Article

Enhanced Dual-Channel Model-Based with Improved Unet++ Network for Landslide Monitoring and Region Extraction in Remote Sensing Images

Junxin Wang [1], Qintong Zhang [1], Hao Xie [1], Yingying Chen [1] and Rui Sun [1,2,*]

1 Faculty of Arts and Sciences, Beijing Normal University, Zhuhai 519087, China;
202111079160@mail.bnu.edu.cn (J.W.); qintongz@mail.bnu.edu.cn (Q.Z.);
202111079279@mail.bnu.edu.cn (H.X.); 202111079225@mail.bnu.edu.cn (Y.C.)
2 State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
* Correspondence: sunrui@bnu.edu.cn

Abstract: Landslide disasters pose significant threats to human life and property; therefore, accurate and effective detection and area extraction methods are crucial in environmental monitoring and disaster management. In our study, we address the critical tasks of landslide detection and area extraction in remote sensing images using advanced deep learning techniques. For landslide detection, we propose an enhanced dual-channel model that leverages EfficientNetB7 for feature extraction and incorporates spatial attention mechanisms (SAMs) to enhance important features. Additionally, we utilize a deep separable convolutional neural network with a Transformers module for feature extraction from digital elevation data (DEM). The extracted features are then fused using a variational autoencoder (VAE) to mine potential features and produce final classification results. Experimental results demonstrate impressive accuracy rates of 98.92% on the Bijie City landslide dataset and 94.70% on the Landslide4Sense dataset. For landslide area extraction, we enhance the traditional Unet++ architecture by incorporating Dilated Convolution to expand the receptive field and enable multi-scale feature extraction. We further integrate the Transformer and Convolutional Block Attention Module to enhance feature focus and introduce multi-task learning, including segmentation and edge detection tasks, to efficiently extract and refine landslide areas. Additionally, conditional random fields (CRFs) are applied for post-processing to refine segmentation boundaries. Comparative analysis demonstrates the superior performance of our proposed model over traditional segmentation models such as Unet, Fully Convolutional Network (FCN), and Segnet, as evidenced by improved metrics: IoU of 0.8631, Dice coefficient of 0.9265, overall accuracy (OA) of 91.53%, and Cohen's kappa coefficient of 0.9185 on the Bijie City landslide dataset; and IoU of 0.8217, Dice coefficient of 0.9021, overall accuracy (OA) of 96.68%, and Cohen's kappa coefficient of 0.8835 on the Landslide4Sense dataset. These findings highlight the effectiveness and robustness of our proposed methodologies in addressing critical challenges in landslide detection and area extraction tasks, with significant implications for enhancing disaster management and risk assessment efforts in remote sensing applications.

Keywords: landslide detection; dual-channel model; feature fusion; variational autoencoder; improving Unet++ model

1. Introduction

1.1. Background and Motivation

Landslide detection and region extraction from remote sensing imagery are imperative tasks in geospatial analysis, essential for hazard assessment, disaster management, and environmental conservation [1]. Landslides, characterized by the sudden movement of soil, rock, and debris down a slope, pose significant threats to human life, infrastructure,

and ecosystems [2]. Remote sensing technologies offer invaluable tools for monitoring and assessing landslide occurrences over large spatial scales, enabling timely response and mitigation efforts.

Landslide susceptibility models can be broadly categorized into heuristic models, physically based models, statistical models, and machine learning models.

Heuristic models are based on experience and expertise, establishing landslide susceptibility evaluation criteria by analyzing factors such as terrain, geology, and rainfall. For example, Stanley et al. (2017) used a heuristic fuzzy approach to combine data on slope [3], faults, geology, forest loss, and road networks to create a global landslide susceptibility map, which was evaluated using the Global Landslide Catalog developed by NASA.

Physically based models describe the mechanisms and development processes of landslides through numerical simulations and physical experiments. Representative physical models include the finite element method and discrete element method. For example, Medina et al. (2021) proposed the Fast Shallow Landslide Assessment Model [4], and Guo et al. developed a Python QGIS plugin named FSLAM to simulate regional landslide susceptibility [5].

The significance of landslide detection in remote sensing lies in its ability to provide early warning systems and facilitate risk assessment in landslide-prone areas. Traditional methods for landslide detection primarily rely on manual interpretation of satellite imagery or aerial photographs, which are often labor-intensive, time-consuming, and subjective. Moreover, these methods may lack the spatial resolution and coverage required for comprehensive landslide mapping and monitoring.

In recent years, we have witnessed a paradigm shift in landslide detection, with deep learning techniques, particularly convolutional neural networks (CNNs), emerging as potent tools for automated and data-driven solutions, as evidenced by pioneering studies such as the one by Wu and Liu et al. [6]. Advances in CNN architectures, including Visual Geometry Group (VGG), ResNet, and DenseNet [7,8], along with the incorporation of attention mechanisms highlighted by Gao et al. [9] and the advent of transfer learning, have significantly enhanced the accuracy and robustness of landslide detection models, mitigating data scarcity issues and improving feature extraction capabilities.

Semantic segmentation techniques, pivotal for accurately delineating landslide areas in remote sensing imagery, have made significant advancements through the integration of deep learning methodologies, as evidenced by seminal works such as Cheng et al. [10] and Zhang et al. [11,12], which demonstrate the efficacy of CNNs in pixel-wise classification; further strides have been made by Wang et al. [13] and Ye et al. [14] in addressing challenges such as class imbalance and boundary refinement using UNet, SegNet, and their variants, while recent innovations by Li et al. [15] and Zhou et al. [16] highlight multi-scale feature fusion and Conditional Random Fields (CRFs) to enhance segmentation accuracy and mitigate boundary ambiguities.

Despite significant progress, numerous challenges remain in the field of landslide detection and area extraction from remote sensing data [17,18]. One major issue is the effective integration of heterogeneous data modalities, such as optical images and digital elevation models (DEMs), to enhance classification accuracy and robustness. The fusion of these diverse data sources necessitates complex algorithms capable of handling different resolutions, noise levels, and data formats to extract meaningful information [19]. Additionally, accurately delineating landslide boundaries poses a significant challenge. Differentiating between landslide and non-landslide areas requires high-resolution segmentation techniques to capture complex spatial variations and subtle texture differences. Existing segmentation algorithms often struggle with these fine-grained details, resulting in imprecise boundary definitions and classification errors [13,20–22].

1.2. Contribution

Addressing the current challenges and limitations of deep learning in the field of remote sensing, this paper proposes innovative methods for landslide detection and region extraction, with the main contributions as follows:

(i) Enhanced Dual-Channel Model for Landslide Detection: We developed a novel model integrating optical imagery and Digital Elevation Model (DEM) data, employing EfficientNetB7 for image data and a deep separable convolutional neural network with Transformer modules for DEM data. The model incorporates a spatial attention mechanism and a variational autoencoder (VAE) for feature fusion, demonstrating excellent classification performance across various datasets.

(ii) Improved UNet++ Architecture for Landslide Area Extraction—DCT-Unet++ model: We have innovatively proposed a model named DCT-Unet++. We enhanced the UNet++ architecture by incorporating Dilated Convolution ('D') for multi-scale feature extraction and integrating Channel and Spatial Attention Mechanisms ('C') and Transformer ('T'). Multi-task learning with segmentation and edge detection, along with Conditional Random Fields (CRFs) for post-processing, allows our model to outperform traditional models such as Unet, FCN, and Segnet in various metrics, including IoU, Dice coefficient, Overall Accuracy, and Cohen's Kappa coefficient.

By leveraging cutting-edge deep learning techniques, our contributions offer robust solutions to complex landslide analysis challenges, enhancing disaster management strategies and advancing environmental science and engineering.

In the next section, this paper elaborates on the principles of the two proposed model architectures: the enhanced dual-channel model for landslide recognition and a DCT-Unet++ model for landslide area segmentation. The third section details the experimental design, presenting the performance results of the proposed models compared to more than ten classic models across three landslide datasets for both recognition and segmentation tasks. The discussion section provides a comprehensive and comparative analysis of the models' performance, objectively discussing their strengths and areas for improvement. Finally, the conclusion section summarizes the entire study and highlights several promising directions for future research. The code related to the experiment is stored at https://github.com/xinxin2021110/DCT-UNETplusplus, accessed on 12 August 2024.

2. Methodology

2.1. Enhanced Dual-Channel Model

The dual input model constitutes a groundbreaking advancement in the realm of landslide image detection, harnessing the synergistic potential of optical imagery and Digital Elevation Model (DEM) data to augment feature extraction and elevate classification efficiency [23]. The model innovatively employs a Variational Autoencoder (VAE) to fuse two distinct data modalities (optical images capturing surface attributes and DEM data) [19].

The architecture of the dual input model manifests as a composite entity comprising two distinct pathways dedicated to processing image and DEM data independently, culminating in a fusion stage orchestrating the assimilation of gleaned features (As shown in Figure 1).
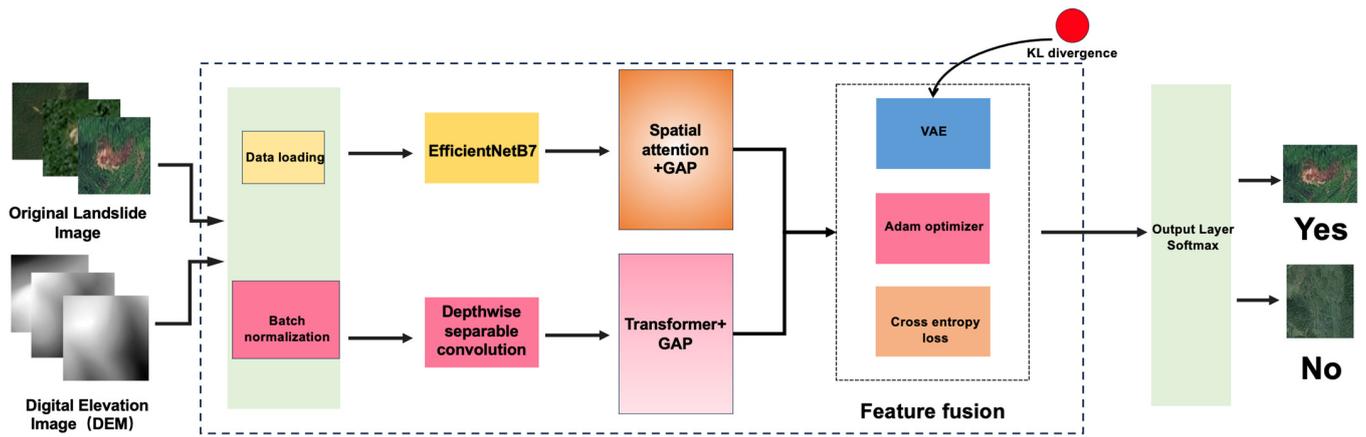
Figure 1. Enhanced dual-channel model framework for realizing the landslide image recognition task.

2.1.1. Image Pathway

The image pathway initiates with the deployment of a pre-trained EfficientNetB7 model to elicit high-order features from the optical imagery [24]. This convolutional cascade engenders a hierarchical abstraction of spatial intricacies and visual characteristics entrenched within the images [25]. Furthermore, the integration of a spatial attention module bolsters the discernment of salient features pervading the image feature maps, formulated as follows (as shown in Figure 2):

$$\text{Spatial Attention} = \text{Conv2D}\big(1, (1,1), \text{activation} = \text{'sigmoid'}\big)(X) \tag{1}$$

$$\text{Enhanced Feature} = \text{Multiply}([X, \text{Spatial Attention}]) \tag{2}$$
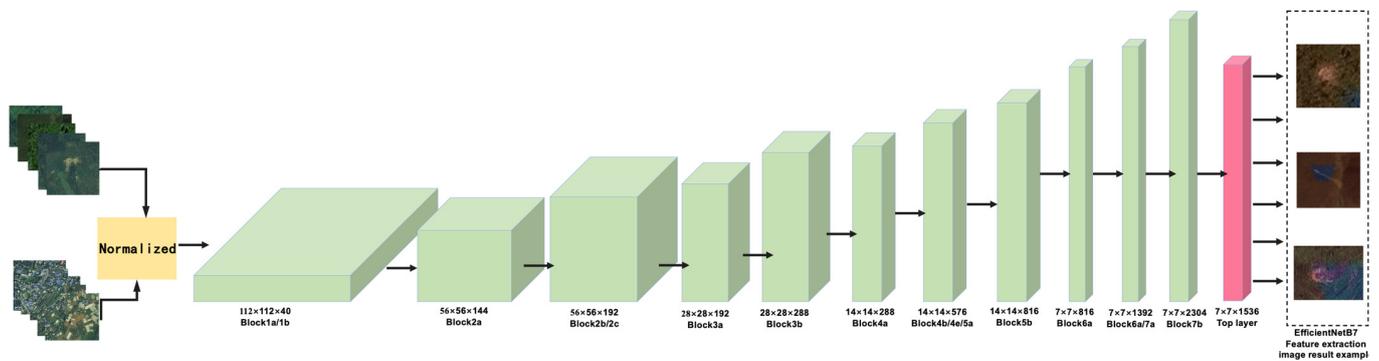
where X denotes the input feature map.

