land

MDPI

Article

Evaluating Training Parameter Impacts on TransU-Net Performance for UAV-Based Landslide Prediction

Wun Puo Lim [1], Shih Yin Ooi [1,2,*], Yee Jian Chew [1,2], Ying Han Pang [1,2], Sheriza Mohd Razali [3] and Yeong Khang Lee [4]

[1] Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Melaka 75450, Malaysia; lim.wun.puo@student.mmu.edu.my (W.P.L.); yhpang@mmu.edu.my (Y.H.P.)
[2] Centre for Advanced Analytics (CAA), COE for Artificial Intelligence, Multimedia University, Jalan Ayer Keroh Lama, Melaka 75450, Malaysia
[3] Laboratory of Sustainable Bioresource Management, Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, Jalan Universiti 1, Serdang 43400, Malaysia; sheriza@upm.edu.my
[4] ViTrox Technologies Sdn. Bhd., Industrial Park, 746, Persiaran Cassia Selatan 3, Batu Kawan 14110, Malaysia; yeong-khang.lee@vitrox.com
* Correspondence: syooi@mmu.edu.my

Abstract

Landslides are among the most destructive geological hazards in Malaysia, especially in mountainous and forested areas. Unmanned aerial vehicle (UAV) imagery offers high spatial resolution and flexible data capture, but deep learning performance is highly sensitive to training hyperparameters. In this study, the TransU-Net model for UAV-based landslide detection was adopted and a systematic ablation study on learning-rate and epoch settings using a coarse-to-fine tuning strategy. The Berembun Forest Reserve dataset was first used to determine the optimal training configuration. Then, the optimised configuration was tested on multiple UAV sub-datasets in the CAS Landslide dataset to evaluate performance stability under different terrain properties and spatial resolutions. The optimised configuration yielded the best F1-score (0.9598) and IoU of 0.9507 on the Berembun Forest Reserve dataset, and consistently high F1-scores across the evaluated CAS Landslide sub-datasets. Qualitative visualisation analysis also revealed good spatial correspondence between the predicted segmentation masks and the ground-truth annotations. Variations in Intersection over Union (IoU) values were mainly associated with boundary delineation uncertainty rather than severe misclassification. Overall, the results show that the performance of UAV-based landslide segmentation can improve by systematic hyperparameter tuning, and the optimised TransU-Net configuration under the evaluated terrain conditions yields promising results.

Keywords: TransU-Net; UAV imagery; landslide detection; learning rate; epochs

Academic Editor: Fabrizio Terenzio Gizzi

Received: 13 April 2026
Revised: 24 May 2026
Accepted: 25 May 2026
Published: 28 May 2026

1. Introduction

Landslides are among the most destructive natural hazards worldwide and caused by a combination of several factors such as geological conditions, topography, precipitation, earthquakes, and human activities [1,2]. In Malaysia, landslides are particularly prevalent in mountainous and tropical regions [3]. They are also known to have long-term and serious effects on the infrastructure, lives, and the environment. Therefore, landslide detection and terrain monitoring are important to mitigate their impacts, particularly in remote and underdeveloped regions where large-scale monitoring infrastructure may be unavailable [1,4].

Traditional approaches for landslide detection, such as field surveys and satellite-based remote sensing, are usually limited in terms of their accessibility, delayed data acquisition, and spatial resolution. While satellite imagery has advantages for providing an overview of a region, it may not be as effective in detecting small-scale or new landslide features. On the other hand, Unmanned Aerial Vehicles (UAVs) are quickly deployable, can be deployed repeatedly for data acquisition, and can deliver very high-resolution imagery for detailed surface analysis. UAV imagery is particularly useful for identifying geomorphological features associated with landslides, such as landslide scars, exposed soil, vegetation disturbance, and surface deformation. Recent studies have also shown that use of UAV-based technologies for hazardous slope characterisation and terrain heterogeneity evaluation in complex mountainous environments is effective [5]. Furthermore, UAV monitoring systems are a cost-effective solution for initial preliminary landslide monitoring in rural or mountainous areas where more sophisticated geotechnical monitoring systems may not be economically viable.

With the recent growth of high-resolution UAV data, the use of deep learning for landslide detection and segmentation has increased. Among these approaches, U-Net-based encoder–decoder networks have demonstrated strong performance in image segmentation tasks [6]. However, traditional convolution-based models might lack the ability to capture the long-range spatial dependencies due to their mainly local convolutional operations [7]. Recent hybrid approaches combining Convolutional Neural Networks (CNNs) and Transformer-based encoders have shown improved capability in capturing both local and global context from imagery data [8]. One of these hybrid models is TransU-Net, which consists of convolutional feature extraction and Transformer-based global representation learning. Previous studies such as S.-N. Tsai and Tsai (2024) [9] and F. Tsai and Tsai (2025) [10] have reported good results for landslide mapping using satellite images with TransU-Net, but the application of this model to UAV images remains limited. Moreover, most of the studies in the literature use fixed hyperparameters of training and perform limited analysis of the impact of learning rate and the number of epochs on the convergence of the model, the stability of training and the segmentation performance. Furthermore, current research on the training configurations has only been conducted on a limited number of datasets with varying terrain features and spatial resolution.

To address these research gaps, this study examines the effects of the learning-rate and epoch set on the segmentation performance, training stability, and cross-dataset transferability of the TransU-Net model for landslide detection and preliminary hazard assessment based on high-resolution UAV imagery. To systematically assess the impact of learning rate and epoch, an ablation study was performed. The experiments were conducted to evaluate in-domain performance and cross-dataset transferability for the evaluated experimental conditions with UAV imagery collected in the Berembun Forest Reserve and with other sub-datasets of the CAS Landslide dataset. Furthermore, the tuned hyperparameters demonstrated promising performance across the evaluated datasets, but may perform differently on other terrain types, different resolutions, and different types of datasets. The outcomes of this study provide insights into the performance of a fine-tuned Transformer-based segmentation model for high-resolution UAV-based landslide detection for hazard assessment at low cost to support landslide risk management and mitigation.

2. Related Work

2.1. Landslide Detection Using UAV Imagery

Landslide detection using unmanned aerial vehicle (UAV) imagery has become increasingly popular in recent years. B. Chen et al. (2025) stated a few points on the UAV becoming very popular and common in recent years for assessing landslides [1]. The first reason is UAV allow near real-time and flexible survey planning, enabling researchers to quickly respond after landslide events and adapt to changing conditions. In addition, UAV can achieve centimetre- to millimetre-level resolution, which is crucial for detail mapping, surface change detection, and 3D modelling of post-landslide terrain. Moreover, it is cheaper than LiDAR flights or satellite missions while maintaining high-quality outputs and integrates well with machine learning and deep learning. Lastly, B. Chen et al. (2025) also mentioned that UAV bridge the gap between coarse satellite imagery and localised ground measurements and enhance overall landslide analysis [1].

Consequently, numerous studies have adopted UAV imagery for landslide investigation and monitoring. For example, Choi et al. (2023) conducted a study on the possibilities of UAV-LiDAR which could offer rather dense and accurate point clouds in a forest configuration [11]. They discovered that even with the canopy cover, the UAV was able to record topographic variation at the millimetre range which was more superior to the traditional ground methods. In another study, Mercuri et al. (2024) conducted a survey in the Calabrian region of Italy that had already been affected by landslides [12]. Due to the inability to conduct manual surveys, the researchers deployed drones equipped with RGB cameras to determine the landslide area and volume changes, and to make overall inferences. Their study demonstrated that drones could provide reliable information very effectively under relatively limited time and challenging conditions. Sun et al. (2024) also achieved success in the creation of the beneficial digital elevation models (DEMs) with the help of UAV photogrammetry and moving structure reconstruction (SfM) algorithm together with ground control points [4]. These models are mainly applied in the analysis of time-varying terrain dynamics and landslide process dynamics. Additionally, Z. H. Li et al. (2024) carried out a thorough landslide monitoring study with UAV images, GNSS, and fieldwork and proved that the UAV-based SPOT analysis could be useful to record the rapid deformation of landslides at the initial stages when the landslides were caused by the reservoir [13].

In addition, B. Chen et al. (2025) designed a landslide detection model, which relies on UAV-based high-resolution remote sensing images with a backpropagation neural network (BPNN) and multi-feature fusion [1]. Their method initially locates potential landslide zones with imagery of UAVs based on spectral data. Then refines the accuracy by combining spatial shape metrics and change metrics based on the pre- and post-landslide images. The research indicates that the UAV imagery offers accurate data in emergency response scenarios when there is need to check landslides rapidly and accurately. Furthermore, Ya'acob et al. (2025) proposed the application of UAVs imagery to examine the distribution of moisture and temperature variations of the ground on a steep slope of a landslide after heavy rainfall [3]. Their research could predict the beginning of saturation ahead with the help of RGB and thermal indicators, which may cause the stability of the slope. Pan and Chang (2025) used UAVs to capture high-resolution aerial images of mountainous slopes to identify landslide-related anomalies which include cracks, erosion, sedimentation, and slope collapses [14]. Their study integrates a ResNet-50 deep-learning model with HSB-based segmentation to automatically classify slope failures and accurately estimate the affected area. Another related study by Pan et al. (2025) also deployed the UAV imagery to identify landslide signs [15]. Nevertheless, their methodology is not limited to the detection of anomalies. They used a CNN-based recognition model along with a fuzzy inference and ANP decision-support system to classify seriousness of landslide-related defects, namely collapses, scouring, erosion, and sediment build up, and to give a structured maintenance recommendation. He et al. (2025) created a real-time deep-learning framework with a UAV to identify different geological hazards such as landslides in the mountainous regions [16].

They employed high-resolution aerial images obtained with the help of UAVs to make a unique dataset of more than 23,000 labelled instances with landslide categories.

Overall, previous studies have demonstrated that UAV imagery provides flexible and high-resolution data acquisition for landslide detection, monitoring, and terrain analysis under complex environmental conditions. They offer a convenient and scalable way of getting high-resolution and site-specific information that could be applied to emergency response and long-term monitoring. As sensor technology continues to evolve, data analysis methods and UAVs are expected to play an increasingly important role in disaster landslide monitoring and prevention. Therefore, this study applies UAV imagery and deep-learning techniques for landslide detection and preliminary identification of landslide-prone areas.

2.2. Deep Learning for Landslide Detection

Recently, deep learning has become a powerful method for landslide detection because of its ability to learn complex spatial features from remote sensing images automatically. Deep-learning models have the potential to enhance segmentation performance and feature representation in high-resolution UAV imagery, outperforming traditional machine learning methods. In recent years, several deep-learning models have been proven to be successful for landslide segmentation and mapping, such as U-Net, YOLO, Mask R-CNN, and TransU-Net.

