of the samples obtained through optimal sampling closely resembled the natural distribution of the simply randomly selected samples. The Moran's I complexity was concentrated in the high-value region of [0.85, 1.00]. In comparison to the training samples obtained through simple random sampling, those acquired through optimal sampling exhibited a greater number of patches with gray-based (Moran's I) complexity scores falling within the range of [0.85, 1.00]. Specifically, the training samples from multi-score optimal sampling contained more patches with Moran's I complexity scores within the range of [0.85, 0.95] compared to the other training samples. This suggested that the training samples contained more feature information than the others because a high level Moran's I complexity reflects a high spatial correlation, indicating a simpler (less random) distribution of geo-objects with fewer land classes.

(a) Distributions of the average entropy-based complexity scores



(b) Distributions of the average gray-based (Moran's I) complexity scores

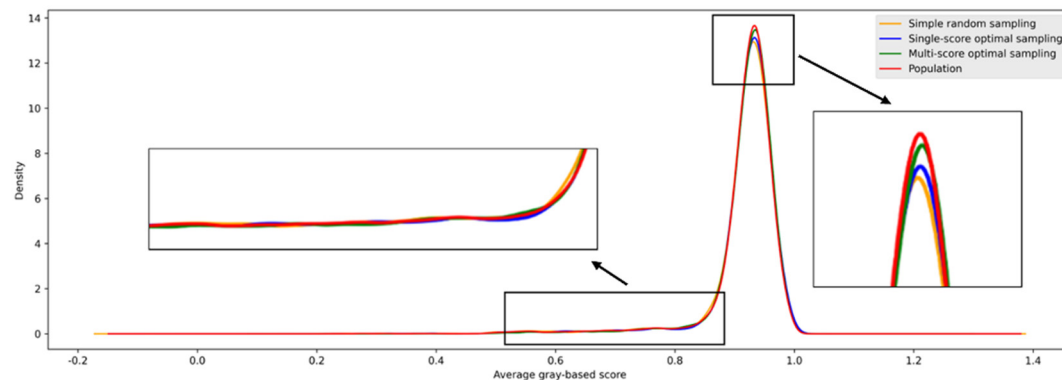


Figure 6. Distributions of the average complexity scores in the training samples selected using optimal sampling methods vs. simple random sampling.

As shown in the results, when the statistical distribution (class proportions) of the samples remains relatively stable, their spatial distribution may present considerable spatial heterogeneity, potentially posing a challenge for the selection of training samples to represent the population feature. The proportion of categories was counted to find the difference in the training samples selected by the three sampling methods (Figure 7). Six classes accounted for more than 5%, of which irrigated fields accounted for more than 35%. Parks, snowy areas, stadiums, squares, railway stations, and airports did not appear in the training samples, with a ratio of 0%. Except for irrigated fields and lakes, the ratio of the classes was roughly the same in the training samples generated in the three ways. For the result (Figure 7), the multi-score optimal sampling seemed to slightly decrease the

proportion of irrigated field and lake areas, compared with the other sampling methods; utilizing single-score optimal sampling led to a slight decrease in the proportion of irrigated field areas and a slight increase in the proportion of lake areas.

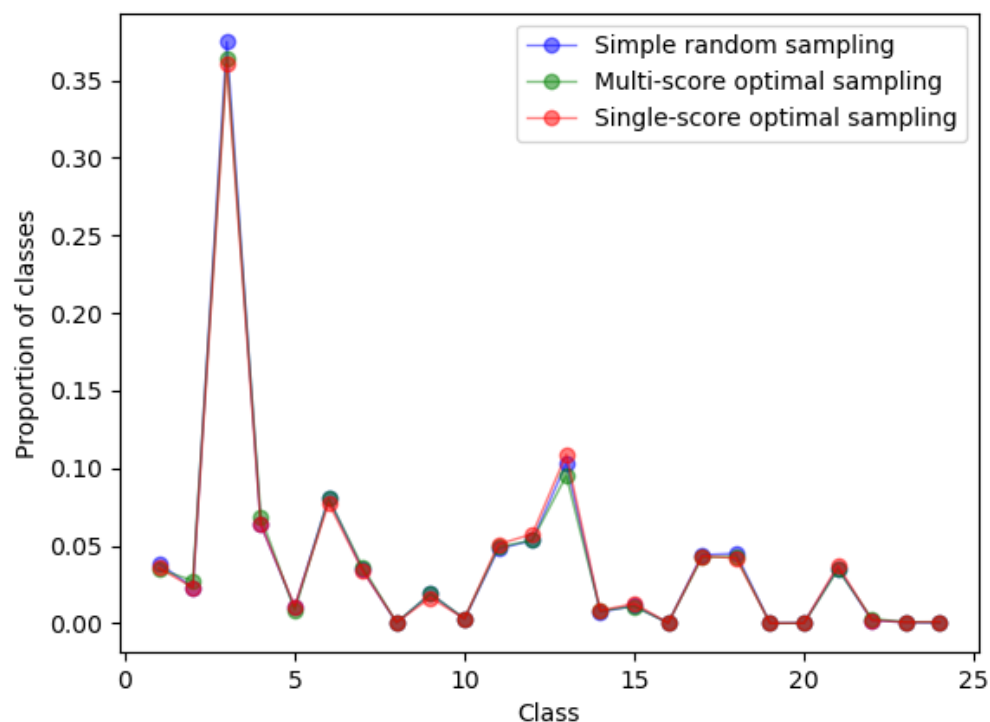


Figure 7. The proportion of class in training samples. (Class index: 1—industrial area, 2—paddy field, 3—irrigated field, 4—dry cropland, 5—garden, 6—arbor forest, 7—shrub forest, 8—park, 9—natural meadow, 10—artificial meadow, 11—river, 12—urban residential, 13—lake, 14—pond, 15—fish pond, 16—snow, 17—bareland, 18—rural residential, 19—stadium, 20—square, 21—road, 22—overpass, 23—railway station, and 24—airport).

3.2.3. Land Cover Segmentations

As shown in the results, the optimal sampling method was able to enhance the multi-class segmentation of remote sensing images, evidenced by the performance across seven representative deep learning models (Table 1) and two machine learning algorithms. The result showed these models had a great performance in multiclass segmentation measured by the ACC, which was more than 96% when using the deep learning models. Compared to simple random sampling, the result showed that single-score optimal sampling improved the performance of FCN-ResNet (increasing the ACC by 0.2% and MIoU by 1.9%), DeepLab-V3 (increasing the ACC by 0.3% and MIoU by 5.5%), SegNet (increasing the ACC by 0.2% and MIoU by 3.4%), SegFormer (increasing the ACC by 0.1% and MIoU by 2.3%), UNet (increasing the ACC by 0.2% and MIoU by 3.0%), Global CNN (increasing the ACC by 0.1% and MIoU by 1.8%), and UperNet (increasing the ACC by 0.1% and MIoU by 1.0%). Multi-score optimal sampling improved the performance of DeepLab-V3 (increasing the MIoU by 0.5%), UNet (increasing the ACC by 0.2% and MIoU by 2.2%), SegFormer (increasing the ACC by 0.2% and MIoU by 2.4%), Global CNN (increasing the MIoU by 1.1%), and UperNet (increasing the MIoU by 0.6%). For the sensitivity experiment using random forest and XGBoost, the result showed that single-score optimal sampling improved the performance of the random forest model (increasing the ACC by 0.2% and MIoU by 0.2%) and XGBoost (increasing the ACC by 0.3% and MIoU by 0.2%), and multi-score optimal sampling improved the performance of the random forest model (increasing the ACC by 0.6% and MIoU by 0.4%) and XGBoost (increasing the ACC by 0.9% and MIoU by 0.4%). As the training samples change, the improvement in MIoU can be used as an indirect indicator of the reduction in sampling bias during model training.