Figure 2. Network architecture diagram of EfficientNetB7 to implement image data characteristics.

2.1.2. DEM Pathway

The DEM pathway is predicated upon a bespoke CNN architecture tailored to the exigencies of terrain data processing [26]. Herein, separable convolutional layers take center stage to cull features from the DEM, with a specific focus on delineating terrain attributes such as slope, elevation, and aspect [27]. Moreover, the infusion of a transformer block fortifies the DEM pathway with self-attention mechanisms (as shown in Figure 3), endowing the model with the acumen to discern pertinent spatial relationships intrinsic to the terrain data, formulated as follows:

$$\text{Transformer Block Output} = \text{TransformerBlock}(X) \tag{3}$$

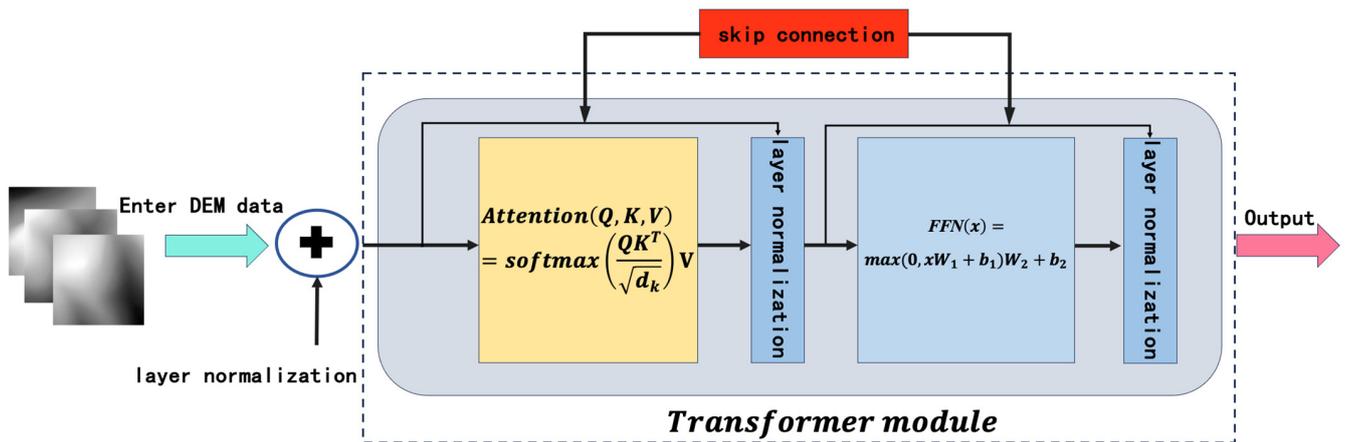where X denotes the input feature map.

Figure 3. Transformer module implements the principle of DEM data feature enhancement.

2.1.3. Feature Fusion

After extracting features from both image and DEM paths, the features from these two sources are combined. The combined feature map is then processed using the encoder block of a variational autoencoder to mine the latent feature vectors. These latent feature vectors are subsequently input into the classifier of the model to complete the region classification task. This fusion approach leverages features from both modalities, preserving the inherent information from each data modality while enhancing the model's ability to capture the synergistic correlations between optical and terrain attributes.

The VAE module (as shown in Figure 4) is depicted as follows:

(1)　Encoder Module:

At the helm of the feature fusion endeavor lies the encoder module, which can distill high-dimensional input features into compact latent representations [28,29]. Mathematically, the encoder module comprises a hierarchical stack of dense layers, culminating in the computation of latent mean (μ) and log-variance ($\log(\sigma^2)$) vectors, encapsulating the essence of the input feature manifold.

$$z_{\text{mean}} = \text{Dense}(x) \tag{4}$$

$$z_{\log(\text{var})} = \text{Dense}(x) \tag{5}$$

Here, x represents the input feature tensor.

(2)　Latent Space Sampling:

Leveraging the Gaussian reparameterization trick, the latent space sampling mechanism imbues the VAE with stochasticity, facilitating the synthesis of diverse yet semantically meaningful feature representations. By sampling from a latent space governed by μ and $\log(\sigma^2)$, the VAE engenders a diverse repertoire of feature representations conducive to robust classification performance.

$$z = \mu + \exp\left(\frac{1}{2} \cdot \log\left(\sigma^2\right)\right) \cdot \epsilon \tag{6}$$

Here, ϵ signifies a random noise vector sampled from a standard normal distribution. σ represents the standard deviation of the latent space distribution.
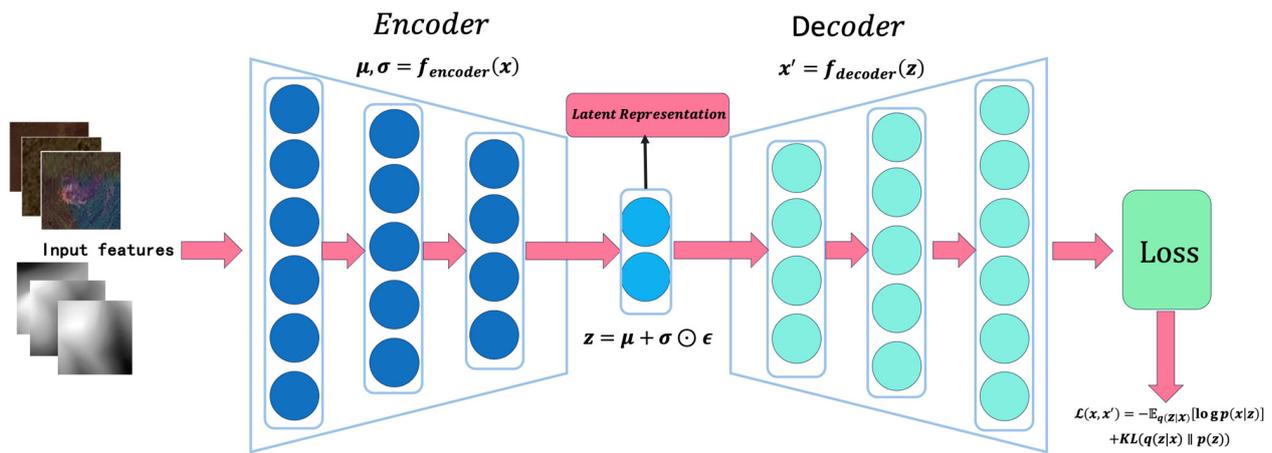
Figure 4. Application principle of the variational autoencoder module.

2.2. Improving the Unet++ Network for Image Region Extraction: DCT-Unet++ Model

To address the challenge of mountain landslide region extraction, we have meticulously crafted a novel image segmentation model (as shown in Figure 5). This model improves upon the traditional Unet++ architecture by replacing the convolutional kernels with dilated convolutions, thereby expanding the receptive field and enabling multiscale feature extraction. Additionally, we integrate the Convolutional Block Attention Module and Transformer attention mechanisms to enhance the model's focus on critical features. As a result, this model is referred to as the DCT-Unet++ model.
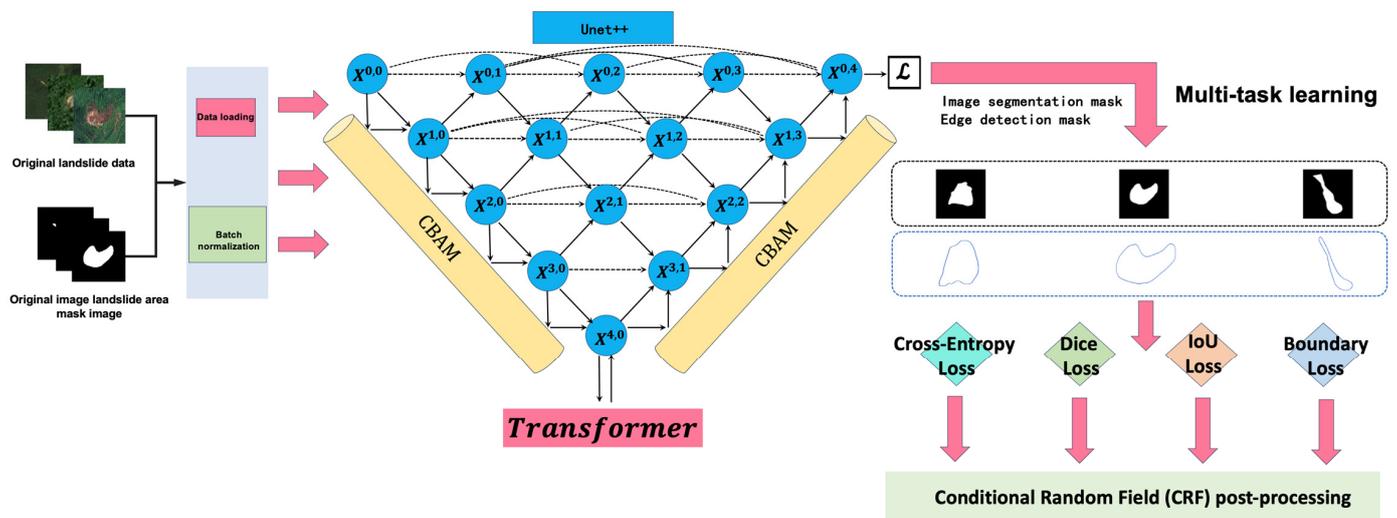
Figure 5. An improved Unet++ network framework for realizing the landslide area segmentation task—DCT-Unet.

Furthermore, at the top-level output of the decoder, the model employs multitask learning, including segmentation and edge detection tasks. This efficient approach enables the effective extraction and further refinement of mountain landslide regions.

2.2.1. Encoder

The encoder module is composed of five convolutional blocks, each integrating dilated convolutions and CBAM for feature extraction and enhancement [30,31]. Dilated convolutions, with increasing dilation rates, are strategically employed to expand the receptive field and capture multi-scale contextual information. The CBAM module dynamically

recalibrates feature maps, emphasizing relevant spatial and channel-wise features. The encoder can be expressed as:

$$\text{Encoder} = \left\{ \text{CBAM}_{d_1}^{(1)} \circ \text{DilatedConv}_{k_1}^{(1)} \circ \text{MaxPooling}^{(1)}, \dots, \text{CBAM}_{d_5}^{(5)} \circ \text{DilatedConv}_{k_5}^{(5)} \right\} \tag{7}$$

where $\text{CBAM}_{d_i}^{(i)}$ represents the Channel and Spatial Attention Module tailored to each convolutional block, $\text{DilatedConv}_{k_i}^{(i)}$ signifies dilated convolutions with kernel size k_i, and $\text{MaxPooling}^{(i)}$ denotes max-pooling operations following each convolutional block.

2.2.2. Convolutional Blocks

Each block comprises two convolutional layers followed by a CBAM module, enhancing feature representation and spatial awareness.