For example, Fu et al. (2022) utilised Mask R-CNN and Swin Transformer on UAV images of landslides on the Wenchuan earthquake in China [17]. The results achieved using this model surpass those achieved by using other commonly used backbone networks, such as ResNet-50 and ResNet-101 in segmentation accuracy and object detection. The findings showed that Swin Transformer architecture had achieved accuracy 0.9328 and F1-score 0.9025. X. Liu et al. (2024) proposed an approach to detect landslides based on Mask R-CNN with a Multiscale Samples based Complex Background Enhancement (MSSCBE) [18]. This method reinforces training samples to enable the model to differentiate landslides and other backgrounds better as well as detect small landslides more precisely. The model with MSSCBE to Jiuzhaigou County improved landslide detection performance compared to the original samples and showed good generalisation capability on the Bijie Landslide dataset.

Moreover, Zhang et al. (2024) used LS-YOLO to augment the functionality of the network to recognise the size of multiscale landslides in high-resolution remote sensing images and was an adapted version of YOLOv5 [19]. In addition, their experiment demonstrated superior detection performance compared to existing models such as YOLOv7 and YOLOX. Q. Liu et al. (2023) also introduced a type of model called SE-YOLOv7 in the UAV-based landslide detection case and introduced the Squeeze-and-Excitation (SE) attention blocks to the YOLOv7 framework [20]. They used high-resolution UAV orthophotos collected in the mountains of China and demonstrated that the better model increased the feature representation as the important features were boosted and the noise reduced. This model achieved strong performance in identifying landslide areas from aerial imagery.

In addition, U-Net is one of the most widely used deep-learning architectures for landslide detection, and several studies have successfully applied it to landslide segmentation tasks. For an example, Vega and Hidalgo (2024) used a U-Net that aided in detecting landslides in the Colombian tropical mountains through transfer learning [21]. Their model got the robustness of the U-Net in high-vegetation and complicated terrain environments. Chew et al. (2024) conducted an experiment in the Berembun Forest Reserve located in Malaysia and trained a typical U-Net with ResNet34 backbone to locate the land degradation due to landslides [22]. Their result demonstrated a good level of model generalisation on high-resolution UAV imagery. Next, Sandric et al. (2024) applied U-Net to the data of

UAVs to detect landslides cracks in Romania and demonstrating effective detection of landslide cracks from UAV imagery [23]. Therefore, U-Net demonstrates strong generalisation capability for detecting complex terrain features and landslide cracks from aerial imagery.

Moreover, TransU-Net is a hybrid architecture that combines the strengths of Convolutional Neural Networks (CNNs) and Transformer-based feature extraction for semantic segmentation tasks [8]. TransU-Net has also demonstrated promising performance in landside detection applications. For example, S.-N. Tsai and Tsai (2024) explored the potential of the TransU-Net model in Taiwan on the Laonong River Basin of Taiwan through the usage of Spot7 satellite images to classify the landslides [9]. The analysis combined the multispectral information, panchromatic images, digital surface models, slope, NDVI, and texture features to create a 16-band dataset to be used to train the models. TransU-Net was tested in a two-class scheme, which was between landslide and non-landslide region, and a five-class scheme which classified between unchanged region, non-landslides, old landslides, new landslides, and vegetation reclaim. The two-class model delivered reliable landslide boundary detection performance. The five-class model categorised various classes of landslides as well, but it was less precise with smaller classes because of a small sample size. The authors identified the potential of TransU-Net in mapping landslides and recommended that such a potential should be improved by adopting a two-stage training framework and incorporating more deep-learning models in the future. Similarly, F. Tsai and Tsai (2025) proposed a deep-learning method based on TransU-Net model to recognise and categorise landslide variations based on multi-temporal satellite data of southern Taiwan [10]. The analysis covered the issues of recurring landslides, the recovery of vegetation covering the old landslides, and the problem of manual analysis of large amounts of satellite data. Their approach used a two-stage training strategy and a data-stacking strategy with the help of the transfer learning of three established landslide datasets including Landslide4Sense, HR-GLDD, and the Bijie dataset. The findings indicated that TransU-Net performed well on high-resolution imagery. The model was able to differentiate landslides impacted regions such as old landslides and new landslides and areas where vegetation had already taken over past failures.

Table 1 shows that TransU-Net has demonstrated strong performance in satellite-imagery-based landslide detection, but its application to UAV imagery remains relatively limited. Previous studies have applied CNN-based, U-Net-based, YOLO-based, and Transformer-based frameworks for UAV-based landslide segmentation with generally promising performance. These methods demonstrated improvements in feature extraction, boundary detection, and segmentation accuracy under different terrain conditions. However, many existing studies mainly focus on architectural improvements while providing limited investigation into the influence of training hyperparameters such as learning rate and epoch on segmentation stability and performance. In addition, a small number of studies provide a systematic assessment of how well a tuned hyperparameter setting can be transferred to UAV datasets that include different vegetation cover, spatial resolution, and terrain type. Therefore, further research is needed to better grasp how hyperparameter configurations relate to landslide segmentation performance of a UAV system under different environmental conditions. Accordingly, this study investigates the use of UAV imagery as input to the TransU-Net framework for landslide segmentation and systematically tests the effect of learning-rate and epoch values on segmentation stability and performance on various terrain conditions and spatial resolutions. This approach aims to improve landslide segmentation performance by combining the high spatial resolution of UAV imagery with the feature extraction capability of TransU-Net.

https://doi.org/10.3390/land15060926

Table 1. Summary table of existing deep-learning model in landslide detection.

Authors	Research Work	DL Model	Dataset Type
Fu et al. (2022) [17]	Fast Seismic Landslide Detection Based on Improved Mask R-CNN	Mask R-CNN + Swin Transformer	UAV imagery
X. Liu et al. (2024) [18]	Landslide Detection with Mask R-CNN Using Complex Background Enhancement Based on Multiscale Samples	Mask R-CNN + MSSCBE	Satellite imagery
Q. Liu et al. (2023) [20]	SE-YOLOv7 Landslide Detection Algorithm Based on Attention Mechanism and Improved Loss Function	SE-YOLOv7 (YOLOv7 + SE attention + Varifocal Loss)	UAV imagery
Chew et al. (2024) [22]	Enhancing Land Management through U-Net Deep Learning: A Case Study on Climate-Related Land Degradation in Berembun Forest Reserve in Malaysia	U-Net (ResNet34 backbone)	UAV imagery
Sandric et al. (2024) [23]	Using High-Resolution UAV Imagery and Artificial Intelligence to Detect and Map Landslide Cracks Automatically	U-Net	UAV imagery
S.-N. Tsai & Tsai (2024) [9]	Exploring the Capability of TransU-Net Model for Landslide Classification	TransU-Net (CNN + Transformer)	Satellite images
Zhang et al. (2024) [19]	LS-YOLO: A Novel Model for Detecting Multiscale Landslides With Remote Sensing Images	YOLOv5 (LS-YOLO)	Satellite images
F. Tsai & Tsai (2025) [10]	Landslide Change Detection from Satellite Images with Deep-Learning Classification	TransU-Net (multi-temporal)	Satellite images

Deep-learning models have shown excellent performance in landslide detection and segmentation, but there are still some factors that affect their performance, including feature extractors, attention mechanism, fusion strategy, training configuration, and loss function. New models are often proposed in different studies, or the existing architectures are modified. However, it is not always obvious which elements of the model have been improved. The model may be more effective, but its improvements are not always obvious. Therefore, scientists often need a way to determine which elements are beneficial, which are redundant, and even which may be detrimental to performance. Although many deep-learning architectures have demonstrated promising segmentation performance, comparatively fewer studies investigate how training configurations influence model stability and segmentation accuracy. Therefore, ablation studies are important for understanding the contribution and sensitivity of different model components and training configurations.

2.3. Overview of Ablation Study

An ablation study is a systematic experimental approach and design to evaluate individual contribution and importance of different components within a complex system or model [24]. In addition to that, the primary method of an ablation study is to isolate and quantify by either removing or manipulating content of variables and monitoring the other variables of system behaviour. Its main objective is to determine which components are essential, redundant, or compensatory to a system or model. In other words, an ablation study is a diagnostic approach that breaks down complex architectures into their individual components to reveal the role and necessity of each element. There are several types of ablation studies, such as model architecture ablation, scheduler ablation, normalisation method ablation, hyperparameter tuning ablation, and others.

https://doi.org/10.3390/land15060926

Hyperparameter tuning ablation [25] seeks to examine how manipulated fixed training settings affect model accuracy, convergence speed, and optimisation efficiency. The learning rate, batch size, dropout probability, context window size, choice of optimiser, number of layers, width of each layer, and weight initialisation strategy are examples of hyperparameters that must be specified prior to training and are fixed in every trial. Moreover, learning rate is a vital hyperparameter since it dictates the size of each update of weight and heavily contributes to the smooth and smooth convergence of the model [26]. The batch size balances the randomness of gradient updates and is also interacting with the normalisation layer whereas the probability of dropout determines the strength of regularisation by avoiding co-adaptation between neurons. Architectural hyperparameters, including depth and width, have direct impact on the capacity of the model, whereas the choice of an optimiser such as SGD, Adam, and RMSProp has an impact on the gradient descent path. Weight initialisation also determines the initial state of optimisation behaviour because it assists to prevent the occurrence of both gradient vanishing and gradient explosion. With a set of hyperparameter changes, the ablation experiments can find the best hyperparameter settings and show how the model is sensitive to various training environments.

However, there are relatively few systematic studies that analyse the learning-rate and epoch that affect stability of segmentation results, convergence behaviour, robustness, and cross-dataset generalisation for UAV-based landslide segmentation tasks. Therefore, this study focuses on learning rate and epoch ablation to analyse their effects on model convergence, training stability, and segmentation accuracy.

2.4. Hyperparameter Tuning in Deep Learning

Hyperparameter tuning [27] is a crucial step in deep learning because hyperparameters govern the overall learning behaviour of the model and significantly influence its accuracy, stability, and ability to generalise to unseen data. Unlike learnable parameters such as weights and biases, hyperparameters are predefined before training begins and remain fixed throughout the training process [28]. The common examples include the learning rate, batch size, number of epochs, dropout probability, optimiser type, network depth, network width, and the level of data augmentation applied during preprocessing [27]. In practice, hyperparameter tuning refers to finding the most effective combination of these settings so that the model performs well on validation data [28]. When done properly, tuning helps the model converge more smoothly, reduces issues of underfitting and overfitting, and enhances its ability to capture complex structures within large datasets [26,27]. The hyperparameters are normally categorised based on functionality. The first category is optimisation-related hyperparameters, which entails the learning rate, learning rate schedule, momentum, and weight decay, which are the aspects that directly affect the gradient descent process and the rate at which a model can achieve a good solution [28]. The second type is training-process hyperparameters, such as batch size and the number of epochs, which control how often weights are updated and the amount of computation is required to perform these updates [28]. The third category refers to regularisation and hyperparameters of the architecture, such as dropout rate, number of layers, number of lateral connections, and intensity of the data augmentation, which can be used to trade off the capacity of the model and its generalisation [27]. The learning rate should be optimised to ensure stable convergence, and the number of epochs should be optimised to avoid underfitting or overfitting in the segmentation training process. Therefore, this study conducts an ablation study on the learning rate and the number of epochs

Learning Rate and Epoch Ablation

Learning rate and epochs are important hyperparameters that directly influence the training behaviour and convergence performance of deep-learning models. A learning rate dictates the magnitude of weight updates in each iteration of the optimisation process, and thus the degree of aggressiveness in the model's descent into the loss landscape [26]. If the learning rate is excessively high may result in unstable gradients, oscillatory behaviour, or training divergence. On the other hand, a learning rate that is excessively low may slow down the convergence process, makes the training process significantly longer, and increase the risk to the optimiser becoming stuck in bad local minima [27]. Due to its direct impact on optimisation dynamics, the learning rate is believed to be the most important hyperparameter in the training of deep learning [29]. Next, the number of epochs means how many times the model is repeatedly trained on the entire training data. An underfit model could also result from not enough epochs, as seen in more complex segmentation problems like landslide detection, where the model needs to capture detailed spatial patterns and the boundaries of the terrain are irregular. On the other hand, too many epochs may lead to overfitting, and the model may remember the properties of the training set and poorly predict new and unexplored regions. Thus, to have a good generalisation, it is important to strike a balance adequate between sufficient learning time and avoiding overfitting.