Table 1. The testing results on the Five-Billion-Pixels dataset.

Model	Simple Random Sampling		Single-Score Optimal Sampling			Multi-Score Optimal Sampling		
	ACC	MIoU	ACC	MIoU	Improved ^a	ACC	MIoU	Improved ^a
FCN-ResNet	0.968	0.428	0.970	0.447	1.9%	0.968	0.427	-
DeepLab-V3 (Encoder:Resnet101)	0.972	0.419	0.975	0.474	5.5%	0.972	0.424	0.5%
SegNet	0.971	0.458	0.973	0.492	3.4%	0.971	0.453	-
Segformer (backbone:B0)	0.974	0.506	0.976	0.529	2.3%	0.976	0.530	2.4%
UNet	0.975	0.520	0.977	0.550	3.0%	0.977	0.542	2.2%
Global CNN	0.981	0.600	0.982	0.618	1.8%	0.981	0.611	1.1%
UperNet (backbone:Resnet101)	0.982	0.618	0.982	0.628	1.0%	0.982	0.624	0.6%

^a: The percentage of the improvement in MioU for semantic segmentation utilizing optimal sampling methods. (The bold numbers indicate the best performance or improvement in each column of the table.)

The predictions, made by UNet, were displayed to demonstrate the improved detail of the semantic segmentation by the optimal sampling strategy (Figure 8). Various typical scenes, including agricultural, urban, rural, urban–rural transition, river, and coastal areas, were employed to evaluate the effectiveness of the optimal sampling and model. These figures illustrated that the models trained using optimal sampling exhibited a more representative selection of training samples, contributing to improvements in semantic segmentation, even when utilizing the same number of training samples, which was 5000 in our experiment. In urban, rural, and urban–rural areas, roads and urban residential areas were classified more precisely, and the boundaries between the classes were also clearer with the optimal sampling, within which single-score optimal sampling had a better effect than that of multi-score optimal sampling. In agricultural areas and river areas, the spatial details of the optimal sampling predictions were more abundant. The optimal sampling method enabled the model to learn more representative features to better cope with classification in coastal areas with a high level of diversity and a dense distribution of features.

The predictions generated by Global CNN (Figure 9) and UperNet (Figure 10) are showcased to illustrate the enhanced level of detail in semantic segmentation achieved through the optimal sampling strategy. Both Global CNN and UperNet demonstrated superior segmentation accuracies compared to that of UNet, particularly in the precise classification of ground objects and the recognition of boundaries. Specifically, single-score optimal sampling with Global CNN exhibited a superior effectiveness compared to the other methods in urban and rural scenarios. It demonstrated more accurate classifications of roads and buildings, along with clearer boundaries. Additionally, in urban–rural regions, the result of multi-score optimal sampling classified urban residential areas and irrigated fields more accurately compared to other methods. For UperNet, the segmentation results obtained by optimal sampling displayed continuous and complete roads and clear building boundaries. Compared to the simple random sampling method, the classification results of the optimal sampling methods showed fewer misclassification instances.

For SegFormer (Figure 11), the optimal sampling method significantly enhanced the segmentation performance of the model. The predictions obtained through optimized sampling were notably more accurate in identifying paddy fields and irrigated fields in agricultural areas and rural areas compared to simple random sampling. In urban areas and urban–rural areas, the segmentation of roads was continuous, and the boundaries between urban residential areas and irrigated fields were clearly identified. Additionally, in river areas, the predictions of riverways obtained by multi-score optimal sampling outperformed those of the other methods.

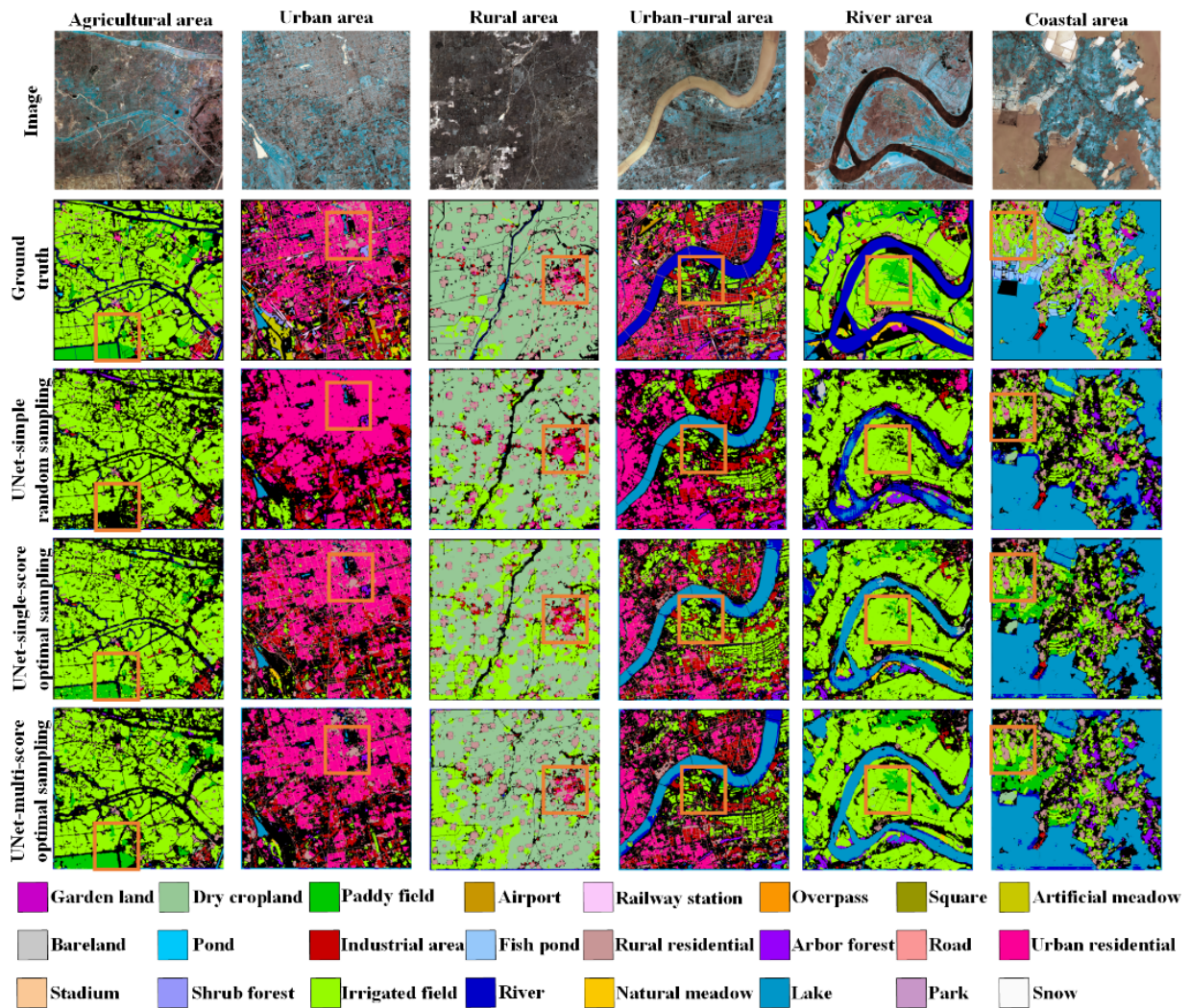


Figure 8. Segmentation results of three sampling methods generated by UNet.

