# Metadata of the chapter that will be visualized in SpringerLink

| | |
|---|---|
| Book Title | Artificial Intelligence XLI |
| Series Title | |
| Chapter Title | Semantic Segmentation for Landslide Detection Using Segformer |
| Copyright Year | 2025 |
| Copyright HolderName | The Author(s), under exclusive license to Springer Nature Switzerland AG |

| | | |
|---|---|---|
| Corresponding Author | Family Name | **Syed** |
| | Particle | |
| | Given Name | **Hasnain Murtaza** |
| | Prefix | |
| | Suffix | |
| | Role | |
| | Division | |
| | Organization | Anglia Ruskin University |
| | Address | East Road, Cambridge, CB1 1PT, UK |
| | Email | hs1008@student.aru.ac.uk |
| | ORCID | http://orcid.org/0000-0003-2697-8288 |
| Author | Family Name | **Oghaz** |
| | Particle | |
| | Given Name | **Mahdi Maktabdar** |
| | Prefix | |
| | Suffix | |
| | Role | |
| | Division | |
| | Organization | Anglia Ruskin University |
| | Address | East Road, Cambridge, CB1 1PT, UK |
| | Email | mahdi.maktabdar@aru.ac.uk |
| | ORCID | http://orcid.org/0000-0002-5263-596X |
| Author | Family Name | **Saheer** |
| | Particle | |
| | Given Name | **Lakshmi Babu** |
| | Prefix | |
| | Suffix | |
| | Role | |
| | Division | |
| | Organization | Anglia Ruskin University |
| | Address | East Road, Cambridge, CB1 1PT, UK |
| | Email | lakshmi.babu-saheer@aru.ac.uk |

Abstract

Landslides pose a significant threat to human life and infrastructure which urges the need for efficient techniques for identifying and categorising them. The advent of deep segmentation models such as the Segformer has shown a remarkable empirical performance for semantic segmentation tasks on well-known benchmark datasets, such as ADE20k and Cityscapes. Therefore, this research proposes utilising

Segformer on the benchmark Chinese Academy of Sciences (CAS) Landslide Dataset, which features high-quality aerial images of areas impacted or prone to landslides. Taking advantage of the multi-scale attention mechanism and long-range dependency modeling characteristics of the Segformer architecture, this research aims to achieve state-of-the-art results for landslide segmentation using aerial images. Experimental results show the advantage of the Segformer model in segmenting landslide areas, with the largest Segformer variant achieving an Intersection over Union (IoU) score of 87.795% on the Unmanned aerial vehicle (UAV) dataset, surpassing the previous state-of-the-art model, Multiscale Feature Fusion and Enhancement Network (MFFENet), by 3.4%. On the Satellite (SAT) dataset, Segformer attained an IoU score of 79.300%, outperforming the previous best model, DeepLabv3+, by 11.163%. For the combined UAV&SAT dataset, Segformer achieved an IoU score of 85.157%, surpassing DeepLabv3+, the best previous model by 5.032%.

| | |
|---|---|
| Keywords<br>(separated by '-') | segformer - landslide detection - semantic segmentation |

# Semantic Segmentation for Landslide Detection Using Segformer

Hasnain Murtaza Syed(✉), Mahdi Maktabdar Oghaz, and Lakshmi Babu Saheer

Anglia Ruskin University, East Road, Cambridge CB1 1PT, UK
hs1008@student.aru.ac.uk, {mahdi.maktabdar,lakshmi.babu-saheer}@aru.ac.uk

**Abstract.** Landslides pose a significant threat to human life and infrastructure which urges the need for efficient techniques for identifying and categorising them. The advent of deep segmentation models such as the Segformer has shown a remarkable empirical performance for semantic segmentation tasks on well-known benchmark datasets, such as ADE20k and Cityscapes. Therefore, this research proposes utilising Segformer on the benchmark Chinese Academy of Sciences (CAS) Landslide Dataset, which features high-quality aerial images of areas impacted or prone to landslides. Taking advantage of the multi-scale attention mechanism and long-range dependency modeling characteristics of the Segformer architecture, this research aims to achieve state-of-the-art results for landslide segmentation using aerial images. Experimental results show the advantage of the Segformer model in segmenting landslide areas, with the largest Segformer variant achieving an Intersection over Union (IoU) score of 87.795% on the Unmanned aerial vehicle (UAV) dataset, surpassing the previous state-of-the-art model, Multiscale Feature Fusion and Enhancement Network (MFFENet), by 3.4%. On the Satellite (SAT) dataset, Segformer attained an IoU score of 79.300%, outperforming the previous best model, DeepLabv3+, by 11.163%. For the combined UAV&SAT dataset, Segformer achieved an IoU score of 85.157%, surpassing DeepLabv3+, the best previous model by 5.032%.