Due to their influence on optimisation behaviour and model stability, learning-rate and epoch ablation studies have become increasingly common in deep-learning research. This type of ablation systematically changes learning-rate values, schedules and training lengths to follow their impact on convergence, stability and ultimate accuracy. Indicatively, Kalra and Barkeshli (2024) showed that the use of warm-up strategy followed by a learning rate decay schedule results in instability reduction in the initial training phases, which allows the model to make use of a higher peak learning rate [29]. The optimisation approach has shown optimal results for both Transformer-based architecture and Convolutional Neural Network (CNN) architecture. Similarly, Al-Ghanimi and Lakizadeh (2025) demonstrated that cosine annealing learning-rate schedules with warm-up phases produced smoother optimisation curves and improved segmentation performance compared to fixed learning-rate training approaches [30].

Recent landslide studies typically employ fixed learning-rate and epoch settings, but systematic investigations into the sensitivity of these training configurations remain limited. Most existing studies primarily focus on architectural improvements or optimisation strategies without comprehensively analysing how different learning-rate and epoch combinations influence segmentation stability and performance. For example, Yang et al. (2022) proposed a variant of TransU-Net called CTransUNet, which includes CBAM (C) and combines ResNet-50 with transformer modules and the model has been trained with SGD, starting with an initial learning rate of 0.1, divided by 20 every 60 entire-epochs [31]. However, the study does not investigate variants of learning-rate or epochs. Similarly, Kariminejad et al. (2024) evaluated half a dozen deep-learning models, such as DeepLabv3+, LinkNet, MA-Net, PSPNet, ResU-Net, and SQ-Net, and measured them on a fixed Adam learning rate of 0.001 across 400 epochs and did not conduct structured learning-rate or epoch sensitivity experiments [32]. Moreover, Zhou et al. (2024) proposed Multiscale Attention Segment Network (MsASNet) model and trained on Adam with a learning rate of 0.001 over 200 epochs [33]. Their ablation results were only in terms of architectural modules and not training hyperparameters. In addition, Wu et al. (2024) proposed a hybrid CNNTransformer model called SCDUNet++ and used AdamW with an initial learning rate of 1×10^{-4} based on cosine decay, but its ablation study did not consider any other learning-rate or epoch schedules [24]. Moreover, Z.-H. Li et al. (2024) applied the UAV-based DeepLab4LS framework in landslide studies and they stated their optimisation and learning-rate parameters but did not include any experiment by varying their learning rate or training duration [34]. Likewise, H. Xu et al. (2025) proposed improved encoder–decoder semantic segmentation network (IEDSSNet) to segment the surface cracks from images with complex backgrounds and trained on AdamW using warm-up, cosine annealing learning-rate schedule on 150 epochs and limited ablation to architecture constituents [35].

2.5. Summary of Findings

Existing research has shown the usefulness of UAV imagery for acquiring flexible and high-resolution data which are relevant for landslide analysis, in the case of complex terrain environments. Previous studies have introduced some deep-learning models such as Mask R-CNN, YOLO, U-Net, and TransU-Net models which have achieved good results in landslide segmentation tasks. However, relatively limited research has investigated the application of TransU-Net for UAV-based landslide imagery. While previous ablation studies tend to focus on architectural components and network optimisation, few studies systematically explore the impact of learning rate and epoch configurations on stability, robustness, and cross-dataset performance in landslide segmentation for UAV applications. In contrast to numerous earlier studies on UAV landslide studies which primarily focused on new network architectures or enhancements of segmentation modules, this study specifically investigates the impact of learning-rate and epoch settings on segmentation stability, convergence behaviour and cross-dataset robustness in UAV-based landslide segmentation. The main contribution of this study is not the development of a new deep-learning architecture, but a systematic evaluation of training hyperparameter sensitivity for UAV-based landslide segmentation using the TransU-Net framework under different terrain characteristics and spatial resolutions. Overall, UAV imagery and deep-learning techniques continue to demonstrate strong potential for high-resolution landslide detection and segmentation. Therefore, this study performs a systematic hyperparameter ablation study by varying learning rate and epoch configurations within the TransU-Net framework for UAV-based landslide segmentation.

3. Methodology

In this study, the TransU-Net architecture for UAV-based landslide segmentation and conducted a systematic hyperparameter ablation study to investigate the effects of learning rate and epochs on segmentation performance. The process was divided into three parts: (1) coarse hyperparameter search, (2) fine learning rate tuning, and (3) cross-dataset testing to test the robustness of the selected hyperparameters. The optimised setting was determined from the experiments to be a fixed learning rate of 2×10^{-4} and 550 epochs. In this study, the following datasets, TransU-Net architecture, training configuration, and evaluation metrics are described.

3.1. Datasets

3.1.1. Berembun Forest Reserve UAV Dataset

The Berembun Forest Reserve UAV dataset is an unmanned aerial vehicle image dataset first introduced by Chew et al. (2024) for land degradation study in Malaysia [22]. The dataset was collected in collaboration with Jabatan Perhutanan Negeri Sembilan (Negeri Sembilan Forestry Department), with drone flight permission obtained on 8 February 2023. It consists of 15 high-resolution aerial images of Berembun Forest Reserve, Jelebu, Malaysia, captured using a DJI Phantom 4 RTK (DJI, Shenzhen, China) equipped with RGB and multispectral sensors. The resolution of each picture was 5472 × 3648 and offered in-depth visual data on slopes with landslides, vegetation in the forest, and the terrain.

In addition, all images were manually labelled to differentiate between land degradation and non-degraded forest areas. The original large images were divided into overlapping 256 × 256 patches following the procedure used by Chew et al. (2024), to improve training efficiency and reduce the possibility of missing smaller degraded areas [22]. To ensure class balance, patches with large areas of degradation were left in place, and some non-degraded patches were selected. After preprocessing, the dataset contained 1697 image patches, consisting of 1272 training patches and 425 validation patches. This overall workflow is depicted in Figure 1.

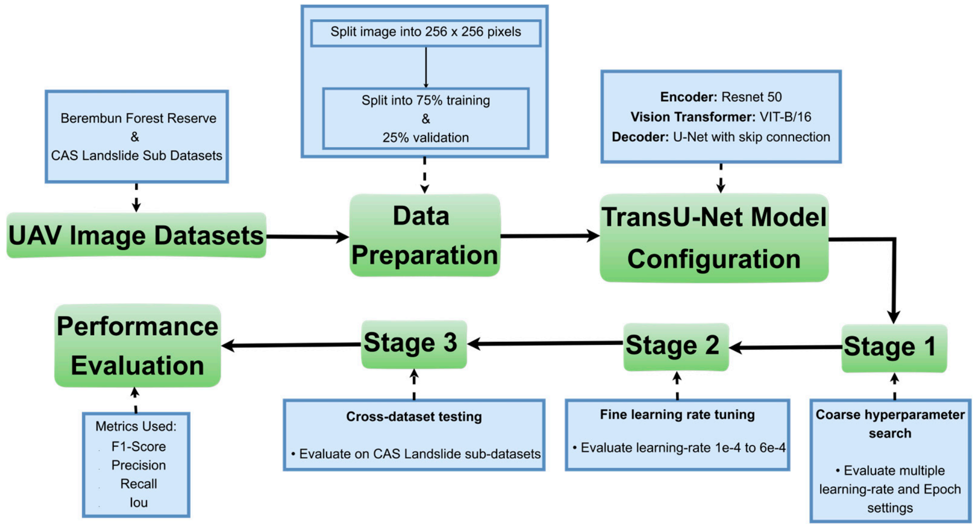

Figure 1. Workflow of the proposed TransU-Net-based UAV landslide segmentation framework.

3.1.2. CAS Landslide Dataset

The second dataset used in this study was the CAS Landslide dataset, which developed by Y. Xu et al. (2023) [36]. The dataset was acquired from 2008 to 2022 and intended for deep-learning landslide detection and segmentation from remote sensing imagery. It consists of over twenty thousand RGB patches each standardised to 512 × 512 pixels and was sampling by nine regions of the mountains with the help of satellite and UAV images platforms. These nine regions are Palu, Hokkaido Iburi-Tobu, Lombok, Tiburon, Peninsula (Haiti), Mengdong Township, Moxitaidi, Moxi Town, Longxi River, and Wenchuan. Although the dataset originated from nine regions, each region may include more than one acquisition type or period, which resulted in a total of sixteen sub-datasets with different spatial resolutions, acquisition dates and imaging sensors. Table 2 have shown the CAS Landslide sub-datasets detail information. Moreover, the binary landslide annotation mask is given to each image patch, with landslide pixels being labelled 1 and non-landslide background pixels being labelled 0. Although the full CAS Landslide dataset includes both satellite and UAV imagery, only UAV-based sub-datasets were used in this study to remain consistent with the research objective. Seven UAV sub-datasets were selected, they are Moxitaidi (UAV-0.6 m), Moxitaidi (UAV-1 m), Moxi Town (0.2 m), Moxi Town (1 m), Longxi River (UAV), Jiuzhai Valley (0.2 m), and Jiuzhai Valley (0.5 m).

Table 2. CAS Landslide dataset with detail information for the sixteen sub-datasets.