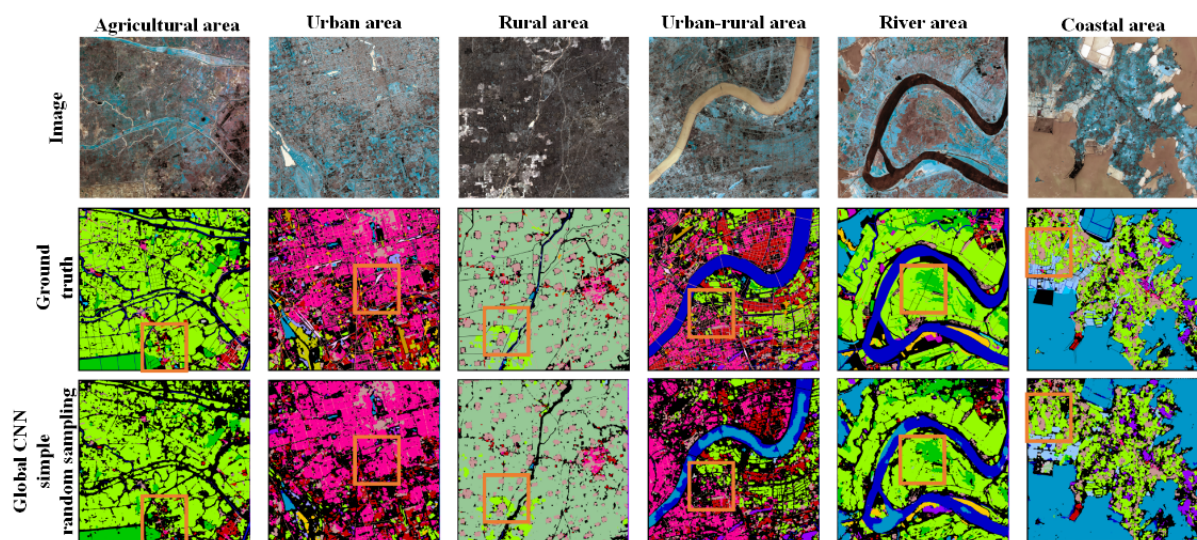


Figure 9. Cont.

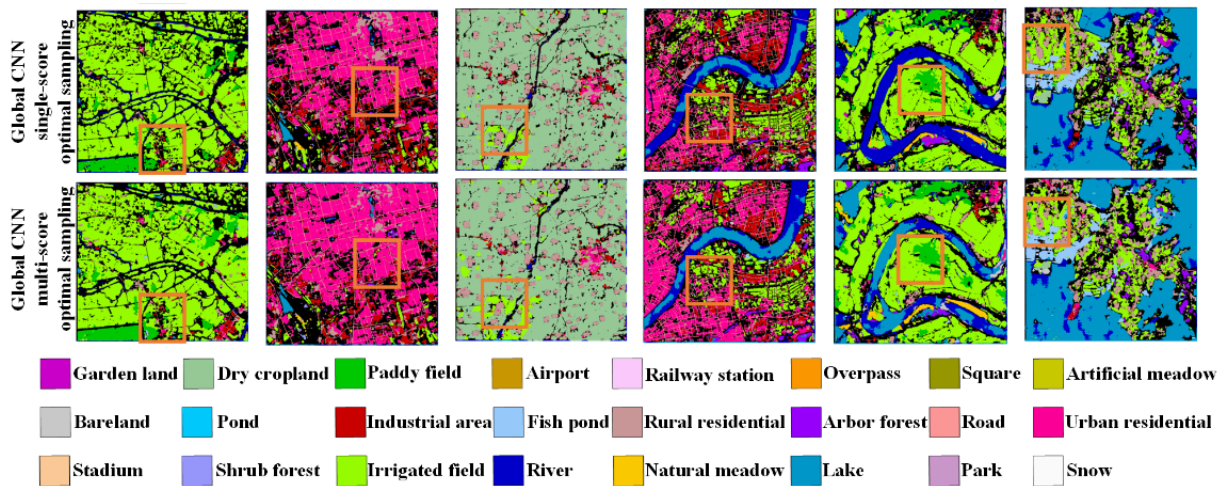


Figure 9. Segmentation results of three sampling methods generated by Global CNN.

