remote sensing

MDPI

Article

Dual-Task Network for Terrace and Ridge Extraction: Automatic Terrace Extraction via Multi-Task Learning

Jun Zhang [1], Jun Zhang [1,*], Xiao Huang [2], Weixun Zhou [3], Huyan Fu [1], Yuyan Chen [1] and Zhenghao Zhan [1]

[1] School of Earth Science, Yunnan University, Kunming 650000, China; zhjun@mail.ynu.edu.cn (J.Z.); fuhuyan@ynu.edu.cn (H.F.); chenyuyan@mail.ynu.edu.cn (Y.C.); zhan-zhenghao@mail.ynu.edu.cn (Z.Z.)
[2] Department of Environmental Sciences, Emory University, Atlanta, GA 30322, USA; xiao.huang2@emory.edu
[3] School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; zhouwx@nuist.edu.cn
* Correspondence: zhjun@ynu.edu.cn; Tel.: +86-137-5957-4184

Abstract: Terrace detection and ridge extraction from high-resolution remote sensing imagery are crucial for soil conservation and grain production on sloping land. Traditional methods use low-to-medium resolution images, missing detailed features and lacking automation. Terrace detection and ridge extraction are closely linked, with each influencing the other's outcomes. However, most studies address these tasks separately, overlooking their interdependence. This research introduces a cutting-edge, multi-scale, and multi-task deep learning framework, termed DTRE-Net, designed for comprehensive terrace information extraction. This framework bridges the gap between terrace detection and ridge extraction, executing them concurrently. The network incorporates residual networks, multi-scale fusion modules, and multi-scale residual correction modules to enhance the model's robustness in feature extraction. Comprehensive evaluations against other deep learning-based semantic segmentation methods using GF-2 terraced imagery from two distinct areas were undertaken. The results revealed intersection over union (IoU) values of 85.18% and 86.09% for different terrace morphologies and 59.79% and 73.65% for ridges. Simultaneously, we have confirmed that the connectivity of results is improved when employing multi-task learning for ridge extraction compared to directly extracting ridges. These outcomes underscore DTRE-Net's superior capability in the automation of terrace and ridge extraction relative to alternative techniques.

Keywords: multi-task learning; terrace information extraction; neural networks; high-resolution remote sensing images

check for updates

Citation: Zhang, J.; Zhang, J.; Huang, X.; Zhou, W.; Fu, H.; Chen, Y.; Zhan, Z. Dual-Task Network for Terrace and Ridge Extraction: Automatic Terrace Extraction via Multi-Task Learning. *Remote Sens.* **2024**, *16*, 568. https://doi.org/10.3390/rs16030568

Academic Editor: John Trinder

Received: 12 December 2023
Revised: 29 January 2024
Accepted: 30 January 2024
Published: 1 February 2024

1. Introduction

Terraced land, recognized as a primary form of cultivated terrain [1], holds considerable importance in enhancing agricultural production diversity, countering soil erosion and degradation, preserving essential agricultural water resources, reducing runoff [2,3], promoting biodiversity [4], and augmenting its ecological and cultural significance. However, the complex and dynamic topography of terraced fields makes traditional manual surveying methods both time-consuming and labor-intensive. As a result, the integration of remote sensing technology with artificial intelligence has been identified as an effective and accurate approach for terrace extraction [5,6].

Terrace mapping through remote sensing primarily utilizes satellite and aerial imagery data. This process involves manually curated feature extraction and classification algorithms based on attributes like color, texture, and shape. Established methodologies include visual interpretation [7,8], texture-spectral techniques [9,10], and object-based classification [11,12]. Although traditional manually crafted features possess clear meanings and interpretable mathematical formulas, these types of features overly rely on the accumulation of expert knowledge, leading to limitations in achieving high recognition performance and excellent generalization capabilities [13]. As spatial resolution heightens, so does the semantic detail related to features, such as boundary definition and the spatial organization of diverse terrain attributes. This enhancement intensifies the data's intricacy, leading to frequent instances of "spectrally similar but distinct objects" [14]. For example, terraced fields, bare lands, and sloping cultivated areas display comparable spectral traits but possess distinct contextual attributes [15]. Furthermore, traditional methodologies are generally tailored to particular scenarios, limiting their adaptability across varied regions and diverse terraced landscapes. This specificity results in challenges like reduced extraction precision and limited repeatability [16]. Such constraints potentially compromise the efficacy and dependability of conventional techniques when applied to high-resolution imagery.

Deep learning, a contemporary learning approach, possesses the inherent ability to autonomously derive complex feature representations directly from unprocessed images [17]. This attribute renders it particularly adept at managing the intricacies and variances inherent in high-resolution datasets. Within the realm of remote sensing image analysis, deep learning demonstrates exceptional adaptability and precision, with extensive research and application focusing on semantic segmentation [18] and object detection [19,20]. In the field of image classification, semantic segmentation models are represented by Convolutional Neural Networks (CNNs). Their versatility has spurred applications across diverse fields, encompassing structures [21,22], road networks [23], agricultural lands [24,25], and aquatic zones [26,27], thereby establishing a robust groundwork for feature discernment and terrace delineation. In contemporary applications, deep learning models proficiently discern the peripheral profiles and overarching attributes of terraced landscapes [28–30]. Academic endeavors have culminated in high-accuracy extraction and categorization of terraced fields through the utilization of these advanced deep learning frameworks [31]. Terraced landscapes, given their intricate geographical nature, often manifest a range of characteristics influenced by diverse factors. These include terraces optimized for water retention or those impacted by drought conditions. Such morphological variations can lead to multiple internal disturbances within terraced fields, such as the presence of vegetation, rocks, and flowing water. These disturbances can considerably hinder the precise extraction capabilities of deep learning models. To mitigate the effects of such internal noise, several researchers have suggested integrating Digital Elevation Models (DEMs) with high-resolution remote sensing imagery [32,33]. However, a persisting challenge arises from terraced fields, bare lands, and sloping cultivated terrains often having analogous spectral properties, yet they exhibit contrasting contextual nuances. Furthermore, the variability in the sizes of terraced fields across different geographical regions introduces significant complexities for deep learning applications in terrace delineation. For instance, the dimensions of terraced fields can differ substantially between regions, adding layers of complexity to their identification. In response to these challenges, scholars have explored a plethora of strategies to amplify the precision and adaptability of deep learning algorithms. Noteworthy among these are the integration of attention mechanisms into conventional CNN structures [34], meticulous parameter optimization techniques [35], and the conceptualization of multi-scale feature extraction frameworks [36]. Collectively, these innovations aim to bolster the extraction accuracy of terraced landscapes.

In terrace research, besides the extraction of the terraces themselves, the pertinent information concerning terrace ridges holds paramount significance. The relationship between terraces and ridges is somewhat symbiotic: the identification of terraced fields dictates the manifestation of field ridges, while ridge extraction accentuates the terraces' geometric features. Nevertheless, the majority of contemporary research predominantly concentrates on the individual extraction of either terraced fields or terrace ridges, neglecting their intrinsic interrelation. This singular focus hampers a comprehensive grasp of the terraced landscape. Within the domain of remote sensing image analysis, multi-task learning, which simultaneously addresses multiple interrelated tasks, has been extensively employed for diverse land feature extractions [37]. For instance, in urban settings, multi-task learning has been pivotal in simultaneously extracting roads and their centerlines [38], with shared information between tasks mitigating challenges arising from scarce road centerline data.

In the context of urban building classification, the utilization of a multi-task learning modeling approach with five interdependent building labels consistently demonstrates superior accuracy and efficiency compared to both single-task learning and classical hard parameter sharing methods [39]. In agrarian contexts, prior studies utilizing multi-task deep convolutional neural networks have showcased marked advancements in delineating agricultural perimeters, field expanses, and cropping patterns [40,41]. Contrastingly, in the realm of terrace research, the potential of multi-task learning remains largely untapped. Hypothetically, by promoting information interchange and parameter consolidation between terraced field identification and ridge extraction, there is an opportunity to curtail the requisite training samples, diminish overfitting tendencies, and amplify the model's overarching adaptability and precision in extraction tasks.

To address the challenges previously highlighted, this research initially curates a dataset representing terraced fields during their fallow phase, utilizing Gaofen-2 (GF-2) satellite remote sensing imagery. Following this, we proposed a dual-task network for terrace and ridge extraction (DTRE-Net), a sophisticated multi-task, multi-scale framework devised explicitly for the dual purpose of terrace detection and ridge extraction.

The primary contributions of this manuscript are delineated as:

(1) The study meticulously assembles two distinct datasets of terraced fields utilizing GF-2 satellite imagery, one showcasing terraced fields and their corresponding ridge samples during the fallow phase in both water-retentive and dry states.

(2) In response to the inherent challenges of internal noise and varied dimensions, we put forth DTRE-Net. This dual-task semantic segmentation model incorporates cavity convolutions, a multi-scale feature fusion module, and a residual correction component. Empirical evaluations underscore DTRE-Net's superior efficacy in terrace detection and ridge extraction relative to contemporary methodologies.

The organization of this manuscript is structured as follows. Section 2 elucidates the methodology behind data procurement and the intricacies of dataset formulation. Section 3 delves into the nuanced architecture of the multi-tasking DTRE-Net framework. Section 4 showcases the experimental outcomes and offers a thorough comparative evaluation against alternative approaches. Section 5 scrutinizes the efficacy of both single- and dual-task performances, critically assessing the proposed modules across diverse scenarios. Section 6 synthesizes the pivotal insights of this research and outlines potential trajectories for subsequent research endeavors and enhancements.

2. Materials

2.1. Experimental Area

The terraced farming cycle is bifurcated into two primary phases: the fallow and planting periods. Throughout the planting phase, terraced fields showcase a plethora of crop covers, distinctive attributes linked to varying crop growth stages, and seasonal fluxes. Contrarily, during the fallow phase, these fields are devoid of crop coverage, streamlining their identification and extraction process. Within this fallow duration, terraced fields predominantly manifest in two configurations: the flat fallow and the water-storing fallow. The former pertains to terraced terrain that undergoes artificial leveling in the fallow span to facilitate soil rejuvenation and uphold its fertility. Conversely, water-storing fallow designates terraced fields purposed for water conservation and irrigation in the fallow phase, aiming to sustain soil hydration and enrich its composition. Notably, neither of these fallow configurations engages in crop cultivation.