AQ1

AQ2

**Keywords:** segformer · landslide detection · semantic segmentation

## 1 Introduction

In recent years there has been a rise in the frequency of landslides causing significant dangers to human lives and infrastructure across the world [1]. Globally, landslides cause an estimated 4,600 deaths annually, highlighting the significant impact of this natural hazard on human lives [6]. Thus, accurate mapping of landslides is essential for emergency response and risk identification. Many studies attempted to use various image segmentation techniques to automate the

M. Bramer and F. Stahl (Eds.): SGAI 2024, LNAI 15447, pp. 1–11, 2025.
https://doi.org/10.1007/978-3-031-77918-3_3

challenging task of landslide mapping using remote sensing and aerial imagery data. Although there has been much research on automated mapping of landslides using remote sensing images, the accuracy is hindered by the quality of the dataset and model performance. Notable studies in this field include: [7] who proposed a fully convolutional network within pyramid pooling (FCN-PP) method for landslide inventory mapping, [3] who developed a You Only Look Once (YOLO) model for detection from satellite images, and [1] who applied U-Net to Landsat 8 satellite images for landslide segmentation. Despite these advances, challenges remain in achieving high accuracy across diverse landslide scenarios. The advent of high-performance segmentation models like Segformer [12] and high-quality datasets such as the CAS [13] offer opportunities to further enhance the accuracy and efficiency of landslide detection and segmentation. Segformer has proven its great performance on similar case studies and datasets like ADE20K [14] and Cityscapes [4] indicating its potential for applications such as landslide detection [12]. This novel research applies the state-of-the-art Segformer architecture to landslide detection using the high-quality CAS landslide dataset. This work is the first to combine Segformer's capabilities with such a comprehensive and well-annotated dataset. By evaluating Segformer across multi-sensor data, this study aims to improve the accuracy and efficiency of automated landslide mapping.

The rest of this paper is structured as follows: Sect. 2 provides an overview of related work in semantic segmentation for landslide detection. Section 3 describes the methodology. Section 4 presents the experimental results and discussion. Finally, Sect. 5 concludes the paper and outlines potential future research directions.

## 2 Related Work

Landslide detection using remote sensing imagery has evolved significantly over the past decade. The advent of deep learning techniques marks a significant leap in detection and segmentation accuracy.

Most automatic landslide detection techniques have recently relied on utilising Convolutional neural networks (CNNs), which have proven highly effective at analyzing image data and extracting relevant features [8]. For instance, [7] conducted a series of experiments to validate their proposed FCN-PP (Fully Convolutional Network with Pyramid Pooling) method for landslide inventory mapping (LIM). [3] conducted experiments to validate their proposed small attentional YOLO model for landslide detection but they faced major issues as small models are efficient but often compromise on performance. [1] proposed using a U-Net with Landsat 8 satellite images but faced issues regarding data imbalance and the quality of available data. Despite these challenges, recent advancements in deep learning have introduced transformer-based models, which have shown good performance on other tasks [12] but there has been little research on their application in detecting landslides.

### 2.1 Transformer Based Models

Vision Transformers (ViT) [5] is the first transformer-based model used to experiment with image classification resulting in state-of-the-art performance. Since then, there has been research utilizing transformers for image segmentation tasks. SegFormer, a Transformer-based semantic segmentation model, was utilized for the identification of landslides by conducting extensive experiments to compare it with other models such as High-Resolution Network (HRNet), DeepLabv3, Attention-UNet, U2Net, and Fast Semantic Segmentation Network (FastSCNN) [11]. Chinese Academy of Sciences [13] developed the CAS Landslide Dataset, a large-scale and multisensory dataset specifically designed for deep learning-based landslide detection. They also compared the results of models such as FCN, U-Net, DeepLabv3+, and MFFENet.

### 2.2 Landslide Datasets

Access to various and well-documented landslide datasets is crucial for progress in landslide detection models. It is worth mentioning that this kind of multisensor, well annotated and large-scale dataset was missing before introducing the CAS (Chinese Academy of Sciences) landslide dataset by [13]. The dataset efficiently overcomes these challenges by providing annotations at a level, for various types of landslide situations, as illustrated in Fig. 1.