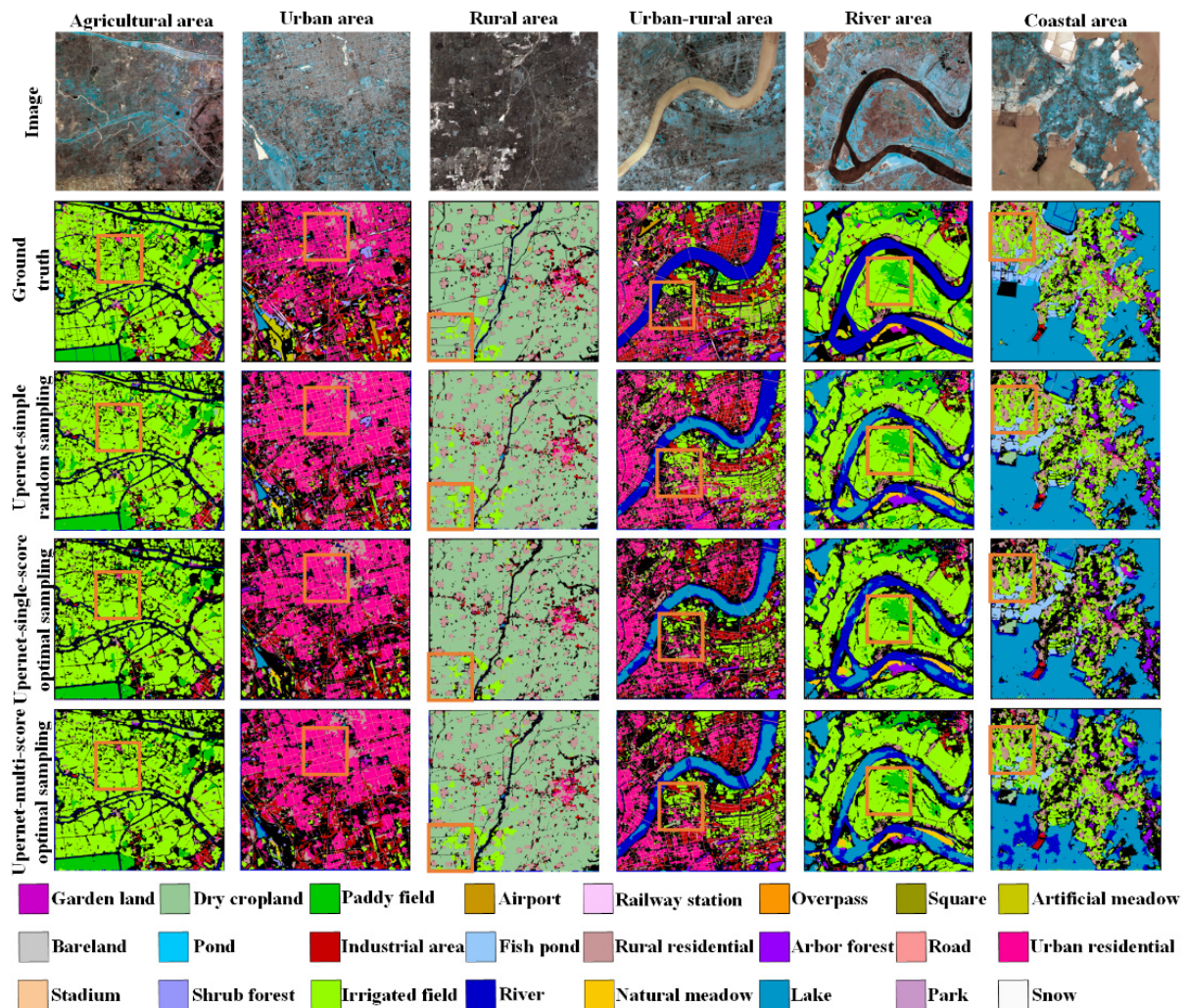


Figure 10. Segmentation results of three sampling methods generated by UperNet.

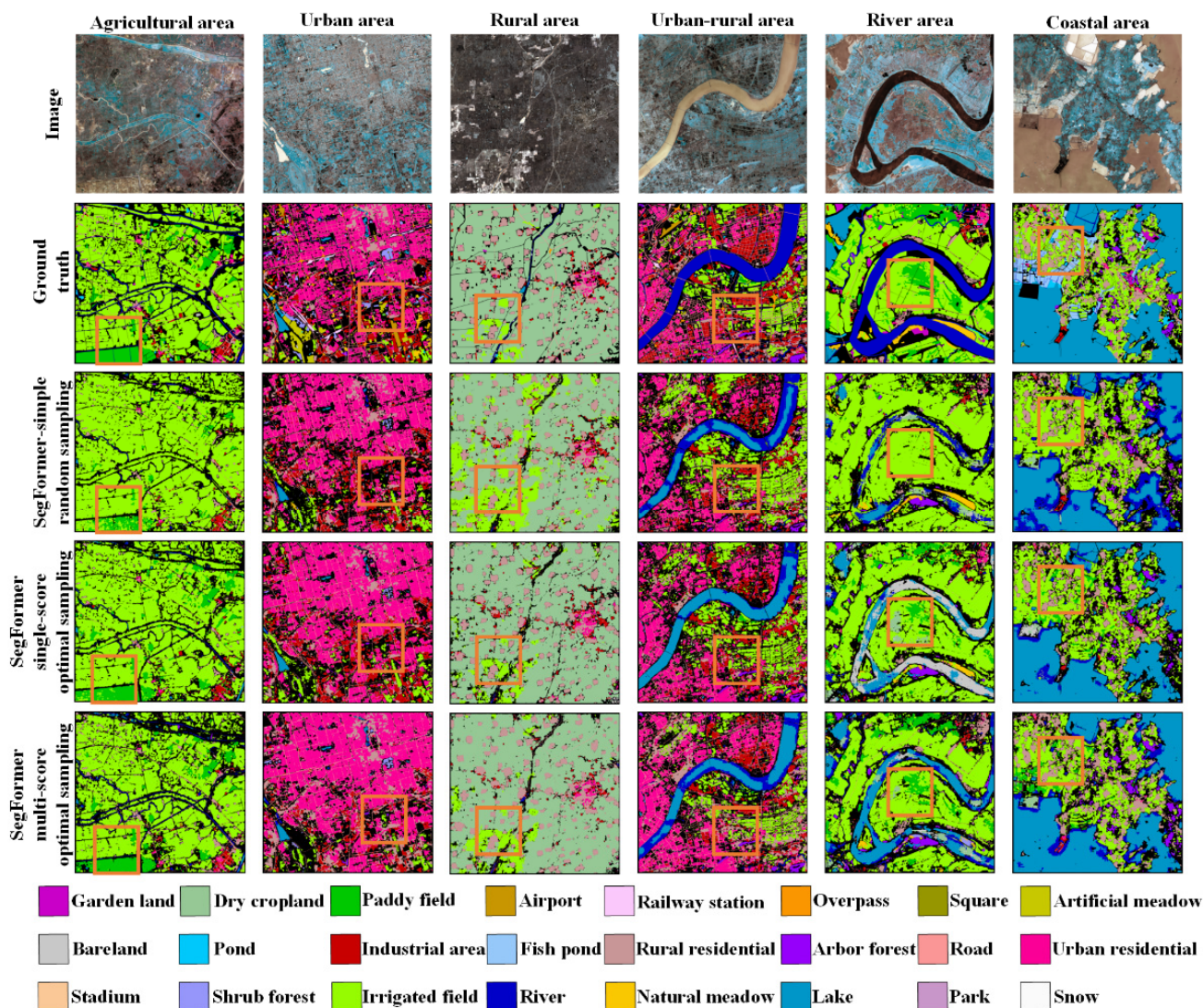


Figure 11. Segmentation results of three sampling methods generated by SegFormer.

4. Discussion

This study proposed two geocomplexity multiclass indicators, derived from the definition of information entropy and Moran’s I, to detect the surface complexity at the pixel level from the ground truth and gray images. A visualization of the complexity scores showed that the entropy-based indicator was sensitive to the boundaries of different classes and the contours of geographical objects, and the gray-based indicator better extracted the spatial structure information of geographical objects in remote sensing images. The ground truth contained the spatial locations and specific category information of geographic objects. The gray images demonstrate the situations of the ground objects through gray values. The entropy-based measure of complexity was based on Shannon’s definition of information entropy. Moran’s I was used to evaluate the spatial autocorrelation of similar and dissimilar values of a variable (gray) observed across space [79]. Within a complex geo-scene, a high entropy score indicates the presence of various types (classes) of geo-objects. The label value of a class remained the same, but its gray image values may vary. Consequently, recognizing the fine-scale spatial structure information of geographical objects using the entropy-based indicator was challenging due to the homogenized expression of the same ground object. In contrast, the gray-based (Moran’s I) indicator proved effective in extracting fine-scale spatial autocorrelation information.