Sub-Dataset	Amount	Source	Sensor	Ground Resolution (m)
Palu	817	Digital Globe Open Data Program, Westminster, CO, USA	WorldView2/3	5
Lombok	436	Digital Globe Open Data Program, Westminster, CO, USA	WorldView2/3	5
Hokkaido Iburi-Tobu	1484	Geospatial information Authority of Japan, Tsukuba, Ibaraki, Japan	SAT	3
Tiburon Peninsula (Sentinel)	606	European Space Agency, Paris, France	Sentinel-2	5
Tiburon Peninsula (Planet)	325	Planet Labs PBC, San Francisco, CA, USA	Planet	4
Mengdong	1155	Beijing Lanyu Fangyuan Technology Co., Ltd., Beijing, China	SuperView-1	0.5
Moxitaidi (SAT)	652	European Space Agency, Paris, France	Sentinel-2	0.6
Moxitaidi (UAV-0.6 m)	984	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	0.6
Moxitaidi (UAV-1 m)	483	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	1
Moxi Town (0.2 m)	1635	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	0.2
Moxi Town (1 m)	160	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	1
Longxi River (SAT)	1769	China Centre for Resources Satellite Data and Application, Beijing, China	GF-1	0.5
Longxi River (UAV)	2504	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	0.5
Jiuzhai valley (0.2 m)	5925	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	0.2
Jiuzhai valley (0.5 m)	1752	Sichuan Geomatics Center, Chengdu, Sichuan, China	UAV	0.5
Wenchuan	178	U.S. Geological Survey, Reston, VA, USA	Landsat	5

3.2. Model Architecture

The official implementation of TransU-Net was used as the primary deep-learning model in this study. TransU-Net is a hybrid semantic segmentation architecture which combines a ResNet-50 convolutional encoder to extract local spatial information and a Vision Transformer (ViT-B/16) module to capture long-range contextual information [19]. A U-Net decoder with multiscale skip connections was used to reconstruct high-resolution segmentation maps. This architecture enables capturing the fine boundary details along with the wider contextual information of the scene, which is crucial for landslide segmentation using UAVs in complex terrain. Figure 2 shows how the TransU-Net model worked in the experiment. Due to GPU memory limitations, the UAV imagery was resized into 256×256 input patches instead of using larger input sizes such as 512×512 pixels. This smaller input size could impact the ability of the Transformer module to learn more information about a very large feature at the boundary level, but it offered a practical balance between computational feasibility and boundary-level feature learning. The UAV images and their corresponding binary masks were divided into 75% training data and 25% validation data. During mask conversion into single-channel label maps, nearest-neighbour interpolation was applied to preserve the original class IDs. Landslide pixels were encoded as 1, while background pixels were encoded as 0. The filename matching was also done to make sure that the UAV images were matched with their corresponding masks The model was trained to follow the official R50-ViT-B/16 TransU-Net configuration with a two-class output of landslide and background classes. Spatial detail recovery was resolved by three skip connections that achieve a balance between detail and computation. To support the ViT 16 × 16 patch embedding, the input resolution was set to 256 × 256, resulting in a 16 × 16 token grid. The pre-trained ImageNet-21k weights (R50+ViT-B_16.npz) were used for convergence stability and faster model training on UAV imagery. It was an apt compromise between the recovery of the boundary and reasoning at the level of the scene for the segmentation of complex terrain.

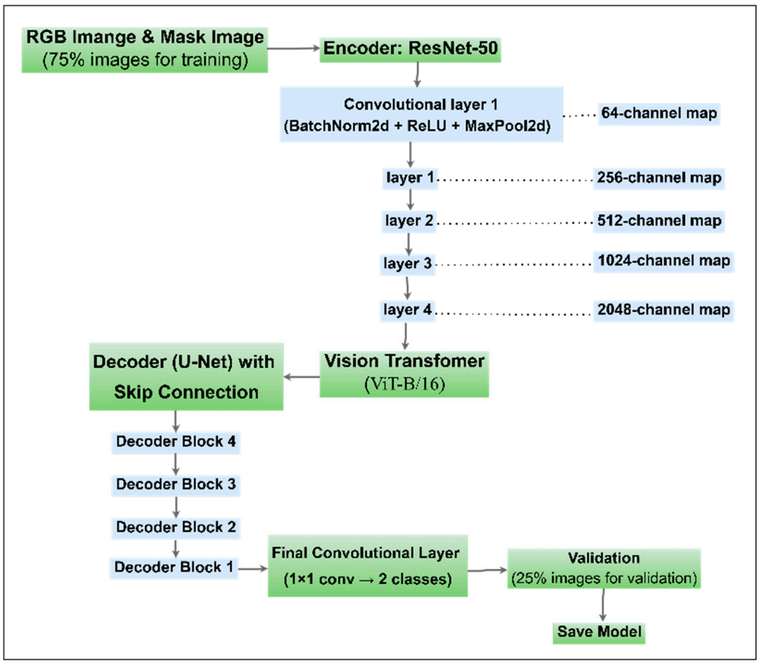

Figure 2. TransU-Net model training and validation stages.

For all experiments, AdamW optimiser was used as it ensures stable convergence and better regularisation in Transformer-based segmentation tasks. The ablation experiments were conducted with a fixed learning-rate configuration so that the effects of the magnitude of the learning-rate and the length of training can be isolated and studied for their direct impact on segmentation performance. The detailed learning-rate and epoch settings are discussed in Section 3.3. In addition, the validation was performed after each epoch. During validation, the model generated per-pixel logits, which were converted into predicted class masks for evaluation. The matrix used for the validation were F1-score and IoU. IoU was used as the main checkpoint selection criterion. The best-performing model was saved as a .pth file when the validation IoU improved, and the final model checkpoint was also saved at the end of training.

3.3. Learning Rate and Epoch Settings

Based on the above studies, learning rate and epochs are one of the primary hyperparameters of the model that affect the convergence, training stability, and accuracy of the final segmentation. Figure 3 shows the three-stage tuning strategy used in this study: coarse learning-rate and epoch exploration, fine learning-rate refinement, and cross-dataset robustness evaluation. This design enabled to explore on a large scale and narrow it down with a fine-tuning approach then confirm the soundness of the chosen hyperparameters with multiple datasets. The first stage conducted an initial coarse sensitivity analysis using the Berembun Forest Reserve dataset. Learning rates of 1×10^{-1}, 1×10^{-2}, 1×10^{-3}, 1×10^{-4}, and 1×10^{-5} were tested with varying training durations from 50 to 600 epochs.

This stage involved a broad range of optimisations, studying how the network converged for various magnitudes of the learning rate and training time, and where stable regions existed for the further optimisation of the network. In Stage 2, the number of epochs was fixed based on the best-performing training duration from Stage 1, and learning rates from 1×10^{-4} to 6×10^{-4} were systematically evaluated to examine optimisation sensitivity under controlled training conditions. In this stage, the sensitivity of the TransU-Net optimisation process was tested under controlled training time by considering smaller variations in learning rate. In the final stage, the optimised learning rate and epoch configuration obtained from the Berembun Forest Reserve dataset was applied to selected UAV-based sub-datasets from the CAS Landslide dataset. The purpose of this stage was not to make sure that the one tuned configuration would provide optimal segmentation results for all the datasets, but to investigate if the same tuned configuration was able to achieve a stable segmentation performance in various terrain characteristics, spatial resolutions, and environmental conditions. To ensure that the observed performance differences were directly associated with the selected learning-rate values and training durations, a fixed learning-rate strategy was intentionally maintained throughout the experiments. Ablation results were not tested with adaptive learning-rate decay schedules like cosine annealing as these schedules add extra optimisation dynamics to the problem which can impact the ablation results. Thus, the fixed learning rate enabled a systematic and controlled investigation of the hyperparameter sensitivities.

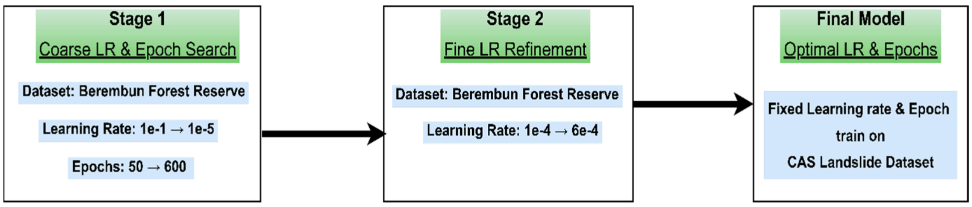

Figure 3. TransU-Net model learning rate and epoch settings stages.

3.4. Evaluation Metrics

This study used two metrics to evaluate the performance: F1-score and Intersection over Union (IoU). These metrics are commonly applied in the field of remote sensing and semantic segmentation due to their complementary nature, offering insights into both pixel classification accuracy and spatial similarity for overlapping areas [17,37]. F1-score is the harmonic mean of precision and recall and is useful for evaluating imbalanced segmentation tasks such as landslide detection [38]. Precision measures the proportion of predicted landslide pixels that are correctly classified, while recall measures the proportion of actual landslide pixels that are correctly detected. Equation (1) shows the mathematical formula for F1-score. IoU is a commonly used metric in computer vision and semantic segmentation for measuring the spatial overlap between predicted and ground-truth regions [39]. It is defined as the ratio of correctly predicted landslide pixels to the union of predicted and actual landslide pixels, as shown in Equation (2). True positive (TP) refers to landslide pixels correctly classified as landslide, false positive (FP) refers to background pixels incorrectly classified as landslide, and false negative (FN) refers to landslide pixels missed by the model. In this study, F1-score was used to evaluate the balance between precision and recall, while IoU was used to assess the spatial agreement between the predicted landslide regions and the ground-truth labels. This distinction is important because in complicated terrain situations, if there is difficulty in defining the boundaries of the model, the F1 score can be high while the IoU values are low.

$$F1score = 2\frac{Precision \times Recall}{Precision + Recall} \tag{1}$$

$$IoU = \frac{True\ Positive}{True\ Positive + False\ Positive + False\ Negetive} \quad (2)$$

4. Experimental

4.1. Coarse Learning Rate and Epochs Analysis

Based on the methodology described in Section 3, the first experimental stage focused on a coarse analysis of learning-rate and epoch sensitivity using the Berembun Forest Reserve UAV dataset. Several experiments were designed to assess the impact of the learning-rate and epoch parameters on the segmentation quality of the Berembun Forest Reserve UAV dataset. Moreover, in this experiment, the model was trained using epoch intervals of 50 from 50 to 600 epochs and learning rate of 1×10^{-1} to 1×10^{-5}. However, due to table formatting and the page limitations, only representative epoch values are reported in Tables 3 and 4.

Table 3. Table of comparison F1-score for epochs 50 to 600 and learning rate 1×10^{-1} to 1×10^{-5} in Berembun Forest Reserve Dataset.

Epochs / **Learning Rate**	**50**	**100**	**200**	**300**	**400**	**500**	**600**
1×10^{-1}/0.1	0.4012	0.7833	0.8654	0.8651	0.8891	0.8896	0.4026
1×10^{-2}/0.01	0.7726	0.8294	0.8683	0.8863	0.9067	0.8724	0.9088
1×10^{-3}/0.001	0.7994	0.8204	0.9008	0.9140	0.9252	0.9289	0.9249
1×10^{-4}/0.0001	0.9232	0.9266	0.9492	0.9491	0.9525	0.9528	0.9523
1×10^{-5}/0.00001	0.8780	0.8888	0.8974	0.9027	0.9112	0.9145	0.9176

Table 4. Table comparison of Intersection over Union (IoU) results for epochs 50 to 600 and learning rate 1×10^{-1} to 1×10^{-5} in Berembun Forest Reserve Dataset.