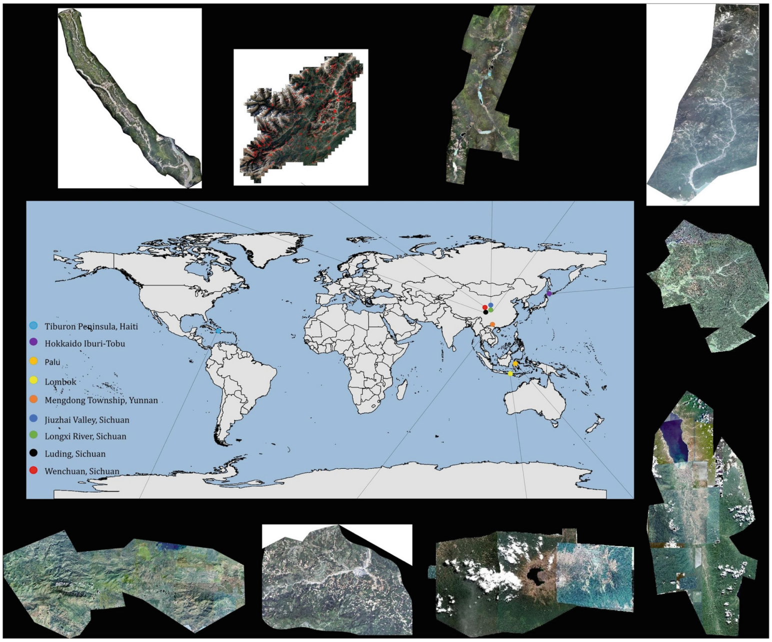

**Fig. 1.** Location map of the study areas.

### 2.3 Research Gaps and Opportunities

While segmentation techniques demonstrate a high detection rate for landslides, there remain significant opportunities for further exploration. One of the main challenges is the lack of high-quality and diverse datasets with comprehensive annotations, which hinders the effective evaluation of segmentation models. The complex and irregular characteristics of landslide features require computational models that can adeptly capture long-range dependencies and manage intricate spatial interactions. Transformer-based models like Segformer have shown potential by outperforming human experts in segmentation tasks, highlighting their capability in the landslide detection process. This paper addresses these research gaps by advancing landslide detection techniques through the use of the Segformer architecture and the CAS landslide dataset.

## 3 Methodology

This section outlines the approach taken in this study to explore how effective the Segformer architecture is, for detecting landslides using the CAS landslide dataset. Starting with the explanation of the dataset and its features then heading towards discussing the Segformer architecture and its key elements. Following that we delve into the specifics of the training and evaluation processes covering aspects such as division, data enhancement methods, and optimization configurations. We then present the setup, which includes three scenarios; UAV, SAT only, and UAV+SAT to evaluate how different data sources and their integration impact landslide detection performance. Finally, implementation specifics like the frameworks and libraries used, along with making our code available, for reliability and further study. The GitHub repository can be found here: www.github.com/syeddhasnainn/landslide-detection-segformer.

### 3.1 Dataset

In this research, the Segformer model was evaluated using the CAS [13] landslide datasets. These datasets include annotations, for high-resolution satellite (SAT) images and unmanned aerial vehicle (UAV) images to aid in landslide detection. The dataset covers a variety of landslide scenarios making it a valuable resource for developing and testing models, for detecting landslides. It contains 19,756 images from nine regions, most from various publicly available datasets and collaborative partners. The image and label were cropped into $512 \times 512$ TIFF format [13]. The dataset was divided into training, test, and validation sets with proportions of 64:20:16, as can be seen in Table 1, to assess the model's performance.

**Table 1.** Distribution of Images Across Datasets in Training, Validation, and Test Sets

| Images Distribution | | | |
|---|---|---|---|
| | Train | Validation | Test |
| UAV | 8603 | 2151 | 2689 |
| SAT | 4040 | 1010 | 1263 |
| UAV + SAT | 12643 | 3161 | 3952 |

### 3.2 Segformer Architecture

The Segformer, first invented by [12], was allegedly designed for this type of semantic segmentation task. It uses a hierarchical architecture with local and global attention which is specialized to extract both low-level details and long-ranged dependencies from the input photo. The architecture structure is an encoder-decoder, with the former taking MiT (the Mix Transformer) backbone and the latter employing a lightweight version of an MLP (Multi-Layer Perceptron) as the decoder, as illustrated in Fig. 2.