Through our results for the typical semantic segmentation networks with different structures, we verified the generalizability and effectiveness of the multiclass complexity-based optimal sampling method. Previous studies [44,53,54,80,81] have shown that the stratified sampling method can obtain training samples from different strata (regions), potentially improving the level of classification accuracy. However, the performance improvement in these studies depended on correctly stratifying (partitioning) the data, as there is no quantified standard indicator to measure the contribution of stratification to performance, and many have overlooked the significant contribution of each individual sample to the model's generalization capability for prediction. In our research, two optimal sampling methods were designed to select training samples with a high level of population representativeness. The distribution (Figure 6) showed the aggregation of the feature information (complexity) within the training sample from the perspective of the information content (entropy) and spatial correlation (Moran's I), respectively. Compared with the training samples acquired by simple random sampling, the proportion of classes and the complexity distribution of the training sample, obtained by optimal sampling, were adjusted so that the training sample reflected information about the population more comprehensively. The entropy-based complexity score was selected as the stratifying factor because its average complexity score distribution greatly reflected the spatially heterogeneity of the samples. Land cover classes and the combination of different features were used as the stratified indicator to acquire high-quality training samples, exploring the impact of the training sample distribution on the accuracy of land cover classification [44]. Compared with our previous work [61], the multiclass scene typically exhibited a more complex geo-object distribution and structure compared to the binary class scene. A combination of geocomplexity indicators was used in our optimal sampling method. The applicability and effectiveness of the multiclass complexity-based optimization sampling method have been demonstrated. Stratified equal random sampling [82] could provide class-level-accurate land use and land cover mapping from remote sensing images, even for minority classes. Compared to the stratified equal random sampling, our optimal sampling method employed stratified weight sampling to ensure a higher probability of selecting samples with a higher complexity or salient features. In scenes with a higher level of complexity, the distribution of geo-objects is more intricate, requiring additional feature information to aid in model learning. Our method could adeptly select samples with rich feature information while simultaneously enhancing the representativeness of training samples. Our research offered a potential solution to address the issue of sample bias in the multiclass semantic segmentation of remote sensing images.

By our extensive experiments involving representative deep learning and machine learning methods, it was observed that the optimal sampling method consistently enhances multiclass semantic segmentation across the models of different structures, with reduced sampling bias. The single-score optimal sampling method exhibited a slight advantage over the multi-score optimal sampling approach in improving the segmentation accuracy of the model. But the performance of multi-score optimal sampling for SegFormer surpassed that of the former. In our experiments, UperNet demonstrated a superior semantic segmentation performance on the GID dataset compared to that of the other models. Furthermore, our research utilized SegFormer-b0 to assess the suitability of the optimal sampling method within the Transformer architecture. An enhanced performance can be achieved with larger SegFormers, such as b1, b2, b3, b4, or even -b5. It is depicted in Figure 6 that the distinction between training samples obtained by three methods was the changed number of patches between different intervals of complexity scores, indicating that the feature information offered by the training samples was different. Consistently, our experimental findings demonstrated that the optimal sampling methods significantly enhanced the prediction performance of models in the multiclass semantic segmentation of land cover, particularly for the models that initially had a low segmentation accuracy (MIoU), such as UNet, DeepLab-V3, and SegFormer-b0. It could be observed that the models with a relatively poor initial performance did not sufficiently extract the relevant feature information when

trained on simply randomly sampled data (Figure 12), since the assumption of simple random sampling is violated in the complex geo-surface scenes. This inadequate feature learning from the simple random training samples resulted in the poor recognition and segmentation of the target regions of interest in urban areas.

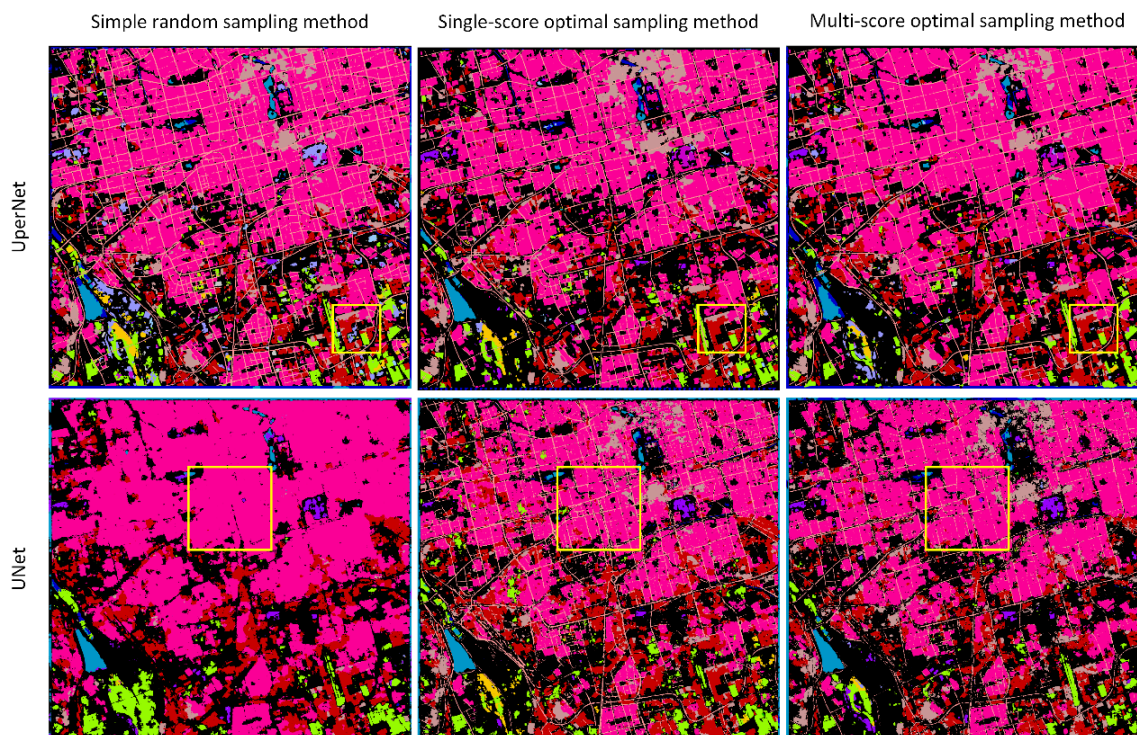


Figure 12. The segmentation of urban areas through UperNet, UNet, and SegNet.

However, upon employing the optimal sampling methods proposed in our research, the selected training samples could contain more representative and informative feature information, satisfying the requirement for the spatiotemporal heterogeneous distribution of the geo-surface, compared to those acquired by simple random sampling. This is because the optimal sampling strategies aim to identify and include samples that better capture the diversity and complexity of the geo-features present in the data. Therefore, by training on these optimally sampled datasets, the models could more effectively mine and leverage the discriminative feature information present in the geo-features. Consequently, their ability to accurately recognize and segment the target regions within urban areas was significantly improved, as evidenced by the substantial performance gains observed in Figure 12. Even without an increased number of training samples, it is illustrated in Figure 12 that the structure and boundaries of roads were accurately identified by UNet when utilizing optimal sampling. Conversely, roads were not recognized by UNet when utilizing simple random sampling.