For the purposes of this research, we designated two experimental terraced sites, each representative of these distinct fallow configurations. The first, located in Potou Township, Jianshui County, Honghe Prefecture, Yunnan Province, exemplifies leveling fallow (subsequently denoted as the T1 area). The second site, situated in Xinjie Township of Yuanyang County, embodies storaged fallow (henceforth labeled as the T2 area). The precise geographical coordinates of these experimental locales, accompanied by their corresponding regional remote sensing depictions, are presented in Figure 1.

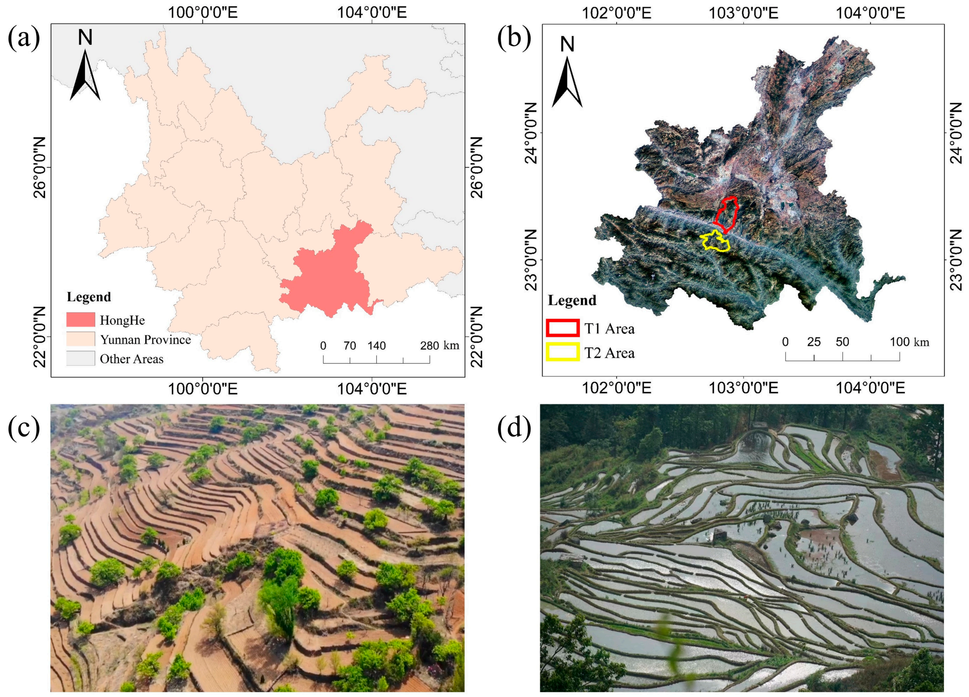

Figure 1. Geographic location and terraced field morphology of the experimental area. (**a**) location map of the experimental area; (**b**) remote sensing image of the experimental area; (**c**) terraced fields in the level fallow state; (**d**) terraced fields in the water storage fallow state.

The T1 region is distinguished by its diverse terraced field typologies, subtle inclinations, pronounced traces of human interventions, and varied land utilizations, encompassing regular farmlands, sloping agricultural terrains, infrastructures, and aquatic expanses. This area stands out as a multifaceted and emblematic subject for investigation. During its fallow phase, the majority of terraces in this vicinity predominantly exhibit a desiccated natural state, although some fields sporadically support scant herbaceous vegetation. The human engagements, coupled with the visual congruence of these terraced fields to sloping cultivated and barren lands, render the task of differentiating terraced fields from other terrains in remote sensing imagery notably intricate. This complexity is further accentuated by the slender dimensions of the field bunds and the terraced fields' muted gradient.

The T2 region is distinguished by its single terraced field type, steep slopes, and small, closely clustered individual field sizes. During the fallow period, the terraced fields in this area undergo artificial irrigation using a canal system, which fosters the proliferation of algae, giving rise to a vivid green appearance in remote sensing imagery. Simultaneously, during the water storage period, the spectral characteristics of the field bunds differ significantly from the terraced fields, with the field bunds in this area typically appearing as white linear features in remote sensing images.

2.2. Dataset Creation

Deep learning models yield optimal accuracy when underpinned by an exhaustive dataset. The model's performance is intrinsically tied to both the volume and precision of the samples [42]. In this research endeavor, we harnessed imagery from the Gaofen-2 (GF-2) satellite, encapsulating two divergent terraced field scenarios, culminating in the creation of multi-task terraced field datasets. These datasets bifurcate into two pivotal sub-datasets:

(1) Terraced field detection datasets, designated for the exploration and appraisal of methodologies extracting terraced field boundaries.

(2) Field bund extraction datasets, tailored for the investigation and assessment of methods pinpointing field bund localities.

The GF-2 satellite, commissioned on 19 August 2014, is outfitted with a state-of-the-art 1-m panchromatic camera and a 4-m multispectral camera. It boasts attributes like high radiometric fidelity, meticulous geolocation, and swift attitude adjustment capabilities, among others. For our analytical pursuits, we cherry-picked two GF-2 imagery segments, captured on 29 December 2019, characterized by stellar data integrity and minimal cloud interference (below 1%). The detailed steps for constructing the sample set are illustrated in Figure 2.

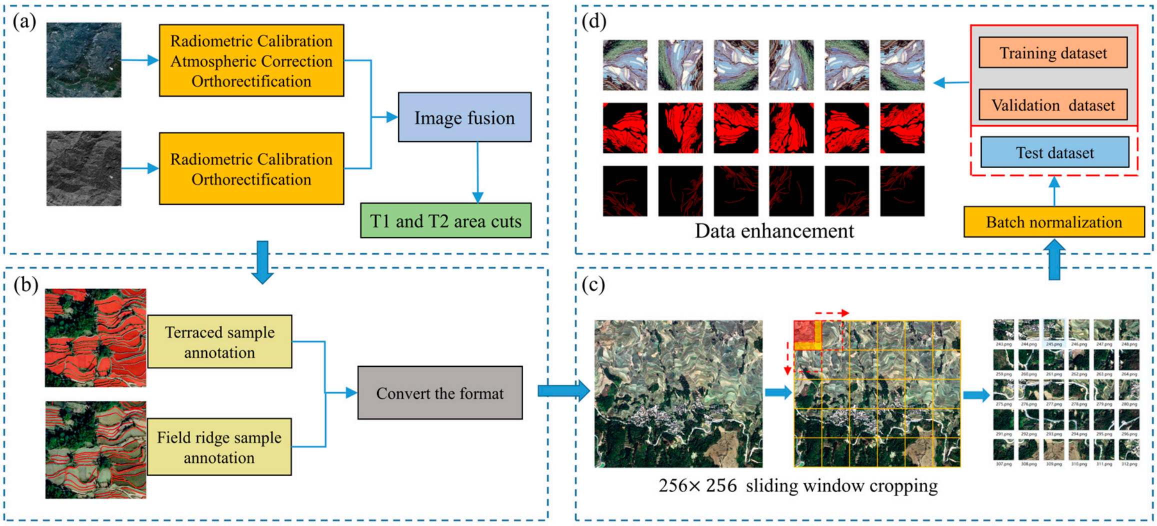

Figure 2. The workflow of data preprocessing and sample set generation: (**a**) image preprocessing and cropping; (**b**) labeling of terraces and ridges; (**c**) sliding window cropping scheme; (**d**) data segmentation and enhancement.

Step 1: Data preprocessing. Before conducting the experiment, image preprocessing was performed using ENVI 5.6 software, which included procedures such as radiometric calibration, atmospheric correction, and orthorectification. Subsequently, the Gram–Schmidt technique was employed to fuse multispectral and panchromatic data, resulting in a four-band image dataset with a spatial resolution exceeding 1 m. To address concerns related to model overfitting and limited generalization ability due to an abundance of non-terraced regions and imbalanced samples, we carefully identified terraced areas within the T1 and T2 regions characterized by high terrace density and diverse backgrounds featuring elements like vegetation, buildings, water bodies, and more. The raster sizes for these areas were 10,324 × 6616 and 7936 × 8455, respectively.

Step 2: Terraced and field bund data annotation. Using ArcGIS 10.7 software, high-resolution remote sensing images were visually interpreted and manually annotated based on the features of terraces and ridges, resulting in the generation of ground truth label images for the corresponding areas. The terrace and ridge identification tasks were treated as one or two binary classification problems, with the target pixel values set to 255 and marked in red (RGB(255, 0, 0)). Simultaneously, background pixel values were set to 0 and marked in black (RGB(0, 0, 0)). Finally, the Feature to Raster tool in ArcGIS 10.7 was employed to convert the annotations into terrace label data.

Step 3: Sample cropping. Considering computer hardware limitations, both images and labels required cropping before input. In this study, we employed a sliding window cropping strategy with a size of 256 × 256 pixels. The horizontal and vertical step sizes were set to 192 pixels.

Step 4: Data partitioning and enhancement. The datasets from the T1 and T2 regions, encompassing terraces and ridges, were systematically segregated into training and validation subsets, adhering to an 8:1:1 ratio. To curtail the potential of overfitting and bolster the model's generalizability, geometric augmentations were executed on both training and validation sets. This was accomplished without altering the intrinsic content or relational dynamics of the features within the images. The augmentation procedures encompassed clockwise rotations (specifically, 90°, 180°, and 270°) and both horizontal and vertical mirror transformations. Post augmentation, the dataset expanded to encompass 10,188 samples from the T1 region and 7128 samples from the T2 region. Concurrently, all image inputs and corresponding labels underwent normalization prior to being fed into the network.

3. Methods

Multi-task learning enables learning multiple related subtasks in parallel while sharing knowledge during the learning process. The relationship between subtasks can improve the model's performance and generalization in comparison to single-task learning. The framework utilized in this study, DTRE-Net, employs multi-task learning as depicted in Figure 3. It maintains the parameter sharing model typical of multi-task learning, where top-level parameters are not independent. Leveraging the inherent interrelation between the two subtasks, DTRE-Net synthesizes supplementary interaction data. In the final stages, separate branches yield the results for both subtasks. Subsequent sections will offer a detailed exposition of the module design nuances and the overall architecture of the DTRE-Net.