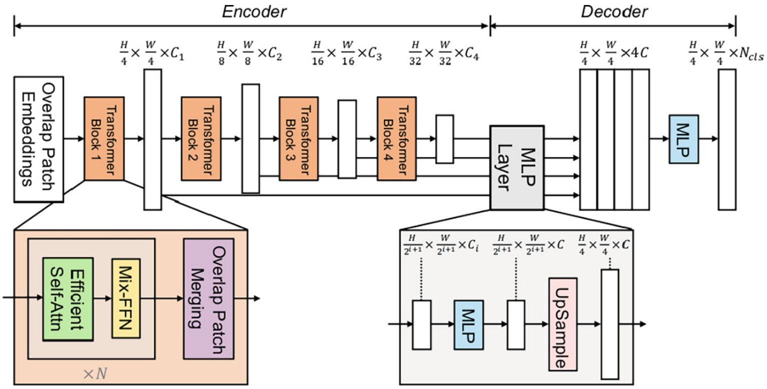
Encoder
Decoder
$\frac{H}{4}\times\frac{W}{4}\times C_1$
$\frac{H}{8}\times\frac{W}{8}\times C_2$
$\frac{H}{16}\times\frac{W}{16}\times C_3$
$\frac{H}{32}\times\frac{W}{32}\times C_4$
$\frac{H}{4}\times\frac{W}{4}\times 4C$
$\frac{H}{4}\times\frac{W}{4}\times N_{cls}$
Overlap Patch Embeddings
Transformer Block 1
Transformer Block 2
Transformer Block 3
Transformer Block 4
MLP Layer
MLP
Efficient Self-Attn
Mix-FFN
Overlap Patch Merging
$\times N$
$\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C_i$
$\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C$
$\frac{H}{4}\times\frac{W}{4}\times C$
MLP
UpSample

**Fig. 2.** A Segformer comprises two components; a Transformer encoder, for capturing both general and detailed features and a compact All MLP decoder, for integrating these various levels of features and forecasting the semantic segmentation mask

The SegFormer models were fine-tuned using the CAS landslide dataset. Different versions of Segformer models were analyzed, ranging from MIT-B0 to MIT-B5 to investigate how the model size and complexity were configured for this landslide detection research.

### 3.3 Training and Evaluation

The Segformer model was modified according to the CAS landslide dataset and UAV and SAT images were used for tuning the model. The training, test, and validation datasets were split in a 64:20:16 ratio. Models were trained using the AdamW optimizer with a learning rate of 0.00006. The performance of the modified models was evaluated using a range of semantic segmentation metrics, including IoU, F1 score, recall, and accuracy.

### 3.4 Experimental Setup

Three sets of experiments were conducted to test the Segformer architecture for landslide detection. The models were trained and evaluated using UAV imagery, then SAT imagery, and finally both UAV and SAT combined. By comparing the results, this research aims to determine which approach provides the most accurate landslide detection.

### 3.5 Implmentation Details

PyTorch [9] which is a popular deep learning framework, was used to carry out all the experiments for this work running on 4 NVIDIA L4 GPUs and it took around 24 h for the full training. The implementation of the Segformer models in the Huggingface library was utilized to establish a common framework for semantic segmentation. The source code including the training and test scripts and dataset preprocessing scripts will be made publicly available for open research and further development in this field.

## 4 Experimental Results and Discussions

This section demonstrates the experimental results using Segformer on the CAS landslide dataset and compares the performance with the latest and semantic segmentation models like FCN, U-Net, DeepLabV3+, and MFFENet [2,8,10,15].

### 4.1 Comparison with State-of-the-Art Methods

The efficacy of the Segformer architecture is proven by comparing it with the leading semantic segmentation approaches. Table 2 compares the results of the Segformer model for UAV, SAT, and UAV+SAT scenarios with those of FCN, U-Net, DeepLabV3+, and MFFENet [12]. In the case where only UAV data is used the Segformer model has obtained an IoU score of 87.795% surpassing the leading model, MFFENet by 3.4%. This shows an enhancement in identifying landslide areas from high-resolution UAV images. The Segformer also excels in precision, recall, F1 score, mean IoU (mIoU), and overall accuracy (OA) compared to models assessed for UAV segmentation. For the SAT dataset, the Segformer model outperforms DeepLabv3+ by 11.163% in terms of IoU score with a score of 79.300% the Segformer model demonstrates its performance in accurately segmenting landslide areas from satellite imagery. In cases where both UAV and SAT datasets are combined the Segformer model achieves a score of 85.157% surpassing DeepLabv3+ by 5.032% as a unique model.