Simultaneously, our proposed optimal sampling approach also yielded significant improvements for models that initially had relatively a high segmentation accuracy (MIoU), such as UperNet and Global CNN. By adjusting and enhancing the feature information present in the training samples through optimal sampling, misclassification errors were further alleviated. Consequently, the structural boundaries and delineation of different land cover classes were more clearly and accurately identified by UperNet when trained on the optimally sampled data, compared to its performance using randomly sampled training data (as evident in Figure 12).

These experimental results demonstrate that our optimal sampling method is capable of obtaining suitable and representative training samples, which can effectively improve the performance of semantic segmentation models for land use and land cover across a range

of architectures and initial accuracy levels. By providing the models with training data that better capture the diverse feature information and spatial heterogeneous distributions present in the geo-features, our approach enables more effective learning and the accurate segmentation of land cover classes in these complex, varying, and high-resolution remotely sensed datasets.

As shown in the results, optimal sampling significantly improved the performance of various deep learning methods, while only providing marginal benefits to other machine learning methods, such as random forest and XGBoost. Deep learning models, such as those employing convolutional or attention mechanisms, can effectively leverage optimal samples by modeling the surrounding context information. This is particularly advantageous because our geocomplexity indicators capture surrounding variation patterns through metrics like entropy and Moran's I, which align well with the context-based learning capabilities of deep learning methods.

In contrast, other machine learning methods like random forest and XGBoost do not possess internal mechanisms to capture contextual information [71,72]. These models rely primarily on the associations between spectral features and target labels for inference and prediction. As a result, the geocomplexity indicators contributed only slightly to the performance improvement in semantic segmentation for these methods. However, while our method's contributions to traditional machine learning techniques like random forest are slight and limited, it significantly enhances deep learning methods. Given the growing dominance of deep learning for the semantic segmentation of land cover using remotely sensed data [7,83], our approach proves especially beneficial for these advanced methods.

5. Conclusions

Our study offered an efficient way for quantifying the complexity of multiclass geo-scenes, sampling representative samples from label datasets, and reducing sampling bias during deep learning model training. This study provides important an reference for quantifying the complexity features of geo-surface scenes and improving the semantic segmentation of high-resolution remote sensing images. Our geocomplexity indicators, including entropy-based and gray-based indicators, could identify the surrounding variation and spatial features on a target pixel or regions by convolution operators. Based on complexity scores, the optimal sampling methods improved the distributions of training samples in model training. In the multiclass segmentation of remote sensing images, the single-score optimal sampling method achieved the greatest improvement in DeepLab-V3 (increasing the ACC by 0.3% and MIoU by 5.5%), and the multi-score optimal sampling method achieved the greatest improvement in SegFormer (increasing the ACC by 0.2% and MIoU by 2.4%). The results indicates that the proposed geocomplexity indicators and the optimal sampling method had promising potential for improving multiclass semantic segmentation tasks on remote sensing imagery. By addressing these challenges of geo-scene complexity and sampling bias, our proposed approach can be used to considerably improve models' performance and generalization capabilities for land use and land cover mapping from remotely sensed data using deep learning.

There were several limitations and prospects for our study. Firstly, the surface complexity was influenced by many factors [25,26], but the entropy-based indicator and gray-based indicator only considered one aspect of complexity. This might have led to the potential impact of other factors in the geo-surface scenes being overlooked. In future research, we will explore more factors to quantify the surface complexity and improve the semantic segmentation of remote sensing images. The combination of various geocomplexity indicators may be an effective measure of enhancing information extraction from remote sensing, and applicable scenarios need to be further explored. Secondly, the result of complexity quantification was not directly involved in model learning. Rich features could improve the performance of information extraction in remote sensing by deep learning models [84]. Surface complexity information has great potential in the field of remote sensing applications. Thirdly, sampling bias will progressively decrease with the expansion of the training

sample size. The relationship between training sample size and sampling bias is a valuable research topic when aiming to obtain the best result with limited samples.

Author Contributions: Conceptualization, W.H. and L.L.; methodology, W.H. and L.L.; software, W.H. and X.G.; validation, W.H. and X.G.; formal analysis, L.L. and W.H.; investigation, L.L.; resources, X.G. and L.L.; data curation, W.H.; writing—original draft preparation, W.H.; writing—review and editing, W.H. and L.L.; visualization, W.H.; supervision, L.L. and W.H.; project administration, L.L.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key Research and Development Program of China (grant number 2021YFB3900501).

Data Availability Statement: Data related to this article are available upon request to the corresponding authors. The data are not publicly available due to [ongoing unpublished studies].

Conflicts of 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.

References

- Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 833–851.
- Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)]
- Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
- Bayoukh, K. A Survey of Multimodal Hybrid Deep Learning for Computer Vision: Architectures, Applications, Trends, and Challenges. *Infin. Fusion* **2024**, *105*, 102217. [[CrossRef](#)]
- Aizenstein, H.; Moore, R.C.; Vahia, I.; Ciarleglio, A. Deep Learning and Geriatric Mental Health. *Am. J. Geriatr. Psychiatry* **2023**, *32*, 270–279. [[CrossRef](#)] [[PubMed](#)]
- Deng, Z.; Wang, T.; Zheng, Y.; Zhang, W.; Yun, Y.-H. Deep Learning in Food Authenticity: Recent Advances and Future Trends. *Trends Food Sci. Technol.* **2024**, *144*, 104344. [[CrossRef](#)]
- Yuan, X.; Shi, J.; Gu, L. A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. *Expert. Syst. Appl.* **2021**, *169*, 114417. [[CrossRef](#)]
- Blaschke, T.; Lang, S.; Lorup, E.J.; Strobl, J.; Zeil, P. Object-Oriented Image Processing in an Integrated GIS/Remote Sensing Environment and Perspectives for Environmental Applications. *Environ. Inf. Plan. Politics Publ.* **2000**, *2*, 555–570.
- Yuan, X.; Sarma, V. Automatic Urban Water-Body Detection and Segmentation From Sparse ALSM Data via Spatially Constrained Model-Driven Clustering. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 73–77. [[CrossRef](#)]
- Yang, S.; Chen, Q.; Yuan, X.; Liu, X. Adaptive Coherency Matrix Estimation for Polarimetric SAR Imagery Based on Local Heterogeneity Coefficients. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6732–6745. [[CrossRef](#)]
- Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [[CrossRef](#)]
- Jadhav, J.K.; Singh, R.P. Automatic Semantic Segmentation and Classification of Remote Sensing Data for Agriculture. *Math. Models Eng.* **2018**, *4*, 112–137. [[CrossRef](#)]
- Rottensteiner, F.; Sohn, G.; Gerke, M.; Wegner, J.D.; Breitkopf, U.; Jung, J. Results of the ISPRS Benchmark on Urban Object Detection and 3D Building Reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 256–271. [[CrossRef](#)]
- Managi, S.; Wang, J.; Zhang, L. Research Progress on Monitoring and Assessment of Forestry Area for Improving Forest Management in China. *For. Econ. Rev.* **2019**, *1*, 57–70. [[CrossRef](#)]
- Li, M. Dynamic Monitoring Algorithm of Natural Resources in Scenic Spots Based on MODIS Remote Sensing Technology. *Earth Sci. Res. J.* **2021**, *25*, 57–64. [[CrossRef](#)]
- Balsamo, G.; Agusti-Panareda, A.; Albergel, C.; Arduini, G.; Beljaars, A.; Bidlot, J.; Blyth, E.; Bousserez, N.; Boussetta, S.; Brown, A.; et al. Satellite and In Situ Observations for Advancing Global Earth Surface Modelling: A Review. *Remote Sens.* **2018**, *10*, 2038. [[CrossRef](#)]
- Fisher, R.A.; Koven, C.D. Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. *J. Adv. Model. Earth Syst.* **2020**, *12*, e2018MS001453. [[CrossRef](#)]
- Kaplan, G.; Avdan, U. Monthly Analysis of Wetlands Dynamics Using Remote Sensing Data. *Int. J. Geo-Inf.* **2018**, *7*, 411. [[CrossRef](#)]
- Wen, J.; Liu, Q.; Xiao, Q.; Liu, Q.; You, D.; Hao, D.; Wu, S.; Lin, X. Characterizing Land Surface Anisotropic Reflectance over Rugged Terrain: A Review of Concepts and Recent Developments. *Remote Sens.* **2018**, *10*, 370. [[CrossRef](#)]
- Mu, X.; Hu, M.; Song, W.; Ruan, G.; Ge, Y.; Wang, J.; Huang, S.; Yan, G. Evaluation of Sampling Methods for Validation of Remotely Sensed Fractional Vegetation Cover. *Remote Sens.* **2015**, *7*, 16164–16182. [[CrossRef](#)]