After the core of the encoder (Conv Block 5), the Transformer module is inserted into our model, a sophisticated mechanism for aggregating contextual information and capturing the long-range dependencies inherent in remote sensing imagery. By integrating Layer Normalization, MultiHead Attention, and Position-wise Feedforward Networks, the Transformer block fosters intricate feature interactions and global context awareness. Mathematically, the Transformer module unfolds as:

$$\begin{aligned} \text{Transformer Block} \quad &= \text{LayerNorm} \rightarrow \text{MultiHeadAttention} \\ &\rightarrow \text{Residual Connection} \rightarrow \text{LayerNorm} \\ &\rightarrow \text{Feedforward Network} \rightarrow \text{Residual Connection} \end{aligned} \tag{8}$$

2.2.3. Decoder

The decoder in DCT-Unet++ orchestrates the restoration of spatial resolution and facilitates feature fusion with encoder layers to refine segmentation accuracy. Employing a combination of upsampling, concatenation, and convolutional operations, the decoder reconstructs detailed structures while preserving contextual information. The decoder's intricate architecture is encapsulated as:

$$\text{Decoder} = \left\{ \text{UpSampling}^{(1)} \circ \text{Concat}^{(1)} \circ \text{Conv}^{(1)}, \dots, \text{UpSampling}^{(4)} \circ \text{Concat}^{(4)} \circ \text{Conv}^{(4)} \right\} \tag{9}$$

where $\text{UpSampling}^{(i)}$ denotes the upsampling operation, $\text{Concat}^{(i)}$ signifies concatenation with corresponding encoder features, and $\text{Conv}^{(i)}$ represents convolutional operations within each decoder block.

2.2.4. CBAM Modules: Feature Recalibration and Attention Mechanisms

After each deconvolution layer of the DCT-Unet++ decoder is output, it is embedded into the CBAM module, serving as pivotal components for feature recalibration and attention mechanisms, accentuating salient features and delineating subtle boundaries. The CBAM module is characterized by:

$$\text{CBAM} = \text{ChannelAttention} \circ \text{SpatialAttention} \tag{10}$$

where ChannelAttention and SpatialAttention represent mechanisms for channel-wise and spatial-wise feature recalibration, respectively.

2.2.5. Composite Loss Function

The composite loss function in the DCT-Unet++ model is a holistic measure that integrates multiple loss components to comprehensively optimize segmentation performance. Each loss component captures distinct aspects of segmentation fidelity, boundary delineation, and spatial overlap between predicted and ground truth masks.

The composite loss function is a weighted sum of the Cross-Entropy, Dice, IoU, and Boundary losses, facilitating comprehensive optimization for semantic segmentation. It is expressed as:

$$\text{Composite Loss} = \text{CrossEntropy} + \text{Dice} + \text{IoU} + \text{Boundary Loss}. \tag{11}$$

Each loss component contributes to capturing different aspects of segmentation fidelity and spatial agreement, ensuring robust performance across diverse remote sensing scenarios.

By amalgamating cutting-edge architectural components and loss formulations, the DCT-Unet++ model transcends conventional segmentation frameworks, demonstrating superior performance in image region extraction tasks. Leveraging advanced mechanisms such as Transformer-based context aggregation and CBAM-guided feature refinement, this model establishes a new paradigm in semantic segmentation.

2.3. Conditional Random Fields (CRFs) for Post-Processing of the DCT-Unet++ Model

The integration of Conditional Random Fields (CRFs) with the DCT-Unet++ model (As shown in Figure 6) represents a sophisticated approach aimed at refining segmentation outputs and enhancing the overall quality of segmentation results [32].
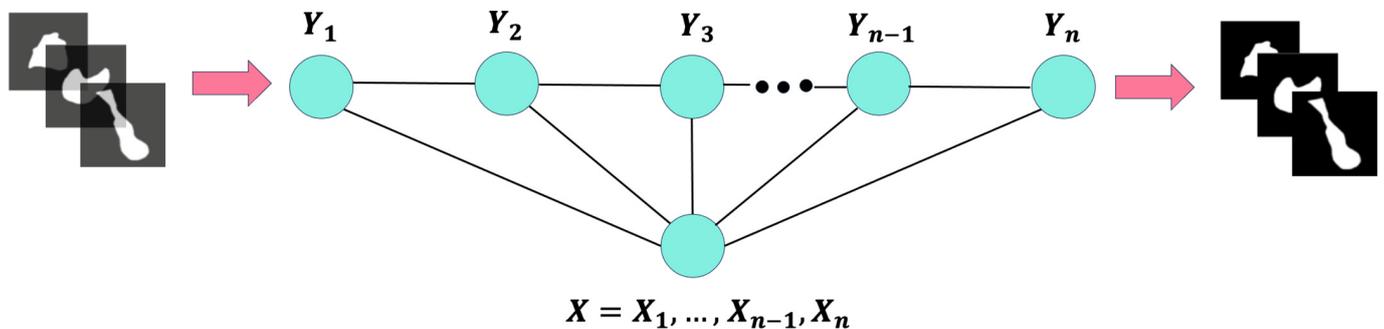

Figure 6. Conditional Random Field (CRF) post-processing application principle.

2.3.1. Probabilistic Refinement

Following the initial segmentation performed by the DCT-Unet++ model, the CRF post-processing step refines the probabilistic output map generated by the model. This refinement is crucial for mitigating noise and inconsistencies inherent in raw segmentation outputs.

$$P(Y|X) = \frac{1}{Z(X)} \exp \left\{ -\sum_i \psi_u(y_i|x_i) - \sum_{i,j} \psi_p(y_i, y_j|x_i, x_j) \right\} \tag{12}$$

In this equation, $P(Y|X)$ represents the conditional probability of the label assignment Y given the observed data X, ψ_u is the unary potential reflecting the likelihood of assigning label y_i to pixel x_i, and ψ_p is the pairwise potential enforcing smoothness by considering the relationship between neighboring pixels (i, j).

2.3.2. Spatial Consistency Enhancement

CRF leverages spatial constraints to enforce consistency among neighboring pixels, thereby fostering smoother transitions between segmented regions. By considering pixel affinities and spatial relationships, CRF facilitates the delineation of coherent and contiguous regions within the segmentation output.

$$E(X, Y) = \sum_i \psi_u(y_i|x_i) + \sum_{i,j} \psi_p(y_i, y_j|x_i, x_j) \tag{13}$$

Here, $E(X, Y)$ represents the energy function to be minimized, where the unary potential ψ_u quantifies the cost of assigning label y_i to pixel x_i, and the pairwise poten-

tial ψ_p enforces spatial consistency by penalizing dissimilar label assignments among neighboring pixels.

The integration of CRF post-processing with the DCT-Unet++ model represents a synergistic approach to semantic segmentation in remote sensing applications. By leveraging spatial constraints and probabilistic refinement, CRF enhances the quality and coherence of segmentation outputs, thereby facilitating more accurate and interpretable results in various remote sensing tasks.

3. Experimental Results and Analysis

3.1. The Design of Experiments

3.1.1. Selection of Models for Comparison

(1) Models on the landslide image recognition task

In selecting the comparative models for our experimental comparison in image recognition, we aimed to cover a wide spectrum of architectures and combinations to comprehensively evaluate the performance of our proposed model. We chose ResNet50, VGG16, MobileNetV2, DenseNet121, InceptionV3, Xception, EfficientNetB0, ResNet152V2, and NASNetMobile due to their popularity and widespread use in the field of deep learning [33–41]. These models represent a diverse range of architectures, including traditional convolutional neural networks (CNNs) such as VGG16 and ResNet50, lightweight models such as MobileNetV2 optimized for mobile and embedded devices, densely connected networks such as DenseNet121, and advanced architectures such as InceptionV3 and Xception, which incorporate different types of convolutional operations. Additionally, we combined these models with each other to explore the potential benefits of ensemble methods and hybrid architectures. This comprehensive selection allows us to conduct a thorough comparison of our proposed model against a diverse set of state-of-the-art baselines, providing insights into its effectiveness and robustness in image recognition tasks.

(2) Models on image segmentation task

In designing our experimental comparison for the image segmentation task, we meticulously selected a diverse set of models to comprehensively evaluate the performance of our proposed methodology. We included well-established architectures such as UNet, FCN, Segnet, and Linknet [42–50], which have been widely used as benchmarks in the field of semantic segmentation. These models represent different approaches to semantic segmentation, ranging from fully convolutional networks such as FCN to encoder–decoder architectures such as UNet. Additionally, we incorporated advanced models such as Deeplabv3, which utilizes atrous convolution and multi-scale features, and PSPNet, which leverages pyramid pooling modules for capturing contextual information at different scales. Furthermore, we explored hybrid architectures and ensemble methods by combining VGG16 with FCN and incorporating attention mechanisms in UNet, aiming to investigate the potential improvements in segmentation performance. By including this diverse set of models, we aimed to conduct a comprehensive evaluation of our proposed methodology against state-of-the-art baselines, providing insights into its effectiveness and generalizability across different segmentation architectures and approaches.

3.1.2. Selection of Indicators for Model Evaluation

In our experimental design, the selection of appropriate evaluation metrics is crucial for accurately assessing the performance of our proposed methodologies in both image recognition and segmentation tasks.

For image recognition, precision, recall, F1 score, and support were selected to gauge the model's ability to classify images accurately while considering class distribution. Conversely, for image segmentation, we opted for metrics such as IoU, Dice coefficient, accuracy, precision, recall, F1 score, overall accuracy (OA), and kappa coefficient to assess the model's performance in delineating object boundaries and accurately segmenting images. These

metrics offer a comprehensive evaluation, capturing various aspects of model performance essential for both tasks.

3.1.3. Experimental Conditions and Hyperparameter Tuning

In this project, the landslide remote sensing image recognition model and its comparison model were trained under the same conditions and number of epochs. Specifically, the models were trained on a computer equipped with an RTX 4090 GPU, using TensorFlow 2.9.0 and Python 3.8. The training consisted of 20 epochs with a batch size of eight, utilizing the Adam optimizer (learning rate = 0.0001) and a categorical crossentropy loss function. Similarly, the landslide remote sensing image segmentation model and its comparison model were trained under the same conditions. These models were also trained on a computer with an RTX 4090 GPU, using TensorFlow 2.9.0 and Python 3.8, for 100 epochs with a batch size of four. Above is a detailed summary of the hyperparameter tuning results for the two models (as shown in Table 1).

Table 1. Summary table of hyperparameter design for two models.

Parameter	Enhanced Dual-Channel Model	DCT-Unet++ Model
Input Image Size	224×224	256×256
Batch Size	8	8
Number of Epochs	20	100
Optimizer	Adam (learning rate = 0.0001)	Adam (learning rate = 0.0001)
Data Split	Training = 80:20	Training = 70:30
Dropout Rate	0.5	None
Number of Heads in Multi-Head Attention	4	8
Key Dimension in Multi-Head Attention	64	128
Intermediate Layer Dimension	512	$4\times$ key dimension (512)
Number of Transformer Attention Layers	1	1
Activation Function in Classification Head	Softmax	Sigmoid
Loss Function	Categorical Crossentropy	Composite loss function (cross-entropy + Dice + IoU)

3.1.4. Datasets

(1) Bijie Landslide Dataset

The Bijie Landslide Dataset, meticulously curated by Ji et al. [25]. from Wuhan University, represents a vital resource for evaluating landslide detection algorithms. This dataset was downloaded from the URL http://study.rsgis.whu.edu.cn/pages/download/, accessed on 12 August 2024, comprising 770 landslide images and 2003 non-landslide images. Cropped from TripleSat satellite imagery, the dataset encompasses a diverse array of terrain types and land cover classes (as shown in Figure 7). Each sample, with varying sizes and resolutions, is annotated with corresponding shapefiles delineating landslide boundaries, facilitating accurate model evaluation.