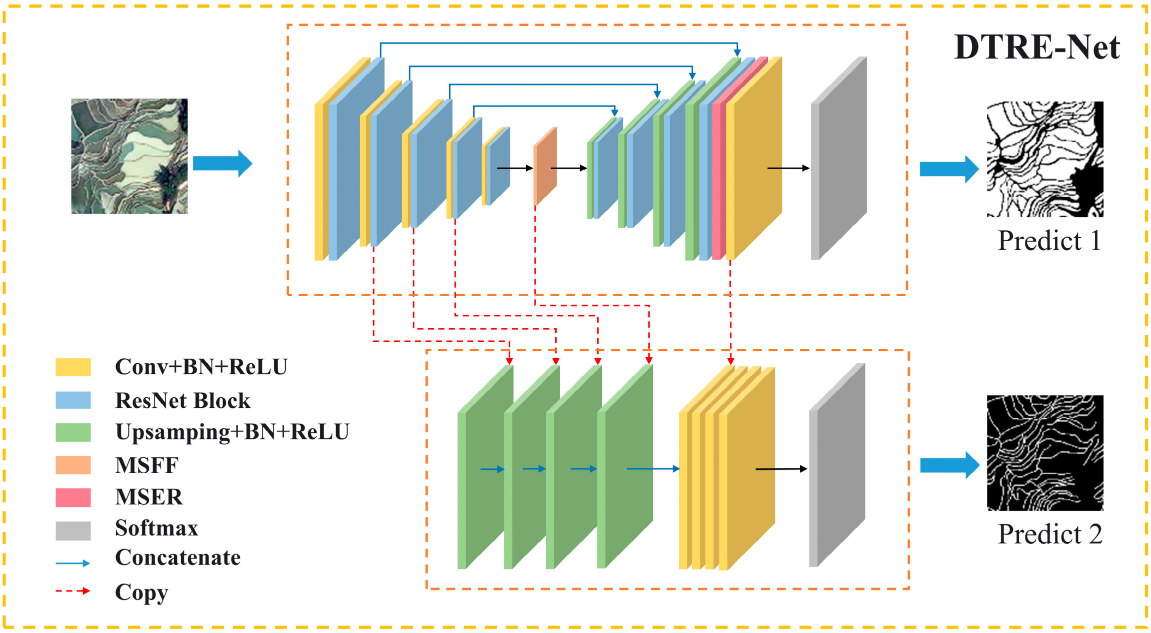

Figure 3. DTRE-Net structure.

3.1. DTRE-Net Architecture

Figure 3 illustrates the architectural configuration of the DTRE-Net, comprising two networks: the terrace detection network in the upper segment and the ridge extraction network in the lower segment. These networks are rigorously synchronized during the training process. The encoder component of the terrace detection network is characterized by the presence of five sets of alternately interleaved convolutional and residual layers, establishing a profound and high-performance neural network. Employing a 256 × 256 × 4 input image, a 32-channel feature map is crafted through the utilization of 32 convolutional layers, employing 1 × 1 convolutional kernels. Subsequent to this, a sequence of four convolutional operations is executed, employing 3 × 3 convolution kernels, effectively doubling the quantity of convolution kernels relative to the preceding convolutional layer. Notably, these operations employ a stride of 2 in both horizontal and vertical directions, resulting in feature maps with dimensions halved in comparison to the input feature maps. Following each convolutional operation, a vital Batch Normalization (BN) and

Rectified Linear Unit (ReLU) activation function process is enacted to expedite network convergence. Deeper within the network, a concerted endeavor to distill profound terraced features ensues, culminating in a feature map measuring 16 × 16, enriched with 512 feature channels, a feat accomplished after five successive convolution and residual operations. The decoder component, composed of four sets of up-sampling and residual modules, meticulously orchestrates the gradual restoration of low-resolution feature maps to their original dimensions. This is achieved through a series of designed up-sampling and feature fusion operations. Concurrently, the corresponding residual modules acquire the capability to discern residual maps generated during up-sampling, consequently refining the low-resolution feature maps into higher-resolution counterparts. This holistic approach preserves intricate details, thereby enhancing the performance of the network in segmentation and generation tasks.

The terrace extraction network constitutes the second half of the network. After performing an up-sampling operation, the output of the Multi-Scale Feature Fusion Module (MSFF) module and the output of the middle three residual layers of the encoder within the terrace detection network are scaled up to match the size of the original image, before being combined with the last convolutional layer in the decoder part. Then, following four convolution operations, each of which is succeeded by a BN layer and a ReLU operation, the last convolution is executed via a solitary 1 × 1 output channel to deduce the conclusive projection of ridge extraction. The connection serves to prevent the ridge extraction from relying too heavily on the terrace detection findings. The initial three residual layers are integrated to make up for any low-level intricacies that may have been overlooked in the terrace detection findings. The merging of various intermediate features at different scales enhances the ridge extraction outcomes. The ultimate ridge extraction network integrates varying levels of features utilizing functions to effectively and flexibly merge local and global information, resulting in superb network performance.

3.2. ResBlock

In pursuit of mitigating the potential loss of vital spatial information ensuing from the reduction in image dimensions following pooling, we incorporated a zero-convolution operation, expanding the receptive field of feature extraction without surging the parameter count or modifying the convolution center's positioning [43]. Drawing inspiration from the residual module concept [44], we crafted a residual module encompassing three parallel zero convolution branches, as delineated in Figure 4. The foremost branch deploys a convolution operation with an expansion factor of one, mirroring the operations of a standard 3 × 3 convolution, to capture localized features. In contrast, the succeeding branches employ sequential zero convolutions with expansion rates of three and nine, aimed at augmenting the receptive field of the convolution kernel. Such zero convolutions, characterized by diverse expansion rates within the residual module, are adept at extracting multi-scaled features. This design enhances the efficiency of terraced feature recognition and extraction, while circumventing superfluous data redundancies.

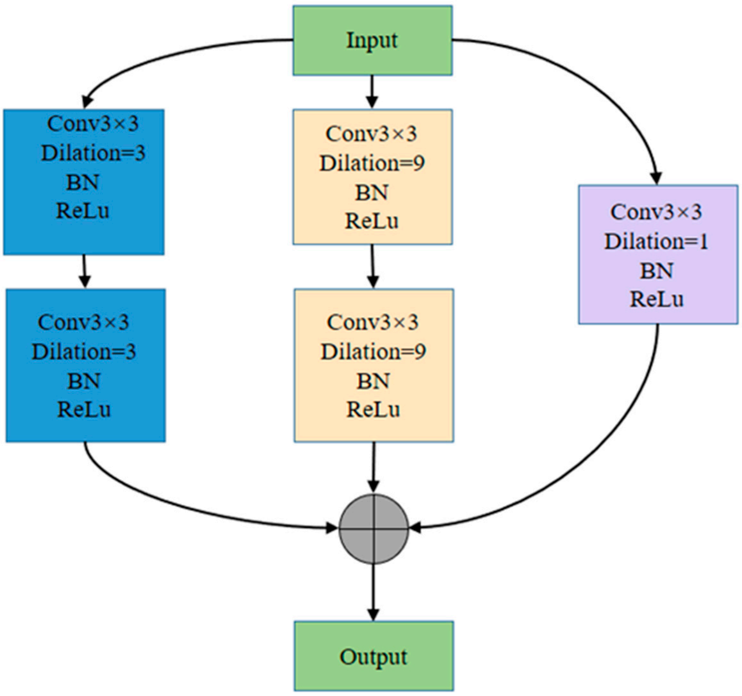

Figure 4. Structure of the ResBlock module.

3.3. Multi-Scale Feature Fusion Module

Terraces typically encompass a range of scale features, including large-scale structures, medium-scale fields or canals, and small-scale shapes and textures. To achieve precise terrace recognition and extraction, the model must accurately perceive and process these multi-scale features. The DTRE-Net encoder and decoder address this challenge by incorporating a MSFF in the central bridge connection, as illustrated in Figure 3. This module aggregates multi-scale and multi-level features from the encoder's output feature map, enhancing contextual information. As shown in Figure 5, this module achieves multi-scale feature fusion using null convolution and pooling operations with varying expansion rates, thereby improving network performance. Specifically, feature maps generated by the encoding network serve as inputs through four parallel branching network structures. The first branch employs three atrous convolutions [45] with expansion rates of 1, 2, and 3 to capture information at smaller scales. The second branch employs three atrous convolutions with expansion ratios of 1, 6, and 12 to expand the scope of feature information. The third and fourth branches utilize Average Pooling and Max Pooling, respectively, to gather global and local information, while up-sampling is employed to restore the input image size. Subsequently, the fused features are recombined, and the number of output feature maps is adjusted using a 1 × 1 convolutional layer. The integration of the multi-scale feature fusion module between the encoder and decoder enhances the network's efficiency in extracting terracing patterns. This integration further reinforces the network's ability to handle multi-scale terraced terrain features while mitigating information loss.

The MSFF module draws inspiration from the ASPP module. In contrast, concatenating multiple dilated convolutions with different rates proves more effective in acquiring diverse multi-scale features compared to a single dilated convolution layer. Simultaneously, this approach enlarges the receptive field without increasing the model parameter count, thereby enhancing the model's expressive capacity.

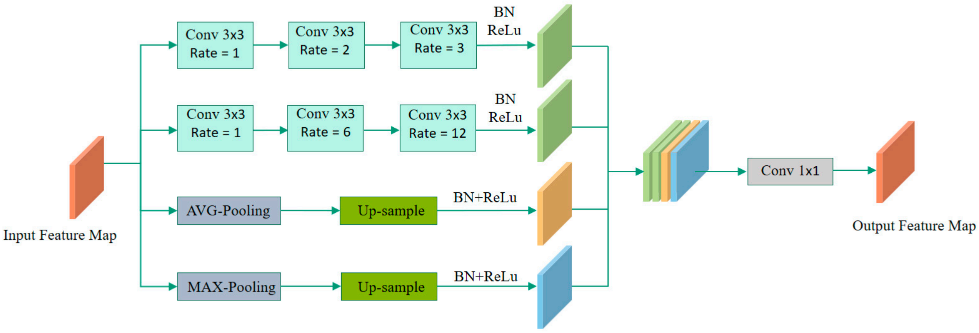

Figure 5. Structure of the MSFF module.