**Table 2.** Performance metrics

| UAV | | | | | | |
|---|---|---|---|---|---|---|
| Model | Precision | Recall | IoU | F1 score | mIoU | OA |
| FCN [13] | 75.045% | 84.016% | 65.057% | 86.724% | 77.456% | 91.468% |
| Unet [13] | 73.694% | 86.394% | 65.991% | 87.136% | 78.019% | 91.658% |
| DeepLabv3+ [13] | 89.289% | 93.739% | 84.261% | 94.715% | 90.142% | 96.721% |
| MFFENet [13] | 89.326% | 93.839% | 84.375% | 94.756% | 90.214% | 96.746% |
| Segformer | 96.178% | 95.923% | **87.795%** | 96.050% | 92.516% | 97.695% |
| SAT | | | | | | |
| FCN [13] | 62.981% | 84.142% | 55.716% | 84.391% | 75.173% | 94.972% |
| Unet [13] | 61.795% | 78.550% | 51.179% | 82.316% | 72.619% | 94.410% |
| DeepLabv3+ [13] | 74.275% | 89.187% | 68.137% | 89.675% | 82.397% | 96.881% |
| MFFENet [13] | 74.141% | 89.141% | 67.998% | 89.621% | 82.318% | 96.862% |
| Segformer | 94.118% | 93.333% | **79.300%** | 93.720% | 88.646% | 98.135% |
| UAV&SAT | | | | | | |
| FCN [13] | 70.847% | 84.014% | 61.757% | 85.864% | 76.515% | 92.848% |
| Unet [13] | 67.479% | 82.360% | 60.115% | 85.311% | 75.697% | 92.653% |
| DeepLabv3+ [13] | 86.128% | 92.013% | 80.125% | 93.563% | 88.316% | 96.687% |
| MFFENet [13] | 86.133% | 92.121% | 80.088% | 93.608% | 88.299% | 96.754% |
| Segformer | 95.483% | 95.147% | **85.157%** | 95.314% | 91.242% | 97.681% |

The research findings show the efficiency of the Segformer architecture in exactly identifying the affected areas by landslides. The transformer-based architecture of Segformer is characterized by the multi-scale attention mechanism and the long-range dependency modeling that are superior to traditional convolutional neural networks. This has relevance to disaster management and mitigation giving the Segformer the potential to be a valuable tool for accurate landslide detection from high-resolution aerial and satellite imagery. In general, this investigation helps the progress of remote sensed imagery for the landslide segmentation.

### 4.2 Qualitative Results

To visually assess the performance of the Segformer models, the ground truth is used to compare with our prediction to investigate the power of the Segformer model in segmenting the satellite imagery datasets (SAT, UAV, and UAV+SAT). The final prediction closely matches the ground truth masks and indicates how well the model could differentiate between landslide and non-landslide areas. The fact that the Segformer model can accurately detect and segment the different features in the aerial visuals is an indication that it can be very reliable for landslide mapping and detection.