Epochs / **Learning Rate**	**50**	**100**	**200**	**300**	**400**	**500**	**600**
1×10^{-1}/0.1	0.6699	0.7687	0.8463	0.8487	0.8701	0.8692	0.6696
1×10^{-2}/0.01	0.7458	0.8127	0.8508	0.8688	0.8881	0.8536	0.8919
1×10^{-3}/0.001	0.7782	0.8156	0.8842	0.9013	0.9107	0.9164	0.9116
1×10^{-4}/0.0001	0.9032	0.9122	0.9361	0.9370	0.9406	0.9410	0.9422
1×10^{-5}/0.00001	0.8526	0.8684	0.8758	0.8800	0.8916	0.8958	0.8987

4.1.1. Learning Rate Analysis

Table 3 shows the F1-score results of 50 to 600 epochs for learning rate 1×10^{-1} to 1×10^{-5} by using Berembun Forest Reserve dataset. At the early stage of training with 50 epochs, the learning rate of 1×10^{-4} achieved an F1-score exceeding 0.9232 which is significantly higher than the learning rate of 1×10^{-3} (0.7994) and 1×10^{-2} (0.7726). The large performance gap suggests that the 1×10^{-4} learning rate enabled more stable feature optimisation and improved class separation during training, which allows it to effectively classify landslides and non-landslides with a small number of training steps. Moreover, the 1×10^{-4} learning rate remained stable and steadily improved as the training went on. It reached approximately 0.95 by 200 epochs which indicates a balanced optimisation between the recall and the precision. This behaviour indicates a balance between precision and recall, with an equal reduction of false positive and false negative predictions. The 1×10^{-2} and 1×10^{-3} learning rate also gradually increased the classification performance, although their performance was significantly lower than 1×10^{-4}, which means that they could not perform the task of refining the class separation and providing a reliable predictive performance. Furthermore, as the training time was prolonged, the advantage of the learning rate of 1×10^{-4} was more pronounced. The F1-score was consistently high for 300–600 epochs which shows that the training was stable and converged. On the other hand, the learning rate of 1×10^{-1} had unstable behaviour. Though middle performance was noted at the intermediate epochs, the F1-score decreased drastically at 600 epochs, which denoted divergence and poor learning. The results indicated that high learning rate generates unstable optimisation behaviour, which affected the convergence and decreased the reliability of the segmentation.

Moreover, Table 4 shows the detailed IoU results and provides clear numerical evidence of the impact of learning rate and number of iterations on spatial overlap and segmentation quality. At the initial training phase at 50 epochs, the learning rate of 1×10^{-4} reached a significantly high IoU score of 0.9032 compared to 0.7782 for learning rate 1×10^{-3} and 0.7458 for 1×10^{-2}. This finding shows that the estimated slide areas with the learning rate of 1×10^{-4} already show strong spatial correspondence with the actual masks in the ground-truth, and thus covered most of the target area region at a very early point in the training process. As training continued to 200 epochs, spatial overlap continued to increase with the learning rate of 1×10^{-4}, and the IoU score maximised to 0.9361. Comparatively, the 1×10^{-3} and 1×10^{-2} learning rate achieved an IoU of 0.8842 and 0.8508, respectively, which showed bigger spatial errors and smaller boundary alignment accuracy. These numerical variations indicate that the 1×10^{-4} learning rate was in a better position to refine spatial features in a better way and the predicted segmentation maps were much closer to the shape and size of the landslide regions.

In addition, the learning rate of 1×10^{-4} also continued to demonstrate improved segmentation performance at higher levels of the epoch. The IoU score was 0.9406 and 0.9410 at 400 and 500 epochs, respectively. The spatial overlap was the greatest at 600 epochs with IoU score of 0.9422. The values of IoU were steadily close to at least 0.94 throughout this range, which means that the performance of the segmentation is stable and of high quality. Conversely, the learning rate of 1×10^{-3} generated an IoU of 0.9107 when using 400 epochs and 0.9164 in 500 epochs, while the learning rate of 1×10^{-2} remained below 0.89 even at extended training durations. On the other hand, spatial overlap was severely degraded with an increase in the number of epochs at the learning rate of 1×10^{-1}. Though, the IoU score of 0.8701 was recorded at 400 epochs, at 600 epochs the score significantly decreased to 0.6696 which showed fragmented segmentation results and considerable loss of spatial coherence. This instability shows that too large learning rate are unable to achieve reliable boundary alignment in the long run of training.

Overall, the results of the F1-score and the IoU analyses indicate that there exists a strong and steady connection between the accuracy of classification and the quality of spatial segmentation used. Learning rates that resulted in higher F1-scores also gave better values of IoU, which implies that correct predictions at the class-level directly lead to better spatial overlap and boundary delineations. Specifically, a learning rate of 1×10^{-4} was found to balance optimisation that maximised both accuracy-recall rate and spatial consistency, leading to a solid identification and accurate definition of the area of landslides. Learning rates that were related to unstable or suboptimal classification accuracy, on the other hand, resulted in discontinuous segmentation output and low spatial overlap. Such consistency in the classification performance with the spatial quality indicates the significance of the joint optimisation of predictive performance and spatial quality when learning hyperparameters in the landslide segmentation exercises.

4.1.2. Epochs Analysis

Table 5 shows the summarised F1-score and IoU of training with a fixed learning rate of 1×10^{-4} in epochs of 50 to 600 at 50-epoch increments. These findings suggest that the overall classification accuracy and the quality of spatial segmentation usually increased with the number of epochs which means that the features are learned and optimised through time. During the initial training phase of 50 epochs, the model obtained F1-score 0.9232 and IoU 0.9032 and indicated that the TransU-Net could learn meaningful features of landslides with just a few iterations of training. Next, with the continuation of the training from 100 to 300 epochs, the F1-score and the IoU steadily increased, which suggests ongoing enhancement of discriminative feature learning and spatial boundary optimisation. These improvements indicate that with additional training able to make the model more accurately represent the more intricate spatial properties of landslide areas and less misclassification occurred.

Table 5. Comparison of F1-score and Intersection over Union (IoU) results for different training epochs using a fixed learning rate of 1×10^{-4}.

Learning Rate	Epochs	F1-Score	IoU
1×10^{-4}/0.0001	50	0.9232	0.9032
	100	0.9266	0.9122
	150	0.9412	0.9286
	200	0.9492	0.9361
	250	0.9458	0.9329
	300	0.9491	0.9370
	350	0.9519	0.9413
	400	0.9525	0.9406
	450	0.9549	0.9430
	500	0.9528	0.9410
	550	0.9566	0.9452
	600	0.9523	0.9422

After 400 epochs, the improvement in performance levels off and it can be said that this represents a saturation effect where additional training produces decreasing returns. The best result was obtained on 550 epochs, where the F1-score 0.9566, and the IoU 0.9452. This performance is the best balance between learning ability and generalisation, in which the model obtained its highest classification, and spatial segmentation performance. Unfortunately, at 600 epochs the F1-score and IoU decreased a bit, indicating the beginning of a mild overfitting because of the overtraining. The extended optimisation past the optimal point can result in the model overfitting to the training data, therefore losing generalisation and spatial consistency can be marginally degraded. Comprehensively, the epoch analysis shows that as the number of epochs increases, the model performance reaches its peak at a certain point, and then the gains decrease and the occurrence of overfitting can take place. Due to these observations, it was decided that the best length of the training period was 550 epochs since it gave the best overall performance and the results of segmentation were stable and reliable.

Overall, this analysis ascertains that there is a significant relationship between classification accuracy and the quality of spatial segmentation. The learning rate of 1×10^{-4} and 550 epochs were found to be the best and most reliable combo which achieved consistent convergence, good predictive accuracy, and accurate spatial delineation of landslide regions.

4.2. Learning Rate Ablation

This section discusses a learning rate ablation on Berembun Forest Reserve UAV dataset in the second stage of the experiment. This stage was intended to make further refinements on the optimisation settings found during the coarse analysis. According to the results of Section 4.1, the epochs were set to be 550 epochs, which is the optimal number of epochs in terms of convergence and generalisation ability. Afterwards, the learning rate was altered systematically to analyse how sensitive the TransU-Net model was to refined changes in the optimisation step size. Six learning rate value from 1×10^{-4} to 6×10^{-4} were tested with all the other training settings held constant. The design of the experiment made the observed performance differences directly attributable to change in the learning rate as opposed to training period or data issues. F1-score and IoU, which are the measures of the accuracy of the classification and the quality of the spatial segmentation respectively, were used to perform the evaluation. Table 6 shows the learning rate results from 1×10^{-4} to 6×10^{-4} by using fixed 550 epochs.

Table 6. Comparison of results between 550 epochs and learning rate in Berembun Forest Reserve Dataset.

Epochs	Learning Rate	F1-Score	IoU
550	$1 \times 10^{-4}/0.0001$	0.9566	0.9452
	$2 \times 10^{-4}/0.0002$	0.9598	0.9507
	$3 \times 10^{-4}/0.0003$	0.9563	0.9454
	$4 \times 10^{-4}/0.0004$	0.9500	0.9383
	$5 \times 10^{-4}/0.0005$	0.9446	0.9333
	$6 \times 10^{-4}/0.0006$	0.9421	0.9294

With a learning rate of 1×10^{-4}, this model reached a good baseline performance based on the F1-score of 0.9566 and the IoU of 0.9452. This result confirms the consistency observed in the learning rate at the coarse level analysis. As the learning rate of 2×10^{-4} was used, the model became even better, reaching the highest F1-score of 0.9598 and highest IoU of 0.9507 of all the tested settings. The results at 2×10^{-4} show improved performance, indicating that moderately larger optimisation updates allowed better gradient refinement and better delineation of landslide boundaries. But even further amplifications in the learning rate led to a slow deterioration in performance. At 3×10^{-4}, the F1-score was a little smaller, at 0.9563, and the IoU was smaller, at 0.9454, indicating the start of smaller optimisation efficiency. This downward trend was even greater at 4×10^{-4}, where the F1-score and IoU dropped to 0.9500 and 0.9383, respectively. The deterioration of the performance was more pronounced at higher learning rate of 5×10^{-4} and 6×10^{-4}, where the F1-scores dropped to 0.9446 and 0.9421, and the IoU values dropped to 0.9333 and 0.9294. The results showed that excessively large learning rate caused overly aggressive parameter updates, which reduced segmentation consistency and weakened the model's ability to refine spatial features accurately.

Overall, the learning-rate ablation experiments showed that the 2×10^{-4} configuration achieved the best balance of classification performance and spatial segmentation accuracy of all learning rate explored. Learning rates below this value gave a slightly suboptimal convergence, whereas increasing the learning rate gave unstable optimisation with low segmentation accuracy. According to these results, the learning rate of 2×10^{-4} at 550 epochs was chosen to be the most appropriate to the final experimental stage.