(2) Landslide4Sens Dataset

The Landslide4Sense dataset, curated for the Landslide4Sense competition by the Institute of Advanced Research in Artificial Intelligence (IARAI), was downloaded from the URL https://www.kaggle.com/datasets/tekbahadurkshetri/landslide4sense, accessed on 12 August 2024 [51] and serves as a benchmark for large-scale landslide detection utilizing multi-source satellite remote sensing imagery, spanning various landslide-affected regions globally from 2015 to 2021. In this experiment, we aim to assess the universality and generalizability of our algorithm by comparing its performance against the previously utilized Bijie City dataset.
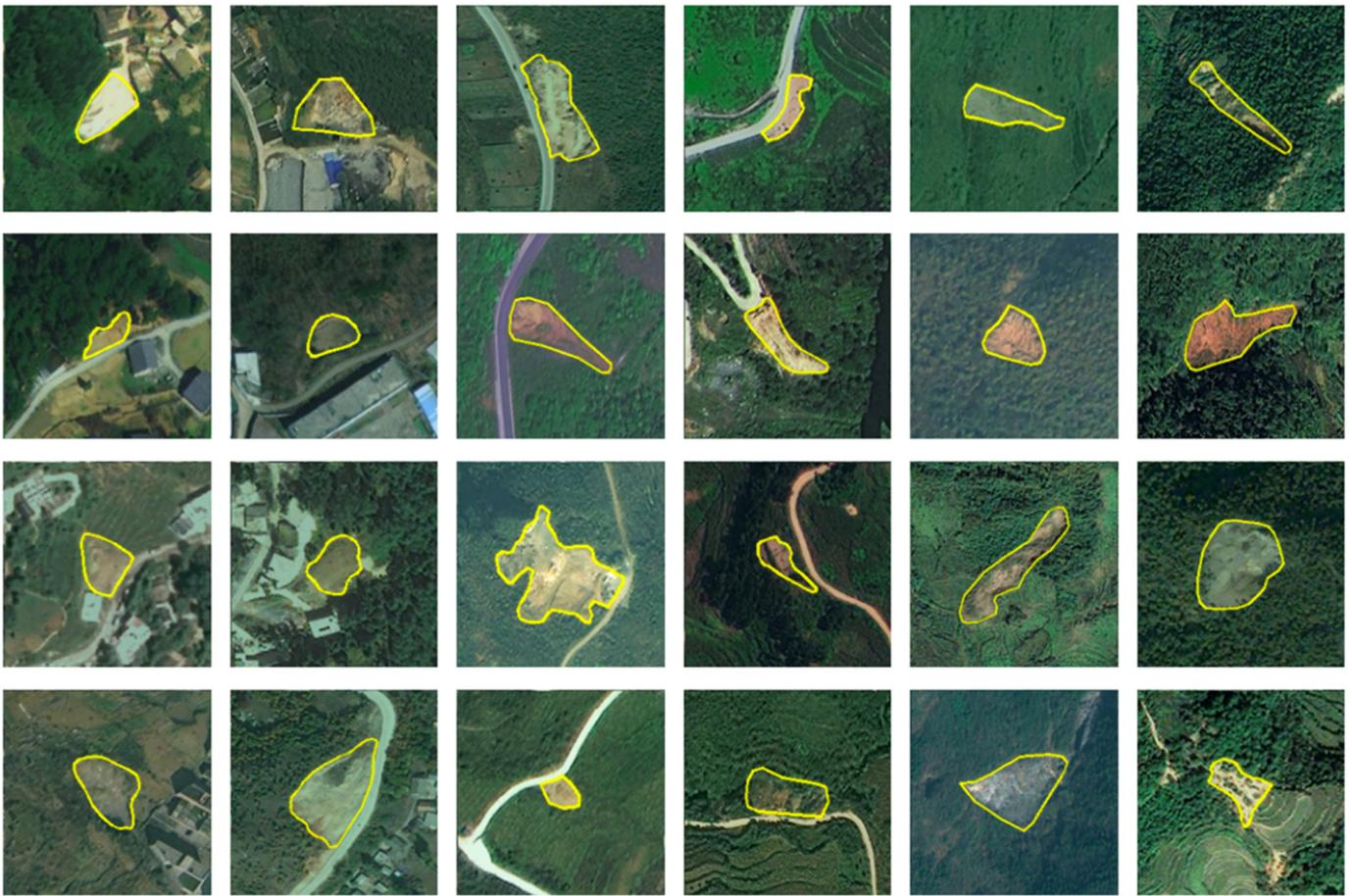
Figure 7. Various landslide instances from the Bijie Landslide Dataset [25].

Comprising training/validation/test subsets, the Landslide4Sense dataset encompasses 3799, 245, and 800 images, respectively. Each image patch integrates 14 bands, including multispectral data from Sentinel-2 (B1–B12), slope data from ALOS PALSAR (B13), and digital elevation model (DEM) from ALOS PALSAR (B14). With all bands standardized to a resolution of approximately 10m per pixel, the image tiles have dimensions of 128×128 pixels and are annotated at the pixel level.

To address data imbalance and ensure robust comparison in training effectiveness, preprocessing steps involve the segmentation processing of the H5 dataset. This includes synthesizing RGB three-channel color images from selected bands and masking data to delineate landslide and non-landslide areas. Additionally, to enhance dataset balance and volume, we preprocessed the dataset. Each sample is subdivided into 3×3 smaller samples, with criteria for categorizing samples as landslide or non-landslide based on the proportion of white areas (as shown in Figure 8).

(3) CAS Landslide Dataset

The CAS Landslide Dataset represents a significant advancement in large-scale multi-sensor landslide detection [52], specifically tailored for deep learning applications. With 20,865 images sourced from nine regions worldwide, this dataset integrates satellite and drone data to address the challenges encountered in landslide identification. Unlike existing datasets, which often exhibit limitations in terms of size, coverage, sensor diversity, and resolution, the CAS Landslide Dataset offers a substantial resource for facilitating rapid and efficient landslide identification endeavors (as shown in Figure 9).
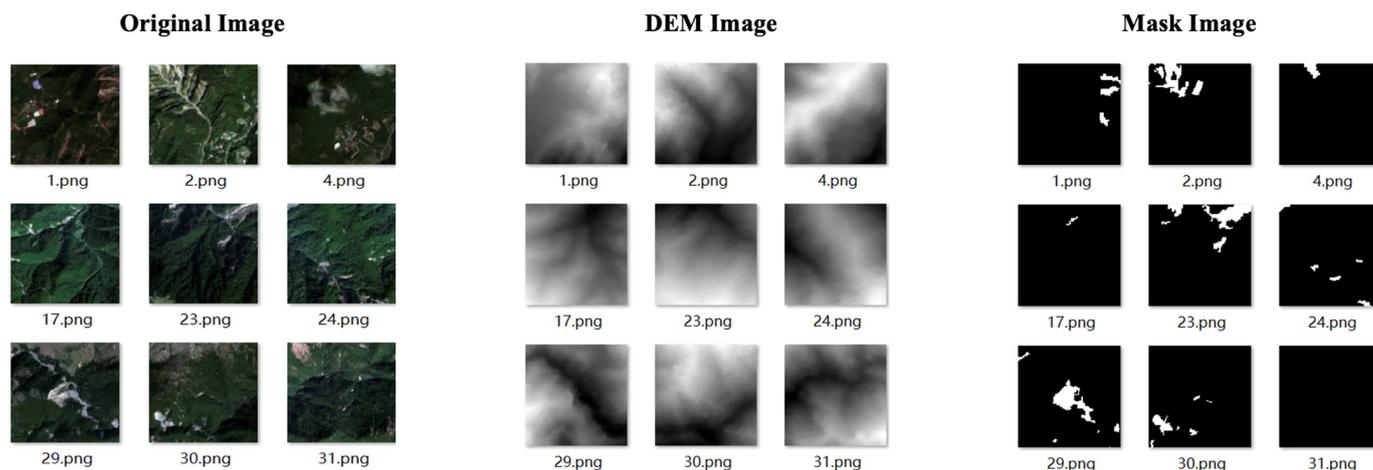
Figure 8. Various landslide instances from the Landslide4Sense Dataset [51].
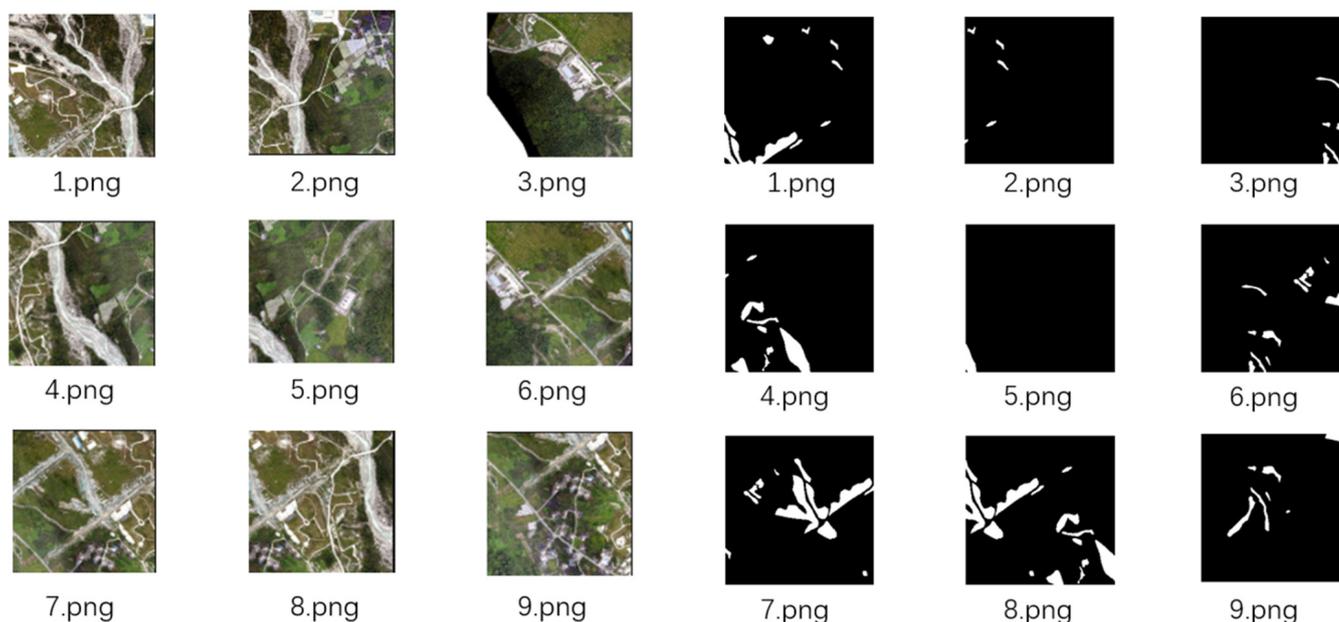

Figure 9. Examples of original images and mask images from the CAS Landslide Dataset [52].

In our study, we utilized the Moxi Platform Unmanned Aerial Vehicle (UAV) 1m dataset, situated in Ganzi County, Sichuan Province, China. This dataset captures the aftermath of a debris flow incident in August 2005, resulting in significant infrastructure damage and economic losses. Preprocessing involved converting pixel values to delineate landslide-affected areas, facilitating subsequent segmentation tasks. However, due to the dispersed nature of landslide occurrences and the absence of Digital Elevation Model (DEM) data, the experiment focused solely on landslide area extraction.

Three different landslide datasets were collected for training the new model, with the model's predictive capabilities varying across these datasets.

(1) Bijie Landslide Dataset: Curated by Ji et al. from Wuhan University, this dataset comprises 770 landslide images and 2003 non-landslide images from TripleSat satellite imagery. It features diverse terrain types and land cover classes, with annotated shapefiles for accurate landslide boundaries. This dataset is characterized by concentrated landslide areas with large landslide regions, making it suitable for both landslide detection and segmentation tasks.

(2) Landslide4Sens Dataset: Created for the Landslide4Sense competition by IARAI, this dataset includes 3799 training images, 245 validation images, and 800 test images, inte-

grating 14 bands (multispectral data, slope data, and DEM) from various global landslide-affected regions between 2015 and 2021. The dataset is balanced through preprocessing, including synthesizing RGB images from selected bands. It represents a middle ground between the Bijie and CAS datasets in terms of landslide area distribution, making it suitable for both detection and segmentation tasks.