3.4. Multi-Scale Edge Residual Correction Model

The residual correction model models and corrects the differences between the predicted values and the actual observed values on the basis of the original model, thereby reducing prediction errors. This research presents the Multi-Scale Edge Residual Correction Model (MSER) with the primary aim of refining the accuracy of terrace extraction. The output from the decoder is channeled as the input for the MSER module. Drawing parallels with MSFF, the MSER utilizes a series of cascaded zero convolutions with 3 × 3 kernels, boasting expansion rates of 3, 6, 12, 24, and 48, as illustrated in Figure 6. Complementing this, residual links are incorporated to amalgamate feature maps across diverse scales. Each convolutional layer within the architecture is sequentially succeeded by BN and ReLU activation functions. When juxtaposed against the direct probability map output, the MSER facilitates a more profound feature extraction from the resultant prediction image. Moreover, the sequential deployment of cascaded convolutions systematically harnesses expansive global information, amalgamating data from multiple scales, which culminates in a notable enhancement in the precision of terrace identification.

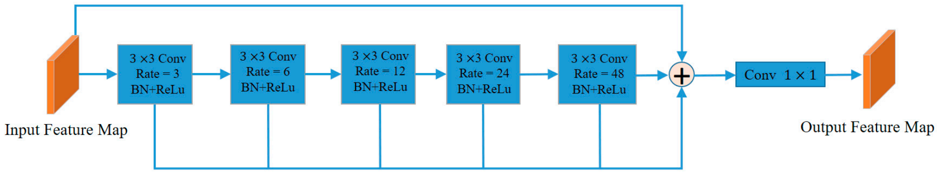

Figure 6. Structure of the MSER module.

The MSER module represents an improvement upon RRM_Lc [46]. It replaces standard convolutions with dilated convolutions, concurrently deepening the network. The gradual increase in the dilation rate allows for the progressive extraction of more global information and the fusion of multiscale information. This approach not only contributes to enhancing the accuracy of edge extraction in terraced fields but also facilitates obtaining more complete results in terraced field extraction.

3.5. Improved Binary Cross-Entropy Loss

Terrace detection and bund extraction constitute a binary semantic segmentation problem. The importance of terraces and bunds during the training process is unequal. To balance the training loss, the loss function is restructured by assigning weights to the losses of positive and negative samples, further improving the binary cross-entropy loss function. The function is defined as follows:

$$\mathrm{L_{BCE}} = \mathrm{W_1} \times (-\mathrm{y})\mathrm{logp}(\mathrm{y}) - \mathrm{W_2} \times (1 - \mathrm{y})\mathrm{log}(1 - \mathrm{p}(\mathrm{y})) \tag{1}$$

where y represents the binary label (0 or 1) and $\mathrm{p}(\mathrm{y})$ signifies the probability associated with the output being classified as label y, the binary cross-entropy loss function is employed. Specifically, when the predicted probability $\mathrm{p}(\mathrm{y})$ tends toward 1 for a given label y, the loss function approaches a value of 0. Conversely, when $p(y)$ approaches 0, the loss function assumes a significantly higher value. This behavior aligns with the inherent characteristics of the logarithmic function, which serves as a robust means of assessing the quality of predictions in a binary classification model, quantifying the extent to which they deviate from the ideal outcome. Here, W_1 and W_2 denote the proportions of target and background instances within the training dataset, respectively, calculated as follows:

$$W_1 = \frac{WS1}{WS1 + WS2} \tag{2}$$

$$W_2 = \frac{WS2}{WS1 + WS2} \tag{3}$$

where WS1 and WS2 represent the count of background pixels and target pixels in all training samples, respectively. This mechanism facilitates the automatic adjustment of weights, taking into account the distribution of positive and negative samples prior to training. Consequently, it ensures a relatively balanced contribution of loss from both classes.

Considering the disparate data distributions and importance between terrace and bund samples in the dataset, directly summing the losses of both tasks and optimizing through backpropagation has certain shortcomings in deriving the loss function for the dual-task model. Therefore, the overall loss function of DTRE-Net is a reweighted sum of losses, and the total loss is described as follows:

$$L_{MTL} = L_1 + \beta \times L_2 \tag{4}$$

$$\beta = \frac{\alpha_1}{\alpha_2} \tag{5}$$

where L_1 and L_2 are the binary cross-entropy loss function for the terrace and the ridge extraction task, respectively; β is the balance weight; α_1 is the number of terrace pixels; α_2 is the number of ridge pixels. The balancing weight is automatically determined through an analysis of positive samples from both tasks prior to training. Consequently, the proposed loss function is well suited for handling unbalanced datasets. Through the optimization of this unified loss function, feature information is systematically shared, thus fostering multi-task learning and enhancing the network's ability to effectively utilize training samples.

3.6. Evaluation Criterion

To evaluate the efficacy of various network modeling techniques, we employed a confusion matrix to juxtapose the terrace recognition outcomes against the ground truth labels. We introduced precision, recall, F1-score, and IoU to assess the performance of different networks on the dataset. The IoU is a commonly employed technique for assessing semantic segmentation results in images, quantifying the degree of overlap between predicted and actual outcomes. It serves as a pivotal tool for evaluating the likeness between predicted and ground truth segmentation results. The calculation of IoU follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{6}$$

The precision rate is defined as the ratio of correctly identified positive samples to all samples predicted as positive, as calculated according to Equation (7). The recall rate, on the other hand, measures the ratio of correctly forecasted positive samples within the entire sample, and its computation is outlined in Equation (8). The F1-score, which amalgamates precision and recall rates into a single metric, signifies the mean value and is illustrated in Equation (9).

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$\text{F1} - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

In the above formula, TP refers to the correct prediction of the positive class. FP indicates the model incorrectly predicting it as the positive class when it is actually the negative class. FN indicates the model incorrectly predicting it as the negative class when it is actually the positive class. These four evaluation metrics provide a comprehensive assessment of the extraction performance from various perspectives, with a maximum achievable value of 1 for all four indices. A higher value indicates a closer resemblance between the extracted terraces and the ground truth, leading to superior model outcomes [17].

4. Results

We established our training infrastructure on TensorFlow 2.4.0, executed on a Windows 11 operating system. The computing environment is anchored by a GV-N3090GAMING graphics card, complemented by 64 GB RAM, and powered by an I9-12900k processor, ensuring optimal performance and steadfast reliability. For the iterative refinement of our neural network's parameters, we adopted the Adam optimizer [47], which converges toward, or in certain instances reaches, the optimal solutions. The batch size for the experiment is set to eight, which is a choice aimed at fully harnessing the parallel processing prowess and efficiency of our GPU. Additionally, the initial learning rate is configured at 0.003, with the IoU serving as our primary metric to gauge discrepancies between our model's predictions and the actual ground truth. After each training cycle, both training and validation losses, along with IoU scores, are computed, and the most competent model across the iterations is preserved.

4.1. Comparison of Terrace Extraction

In this study, we conducted comparative experiments in both T1 and T2 regions to evaluate the performance of various models, including DTRE-Net, FCN [48], PSPNet [43], UNet [49], and DeepLabv3+ [45]. All models underwent training under identical conditions with standard parameters. Figures 7 and 8 provide visual comparisons between DTRE-Net and the other competing models in the T1 and T2 regions, showcasing the segmentation results for the test dataset, which includes typical terraced scenes labeled a–g in these figures.

All investigated models can broadly demarcate the terraced regions. In the T1 area, the simultaneous presence of terraces alongside sloping fields and barren land contributes to distinct instances of both false positives and negatives. Areas with moderate inclinations, exhibiting clearer distinctions and facilitating easier terrace extractions, tend to register fewer false positives, as exemplified in Figure 7e,f. Conversely, in regions with milder slopes, like that in Figure 7c, one can observe sporadic inconsistencies within the terrace plots. Particularly in zones where sloping terrains intersect with terraced fields, as portrayed in Figure 7b, incorrect predictions emerge prominently. In contrast, within the T2 area (Figure 8), once the terraces are inundated, their differentiation from barren and cultivated lands becomes more pronounced, resulting in a generally enhanced performance relative to the T1 zone. However, given the more compact terrace patches in this region, the edge predictions of terraces manifest suboptimal quality, leading to potential inaccuracies in demarcating boundaries against non-terraced zones, as shown in Figure 8b–d.

Across both regions, all five models are able to trace the contours of the terraces. From a visual perspective, the DTRE-Net model exhibits relatively better integrity for the interior of terraces, and there is an improvement in the phenomenon of erroneous extraction. Particularly for terraces in the T1 region, the occurrence of internal "wormhole" artifacts is noticeably reduced.

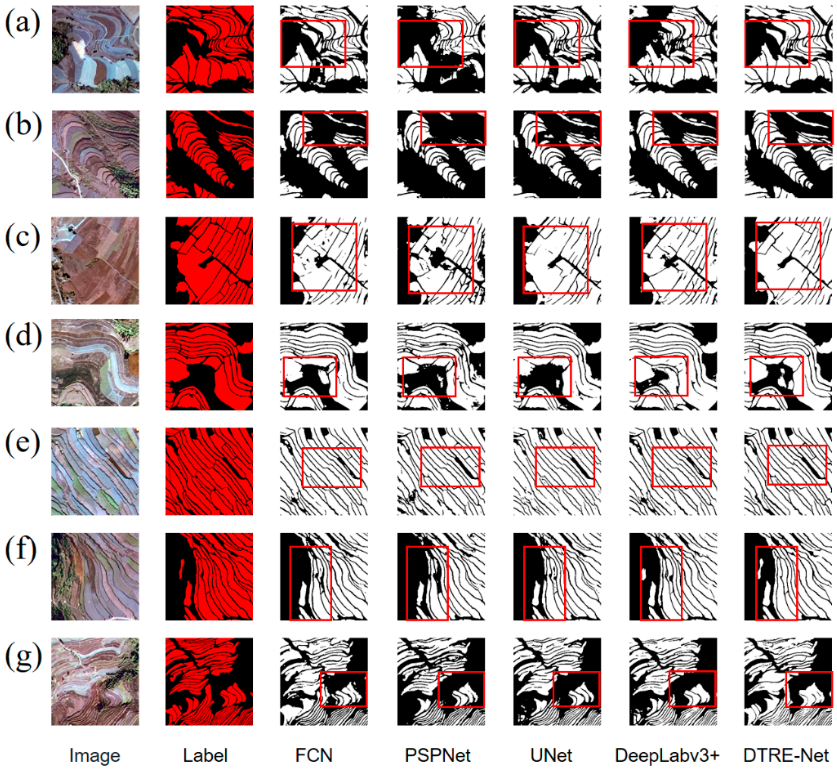

Figure 7. Visual comparison of extraction results from the investigated models in terraced T1 area: (**a**–**g**) showcase typical scenes of terraced fields with smooth fallow cultivation, with the red square highlighting the differences among them.