21. Feng, W.; Boukir, S.; Huang, W. Margin-Based Random Forest for Imbalanced Land Cover Classification. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Yokohama, Japan, 2019; pp. 3085–3088.
22. Yang, Y.; Sun, X.; Diao, W.; Yin, D.; Yang, Z.; Li, X. Statistical Sample Selection and Multivariate Knowledge Mining for Lightweight Detectors in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5626414. [[CrossRef](#)]
23. Suh, N.P. A Theory of Complexity, Periodicity and the Design Axioms. *Res. Eng. Des.* **1999**, *11*, 116–132. [[CrossRef](#)]
24. Steffen, W.; Richardson, K.; Rockström, J.; Schellnhuber, H.J.; Dube, O.P.; Dutreuil, S.; Lenton, T.M.; Lubchenco, J. The Emergence and Evolution of Earth System Science. *Nat. Rev. Earth Environ.* **2020**, *1*, 54–63. [[CrossRef](#)]
25. Cheng, Q. Quantitative Simulation and Prediction of Extreme Geological Events. *Sci. China Earth Sci.* **2022**, *65*, 1012–1029. [[CrossRef](#)]
26. Lovejoy, S. The 2021 “Complex Systems” Nobel Prize: The Climate, with and without Geocomplexity. *AGU Adv.* **2022**, *3*, e2021AV000640. [[CrossRef](#)]
27. Ge, Y.; Jin, Y.; Stein, A.; Chen, Y.; Wang, J.; Wang, J.; Cheng, Q.; Bai, H.; Liu, M.; Atkinson, P.M. Principles and Methods of Scaling Geospatial Earth Science Data. *Earth-Sci. Rev.* **2019**, *197*, 102897. [[CrossRef](#)]
28. Jiang, H.; Shihua, L.; Dong, Y. Multidimensional Meteorological Variables for Wind Speed Forecasting in Qinghai Region of China: A Novel Approach. *Adv. Meteorol.* **2020**, *2020*, 5396473. [[CrossRef](#)]
29. Zhang, X.; Shi, W.; Lv, Z. Uncertainty Assessment in Multitemporal Land Use/Cover Mapping with Classification System Semantic Heterogeneity. *Remote Sens.* **2019**, *11*, 2509. [[CrossRef](#)]
30. Rufino, M.M.; Bez, N.; Brind’Amour, A. Ability of Spatial Indicators to Detect Geographic Changes (Shift, Shrink and Split) across Biomass Levels and Sample Sizes. *Ecol. Indic.* **2020**, *115*, 106393. [[CrossRef](#)]
31. Owers, C.J.; Rogers, K.; Woodroffe, C.D. Identifying Spatial Variability and Complexity in Wetland Vegetation Using an Object-Based Approach. *Int. J. Remote Sens.* **2016**, *37*, 4296–4316. [[CrossRef](#)]
32. Batty, M.; Morphet, R.; Masucci, P.; Stanilov, K. Entropy, Complexity, and Spatial Information. *J. Geogr. Syst.* **2014**, *16*, 363–385. [[CrossRef](#)]
33. Yanovski, R.; Nelson, P.A.; Abelson, A. Structural Complexity in Coral Reefs: Examination of a Novel Evaluation Tool on Different Spatial Scales. *Front. Ecol. Evol.* **2017**, *5*, 27. [[CrossRef](#)]
34. Zhang, Z.; Song, Y.; Luo, P.; Wu, P. Geocomplexity Explains Spatial Errors. *Int. J. Geogr. Inf. Sci.* **2023**, *37*, 1449–1469. [[CrossRef](#)]
35. Xie, H.; Tong, X.; Meng, W.; Liang, D.; Wang, Z.; Shi, W. A Multilevel Stratified Spatial Sampling Approach for the Quality Assessment of Remote-Sensing-Derived Products. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4699–4713. [[CrossRef](#)]
36. Li, G.; Gao, Q.; Yang, M.; Gao, X. Active Learning Based on Similarity Level Histogram and Adaptive-Scale Sampling for Very High Resolution Image Classification. *Neural Netw.* **2023**, *167*, 22–35. [[CrossRef](#)]
37. Buolamwini, J.; Gebru, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018.
38. Krishnan, R.; Sinha, A.; Ahuja, N.A.; Subedar, M.; Tickoo, O.; Iyer, R.R. Mitigating Sampling Bias and Improving Robustness in Active Learning. *arXiv* **2021**, arXiv:Abs/2109.06321.
39. Bhatt, U.; Antorán, J.; Zhang, Y.; Liao, Q.V.; Sattigeri, P.; Fogliato, R.; Melançon, G.; Krishnan, R.; Stanley, J.; Tickoo, O.; et al. Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, 19–21 May 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 401–413.
40. Li, W.; Chen, K.; Chen, H.; Shi, Z. Geographical Knowledge-Driven Representation Learning for Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5405516. [[CrossRef](#)]
41. Li, H.; Li, Y.; Zhang, G.; Liu, R.; Huang, H.; Zhu, Q.; Tao, C. Global and Local Contrastive Self-Supervised Learning for Semantic Segmentation of HR Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5618014. [[CrossRef](#)]
42. Lin, D.; Fu, K.; Wang, Y.; Xu, G.; Sun, X. MARTA GANs: Unsupervised Representation Learning for Remote Sensing Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2092–2096. [[CrossRef](#)]
43. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [[CrossRef](#)]
44. Li, C.; Ma, Z.; Wang, L.; Yu, W.; Tan, D.; Gao, B.; Feng, Q.; Guo, H.; Zhao, Y. Improving the Accuracy of Land Cover Mapping by Distributing Training Samples. *Remote Sens.* **2021**, *13*, 4594. [[CrossRef](#)]
45. Wagenaar, D.; Hermawan, T.; Van Den Homberg, M.J.C.; Aerts, J.C.J.H.; Kreibich, H.; De Moel, H.; Bouwer, L.M. Improved Transferability of Data-Driven Damage Models Through Sample Selection Bias Correction. *Risk Anal.* **2021**, *41*, 37–55. [[CrossRef](#)]
46. Zadrozny, B. Learning and Evaluating Classifiers under Sample Selection Bias. In Proceedings of the Twenty-First International Conference on Machine Learning—ICML’04, Banff, AB, Canada, 4–8 July 2004; ACM Press: Banff, AB, Canada, 2004; p. 114.
47. Boschetti, L.; Stehman, S.V.; Roy, D.P. A Stratified Random Sampling Design in Space and Time for Regional to Global Scale Burned Area Product Validation. *Remote Sens. Environ.* **2016**, *186*, 465–478. [[CrossRef](#)]
48. Wagner, J.E.; Stehman, S.V. Optimizing Sample Size Allocation to Strata for Estimating Area and Map Accuracy. *Remote Sens. Environ.* **2015**, *168*, 126–133. [[CrossRef](#)]
49. Zhang, J.; Zhang, W.; Mei, Y.; Yang, W. Geostatistical Characterization of Local Accuracies in Remotely Sensed Land Cover Change Categorization with Complexly Configured Reference Samples. *Remote Sens. Environ.* **2019**, *223*, 63–81. [[CrossRef](#)]