(3) CAS Landslide Dataset: This large-scale dataset includes 20,865 images from nine global regions, integrating satellite and drone data. Specifically tailored for deep learning applications, it captures the aftermath of a debris flow incident in Ganzi County, Sichuan Province, China. The dataset lacks DEM data and focuses on high-resolution UAV imagery, with landslide areas being more scattered. Therefore, it is suitable only for landslide segmentation tasks.

The differences among these datasets—such as the concentration and size of landslide areas, the inclusion of DEM data, and the distribution of landslide regions—highlight the proposed model's adaptability to various scenarios. The Bijie and Landslide4Sens datasets support both landslide detection and segmentation tasks due to their comprehensive data, while the CAS dataset, lacking DEM data, is restricted to segmentation tasks. This diversity in datasets allows for an in-depth exploration of the model's performance in detecting landslides under different conditions.

3.2. Experiments on Dataset 1: The Bijie Landslide Dataset

3.2.1. Landslide Image Recognition Performance

(1) Enhanced Dual-Channel Model

For the task of landslide image recognition, we partitioned the dataset into training and testing sets in an 8:2 ratio. It can be observed that the model has reached a relatively stable convergence state after 20 training epochs (Figure 10).
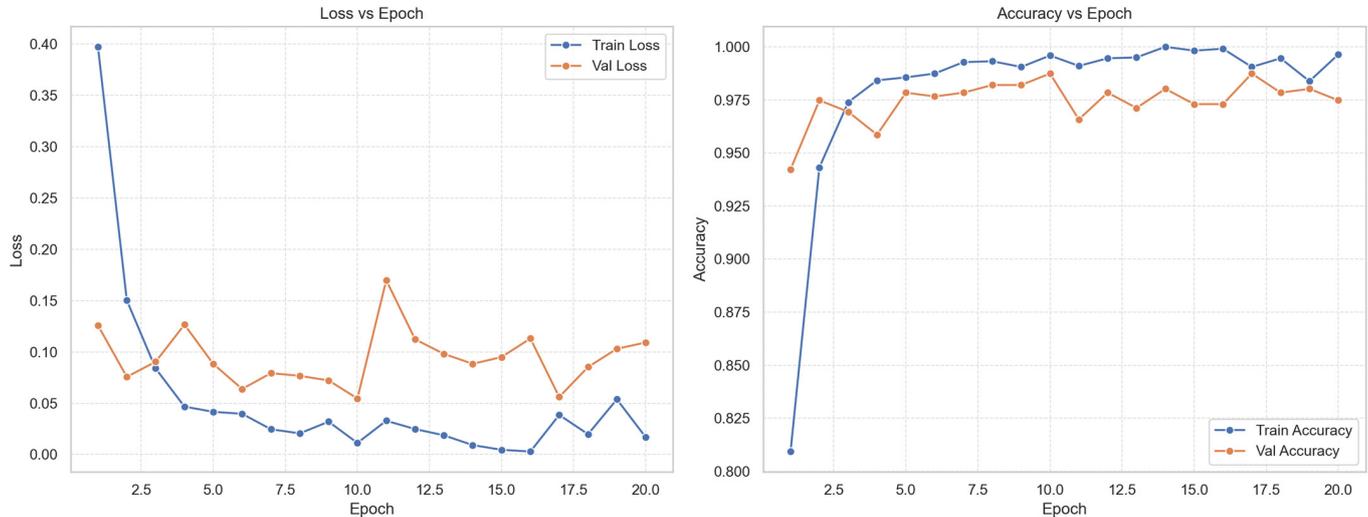
Figure 10. Loss and Accuracy during our model training process.

Our enhanced dual-channel model achieves exemplary performance in classifying landslide and non-landslide images, as evidenced by the precision, recall, and F1 score metrics (Table 1). With precision and recall scores of 0.99 for both classes, the model demonstrates exceptional ability in correctly identifying landslide occurrences while minimizing false positives and negatives. The high F1 score (0.99) underscores its balanced performance in terms of precision and recall, indicating robustness in handling imbalanced class distributions (as shown in Table 2, Figure 10).

Table 2. Classification performance effect of the enhanced dual-channel model on the landslide image recognition task using the Bijie Landslide Dataset.

	Precision	**Recall**	**F1 score**	**Support**
non-landslide	0.99	0.99	0.99	399
landslide	0.98	0.98	0.98	156
Accuracy			0.99	555
macro avg	0.99	0.99	0.99	555
weighted avg	0.99	0.99	0.99	555
non-landslide	0.99	0.99	0.99	399

(2) Comparative Analysis

In comparison to classic models for dual-channel image recognition, our model stands out prominently, exhibiting superior performance across all key evaluation metrics. Notably, while some models achieve competitive results in individual metrics (for example, the combined model of MoibileNet and DenseNet121 and the combined model of EfficientNetB0 and ResNet152V2), none match the comprehensive performance achieved by our proposed approach (as shown in Table 3).

Table 3. Comparison of performance results on image recognition tasks using the Bijie Landslide Dataset.

	Accuracy	**Precision**	**Recall**	**F1 Score**
Restnet50 + CNN	0.8799	0.8894	0.8707	0.8582
VGG16 + InceptionV3	0.9275	0.9288	0.9263	0.9275
ResNet50 + MobileNetV2	0.9275	0.9300	0.9250	0.9275
MobileNetV2 + Xception	0.9495	0.9519	0.9472	0.9459
MoibileNet + DenseNet121	0.9716	0.9744	0.9688	0.9716
MobileNetV2 + InceptionResNetV2	0.8630	0.8785	0.8481	0.8491
DenseNet121 + VGG16	0.8593	0.8694	0.8495	0.8369
InceptionV3 + MobileNetV2	0.9459	0.9459	0.9459	0.9459
VGG19 + MobileNetV2	0.8805	0.8838	0.8772	0.8791
InceptionV3 + DenseNet121	0.9666	0.9675	0.9656	0.9688
MobileNetV2 + Xception	0.7363	0.8382	0.6565	0.6703
EfficientNetB0 + ResNet152V2	0.9538	0.9547	0.9528	0.9559
NASNetMobile + Xception	0.5494	0.5114	0.5936	0.5489
Our model	0.9892	0.9892	0.9892	0.9892

3.2.2. DCT-Unet++ Model Performance in Landslide Image Segmentation

(1) DCT-Unet++ Model

Subsequently, for landslide area extraction, we split the dataset into training and testing sets in a 7:3 ratio and employed our improved DCT-Unet++ model. Our improved DCT-Unet++ model excels in landslide area segmentation (as shown in Figure 11). Our model achieved an IoU of 0.8631, indicating a high degree of overlap between predicted and ground truth regions. Additionally, the Dice coefficient reached 0.9265, further reflecting the model's ability to accurately describe landslide areas with minimal misclassifications. Furthermore, all other metrics of our model also achieved excellent results, further confirming the outstanding performance of our model on this task.
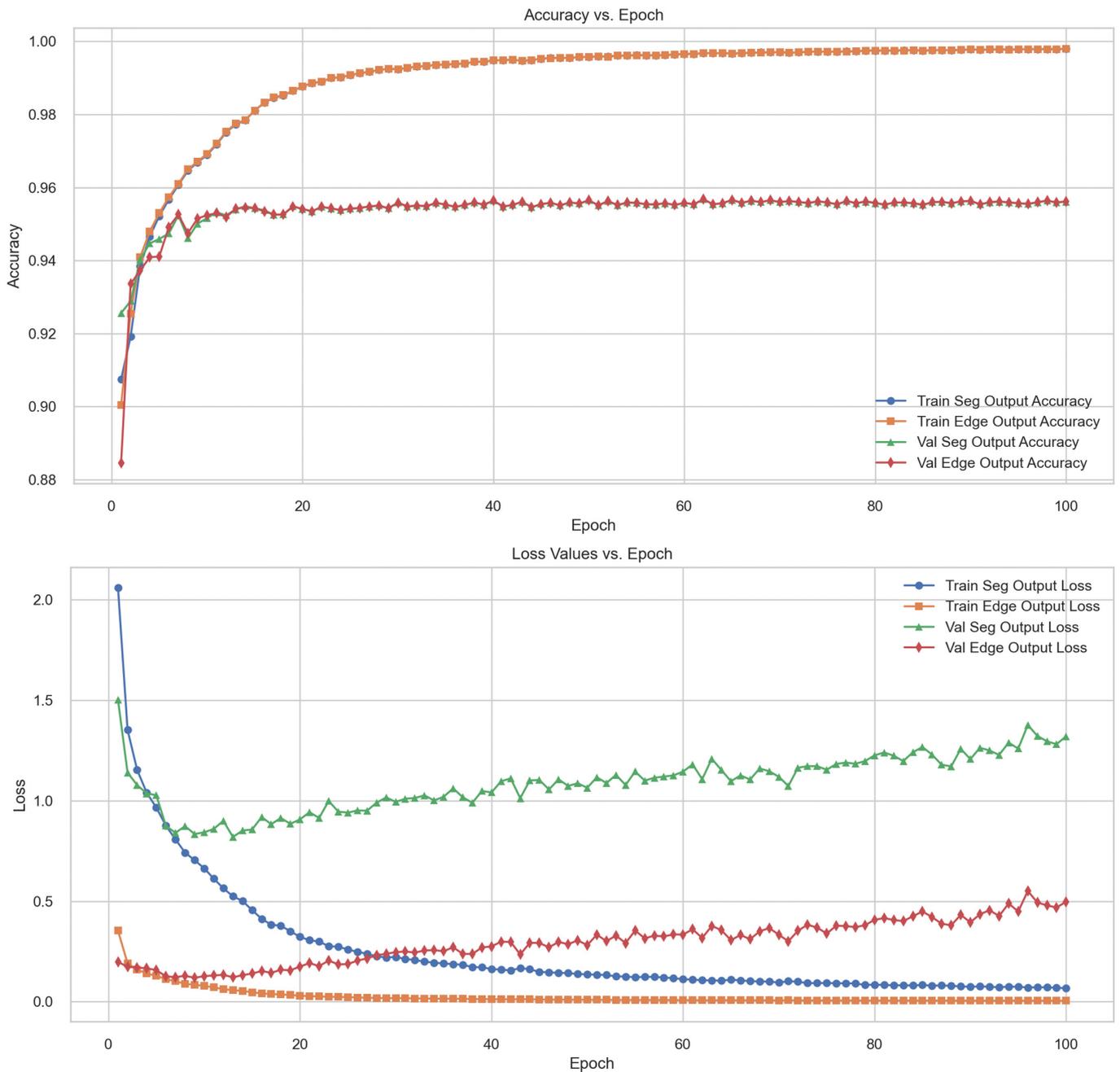
Figure 11. Loss and Accuracy change curve during DCT-Unet++ model training process.

(2)   Comparative Analysis

Comparative evaluation against classic models commonly used for image segmentation tasks reaffirms the superiority of our DCT-Unet++ model. While some models exhibit competitive performance in specific metrics (Unet, PSPNet model), our model consistently outperforms them across a comprehensive range of evaluation criteria (as shown in Figure 12, Table 4).

Our model performed best in terms of IoU (0.8631) and Dice coefficient (0.9265), indicating superior overlap and consistency between predicted and ground truth regions compared to other models. Furthermore, our model demonstrated outstanding performance in accuracy, precision, recall, and F1 score, achieving values of 0.9855, 0.9505, 0.9038, and 0.9265, respectively. This suggests that our model excels in both classification accuracy and the quality of predicting positive and negative samples compared to other models.