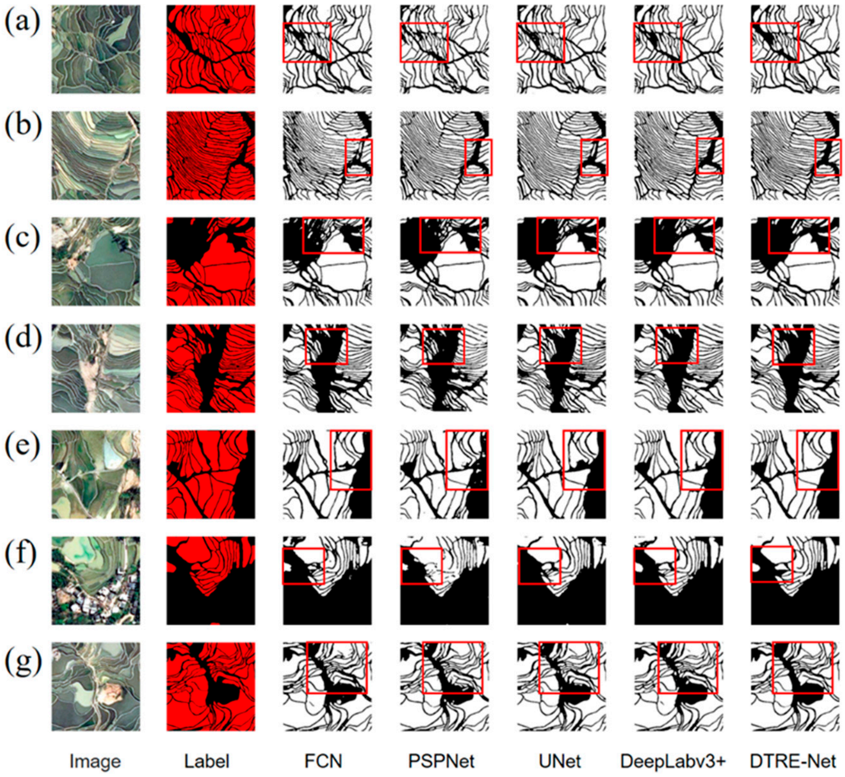

Figure 8. Visual comparison of extraction results from the investigated models in terraced T2 area: (**a**–**g**) showcase typical scenes of terraced fields with water storage and fallow cultivation, with the red square highlighting the variations among them.

Table 1 presents the numerical performance of various models on the T1 and T2 datasets, encompassing metrics such as precision, recall, F1-score, and IoU. Across both T1 and T2 regions, DTRE-Net demonstrates superior performance, with precision, recall, and F1-score all exceeding 90%, and IoU metrics surpassing 85%. FCN and PSPNet exhibit comparable performances, showcasing high precision but relatively lower IoU values. Compared to Deeplabv3+, UNet demonstrates superior performance in the T1 region, while its performance lags behind Deeplabv3+ in the T2 region. Moreover, all networks exhibit higher precision values in the T2 region compared to the T1 region, with DTRE-Net showing a notable improvement in precision compared to other networks. In summary, through a comprehensive analysis of precision evaluation and visual interpretation in both the T1 and T2 regions, DTRE-Net emerges as the standout performer, exemplified by its elevated levels of IoU and recall. DTRE-Net excels in performance across both regions, establishing itself as the overall top-performing model.

Table 1. Terrace classification evaluation for each experimental model in areas T1 and T2.

Area	Methods	Precision (%)	Recall (%)	F1-score (%)	IoU (%)
T1	FCN	88.73	85.84	87.27	77.41
	PSPNet	88.79	84.58	86.63	75.42
	UNet	88.65	90.57	89.60	81.12
	DeepLabv3+	89.98	85.75	87.81	78.27
	DTRE-Net	93.35	91.83	92.58	85.18
T2	FCN	85.21	91.30	88.15	78.81
	PSPNet	85.12	87.04	86.07	75.54
	UNet	86.97	91.74	89.29	80.65
	DeepLabv3+	87.51	92.27	89.82	81.53
	DTRE-Net	91.43	93.65	92.53	86.09

The experimental results on extracting terrace fields in arid level and water-stored level landscapes reveal some notable trends. For arid level terraces, observations indicate a propensity to blend with sloped cultivated land, leading to significant noise interference in the extraction results and consequently resulting in incomplete identification of terraced plots. Conversely, for water-stored terraces, the extraction outcomes exhibit distinct and complete features due to the relatively even water surfaces. Nevertheless, it is worth noting that the extraction results for such terraces may be influenced by the water content of the plots, introducing a certain degree of error. Although the precision metrics for terrace extraction in the T1 and T2 regions show minimal differences, the visual effectiveness is more pronounced in the T2 region.

4.2. Comparison of Field Ridge Extraction

To validate the performance of the DTRE-Net model in terrace bund extraction, Figures 9 and 10 present visual comparisons of the predicted terrace bunds in the T1 and T2 regions, respectively. Upon visual inspection of the T1 region, the intricate environmental backdrop of the terrace fields appears to compromise the overall precision of terrace bund extraction. This suboptimal performance is particularly pronounced in regions characterized by milder inclines or predominantly planar landscapes. Notably, maintaining the continuity of terrace bunds emerges as a significant challenge, as depicted in Figure 9c,d. Among the models evaluated, both UNet and DeepLabv3+ demonstrate a relatively superior capability in preserving bund continuity in specific sectors, as evidenced in Figure 9g. The proposed DTRE-Net model exhibits enhanced proficiency in rectifying prediction errors. This can be attributed to the strategic incorporation of the MSER module during the terminal phases of the network. This module facilitates real-time corrections during the prediction phase, anchored by the ground truth, while also accentuating the network's focus on contextual nuances.

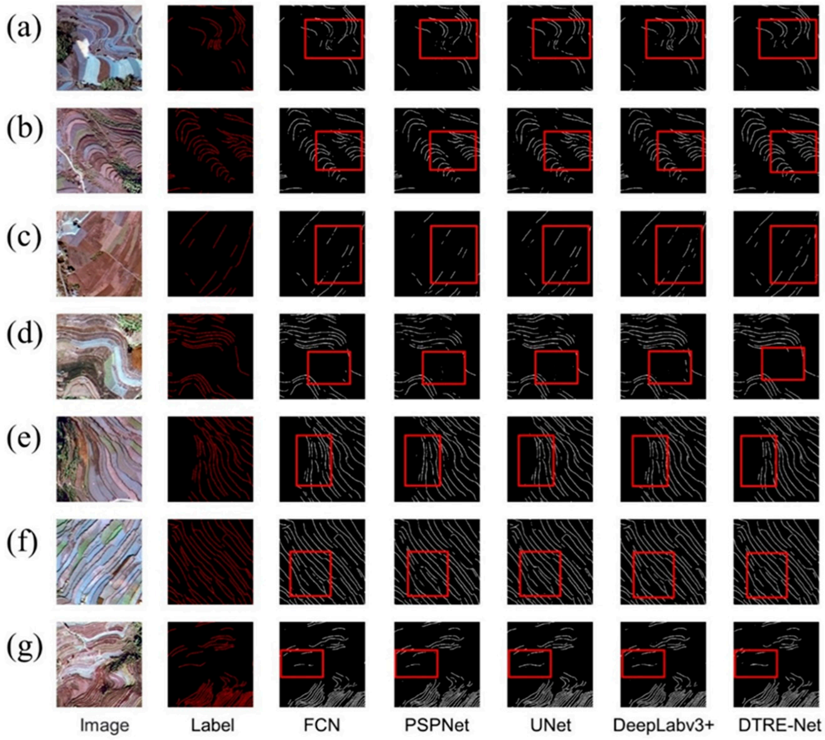

Figure 9. Visual comparison of extraction results of terrace ridge in T1 area under different networks: (**a**–**g**) showcase typical scenes of terraced fields with smooth fallow cultivation, with the red square highlighting the differences among them.

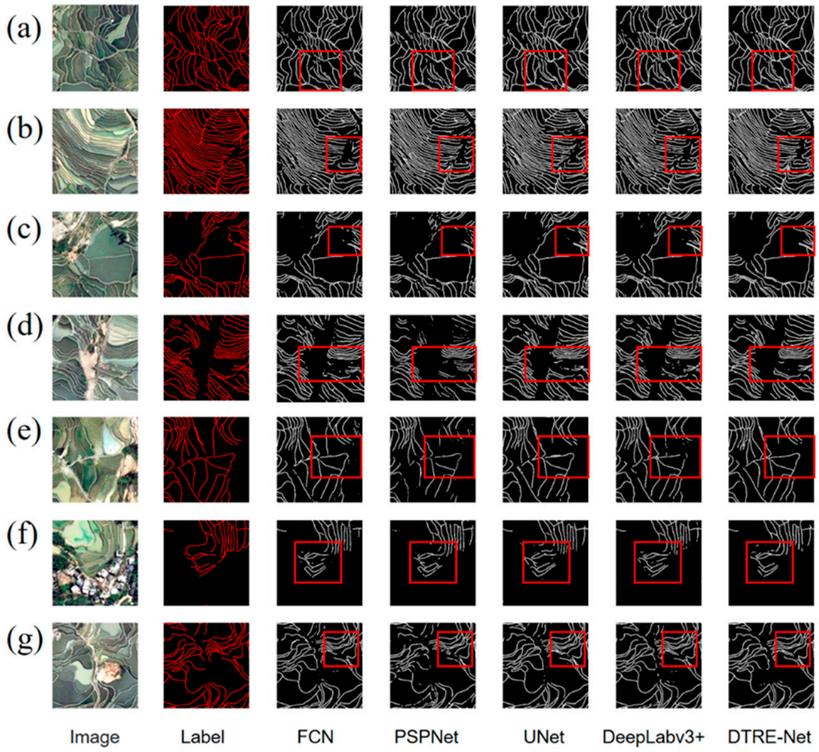

Figure 10. Visual comparison of extraction results of terrace ridge in T2 area under different networks: (**a**–**g**) showcase typical scenes of terraced fields with water storage and fallow cultivation, with the red square highlighting the variations among them.