4.3. Experimental Result

Based on above results, the learning rate of 2×10^{-4} with 550 epochs was chosen as the best setup and used in the final stage. In this stage, the TransU-Net model was retrained using the fixed hyperparameters on the Berembun Forest Reserve UAV dataset and subsequently tested on various CAS Landslide UAV sub-datasets to analyse in-domain segmentation performance and the stability of the hyperparameters under different data conditions.

Table 7 shows results of Berembun Forest Reserve dataset and CAS Landslide sub-datasets. Using the optimised configuration, the model was first retrained on the Berembun Forest Reserve dataset. As shown in the Table 7, this dataset achieved an F1-score of 0.9598 and IoU of 0.9507, which indicates that the model achieved strong pixel-level classification performance and relatively accurate landslide boundary localisation on the evaluated dataset. Moreover, the value of high precision of 0.9653 and recall of 0.9579 indicates that the model was capable of accurately balancing between false positive and false negative outcomes. These findings affirm that the chosen learning rate and epochs give consistent convergence and excellent segmentation performance on the dataset that was employed in the ablation experiments. After that, the same trained configuration was used with the other sub-datasets from the CAS Landslide dataset that were derived from UAV platforms. This stage aimed to evaluate whether the optimised configuration could maintain stable segmentation performance across UAV datasets with different geographical characteristics and spatial resolutions. As shown in the Table 7, the TransU-Net model showed stable high classification performance on each sub-dataset that was tested, and F1-scores were between 0.9503 and 0.9858. Specifically, Jiuzhai Valley datasets with 0.2 m and 0.5 m are the datasets that captured the highest F1-scores of 0.9858 and 0.9806, respectively, which implies that the model was very effective at drawing landslides in very high-resolution UAV images. On the same note, good results were also realised in the Moxi Town data and Moxitaidi data and F1-scores were seen to be greater than 0.95. These results indicate that the chosen configuration achieved a relatively stable segmentation accuracy in the various UAV datasets analysed, regardless of the difference in the terrain characteristics, resolution of the images and size of the landslides.

Table 7. Comparison of segmentation performance using F1-score, precision, recall, and Intersection over Union (IoU) metrics of Berembun Forest Reserve Dataset and CAS Landslide sub-datasets.

Datasets	F1-Score	Precision	Recall	IoU
Berembun Forest Reserve	0.9598	0.9653	0.9579	0.9507
Jiuzhai_0.2m	0.9858	0.9853	0.9864	0.9580
Jiuzhai_0.5m	0.9806	0.9810	0.9811	0.9374
Longxi_river	0.9503	0.9613	0.9441	0.8427
moxi_town_0.2m	0.9770	0.9782	0.9774	0.9317
moxi_town_1m	0.9503	0.9613	0.9441	0.8427
Moxitaidi_0.6m	0.9678	0.9759	0.9626	0.8979
Moxitaidi_1m	0.9543	0.9651	0.9469	0.8515

Furthermore, the IoU values showed more variability among the datasets despite the continually high F1-scores. From Table 7, Berembun Forest Reserve dataset achieved a high IoU of 0.9507, lower IoU values were observed for datasets such as Longxi River (0.8427) and Moxitaidi (1 m) (0.8515). The lower IoU values in some datasets might stem from higher terrain complexities, land-cover diversity, and lower spatial resolution, which can lead to difficulties in precise boundary definition. Additionally, the small spatial differences between prediction regions and ground-truth annotations can be caused by an irregular geometry of the landslide and fragmented boundaries, which can affect the IoU measures disproportionately. However, the consistently high precision and recall values indicate that the model reliably identified landslide regions despite minor boundary discrepancies.

Overall, the analysis of F1-score and IoU combined indicates a significant performance feature of the proposed approach. The high F1-scores are consistent, which means that the model is extremely effective at classifying and discriminating between landslides and non-landslides on the pixel level and can work with different datasets. Conversely, the comparatively smaller values of IoU in some datasets are only indicative of level of uncertainty on the boundaries of objects and not misclassification. This indicates that the model is reliably used to detect landslides, but fine-grained boundary delineation is more challenging in data with a coarser resolution or complicated topography. As a result, the optimised training setup allows the TransU-Net model to reach high and stable detection results and indicates room to further enhance the results by introducing new loss functions that consider boundaries or introducing finer-resolution data integration.

Visual Comparison of Segmentation Results

To further evaluate the segmentation capability of the proposed UAV–TransU-Net framework, qualitative visualisation analysis was conducted using representative samples from the Berembun Forest Reserve dataset and CAS Landslide UAV-based datasets, such as Jiuzhai, Longxi River, Moxi Town, and Moxitaidi. Figures 4–9 show the comparisons of the original UAV image with the ground-truth mask and the predicted segmentation result produced by the proposed model. The examples chosen are exposed slope regions, vegetation-rich areas, broken-up landslide structures, and small-scale degraded areas at different UAV image resolutions. The visualisation results from the proposed approach are qualitative, which showed that the framework could identify landslide regions that had relatively high spatial consistency under various terrain conditions. In most examples, the segmentation masks obtained matched well the morphology and the boundary structure of the ground-truth annotations. This implies that the contextual and local spatial information was preserved during landslide segmentation by combining Transformer-based global feature extraction with U-Net style decoder reconstruction.

For the Berembun Forest Reserve dataset (Figure 4), the predicted segmentation outputs showed strong agreement with the manually annotated ground-truth masks, especially in areas with exposed slopes and obvious texture differences between the surface of the landsliding areas and the vegetated areas. The model was able to capture elongated and irregular landslide boundaries while maintaining relatively smooth segmentation continuity. This observation confirms the quantitative results presented above, with higher F1-score values reflecting good identification of landslide areas.

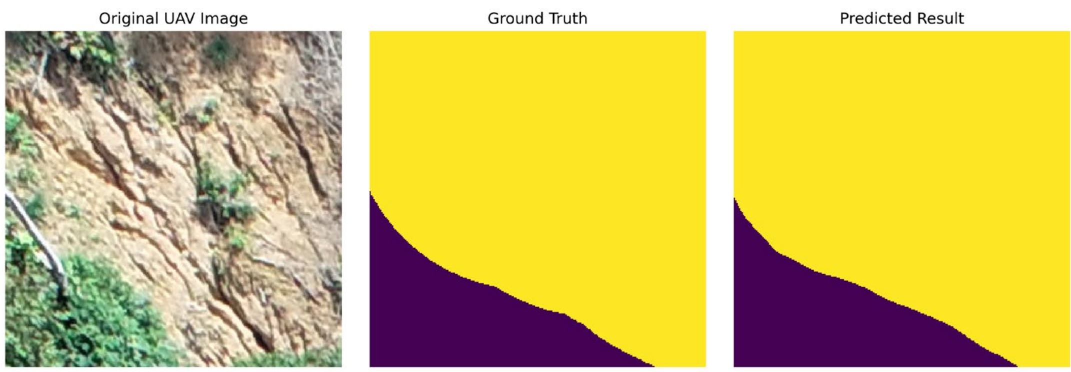

Figure 4. Visual comparison of landslide segmentation results on Berembun Forest Reserve dataset between the original image, ground-truth mask, and the segmented result by the model. Yellow regions represent landslide areas, while purple regions represent background or non-landslide areas.

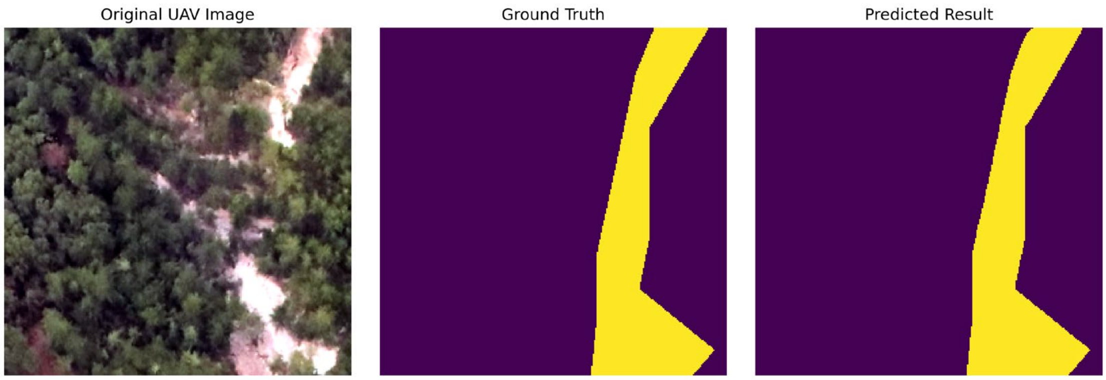

Figure 5. Visual comparison of landslide segmentation results on Jiuzhai Valley UAV (0.2 m) dataset between the original image, ground-truth mask, and the segmented result by the model. Yellow regions represent landslide areas, while purple regions represent background or non-landslide areas.

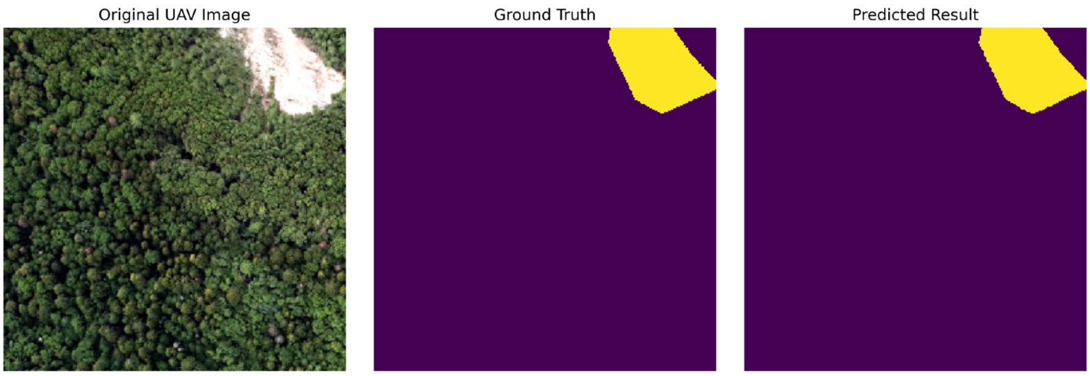

Figure 6. Visual comparison of landslide segmentation results on Jiuzhai Valley UAV (0.5 m) dataset between the original image, ground-truth mask, and the segmented result by the model. Yellow regions represent landslide areas, while purple regions represent background or non-landslide areas.