Regarding overall accuracy (0.9153) and Kappa coefficient (0.9185), our model also achieved the highest scores, indicating superior classification performance and consistency with ground truth across the entire dataset compared to other models.
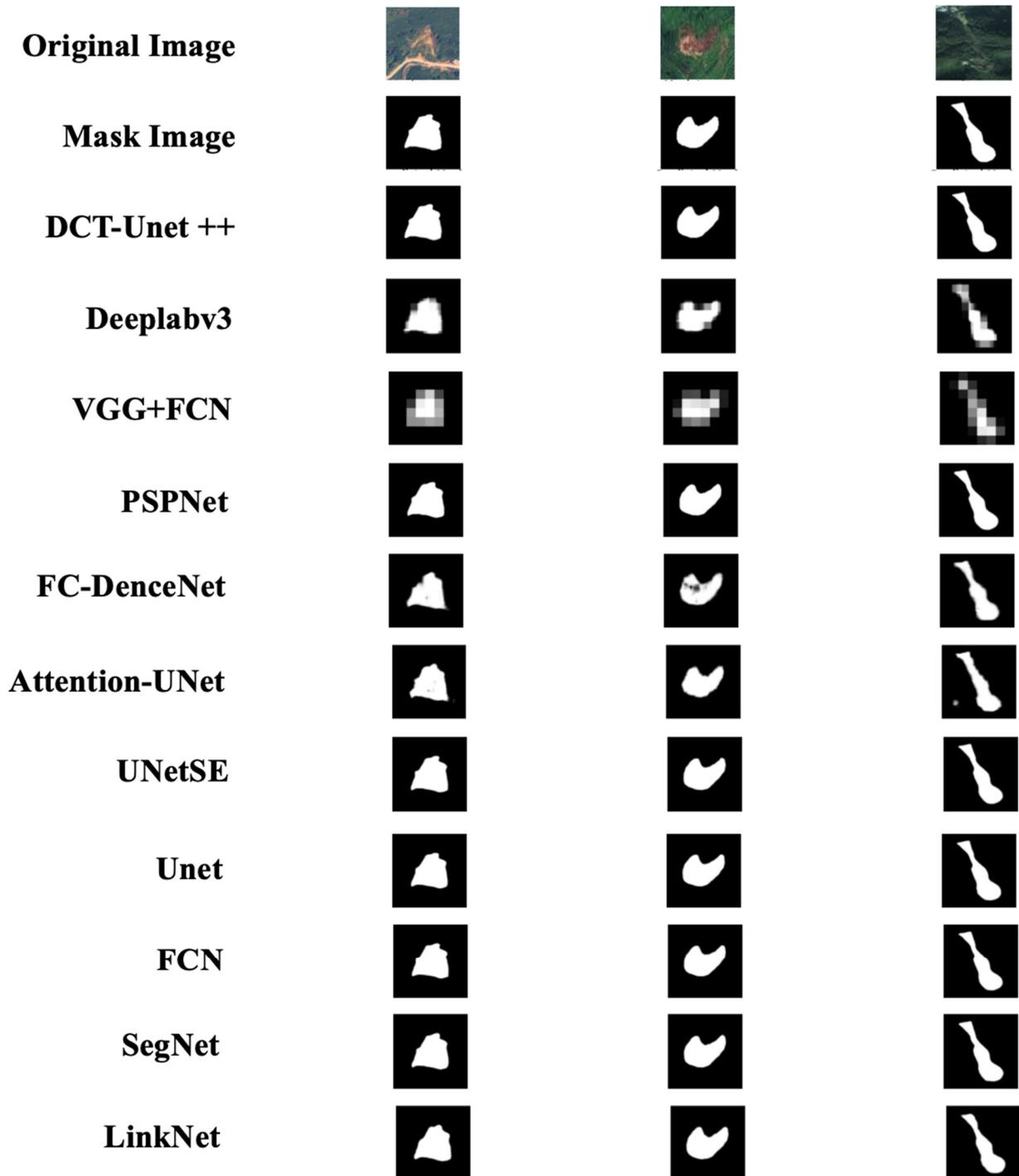

Figure 12. Examples of segmentation results in the Bijie Landslide Dataset.

Table 4. Comparison of performance results on the image segmentation task using the Bijie Landslide Dataset.

	IoU	Dice Coefficient	Accuracy	Precision	Recall	F1 Score	Overall Accuracy (OA)	Kappa Coefficient
UNet	0.8529	0.9206	0.9841	0.9300	0.9115	0.9206	0.9231	0.9118
FCN	0.8436	0.9151	0.9832	0.9377	0.8937	0.9151	0.9050	0.9058
Segnet	0.8185	0.9002	0.9801	0.9155	0.8854	0.9002	0.8966	0.8891
linknet	0.8303	0.9073	0.9815	0.9199	0.8950	0.9073	0.9064	0.8970
Deeplabv3	0.7440	0.8532	0.9716	0.8958	0.8145	0.8532	0.8248	0.8376
VGG16 + FCN	0.6050	0.7539	0.9543	0.8292	0.6911	0.7539	0.6998	0.7289
PSPNet	0.8543	0.9214	0.9844	0.9424	0.9014	0.9214	0.9128	0.9128
FC-DenseNet	0.6962	0.8209	0.9663	0.8883	0.7630	0.8209	0.7726	0.8024
Attention-Unet	0.7267	0.8417	0.9675	0.8308	0.8530	0.8417	0.8638	0.8236
UNetSE	0.8529	0.9206	0.9843	0.9402	0.9018	0.9206	0.9133	0.9119
Our model	0.8631	0.9265	0.9855	0.9505	0.9038	0.9265	0.9153	0.9185

3.3. Experiments on Dataset 2: Landslide4Sense Dataset

3.3.1. Landslide Image Recognition Performance

(1)   Enhanced Dual-Channel Model

Our enhanced dual-channel model achieves commendable performance in landslide image recognition. For the non-landslide image category, the model achieved a high level of accuracy and comprehensiveness in its predictions. For the landslide image category, the model's performance was slightly inferior, but still reached a commendable level. With an overall accuracy of 0.95 (as shown in Table 5), the model showcases robustness in accurately classifying both landslide and non-landslide images, crucial for effective disaster risk management. The macro average (0.93) and weighted average (0.95) further demonstrate the model's balance in handling different classes and its overall accuracy.

Table 5. Classification performance effect of the enhanced dual-channel model on the landslide image recognition task in the Landslide4Sens Dataset.

	Precision	Recall	F1 Score	Support
non-landslide	0.97	0.96	0.96	4434
Landslide	0.90	0.91	0.91	1733
Accuracy			0.95	6167
macro avg	0.93	0.94	0.93	6167
weighted avg	0.95	0.95	0.95	6167
non-landslide	0.97	0.96	0.96	4434

(2)   Comparative Analysis:

Comparative evaluation against classic models for dual-channel image recognition reveals the superior performance of our proposed approach. Our model consistently outperforms other models across key evaluation metrics, exhibiting higher accuracy, precision, and recall, as well as a higher F1 score (as shown in Table 6). Although certain models may have demonstrated outstanding performance on specific metrics, they have not attained the comprehensive performance exhibited by our model.

Table 6. Comparison of performance results in the image recognition task in the Landslide4Sens Dataset.

	Accuracy	Precision	Recall	F1 Score
Restnet50 + CNN	0.8531	0.8641	0.8423	0.8233
VGG16 + InceptionV3	0.9066	0.9044	0.9088	0.9116
ResNet50 + MobileNetV2	0.9194	0.9188	0.9201	0.9173
MobileNetV2 + Xception	0.8995	0.9028	0.8961	0.8931
MoibileNet + DenseNet121	0.9323	0.9315	0.9331	0.9273
MobileNetV2 + InceptionResNetV2	0.8487	0.8561	0.8414	0.8459
DenseNet121 + VGG16	0.8874	0.8633	0.9130	0.8459
InceptionV3 + MobileNetV2	0.9102	0.9101	0.9102	0.9116
VGG19 + MobileNetV2	0.5994	0.5155	0.7160	0.5942
InceptionV3 + DenseNet121	0.9116	0.9086	0.9146	0.9116
MobileNetV2 + Xception	0.9209	0.9231	0.9186	0.9173
EfficientNetB0 + ResNet152V2	0.9315	0.9344	0.9286	0.9301
NASNetMobile + Xception	0.8838	0.8870	0.8806	0.8717
Our model	0.9470	0.9473	0.9470	0.9471

3.3.2. DCT-Unet++ Model Performance in Landslide Area Extraction

(1)   DCT-Unet++ Model

In the landslide area extraction task, our improved DCT-Unet++ model showcases significant performance improvements, achieving an IoU of 0.82, Dice coefficient of 0.90, and overall accuracy of 0.97, with precision and recall scores of 0.91 and 0.90, respectively.

(2)   Comparative Analysis:

In the task of landslide area segmentation, based on dataset 2, our model demonstrates stable and outstanding performance, with most metrics ranking among the top (as shown in Table 7, Figure 13). However, it is noteworthy that on this dataset, the recall rate resides at a moderately elevated level, potentially indicating instances of false negatives. Therefore, there remains room for further enhancement in improving recall. Nonetheless, overall, our model maintains its leading position in the comprehensive evaluation across a range of metrics.

Table 7. Comparison of performance results of the image segmentation task in the Landslide4Sens Dataset.

|	IoU	Dice Coefficient	Accuracy	Precision	Recall	F1 Score	Kappa Coefficient
UNet	0.8197	0.9009	0.9678	0.8932	0.9089	0.9009	0.8817
FCN	0.8241	0.9036	0.9688	0.9008	0.9063	0.9036	0.8850
Segnet	0.8231	0.9030	0.9686	0.8980	0.9080	0.9030	0.8842
linknet	0.8179	0.8998	0.9681	0.9092	0.8906	0.8998	0.8808
Deeplabv3	0.6718	0.8037	0.9381	0.8211	0.7870	0.8037	0.7670
VGG16 + FCN	0.5346	0.6967	0.9110	0.7724	0.6345	0.6967	0.6452
PSPNet	0.8217	0.9021	0.9681	0.8912	0.9133	0.9021	0.8831
FC-DenseNet	0.8153	0.8983	0.9669	0.8900	0.9067	0.8983	0.8785
Attention U-Net	0.8159	0.8986	0.9670	0.8907	0.9067	0.8986	0.8789
UNetSE	0.8270	0.9053	0.9693	0.9002	0.9105	0.9053	0.8870
UNetDC	0.8205	0.9014	0.9688	0.9171	0.8862	0.9014	0.8828
Our model	0.8217	0.9021	0.9686	0.9068	0.8975	0.9021	0.8835
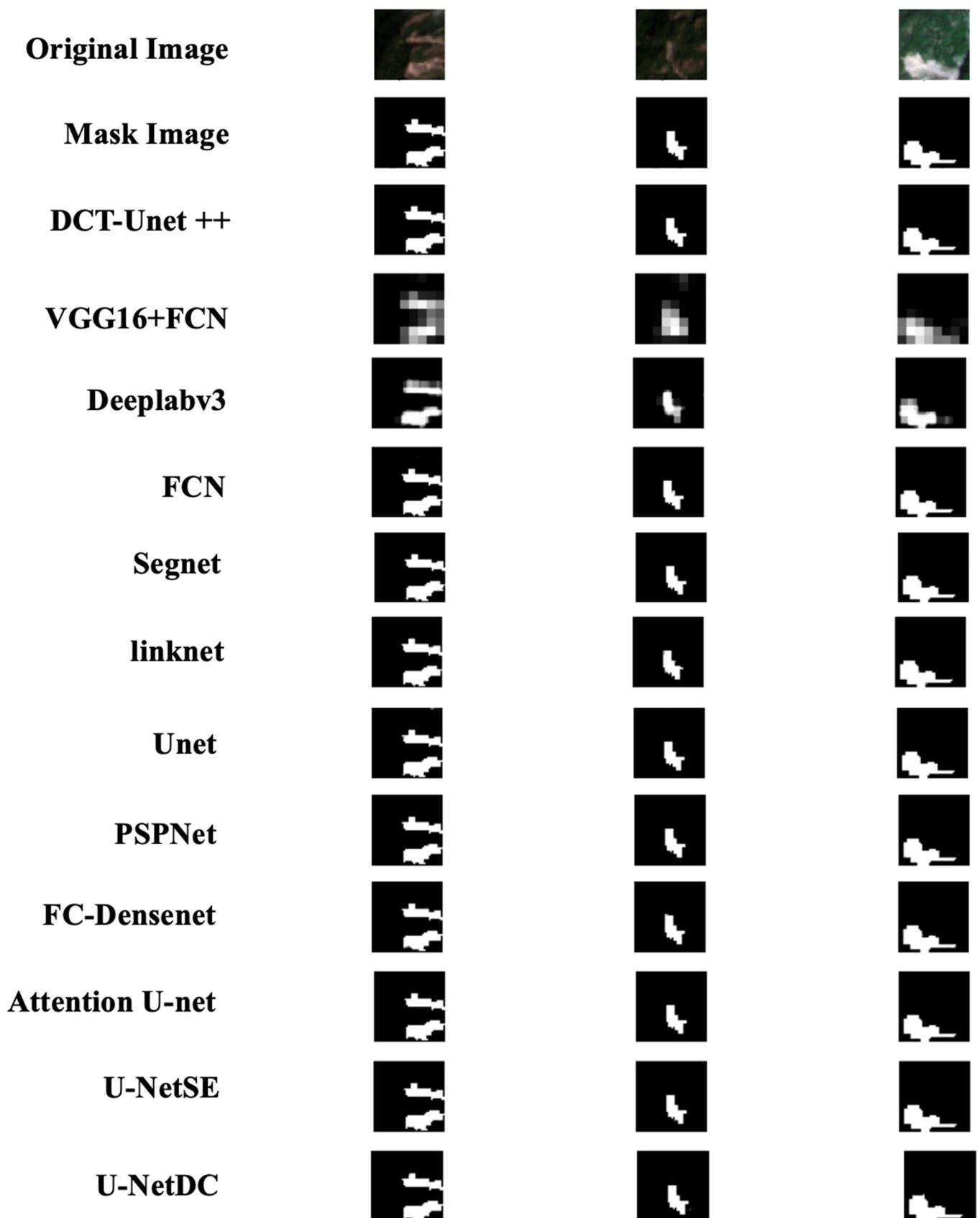
Figure 13. Examples of segmentation results in the Landslide4Sens Dataset.