Within the T2 region, characterized by distinct terrace bund textures, the models generally exhibit enhanced performance. However, in locales where terrace bunds are obfuscated by vegetative shadows, as depicted in Figure 10c, the DTRE-Net model does not classify them as terrace bunds. In contrast, the remaining four networks exhibit a tendency to partially categorize these shadows as bunds. Notably, UNet demonstrates a superior prediction capability in this domain. This can be ascribed to the restricted learning of analogous scenarios facilitated by the MSFF module, thereby compromising its prediction accuracy. Overall, although DTRE-Net may not invariably eclipse other networks across diverse environments, its prowess in achieving superior completeness and precision is notably superior to alternative models.

Table 2 presents the quantitative comparison results of the proposed method in this paper with FCN, PSPNet, UNet, and DeepLabv3+ networks in two different regions. From the experimental metrics, DTRE-Net consistently achieves superior outcomes in terrace testing across both regions. In the T1 region, DTRE-Net demonstrates precision, recall, F1-score, and IoU values of 70.71%, 79.48%, 74.84%, and 59.79%, respectively. In the T2 region, DTRE-Net outperforms with precision, recall, F1-score, and IoU values of 83.72%, 85.96%, 84.83%, and 73.65%. In the T1 region, methods other than DTRE-Net exhibit similar performance, with minimal differences in metric values. In the T2 region, UNet and DeepLabv3+ achieve favorable results, with IoU values of 66.94% and 67.88%, respectively. Overall, in the terrace bund extraction task, when comparing T2 to T1, all metrics for the methods, except for a minor difference in recall, show differences exceeding 10%. This indicates that the type of terracing significantly influences the results of terrace bund extraction.

Table 2. Terrace ridge classification evaluation for each experimental model in areas T1 and T2.

Area	Methods	Precision (%)	Recall (%)	F1-score (%)	IoU (%)
T1	FCN	60.17	75.84	67.10	50.49
	PSPNet	60.18	72.17	65.63	48.85
	UNet	61.74	74.27	67.43	50.87
	DeepLabv3+	63.14	77.01	69.39	51.12
	DTRE-Net	70.71	79.48	74.84	59.79
T2	FCN	75.48	82.12	78.66	64.82
	PSPNet	72.39	79.52	75.79	61.01
	UNet	77.78	82.79	80.20	66.94
	DeepLabv3+	78.36	83.54	80.87	67.88
	DTRE-Net	83.72	85.96	84.83	73.65

For terrace bund extraction, there is a significant disparity observed in the experimental comparison between flat-level and water-stored level terraces. In the T1 and T2 regions, there are evident differences in both the visual contrasts and precision metrics for terrace bund extraction. In the case of flat-level terraces in the T1 region, the terrace bund extraction results are less satisfactory, particularly in areas with gentle slopes, where intermittent discontinuities are observed. Conversely, for water-stored level terraces, the terrace bund extraction outcomes are more comprehensive, effectively distinguishing them from the terraced fields. This indicates that different topographical features, particularly in arid and water-stored scenarios, influence the model's performance in terrace bund extraction tasks.

5. Discussion

5.1. Single Tasking versus Dual Tasking

A comparative analysis was undertaken to compare the performance of single-task and multi-task models. Table 3 delineates the outcomes of this comparison, underscoring that the dual-task network markedly bolsters terrace extraction capabilities. Such enhancement is ascribed to the terrace extraction process, which furnishes pertinent positional and directional data to the branch task, thus endowing the model with a more enriched contextual understanding. When measured against the single-task paradigm, the dual-task learning strategy manifests pronounced augmentations across all evaluative metrics. Specifically, there are increments of 7.47%, 5.5%, 6.65%, and 8.06% in precision, recall, F1-score, and IoU for the T1 region. Correspondingly, there are boosts of 4.36%, 4.77%, 4.56%, and 6.61% for the T2 region. Collectively, across the facets of both datasets, the dual-task model registers enhancements exceeding 4% in precision, recall, F1-score, and IoU. The improvement in IoU is particularly noteworthy, underscoring the dual-task network's proficiency in amplifying the congruence between anticipated and actual outcomes.

Table 3. Comparison of single-task and dual-task.

Area	Methods	Precision (%)	Recall (%)	F1-score (%)	IoU (%)
T1	Single-task	63.24	73.98	68.19	51.73
	Dual-task	70.71	79.48	74.84	59.79
T2	Single-task	79.36	81.19	80.27	67.04
	Dual-task	83.72	85.96	84.83	73.65

Figures 11 and 12 present visual delineations contrasting the proposed methodology with the single-task model in the context of terrace extraction. In the figures, a–e represent the labels and predicted results for several typical terrace bund test data. Analyzing the predictive outcomes reveals that, under identical parameter configurations, the single-task model adeptly extracts terraces for the majority of terraced fields. However, in certain intricate scenarios, it may manifest gaps, oversights, or segmented terraces, which could compromise the model's overarching efficacy and trustworthiness. Conversely, while the dual-task model is not wholly immune to such discrepancies, it showcases enhanced continuity in its extraction endeavors. This implies a more consistent and dependable performance in both precision and thoroughness.

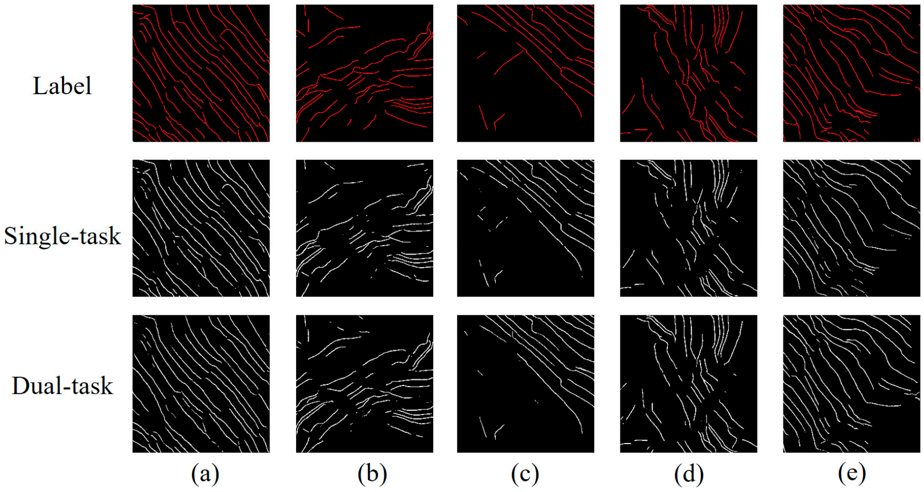

Figure 11. Results of single-task and dual-task terraced ridges in T1 area: (**a**–**e**) showcase typical scenes of terraced fields with smooth fallow cultivation.

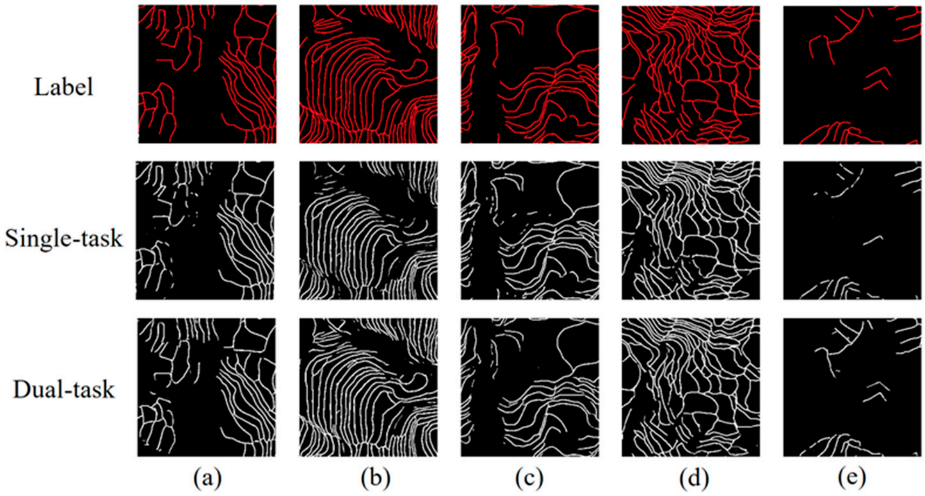

Figure 12. Results of single-task and dual-task terraced fields in T2 area: (**a**–**e**) showcase typical scenes of terraced fields with water storage and fallow cultivation.

5.2. Ablation Experiments

To further investigate the effectiveness of the proposed modules, we conducted experiments on the combined terraced datasets from the T1 and T2 regions. These experiments included ablation studies of the MSFF and MSER modules. The networks compared in the experiment consist of the Base network without the MSFF and MSER modules (Base), the network with the MSFF module added to the Base network (Base + MSFF), the network with the MSER module added to the Base network (Base + MSER), and the network with both the MSFF and MSER modules added to the Base network (DTRE-Net).

Figure A1 presents the loss function trajectories for the four models during their training and validation phases. The graph underscores a swift descent in training loss values as iterations increase, eventually converging near zero. Conversely, the validation loss values, after an initial rapid reduction, tend to stabilize around a specific point. When juxtaposed against the Base model, the loss function curves for Base + MSFF, Base + MSER, and DTRE-Net on the validation set appear more refined and consistent. Notably, in its later epochs, the training loss of DTRE-Net mirrors its validation loss. This suggests that, over the course of its training, DTRE-Net has effectively assimilated generic features and patterns, thereby enhancing its predictive accuracy for unfamiliar datasets.