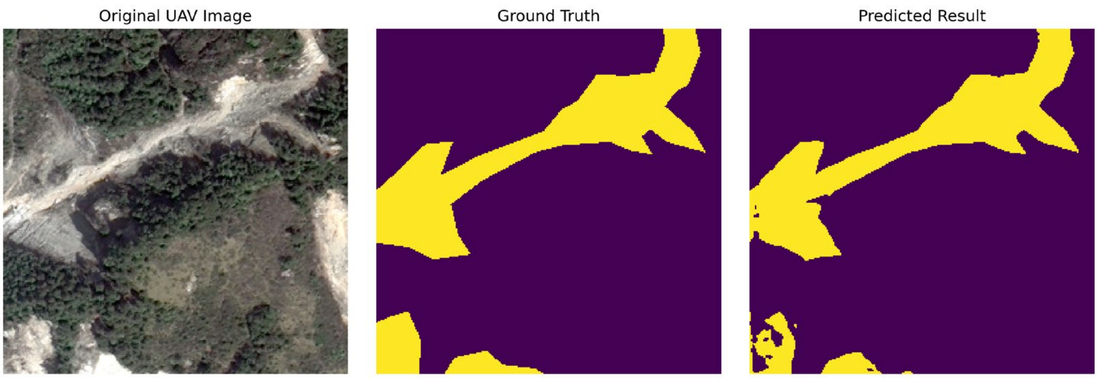

Figure 7. Visual comparison of landslide segmentation results on Longxi River dataset between the original image, ground-truth mask, and the segmented result by the model. Yellow regions represent landslide areas, while purple regions represent background or non-landslide areas.

Next, Figures 5 and 6 present qualitative visualisation results from the Jiuzhai Valley with 0.2 m and 0.5 m resolutions, respectively. The predicted segmentation masks were in good spatial agreement with the ground-truth masks for both datasets, especially in the stretch out boundaries of the landslides and in the exposed degraded areas. Although dense vegetation coverage surrounded several landslide areas, the proposed UAV–TransU-Net framework was still capable of accurately distinguishing landslide regions from the surrounding terrain. A few of the boundary deviations were minor, and the morphology and spatial structure of landslide areas were well maintained. The qualitative observations are further supported by high F1-scores and IoU values obtained during the quantitative evaluation, which suggest the proposed framework achieved stable segmentation performance for various UAV image resolutions and vegetation conditions.

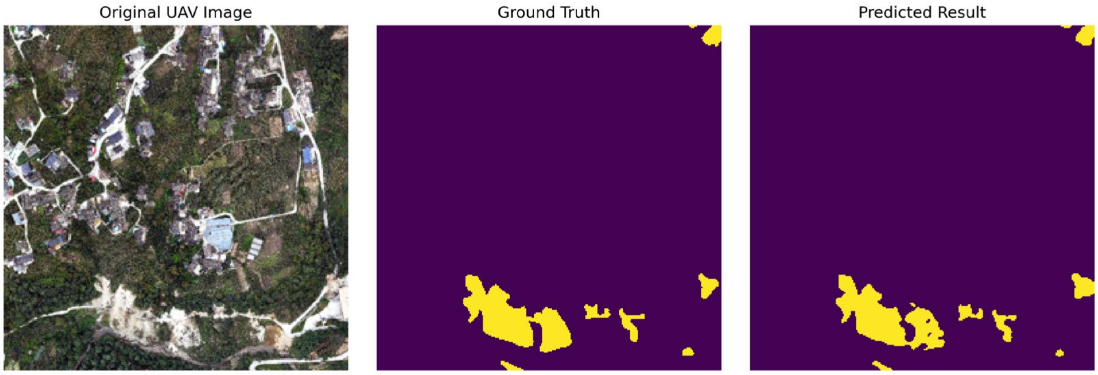

Figure 8. Visual comparison of landslide segmentation results on Moxi Town UAV (1 m) dataset between the original image, ground-truth mask, and the segmented result by the model. Yellow regions represent landslide areas, while purple regions represent background or non-landslide areas.

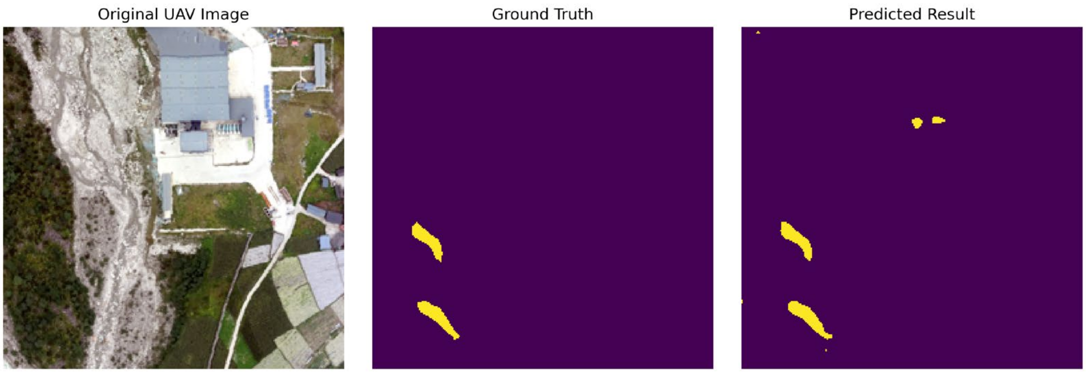

Figure 9. Visual comparison of landslide segmentation results on Moxitaidi UAV (0.6 m) dataset between the original image, ground-truth mask, and the segmented result by the model. Yellow regions represent landslide areas, while purple regions represent background or non-landslide areas.

Figure 7 shows an example of landslide structure segmentation in the Longxi River dataset. The shape of the major degraded areas was well captured with the proposed framework, and the overall spatial continuity of the landslide patterns was maintained. Some over-segmentation artefacts were however observed in the vicinity of intersections of complex boundaries and land regions that are not connected. The results suggest that the irregular shape of landslides and the variable surface roughness of the land are still difficult to use for delineating the boundary accurately. Next, the segmentation results for the Moxi Town dataset (1 m resolution) with several landslide regions that have been segmented into several different parts are provided in Figure 8. This proposed framework was shown to work well in capturing multiple disconnected degraded regions with relative spatial localisation. A few small areas of inconsistent segmentation were identified, but the predicted masks overall maintained the segmentation pattern of the annotated landslides. Lastly, Figure 9 shows qualitative segmentation results from the Moxitaidi 0.6 m resolution dataset. In this example, landslide areas were narrow, located at a low density and these make the accurate segmentation become difficult. Even though the proposed framework targeted a small target, it was still able to detect the main degraded areas with a reasonable spatial agreement. This type of activity is an effective example of the capabilities of the UAV–TransU-Net system to detect small-scale landslide areas from UAV images.

In summary, the qualitative visualisation outcomes shown in Figures 4–9 reveal that the developed UAV–TransU-Net framework was found to be practically viable for landslide segmentation using UAVs under various datasets and terrain contexts. The framework performed relatively well across the exposed slopes, fragmented landslide areas, vegetation-rich areas, and lower resolution UAV images, with reasonable agreement with the ground-truth annotations. This further confirms the quantitative assessment results and proves the applicability of the proposed framework in the initiation of the landslide hazard assessment and monitoring of the terrain by UAV imagery.

5. Conclusions and Future Work

In conclusion, this study proposed the TransU-Net model with UAV imagery framework for landslide-region segmentation collected from the Berembun Forest Reserve dataset and sub-datasets of CAS Landslide datasets. This framework combined the global feature extraction capability of the Transformer with a U-Net based decoder architecture which preserved the contextual and local spatial information during segmentation process. The experimental results showed that the framework achieved a high segmentation rate in various datasets and terrain conditions, such as vegetated areas, fragmented landslide areas, and varying UAV image resolutions. Moreover, the quantitative evaluation results showed that the proposed framework achieved relatively high F1-score and IoU performance for multiple datasets, and the qualitative visualisation analysis also showed that the predicted segmentation mask and the ground-truth annotations had good spatial similarity. The model generally maintained the general shape and boundary delineation of landslide areas with elongated, fragmented, and vegetated topography features. The results indicated that the incorporation of global attention generated by Transformer and spatial reconstruction based on U-Net helped achieve robust segmentation performance in complex UAV imaging scenarios. The results of this study prove the practicality of using UAV-based landslide segmentation in the preliminary hazard assessment and terrain monitoring application. The proposed framework may support more accessible and lower-cost landslide-region identification in areas that are difficult to access using traditional field investigation approaches. This ability could potentially benefit local authorities, developers, and communities in determining regions of degraded slopes and enhance their early observations of the terrain.

Although the segmentation performance obtained in this study is encouraging, there are still a few limitations. Factors that affected the segmentation accuracy include dense vegetation cover, complex terrain structures, varying resolutions of the images, and small-scale fragmented landslide regions. However, minor deviation from the boundaries and spatial inconsistencies between the predicted segmentation masks and ground-truth annotations were still seen in a few difficult samples. These restrictions suggest that the segmentation of landslides from UAV images is a difficult task in heterogeneous environments. While the suggested framework showed good performance over the test sets used, further validation is still required in other geographical regions and in different environmental conditions. In addition, it should also be noted that this study primarily focused on analysing the sensitivity of learning rate and epoch parameters within the TransU-Net framework for UAV-based landslide segmentation rather than conducting a comprehensive benchmarking comparison across multiple segmentation architectures. Therefore, the current investigation did not involve comprehensive comparisons to other deep-learning models like U-Net, DeepLabV3+, or Mask R-CNN. To provide broader benchmarking analysis, future studies can also compare the performance of various segmentation architectures when trained using the same UAV-based training scenarios. Furthermore, the segmentation results in the terrain with dense vegetation and highly heterogeneous surfaces could be further enhanced by considering multiscale feature extraction and adaptive attention mechanisms. Future research could also explore using a greater number of UAV datasets from a variety of geographical locations and environmental conditions to enhance the model's ability to generalise. In addition, future studies could explore the optimisation of hyperparameters within specific datasets as well as the cross-domain sensitivity analysis to further assess the transferability and robustness of the chosen training configuration. Next, the computational constraint in this study meant that the input patch size was set to 256 $\times$ 256 pixels, which could constrain the model's ability to represent larger spatial context in large landslide areas. Future research could explore higher-resolution inputs such as 512 $\times$ 512 pixels with greater computational resources to further improve boundary delineation and contextual feature learning. Lastly, it is possible to test in the future lightweight or real-time versions of TransU-Net and test on other regions or multi-temporal UAV datasets to assist with operational landslide monitoring.

Author Contributions: Conceptualisation, W.P.L. and S.Y.O.; methodology, W.P.L., S.Y.O., and Y.J.C.; software, W.P.L.; validation, W.P.L., S.Y.O., Y.J.C., and Y.H.P.; formal analysis, W.P.L. and Y.K.L.; investigation, W.P.L.; resources, S.M.R.; data curation, W.P.L., S.M.R., and Y.K.L.; writing—original draft preparation, W.P.L.; writing—review and editing, S.Y.O., Y.J.C., Y.H.P., S.M.R., and Y.K.L.; visualisation, W.P.L.; supervision, S.Y.O. and Y.J.C.; project administration, S.Y.O.; funding acquisition, S.Y.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Multimedia University, grant number MMUI/240102 and the APC was funded by Multimedia University.