3.4. Experiments on Dataset 3: CAS Landslide Dataset

In comparison with other models, our model shows its heightened accuracy in pixel-level predictions and improved capability to capture target positions and boundaries. Moreover, in terms of accuracy (0.9319) and Kappa coefficient (0.7996), DCT Unet++ also performs well, signifying a considerable consistency between predicted results and ground truth.

However, in precision (0.8978) and recall (0.7943), DCT Unet++ falls slightly behind some other models, such as Segnet and Linknet, suggesting the presence of misclassifications or missed detections to some extent (as shown in Table 8, Figure 14).
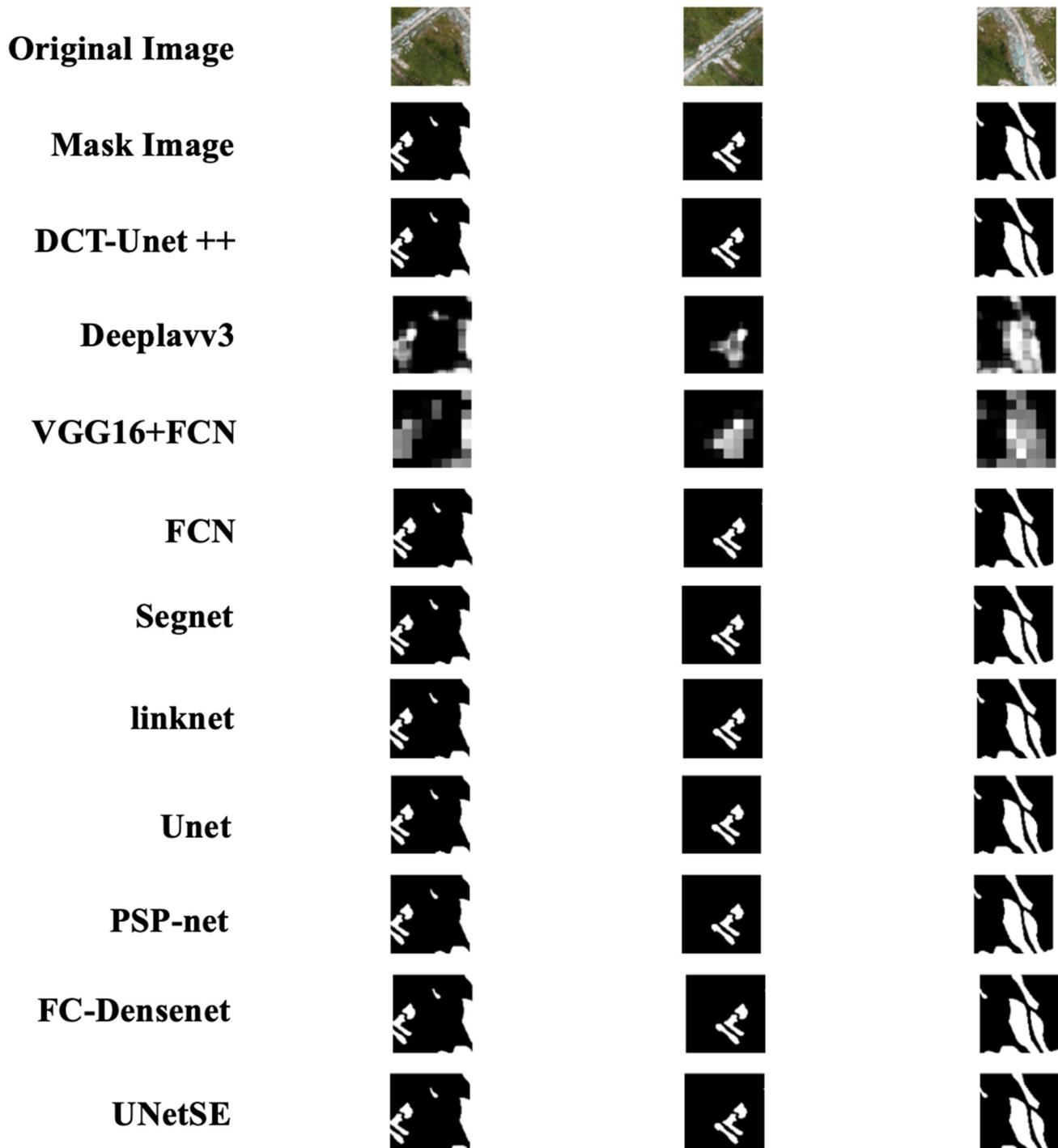

Figure 14. Examples of segmentation results in The CAS Landslide Dataset.

Table 8. Comparison of the performance results of the image segmentation task in the CAS Landslide Dataset.

	IoU	Dice Coefficient	Accuracy	Precision	Recall	F1 Score	Kappa Coefficient
UNet	0.7158	0.8344	0.9293	0.9045	0.7743	0.8344	0.7897
FCN	0.7153	0.8340	0.9287	0.8984	0.7782	0.8340	0.7889
Segnet	0.7468	0.8551	0.9379	0.9242	0.7955	0.8551	0.8159
linknet	0.7452	0.8540	0.9378	0.9286	0.7905	0.8540	0.8148
Deeplabv3	0.6292	0.7724	0.9048	0.8580	0.7023	0.7724	0.7129
VGG16 + FCN	0.5816	0.7355	0.8823	0.7613	0.7113	0.7355	0.6599
PSPNet	0.7177	0.8357	0.9294	0.9001	0.7798	0.8357	0.7910
FC-DenseNet	0.7393	0.8501	0.9358	0.9185	0.7912	0.8501	0.8095
UNetSE	0.7114	0.8314	0.9280	0.9012	0.7716	0.8314	0.7859
Our model	0.7284	0.8429	0.9319	0.8978	0.7943	0.8429	0.7996

Upon analysis, it is believed that the dispersed distribution of landslide areas in this dataset poses a challenge to the model's identification, as DCT Unet++ excels more in detecting concentrated and larger landslide regions. Nonetheless, our model continues to maintain a leading position in overall performance and the accurate identification of areas affected by landslides.

4. Discussion

4.1. Generalizability and Universality of Model Algorithms

The experimental results presented in this study demonstrate the efficacy and robustness of our model algorithms across diverse datasets and geographical regions. The performance analysis on the Bijie Landslide Dataset showcased the exceptional capabilities of our enhanced dual-channel model in landslide image recognition. Achieving precision and recall scores of 0.99 for both landslide and non-landslide instances, the model demonstrated remarkable accuracy and consistency in classifying images. Furthermore, our improved DCT-Unet++ model exhibited superior performance in landslide area extraction, with high IoU, Dice coefficient, accuracy, precision, recall, and F1 score metrics.

By conducting comprehensive evaluations on the other distinct datasets (the Bijie City Dataset, Landslide4Sense Dataset, and CAS Landslide Dataset), the experiment proved the generalizability and universality of our proposed methodologies in landslide detection and segmentation tasks. However, it is important to note that our model's performance on the CAS Landslide Dataset, which includes more dispersed and smaller-scale landslide areas, was slightly lower in precision and recall compared to larger, more concentrated landslides. Specifically, on the CAS dataset, our model achieved a precision of 0.8978 and a recall of 0.7943, which indicates a need for further improvement in these challenging conditions. This performance decline may be due to the difficulty of effectively capturing the features of dispersed and small-scale landslide areas, and future work should focus on improving the model to better identify these complex terrain features.

4.2. Comparative Analysis and Model Superiority

In comparative analyses against classic models for both landslide image recognition and area extraction tasks, our proposed methodologies consistently outperformed alternative approaches. Across various evaluation metrics, including accuracy, precision, recall, F1 score, IoU, and Dice coefficient, our models exhibited superior performance, underscoring their efficacy and reliability. For example, on the Bijie Landslide Dataset, our enhanced dual-channel model achieved an accuracy of 0.9892, surpassing models such as ResNet50 + MobileNetV2 (accuracy of 0.9275) and MobileNetV2 + Xception (accuracy of 0.8995). Similarly, our DCT-Unet++ model demonstrated an IoU of 0.8631 and a Dice coefficient of 0.9265 on the same dataset, outperforming traditional models such as UNet (IoU of 0.8529 and Dice coefficient of 0.9206) and Segnet (IoU of 0.8185 and Dice coefficient of 0.9002).

On the Landslide4Sense dataset, our enhanced dual-channel model's overall accuracy of 0.9470 was higher than MobileNet + DenseNet121 (accuracy of 0.9323) and EfficientNetB0 + ResNet152V2 (accuracy of 0.9315). Additionally, the DCT-Unet++ model achieved an IoU of 0.8217 and a Dice coefficient of 0.9021, outperforming models such as PSPNet (IoU of 0.8217, Dice coefficient of 0.9021) and UNetSE (IoU of 0.8270, Dice coefficient of 0.9053). These results highlight the stability and superior performance of our model across different datasets.

For the CAS Landslide Dataset, our model achieved an IoU of 0.7284 and a Dice coefficient of 0.8429, while Segnet and Linknet achieved slightly higher IoU and Dice coefficients (Segnet: IoU of 0.7468, Dice coefficient of 0.8551; Linknet: IoU of 0.7452, Dice coefficient of 0.8540). Despite these results, our model's accuracy (0.9319) was higher compared to UNet (accuracy of 0.9293) and FCN (accuracy of 0.9287). These findings indicate that while our model excels in most scenarios, further refinement is needed for more complex and dispersed landslide regions.

It is noteworthy that while some models may achieve comparable or nearly equivalent performance to our model on certain evaluation metrics, none can match the overall excellence exhibited by our model. This emphasizes the superiority and effectiveness of our model algorithms in leveraging multi-source satellite remote sensing data for landslide detection and segmentation tasks.

4.3. Future Research Directions

While our study yielded promising results, there are avenues for future research to explore. From the experiments, it is evident that our model excels particularly in the detection of large-scale landslide areas, whereas it may exhibit less precision in segmenting smaller and scattered landslide regions. For instance, in the Landslide4Sense dataset, our model achieved an overall accuracy of 0.9686 but showed a moderate recall rate of 0.8975, indicating instances of false negatives. This phenomenon may be due to the more challenging nature of capturing features of small-scale landslide areas, necessitating more fine-grained feature extraction and enhancement mechanisms to improve the model's recognition capabilities.