50. Meng, X.-L. Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election. *Ann. Appl. Stat.* **2018**, *12*, 685–726. [[CrossRef](#)]
51. Wang, J.; Haining, R.; Zhang, T.; Xu, C.; Hu, M.; Yin, Q.; Li, L.; Zhou, C.; Li, G.; Chen, H. Statistical Modeling of Spatially Stratified Heterogeneous Data. *Ann. Am. Assoc. Geogr.* **2024**, *114*, 499–519. [[CrossRef](#)]
52. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [[CrossRef](#)]
53. Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sens.* **2014**, *6*, 964–983. [[CrossRef](#)]
54. Ghorbanian, A.; Kakooei, M.; Amani, M.; Mahdavi, S.; Mohammadzadeh, A.; Hasanlou, M. Improved Land Cover Map of Iran Using Sentinel Imagery within Google Earth Engine and a Novel Automatic Workflow for Land Cover Classification Using Migrated Training Samples. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 276–288. [[CrossRef](#)]
55. Priyanka, N.S.; Lal, S.; Nalini, J.; Reddy, C.S.; Dell'Acqua, F. DIResUNet: Architecture for Multiclass Semantic Segmentation of High Resolution Remote Sensing Imagery Data. *Appl. Intell.* **2022**, *52*, 15462–15482. [[CrossRef](#)]
56. Wang, X.; Xiong, X.; Ning, C. Multi-Label Remote Sensing Scene Classification Using Multi-Bag Integration. *IEEE Access* **2019**, *7*, 120399–120410. [[CrossRef](#)]
57. Ilunga, M. Shannon Entropy for Measuring Spatial Complexity Associated with Mean Annual Runoff of Tertiary Catchments of the Middle Vaal Basin in South Africa. *Entropy* **2019**, *21*, 366. [[CrossRef](#)]
58. Guo, L.; Du, S.; Haining, R.; Zhang, L. Global and Local Indicators of Spatial Association between Points and Polygons: A Study of Land Use Change. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 384–396. [[CrossRef](#)]
59. Frigg, R.; Werndl, C. Entropy: A Guide for the Perplexed. In *Probabilities in Physics*; Beisbart, C., Hartmann, S., Eds.; Oxford University Press: Oxford, UK, 2011; pp. 115–142. ISBN 978-0-19-957743-9.
60. Li, L. Deep Residual Autoencoder with Multiscaling for Semantic Segmentation of Land-Use Images. *Remote Sens.* **2019**, *11*, 2142. [[CrossRef](#)]
61. Li, L.; Zhu, Z.; Wang, C. Multiscale Entropy-Based Surface Complexity Analysis for Land Cover Image Semantic Segmentation. *Remote Sens.* **2023**, *15*, 2192. [[CrossRef](#)]
62. Li, H.; Calder, C.A.; Cressie, N. Beyond Moran's *I*: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geogr. Anal.* **2007**, *39*, 357–375. [[CrossRef](#)]
63. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
64. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large Kernel Matters—Improve Semantic Segmentation by Global Convolutional Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Honolulu, HI, USA, 2017; pp. 1743–1751.
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; IEEE: Las Vegas, NV, USA, 2016; pp. 770–778.
66. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
67. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
68. Xiao, T.; Liu, Y.; Zhou, B.; Jiang, Y.; Sun, J. Unified Perceptual Parsing for Scene Understanding. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 432–448.
69. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010.
70. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Álvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In Proceedings of the Neural Information Processing Systems, Online, 6–14 December 2021.
71. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
72. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: San Francisco, CA, USA, 2016; pp. 785–794.
73. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [[CrossRef](#)]
74. Tong, X.-Y.; Xia, G.-S.; Zhu, X.X. Enabling Country-Scale Land Cover Mapping with Meter-Resolution Satellite Imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 178–196. [[CrossRef](#)]

75. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
76. Ma, Y.D.; Qing, L.; Qian, Z.B. Automated Image Segmentation Using Improved PCNN Model Based on Cross-Entropy. In Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004; pp. 743–746.
77. Jadon, S. A Survey of Loss Functions for Semantic Segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 27–29 October 2020; IEEE: Via del Mar, Chile, 2020; pp. 1–7.
78. Moran, P. artist Notes on Continuous Stochastic Phenomena. *Biometrika* **1950**, *37*, 17–23. [[CrossRef](#)]
79. Poudyal, N.C.; Butler, B.J.; Hodges, D.G. Spatial Analysis of Family Forest Landownership in the Southern United States. *Landsc. Urban Plan.* **2019**, *188*, 163–170. [[CrossRef](#)]
80. Colditz, R. An Evaluation of Different Training Sample Allocation Schemes for Discrete and Continuous Land Cover Classification Using Decision Tree-Based Algorithms. *Remote Sens.* **2015**, *7*, 9655–9681. [[CrossRef](#)]
81. Liu, Z.; Pontius, R.G., Jr. The Total Operating Characteristic from Stratified Random Sampling with an Application to Flood Mapping. *Remote Sens.* **2021**, *13*, 3922. [[CrossRef](#)]
82. Shetty, S.; Gupta, P.K.; Belgiu, M.; Srivastav, S.K. Assessing the Effect of Training Sampling Design on the Performance of Machine Learning Classifiers for Land Cover Mapping Using Multi-Temporal Remote Sensing Data and Google Earth Engine. *Remote Sens.* **2021**, *13*, 1433. [[CrossRef](#)]
83. Cheng, J.; Deng, C.; Su, Y.; An, Z.; Wang, Q. Methods and Datasets on Semantic Segmentation for Unmanned Aerial Vehicle Remote Sensing Images: A Review. *ISPRS J. Photogramm. Remote Sens.* **2024**, *211*, 1–34. [[CrossRef](#)]
84. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.