Data Availability Statement: The data supporting the findings of this study are available as follows: (1) The CAS Landslide sub-datasets (Jiuzhai valley [0.2 m, 0.5 m], Longxi River, Moxi Town [0.2 m, 1 m], and Moxitaidi [UAV-0.6 m, UAV-1 m]) are publicly available in Zenodo at https://zenodo.org/records/10294997 (accessed on 11 July 2025); (2) The Berembun dataset used in this study contains landscape imagery of a protected forest area in Malaysia and is subject to privacy and ethical restrictions to ensure the conservation and security of the site. Consequently, these data are not publicly available but may be available from the corresponding author (S.Y.O.) upon reasonable request and with the permission of the relevant authorities.

Acknowledgments: During the preparation of this manuscript, the authors used Grammarly (v1.2), ChatGPT (GPT-4), and Gemini (3.1 Pro) for the purposes of improving language clarity and refining the technical phrasing of the manuscript. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Author Yeong Khang Lee was employed by the company ViTrox Technologies Sdn. Bhd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
CNN	Convolutional Neural Network
SE	Squeeze-and-Excitation
MSSCBE	Multiscale Sample Complex Background Enhancement
IoU	Intersection over Union
YOLO	You Only Look Once

References

1. Chen, B.; Maurer, J.; Gong, W. Applications of UAV in Landslide Research: A Review. *Landslides* **2025**, *22*, 3029–3048. [CrossRef]

2. Ehsan, M.; Anees, M.T.; Bakar, A.F.B.A.; Ahmed, A. A Review of Geological and Triggering Factors Influencing Landslide Susceptibility: Artificial Intelligence-Based Trends in Mapping and Prediction. *Int. J. Environ. Sci. Technol.* **2025**, *22*, 17347–17382. [CrossRef]
3. Ya'acob, N.; Rahman, A.A.A.; Yusof, A.L.; Ali, D.M.; Naim, N.F. Landslide Detection Using Analysed UAV Imagery. *J. Adv. Res. Appl. Sci. Eng. Technol.* **2025**, *45*, 168–188. [CrossRef]
4. Sun, J.; Yuan, G.; Song, L.; Zhang, H. Unmanned Aerial Vehicles (UAVs) in Landslide Investigation and Monitoring: A Review. *Drones* **2024**, *8*, 30. [CrossRef]
5. Liu, C.; Bao, H.; Wang, T.; Zhang, J.; Lan, H.; Qi, S.; Yuan, W.; Koshimura, S. Intelligent Characterization of Discontinuities and Heterogeneity Evaluation of Potential Hazard Sources in High-Steep Rock Slope by TLS-UAV Technology. *J. Rock. Mech. Geotech. Eng.* **2026**, *18*, 509–527. [CrossRef]
6. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597. [CrossRef]
7. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual, 3–7 May 2021.
8. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306. [CrossRef]
9. Tsai, S.-N.; Tsai, F. Exploring the Capability of TransUNet Model for Landslide Classification. In *Proceedings of the 45th Asian Conference on Remote Sensing (ACRS 2024), Colombo, Sri Lanka, 17–21 November 2024*; Asian Association on Remote Sensing (AARS): Colombo, Sri Lanka, 2024.
10. Tsai, F.; Tsai, S.-N. Landslide Change Detection from Satellite Images with Deep Learning Classification. In *Proceedings of the European Geosciences Union General Assembly 2025 (EGU25), Vienna, Austria, 27 April–2 May 2025*; European Geosciences Union (EGU): Vienna, Austria, 2025.
11. Choi, S.K.; Ramirez, R.A.; Kwon, T.H. Acquisition of High-Resolution Topographic Information in Forest Environments Using Integrated UAV-LiDAR System: System Development and Field Demonstration. *Heliyon* **2023**, *9*, e20225. [CrossRef]
12. Mercuri, M.; Biondino, D.; Ciurleo, M.; Cofone, G.; Conforti, M.; Gullà, G.; Stellato, M.C.; Borrelli, L. The Use of an Unmanned Aerial Vehicle (UAV) for First-Failure Landslide Detection. *GeoHazards* **2024**, *5*, 683–699. [CrossRef]
13. Li, Z.H.; Jiang, N.; Shi, A.C.; Zhao, L.Y.; Xian, Z.; Luo, X.L.; Li, H.B.; Zhou, J.W. Reservoir Landslide Monitoring and Mechanism Analysis Based on UAV Photogrammetry and Sub-Pixel Offset Tracking: A Case Study of Wulipo Landslide. *Front. Earth Sci.* **2024**, *11*, 1333815. [CrossRef]
14. Pan, N.H.; Chang, B.Y. An Integrated UAV and Deep Learning Framework With HSB-Based Segmentation for Automated Slope Anomaly Detection in Mountainous Soil and Water Conservation Sites. *Adv. Civ. Eng.* **2025**, *2025*, 5580438. [CrossRef]
15. Pan, N.H.; Chang, H.C.; Chen, K.Y. A Hybrid Fuzzy Decision Support Framework for UAV-Based Durability Assessment of Civil Infrastructure. *IEEE Access* **2025**, *13*, 186408–186423. [CrossRef]
16. He, L.; Zhou, Y.; Lin, S.; Hong, Y. Real-Time Uav-Based Deep Learning for Multi-Class Geohazard Detection and Infrastructure Risk Monitoring in Mountainous Terrain. 2025. Available online: https://ssrn.com/abstract=5363637 (accessed on 16 November 2025).
17. Fu, R.; He, J.; Liu, G.; Li, W.; Mao, J.; He, M.; Lin, Y. Fast Seismic Landslide Detection Based on Improved Mask R-CNN. *Remote Sens.* **2022**, *14*, 3928. [CrossRef]
18. Liu, X.; Xu, L.; Zhang, J. Landslide Detection with Mask R-CNN Using Complex Background Enhancement Based on Multi-Scale Samples. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2300823. [CrossRef]
19. Zhang, W.; Liu, Z.; Zhou, S.; Qi, W.; Wu, X.; Zhang, T.; Han, L. LS-YOLO: A Novel Model for Detecting Multiscale Landslides With Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4952–4965. [CrossRef]
20. Liu, Q.; Wu, T.; Deng, Y.; Liu, Z. SE-YOLOv7 Landslide Detection Algorithm Based on Attention Mechanism and Improved Loss Function. *Land* **2023**, *12*, 1522. [CrossRef]
21. Vega, J.; Hidalgo, C. Evaluation of U-Net Transfer Learning Model for Semantic Segmentation of Landslides in the Colombian Tropical Mountain Region. *MATEC Web Conf.* **2024**, *396*, 19002. [CrossRef]
22. Chew, Y.J.; Ooi, S.Y.; Mohd-Razali, S.; Pang, Y.H.; Lim, Z.Y. Enhancing Land Management through U-Net Deep Learning: A Case Study on Climate-Related Land Degradation in Berembun Forest Reserve in Malaysia. *JOIV Int. J. Inform. Vis.* **2024**, *8*, 2419–2427. [CrossRef]
23. Sandric, I.; Chitu, Z.; Ilinca, V.; Irimia, R. Using High-Resolution UAV Imagery and Artificial Intelligence to Detect and Map Landslide Cracks Automatically. *Landslides* **2024**, *21*, 2535–2543. [CrossRef]
24. Wu, L.; Liu, R.; Ju, N.; Zhang, A.; Gou, J.; He, G.; Lei, Y. Landslide Mapping Based on a Hybrid CNN-Transformer Network and Deep Transfer Learning Using Remote Sensing Images with Topographic and Spectral Features. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *126*, 103612. [CrossRef]

25. Meyes, R.; Lu, M.; de Puiseau, C.W.; Meisen, T. Ablation Studies in Artificial Neural Networks. *arXiv* **2019**, arXiv:1901.08644. [CrossRef]
26. Smith, L.N. A Disciplined Approach to Neural Network Hyper-Parameters: Part 1—Learning Rate, Batch Size, Momentum, and Weight Decay. *arXiv* **2018**, arXiv:1803.09820. [CrossRef]
27. Yu, T.; Zhu, H. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *arXiv* **2020**, arXiv:2003.05689. [CrossRef]
28. Bischl, B.; Binder, M.; Lang, M.; Pielok, T.; Richter, J.; Coors, S.; Thomas, J.; Ullmann, T.; Becker, M.; Boulesteix, A.L.; et al. Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2021**, *13*, e1484. [CrossRef]
29. Kalra, D.S.; Barkeshli, M. Why Warmup the Learning Rate? Underlying Mechanisms and Improvements. *arXiv* **2024**. [CrossRef]
30. Al-Ghanimi, H.H.O.; Lakizadeh, A. A CNN-Based Framework for Enhancing Image Segmentation and Scene Interpretation in Fake Colorized Images: Addressing Challenges in Image Analysis and Understanding. *Ing. Des. Syst. D'inf.* **2025**, *30*, 517–531. [CrossRef]
31. Yang, Z.; Xu, C.; Li, L.; Yang, Z.; Xu, C.; Li, L. Landslide Detection Based on ResU-Net with Transformer and CBAM Embedded: Two Examples with Geologically Different Environments. *Remote Sens.* **2022**, *14*, 2885. [CrossRef]
32. Kariminejad, N.; Shahabi, H.; Ghorbanzadeh, O.; Shafaie, V.; Hosseinalizadeh, M.; Homayouni, S.; Pourghasemi, H.R. Evaluation of Various Deep Learning Algorithms for Landslide and Sinkhole Detection from UAV Imagery in a Semi-Arid Environment. *Earth Syst. Environ.* **2024**, *8*, 1387–1398. [CrossRef]
33. Zhou, N.; Hong, J.; Cui, W.; Wu, S.; Zhang, Z.; 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]
34. Li, Z.-H.; Shi, A.-C.; Xiao, H.-X.; Niu, Z.-H.; Jiang, N.; Li, H.-B.; Hu, Y.-X.; Li, Z.-H.; Shi, A.-C.; Xiao, H.-X.; et al. Robust Landslide Recognition Using UAV Datasets: A Case Study in Baihetan Reservoir. *Remote Sens.* **2024**, *16*, 2558. [CrossRef]
35. Xu, H.; Wang, L.; Shu, B.; Zhang, Q.; Li, X.; Xu, H.; Wang, L.; Shu, B.; Zhang, Q.; Li, X. Automatic Detection of Landslide Surface Cracks from UAV Images Using Improved U-Network. *Remote Sens.* **2025**, *17*, 2150. [CrossRef]
36. 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]
37. Pham, T.M.; Do, N.; Pham, H.T.T.; Bui, H.T.; Do, T.T.; Hoang, M.V. CResU-Net: A Method for Landslide Mapping Using Deep Learning. *MLST* **2024**, *5*, 035008. [CrossRef]
38. F1 Score in Machine Learning: Intro & Calculation. Available online: https://www.v7labs.com/blog/f1-score-guide (accessed on 7 May 2025).
39. Deval Shah Intersection over Union (IoU): Definition, Calculation, Code. Available online: https://www.v7labs.com/blog/intersection-over-union-guide (accessed on 7 May 2025).