Therefore, further refinement of model architecture and optimization strategies to enhance model generalization represents a crucial avenue for future research, enabling the model to achieve superior performance on datasets with unique characteristics. Furthermore, investigations into transfer learning and domain adaptation techniques could facilitate the deployment of our models in diverse geographical regions with varying terrain and environmental conditions. Additionally, the incorporation of additional region-specific feature data obtained through remote sensing, such as slope data, holds promise for significantly improving detection accuracy. Specifically, utilizing high-resolution remote sensing data and more advanced image processing techniques can more accurately capture the subtle changes in landslide areas, thereby improving the detection and segmentation accuracy of the model. Additionally, developing more efficient post-processing algorithms, such as Conditional Random Fields (CRFs), can further refine segmentation boundaries and enhance the overall performance of the model.

By addressing these challenges, future research endeavors can continue to advance the state-of-the-art methods in landslide detection and contribute to the development of more effective disaster management solutions.

5. Conclusions

This paper introduces innovative algorithms for landslide detection and segmentation using multi-source satellite remote sensing imagery. Our methodologies, applied across three distinct datasets—the Bijie City Dataset, Landslide4Sense Dataset, and CAS Landslide Dataset—demonstrate consistent effectiveness and robustness.

Our enhanced dual-channel model, which leverages EfficientNetB7 for feature extraction and incorporates spatial attention mechanisms (SAMs), excels in recognizing landslide

images with impressive precision and recall scores across all three datasets. Additionally, the integration of a deep separable convolutional neural network with a Transformers module for feature extraction from digital elevation data (DEM) and subsequent feature fusion using a variational autoencoder (VAE) represents a significant innovation in the field. This approach not only improves the accuracy of landslide detection but also enhances the model's ability to mine potential features from diverse data sources.

Similarly, our improved DCT-Unet++ model, which incorporates Dilated Convolution to expand the receptive field and enables multi-scale feature extraction, shows superior performance in accurately delineating areas affected by landslides. The integration of Transformer and CBAM attention modules further enhances feature focus, while the introduction of multi-task learning for segmentation and edge detection tasks significantly refines landslide area extraction. Post-processing with conditional random fields (CRFs) further optimizes segmentation boundaries, demonstrating our model's advanced capability compared to traditional segmentation models such as Unet, FCN, and Segnet.

The practical implications of our research are profound for disaster management and risk assessment. By enabling precise identification and detailed mapping of landslide occurrences, our models provide crucial support for early warning systems, urban planning and infrastructure development, insurance and financial services, environmental and land-use management, research and development in geosciences, and many other social aspects.

Looking ahead, future research may focus on refining model architectures, exploring transfer learning techniques, and addressing challenges associated with diverse geographical regions and dataset characteristics. Advancements in landslide detection and segmentation will not only lead to more effective disaster management solutions but will also significantly enhance community safety and resilience in landslide-prone areas.

Author Contributions: Conceptualization, R.S.; Methodology, J.W.; Validation, J.W., Q.Z., H.X. and Y.C.; Data curation, Q.Z.; Writing—original draft, J.W.; Writing—review & editing, R.S.; Visualization, Q.Z., H.X. and Y.C.; Supervision, R.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are openly available in https://github.com/xinxin2021110/DCT-UNETplusplus, accessed on 12 August 2024.

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

References

1. Wagner, C.S. Mental models of flash floods and landslides. *Risk Anal.* **2007**, *27*, 671–682. [CrossRef] [PubMed]
2. Scaioni, M.; Longoni, L.; Melillo, V.; Papini, M. Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. *Remote Sens.* **2014**, *6*, 9600–9652. [CrossRef]
3. Stanley, T.; Kirschbaum, D.B. A heuristic approach to global landslide susceptibility mapping. *Nat. Hazards* **2017**, *8*, 145–164. [CrossRef] [PubMed]
4. Medina, V.; Hürlimann, M.; Guo, Z.; Lloret, A.; Vaunat, J. Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *Catena* **2021**, *201*, 105213. [CrossRef]
5. Guo, Z.; Torra, O.; Hürlimann, M.; Abancó, C.; Medina, V. FSLAM: A QGIS plugin for fast regional susceptibility assessment of rainfall-induced landslides. *Environ. Model. Softw.* **2022**, *150*, 105354. [CrossRef]
6. Wu, L.; Liu, R.; Li, G.; Gou, J.; Lei, Y. Landslide Detection Methods Based on Deep Learning in Remote Sensing Images. In Proceedings of the 2022 29th International Conference on Geoinformatics, Beijing, China, 15–18 August 2022; pp. 1–4. [CrossRef]
7. Bui, T.-A.; Lee, P.-J.; Lum, K.; Loh, C.; Tan, K. Deep Learning for Landslide Recognition in Satellite Architecture. *IEEE Access* **2020**, *8*, 143665–143678. [CrossRef]
8. Neupane, B.; Horanont, T.; Aryal, J. Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis. *Remote Sens.* **2021**, *13*, 808. [CrossRef]
9. Gao, O.; Niu, C.; Liu, W.; Li, T.; Zhang, H.; Hu, Q. E-DeepLabV3+: A Landslide Detection Method for Remote Sensing Images. In Proceedings of the 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 17–19 June 2022; Volume 10, pp. 573–577. [CrossRef]

10.   Cheng, Y.; Li, Y.; Cui, P.; Liang, L.; Pirasteh, S.; Marcato, J.; Gonçalves, W.; Li, J. Accurate landslide detection leveraging UAV-based aerial remote sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *12*, 5047–5060. [CrossRef]

11.   Lissak, C.M.; Leister, A.; Zakharov, I.A. Remote Sensing for Assessing Landslides and Associated Hazards. *Int. J. Remote Sens.* **2020**, *41*, 1391–1435. [CrossRef]

12.   Zhang, W.; Liu, Z.; Yu, H.; Zhou, S.; Jiang, H.; Guo, Y. Comparison of landslide detection based on different deep learning algorithms. In Proceedings of the 2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS), Zhoushan, China, 22–24 April 2022; pp. 158–162. [CrossRef]

13.   Wang, K.; Han, L. A Study of High-Resolution Remote Sensing Image Landslide Detection with Optimized Anchor Boxes and Edge Enhancement. *Eur. J. Remote Sens.* **2023**, 2289616. [CrossRef]

14.   Ye, C.; Li, Y.; Cui, P.; Liang, L.; Pirasteh, S.; Marcato, J.; Gonçalves, W.; Li, J. Landslide Detection of Hyperspectral Remote Sensing Data Based on Deep Learning with Constraints. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 5047–5060. [CrossRef]

15.   Li, H.; He, Y.; Xu, Q.; Deng, J.; Li, W.; Wei, Y.; Zhou, J. Semantic segmentation of loess landslides with STAPLE mask and fully connected conditional random field. *Landslides* **2023**, *20*, 367–380. [CrossRef]

16.   Zhou, N.; Hong, J.; Cui, W.; Wu, S.; Zhang, Z. A Multiscale Attention Segment Network-Based Semantic Segmentation Model for Landslide Remote Sensing Images. *Remote Sens.* **2024**, *16*, 1712. [CrossRef]

17.   Piralilou, S.T.; Shahabi, H.; Jarihani, B.; Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Aryal, J. Landslide Detection Using Multi-Scale Image Segmentation and Different Machine Learning Models in the Higher Himalayas. *Remote Sens.* **2019**, *11*, 2575. [CrossRef]

18.   Soares, L.P.; Dias, H.C.; Garcia, G.P.B.; Grohmann, C.H. Landslide Segmentation with Deep Learning: Evaluating Model Generalization in Rainfall-Induced Landslides in Brazil. *Remote Sens.* **2022**, *14*, 2237. [CrossRef]

19.   Liu, X.; Peng, Y.; Lu, Z.; Li, W.; Yu, J.; Ge, D. Feature-Fusion Segmentation Network for Landslide Detection Using High-Resolution Remote Sensing Images and Digital Elevation Model Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4500314. [CrossRef]

20.   Mohan, A.; Singh, A.K.; Kumar, B.; Dwivedi, R. Review on remote sensing methods for landslide detection using machine and deep learning. *Trans. Emerg. Telecommun. Technol.* **2020**, *32*, e3998. [CrossRef]

21.   Jiang, W.; Xi, J.; Li, Z.; Zang, M.; Chen, B.; Zhang, C.; Liu, Z.; Gao, S.; Zhu, W. Deep Learning for Landslide Detection and Segmentation in High-Resolution Optical Images along the Sichuan-Tibet Transportation Corridor. *Remote Sens.* **2022**, *14*, 5490. [CrossRef]

22.   Chen, X.; Liu, M.; Li, D.; Jia, J.; Yang, A.; Zheng, W.; Yin, L. Conv-trans dual network for landslide detection of multi-channel optical remote sensing images. *Front. Earth Sci.* **2023**, *11*, 1182145. [CrossRef]

23.   Liu, Y.; Zhang, W.; Chen, X.; Yu, M.; Sun, Y.; Meng, F.; Fan, X. Landslide detection of high-resolution satellite images using asymmetric dual-channel network. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4091–4094. [CrossRef]

24.   Shahabi, H.; Rahimzad, M.; Piralilou, S.T.; Ghorbanzadeh, O.; Homayouni, S.; Blaschke, T.; Lim, S.; Ghamisi, P. Unsupervised Deep Learning for Landslide Detection from Multispectral Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 4698. [CrossRef]

25.   Ji, S.; Yu, D.; Shen, C.; Li, W.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [CrossRef]

26.   Travelletti, J.; Delacourt, C.; Allemand, P.; Malet, J.; Schmittbuhl, J.; Toussaint, R.; Bastard, M. Correlation of multi-temporal ground-based optical images for landslide monitoring: Application, potential and limitations. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 39–55. [CrossRef]

27.   Rau, J.; Jhan, J.; Rau, R. Semiautomatic Object-Oriented Landslide Recognition Scheme from Multisensor Optical Imagery and DEM. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1336–1349. [CrossRef]

28.   Dong, C.; Xue, T.; Wang, C. The feature representation ability of variational autoencoder. In Proceedings of the 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, China, 18–21 June 2018; pp. 680–684. [CrossRef]

29.   Wiewel, F.; Yang, B. Continual Learning for Anomaly Detection with Variational Autoencoder. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019 . [CrossRef]

30.   Che, L.; Yang, X.; Wang, L. Text feature extraction based on stacked variational autoencoder. *Microprocess. Microsyst.* **2020**, *76*, 103063. [CrossRef]

31.   Xie, R.; Jan, N.M.; Hao, K.; Chen, L.; Huang, B. Supervised variational autoencoders for soft sensor modeling with missing data. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2820–2828. [CrossRef]

32.   Lafferty, J.; McCallum, A.; Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML '01), Williamstown, MA, USA, 28 June–1 July 2001; pp. 282–289.

33.   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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

34.   Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015. [CrossRef]

35. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.

36. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [CrossRef]

37. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

38. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]

39. Tan, M.; Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.

40. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision (ECCV), Las Vegas, NV, USA, 27–30 June 2016; pp. 630–645.

41. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 8697–8710. [CrossRef]

42. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.

43. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

44. 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]

45. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4. [CrossRef]

46. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]

47. 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 (CVPR), St. Petersburg, FL, USA, 10–13 December 2017; pp. 2881–2890. [CrossRef]

48. Jegou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1175–1183. [CrossRef]

49. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999. [CrossRef]

50. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141. [CrossRef]

51. Ghorbanzadeh, O.; Xu, Y.; Zhao, H.; Wang, J.; Zhong, Y.; Zhao, D.; Zang, Q.; Wang, S.; Zhang, F.; Shi, Y.; et al. The outcome of the 2022 Landslide4Sense competition: Advanced landslide detection from multisource satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9927–9942. [CrossRef]

52. Xu, Y.; Ouyang, C.; Xu, Q.; Wang, D.; Zhao, B.; Luo, Y. CAS Landslide Dataset: A Large-Scale and Multisensor Dataset for Deep Learning-Based Landslide Detection. *Sci. Data* **2024**, *11*, 12. [CrossRef] [PubMed]