The experiment analyzed the IoU values during both the training and validation processes for four network models, as illustrated in Figure A2. On the validation set, the Base network experiences eight significant fluctuations, with a notable fluctuation still evident around the 98th epoch, reaching a local value of 0.7378. In contrast, Base + MSER and Base + MSFF exhibit noticeable improvements throughout the process but still manifest four significant fluctuations. DTRE-Net encounters a significant drop around the eighth epoch, with an IoU value of 0.6063 at that point. The experiments revealed that, in terms of the IoU metric, while all four models ultimately converge around 0.82, DTRE-Net, benefiting from the integration of MSFF and MSER, demonstrates greater stability and consistent accuracy in predicting results across diverse scenarios throughout the entire process.

Table 4 presents the accuracy, recall, F1-score, and IoU values for models with different module combinations, clearly indicating that models incorporating additional feature extraction modules on top of the Base network exhibit improved segmentation performance. Upon integrating the MSFF module into the Base network, all metrics show enhancements. In the integration of MSER into the foundational Base network, a marginal decrease in recall is observed, accompanied by an increase in precision, IoU, and F1-score. DTRE-Net demonstrates improvements in all four metrics, with the IoU increasing by 1.68% compared to the Base network. This highlights the effectiveness of our approach, which introduces the

MSFF and MSER modules, in enhancing terrace extraction performance when compared to models without these modules.

Table 4. Different modules and basic network combination experiments.

Methods	Precision (%)	Recall (%)	F1-Score (%)	IoU (%)
Base	89.96	90.19	90.08	81.95
Base + MSFF	90.83	90.51	90.68	82.94
Base + MSER	91.12	89.75	90.55	82.73
DTRE-Net	91.13	91.04	91.09	83.63

The differences between the global IoU in each dataset were evaluated using the t-student hypothesis test. The calculated value of the t-student test was compared to the critical value tc. The null hypothesis is rejected if or $t \geq tc$ or $t \leq -tc$. In the first case, the mean value is considered significantly higher, and, in the second case, significantly lower. In this study, a confidence level of 95% was set, with 474 degrees of freedom, corresponding to a critical value of tc = 1.965. Comparing the results between DTRE-Net and Base, a t-value of 3.411 was obtained. Comparing the results between DTRE-Net and Base + MSFF, a t-value of 2.405 was obtained. Comparing the results between DTRE-Net and Base + MSER, a t-value of 2.703 was obtained. These values are statistically significant, indicating that the DTRE-Net network effectively improves terrace extraction results.

6. Conclusions

In this study, we introduce DTRE-Net, an innovative multi-task framework adept at concurrently executing terrace and ridge extraction tasks from high-resolution remote sensing images, providing a holistic view of terrace extractions. Notably, while the bund extraction network and terrace detection network share mutual information, they maintain a degree of autonomy, ensuring that they do not solely depend on the output from terrace detection. When benchmarked against prevailing state-of-the-art semantic segmentation techniques, our proposed DTRE-Net architecture showcases its ability to discern finer details, produce consistent outcomes, and accurately pinpoint a variety of terraces even in intricate terrains. Furthermore, we've curated two terrace datasets from distinct regions using high-resolution satellite imagery and embarked on rigorous comparative experiments, underscoring the superior efficacy of our proposed methodology. The experimental results demonstrate that the proposed DTRE-Net outperforms traditional networks in terms of both accuracy and visual effects. In comparison to single-task networks, the two mutually dependent label learning strategies can enhance accuracy through parameter sharing. Additionally, the introduced MSFF module and MSER module, as suggested in the paper, contribute to the stability of the model.

In the present study, the multi-task learning framework demonstrates a reduction in the overall training time for two tasks compared to all single-task learning frameworks, positioning it favorably in terms of efficiency. However, it is noteworthy that, in order to capture multi-scale features effectively, the network introduces additional modules, leading to a substantial increase in both the parameter and floating-point operations (FLOPs). This results in heightened dependence of the model on computer resources, including increased storage and computational capacity. Consequently, in resource-constrained environments, the escalated complexity introduced by these additions may pose a limitation. Future research endeavors may need to explore methodologies for optimizing network structures to mitigate the computational burden while retaining the advantages of multi-task learning. Moreover, owing to the utilization of consumer-grade graphics cards during our experiments, we deliberately capped the number of training epochs at 100, aiming to conserve computational resources and expedite the process. While this strategy curtailed potential overfitting, it concurrently constrained the model's ability to discern more nuanced patterns and features. As we look ahead, re-evaluating the epoch count emerges as a pivotal consideration.

Author Contributions: J.Z. (Jun Zhang 1): Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing—original draft, Visualization, Project administration, Funding acquisition. J.Z. (Jun Zhang 2): Conceptualization, Formal analysis, Resources, Writing—review and editing, Project administration, Supervision, Funding acquisition. X.H.: Validation, Writing—review and editing. W.Z.: Validation, Writing—review and editing. H.F.: Investigation, Writing—review and editing. Y.C.: Writing—review and editing, Visualization. Z.Z.: Writing—review and editing, Visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the High-resolution Special Project: Yunnan Provincial Government Comprehensive Governance Deep Application and Scale Industrialization Demonstration Project (89-Y50G31-9001-22/23), the Graduate Innovative Talent Training Project of Yunnan University—Quality Improvement Plan for Graduate Teaching Materials Construction (HXKC202112), and the Postgraduate Research and Innovation Foundation of Yunnan University (KC–22222840).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: All authors sincerely thank the reviewers for their beneficial, careful, and detailed comments and suggestions for how we could improve the paper. We are also grateful to the Gaofen-2 satellite image data provided by the Gaofen Center in Yunnan Province.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

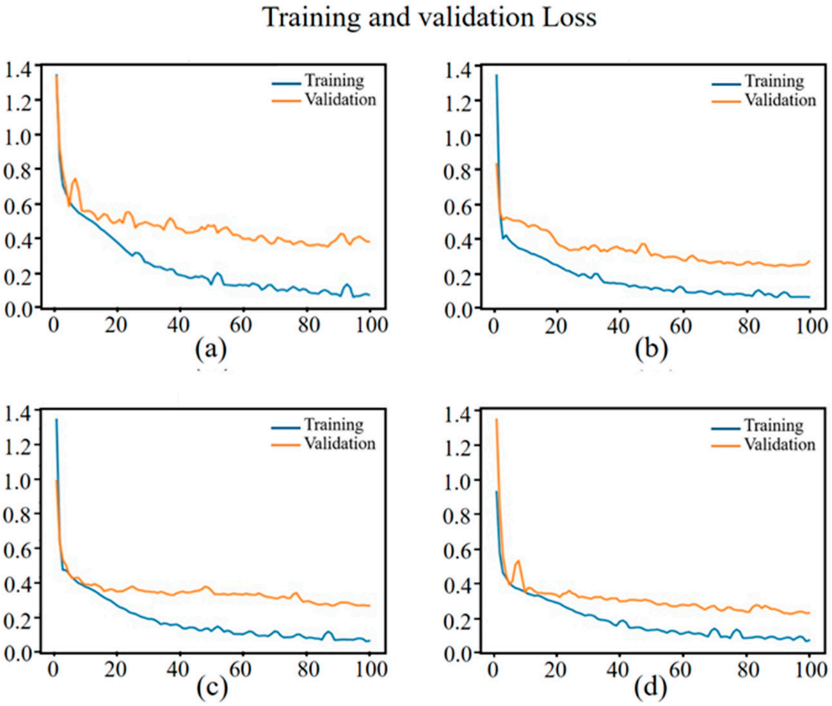

Figure A1. Training and validation loss graphs. (**a**) Training and validation loss curves for the Base model; (**b**) training and validation loss curves for the Base + MSFF model; (**c**) training and validation loss curves for the Base + MSER model; (**d**) training and validation loss curves for the DTRE-Net model.

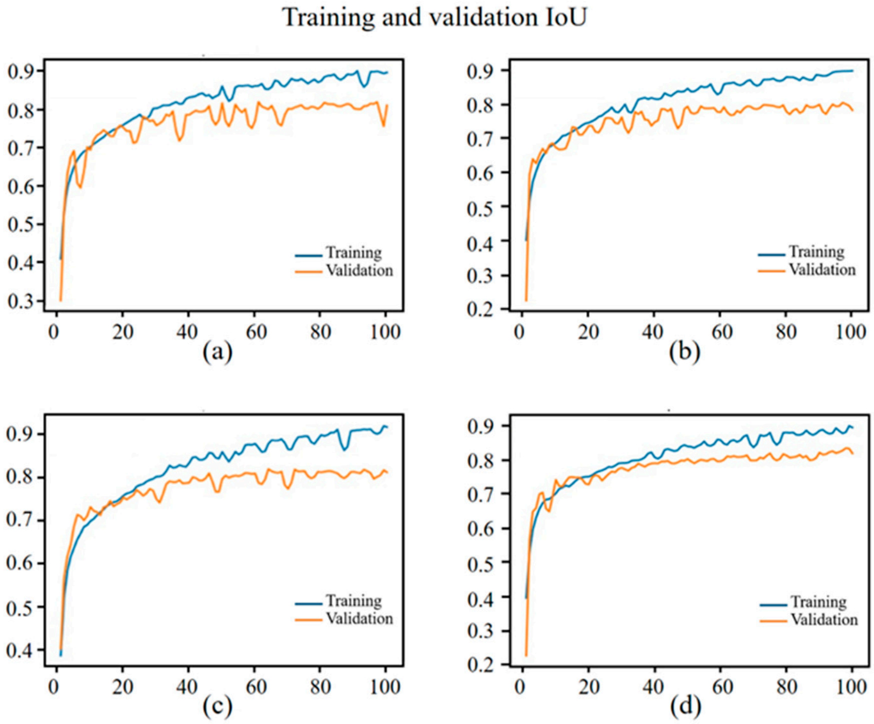

Figure A2. Training and validation IoU graphs. (**a**) IoU curves for training and validation of the Base model; (**b**) IoU curves for training and validation of the Base + MSFF model; (**c**) IoU curves for training and validation of the Base + MSER model; (**d**) IoU curves for training and validation of the DTRE-Net model.

References

1. Yu, M.; Li, Y.; Luo, G.; Yu, L.; Chen, M. Agroecosystem composition and landscape ecological risk evolution of rice terraces in the southern mountains, China. *Ecol. Indic.* **2022**, *145*, 109625. [CrossRef]
2. Mishra, P.K.; Rai, A.; Rai, S.C. Indigenous knowledge of terrace management for soil and water conservation in the Sikkim Himalaya, India. *Indian J. Tradit. Know.* **2020**, *19*, 475–485. [CrossRef]
3. Xu, Q.; Wu, P.; Dai, J.; Wang, T.; Li, Z.; Cai, C.; Shi, Z. The effects of rainfall regimes and terracing on runoff and erosion in the Three Gorges area, China. *Environ. Sci. Pollut. Res.* **2018**, *25*, 9474–9484. [CrossRef] [PubMed]
4. Zhao, X.; Zhu, H.; Dong, K.; Li, D. Plant Community and Succession in Lowland Grasslands under Saline–Alkali Conditions with Grazing Exclusion. *Agron. J.* **2017**, *109*, 2428–2437. [CrossRef]
5. Dai, W.; Na, J.; Huang, N.; Hu, G.; Yang, X.; Tang, G.; Xiong, L.; Li, F. Integrated edge detection and terrain analysis for agricultural terrace delineation from remote sensing images. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 484–503. [CrossRef]
6. Diaz-Gonzalez, F.A.; Vuelvas, J.; Correa, C.A.; Vallejo, V.E.; Patino, D. Machine learning and remote sensing techniques applied to estimate soil indicators—Review. *Ecol. Ind.* **2022**, *135*, 108517. [CrossRef]
7. Martínez-Casasnovas, J.A.; Ramos, M.C.; Cots-Folch, R. Influence of the EU CAP on terrain morphology and vineyard cultivation in the Priorat region of NE Spain. *Land Use Policy* **2010**, *27*, 11–21. [CrossRef]
8. Agnoletti, M.; Cargnello, G.; Gardin, L.; Santoro, A.; Bazzoffi, P.; Sansone, L.; Pezza, L.; Belfiore, N. Traditional landscape and rural development: Comparative study in three terraced areas in northern, central and southern Italy to evaluate the efficacy of GAEC standard 4.4 of cross compliance. *Ital. J. Agron.* **2011**, *6*, e16. [CrossRef]
9. Zhang, Y.; Shi, M.; Zhao, X.; Wang, X.; Luo, Z.; Zhao, Y. Methods for Automatic Identification and Extraction of Terraces from High Spatial Resolution Satellite Data (China-GF-1). *Int. Soil Water Conserv. Res.* **2017**, *5*, 17–25. [CrossRef]
10. Hellman, I.; Heinse, R.; Karl, J.W.; Corrao, M. Detection of terracettes in semi-arid rangelands using Fourier-based image analysis of very-high-resolution satellite imagery. *Earth Surf. Process. Landf.* **2020**, *45*, 3368–3380. [CrossRef]
11. Luo, L.; Li, F.; Dai, Z.; Yang, X.; Liu, W.; Fang, X. Terrace extraction based on remote sensing images and digital elevation model in the loess plateau, China. *Earth Sci. Inform.* **2020**, *13*, 433–446. [CrossRef]
12. Zhao, H.; Fang, X.; Ding, H.; Josef, S.; Xiong, L.; Na, J.; Tang, G. Extraction of terraces on the Loess Plateau from high-resolution DEMs and imagery utilizing object-based image analysis. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 157. [CrossRef]
13. Zhang, X.; Feng, S.; Zhao, C.; Sun, Z.; Zhang, S.; Ji, K. MGSFA-Net: Multi-Scale Global Scattering Feature Association Network for SAR Ship Target Recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**. *early access*. [CrossRef]
14. Zhao, W.; Du, S. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *113*, 155–165. [CrossRef]

15. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-Use Classification Using Taxi GPS Traces. *IEEE Trans. Intell. Transp. Syst.* **2012**, *14*, 113–123. [CrossRef]
16. Xiong, L.; Tang, G.; Yang, X.; Li, F. Geomorphology-oriented digital terrain analysis: Progress and perspectives. *J. Geogr. Sci.* **2021**, *31*, 456–476. [CrossRef]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
18. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 60–77. [CrossRef]
19. Sun, Z.; Dai, M.; Leng, X.; Leng, Y.; Xiong, B.; Ji, K.; Kuang, G. An anchor-free detection method for ship targets in high-resolution SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7799–7816. [CrossRef]
20. Zheng, J.; Yuan, S.; Wu, W.; Li, W.; Yu, L.; Fu, H.; Coomes, D. Surveying coconut trees using high-resolution satellite imagery in remote atolls of the Pacific Ocean. *Remote Sens. Environ.* **2023**, *287*, 113485. [CrossRef]
21. Hui, J.; Du, M.; Ye, X.; Qin, Q.; Sui, J. Effective Building Extraction from High-Resolution Remote Sensing Images with Multitask Driven Deep Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 786–790. [CrossRef]
22. Yu, M.; Rui, X.; Xie, W.; Xu, X.; Wei, W. Research on Automatic Identification Method of Terraces on the Loess Plateau Based on Deep Transfer Learning. *Remote Sens.* **2022**, *14*, 2446. [CrossRef]
23. Luo, L.; Li, P.; Yan, X. Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review. *Energies* **2021**, *14*, 7982. [CrossRef]
24. Tian, H.; Wang, P.; Tansey, K.; Han, D.; Zhang, J.; Zhang, S.; Li, H. A deep learning framework under attention mechanism for wheat yield estimation using remotely sensed indices in the Guanzhong Plain, PR China. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102375. [CrossRef]
25. Liu, Z.; Li, N.; Wang, L.; Zhu, J.; Qin, F. A multi-angle comprehensive solution based on deep learning to extract cultivated land information from high-resolution remote sensing images. *Ecol. Indic.* **2022**, *141*, 108961. [CrossRef]
26. Li, M.; Wu, P.; Wang, B.; Park, H.; Yang, H.; Wu, Y. A Deep Learning Method of Water Body Extraction from High Resolution Remote Sensing Images with Multisensors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3120–3132. [CrossRef]
27. Yao, J.; Sun, S.; Zhai, H.; Feger, K.-H.; Zhang, L.; Tang, X.; Li, G.; Wang, Q. Dynamic monitoring of the largest reservoir in North China based on multi-source satellite remote sensing from 2013 to 2022: Water area, water level, water storage and water quality. *Ecol. Indic.* **2022**, *144*, 109470. [CrossRef]
28. Yu, B.; Yang, A.; Chen, F.; Wang, N.; Wang, L. SNNFD, spiking neural segmentation network in frequency domain using high spatial resolution images for building extraction. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102930. [CrossRef]
29. Wang, Y.; Kong, X.; Guo, K.; Zhao, C.; Zhao, J. Intelligent Extraction of Terracing Using the ASPP ArrU-Net Deep Learning Model for Soil and Water Conservation on the Loess Plateau. *Agriculture* **2023**, *13*, 1283. [CrossRef]
30. Lu, Y.; Li, X.; Xin, L.; Song, H.; Wang, X. Mapping the terraces on the Loess Plateau based on a deep learning-based model at 1.89 m resolution. *Sci. Data* **2023**, *10*, 115. [CrossRef] [PubMed]
31. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556. [CrossRef] [PubMed]
32. Zhao, F.; Xiong, L.-Y.; Wang, C.; Wang, H.-R.; Wei, H.; Tang, G.-A. Terraces mapping by using deep learning approach from remote sensing images and digital elevation models. *Trans. GIS* **2021**, *25*, 2438–2454. [CrossRef]
33. Sofia, G.; Bailly, J.S.; Chehata, N.; Tarolli, P.; Levavasseur, F. Comparison of pleiades and LiDAR digital elevation models for terraces detection in farmlands. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1567–1576. [CrossRef]
34. Shi, H.; Fan, J.; Wang, Y.; Chen, L. Dual Attention Feature Fusion and Adaptive Context for Accurate Segmentation of Very High-Resolution Remote Sensing Images. *Remote Sens.* **2021**, *13*, 3715. [CrossRef]
35. Zhang, D.; Pan, Y.; Zhang, J.; Hu, T.; Zhao, J.; Li, N.; Chen, Q. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sens. Environ.* **2020**, *247*, 111912. [CrossRef]
36. Shen, X.; Weng, L.; Xia, M.; Lin, H. Multi-Scale Feature Aggregation Network for Semantic Segmentation of Land Cover. *Remote Sens.* **2022**, *14*, 6156. [CrossRef]
37. Wang, H.; Yao, Z.; Li, T.; Ying, Z.; Wu, X.; Hao, S.; Liu, M.; Wang, Z.; Gu, T. Enhanced open biomass burning detection: The BranTNet approach using UAV aerial imagery and deep learning for environmental protection and health preservation. *Ecol. Indic.* **2023**, *154*, 110788. [CrossRef]
38. Shao, Z.; Zhou, Z.; Huang, X.; Zhang, Y. MRENet: Simultaneous Extraction of Road Surface and Road Centerline in Complex Urban Scenes from Very High-Resolution Images. *Remote Sens.* **2021**, *13*, 239. [CrossRef]
39. Pelizari, P.A.; Geiß, C.; Groth, S.; Taubenböck, H. Deep multitask learning with label interdependency distillation for multicriteria street-level image classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *204*, 275–290. [CrossRef]
40. Xu, L.; Yang, P.; Yu, J.; Peng, F.; Xu, J.; Song, S.; Wu, Y. Extraction of cropland field parcels with high resolution remote sensing using multi-task learning. *Eur. J. Remote Sens.* **2023**, *56*, 2181874. [CrossRef]
41. Li, M.; Long, J.; Stein, A.; Wang, X. Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2023**, *200*, 24–40. [CrossRef]

42. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018. [CrossRef]
43. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
45. 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 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
46. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large Kernel Matters–Improve Semantic Segmentation by Global Convolutional Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4353–4361.
47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
48. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
49. 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; Volume 9351, pp. 234–241.

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.