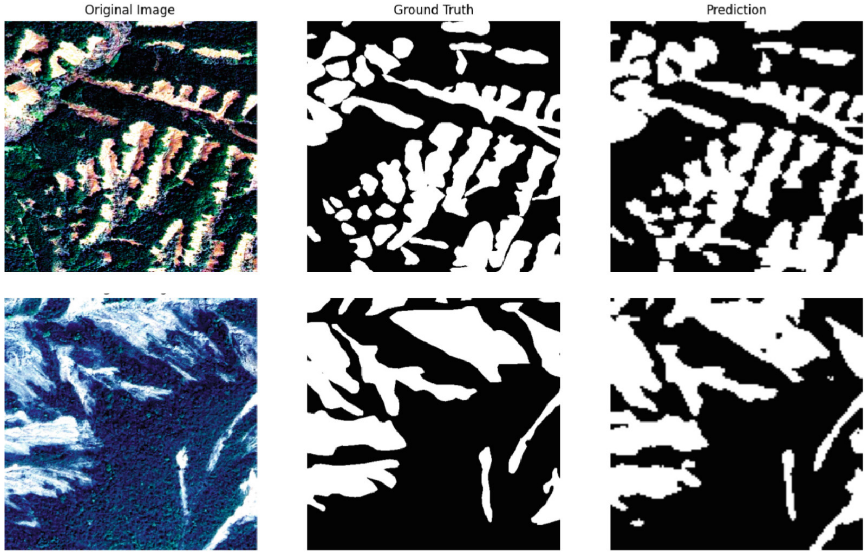

**Fig. 3.** Inaccurate Predictions

We noticed that the images that display high contrast and clear distinctions between different terrain features demonstrate very strong performance. The ground truth segmentation reflects these clear boundaries, and the model's predictions closely align with this ground truth as shown in Fig. 3. The success in these cases can be attributed to the well-defined patterns in the original image, which facilitate accurate segmentation. The distinct differences in color and texture provide the model with clear cues for differentiation and making accurate predictions.

On the other hand, the images that show low contrast and dark regions were challenging to predict accurately. The subtle features and dark, noisy backgrounds are likely misinterpreted by the model as significant features, leading to incorrect segmentation. The ground truth indicates that there are no significant segmented regions, suggesting that the features to be detected are either very subtle or entirely absent. However, the model's prediction includes several segmented regions that are not present in the ground truth, resulting in false positives (see Fig. 4).

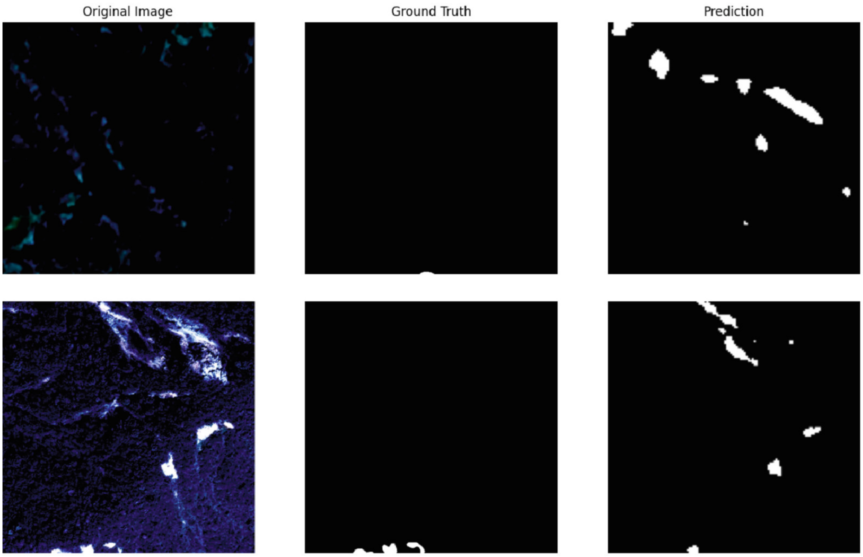

**Fig. 4.** Inaccurate Predictions

### 4.3 Discussion and Future Work

The data-based results presented in this section demonstrate the success of the Segformer architecture for landslide detection using high-resolution UAVs and satellite imagery. The Segformer models outperform other latest semantic segmentation methods, such as FCN, U-Net, DeepLabV3+, and MFFENet, across various evaluation metrics and scenarios. The superior performance of Segformer can be attributed to its transformer-based design, which enables it to capture long-range dependencies and model both local and global features effectively. The hierarchical structure of Segformer, combining the MiT backbone and the All-MLP head, allows it to learn multi-scale representations and efficiently merge information from different scales for accurate landslide detection. Nevertheless, the areas of noticeable weak points, as well as prospects for future research, should be mentioned.

The study utilizes a dataset based on the CAS landslide dataset, which covers a wide range of geographic areas. Future research could address common challenges in remote sensing data, such as insufficient content in cropped images due to boundary issues, low proportion of target objects, obstruction by cloud cover, and discontinuities from image stitching.

## 5 Conclusion

While our study has made significant strides in landslide detection using the Segformer architecture and the CAS landslide dataset, it also highlights several areas for future research. One key challenge in this field remains the scarcity of high-quality, diverse datasets with comprehensive annotations, which limits the thorough evaluation of segmentation models. The complex and irregular nature of landslide features necessitates advanced computational models capable of capturing long-range dependencies and managing intricate spatial interactions effectively.

Our research demonstrates the potential of transformer-based models like Segformer in addressing these challenges, showcasing their ability to outperform human experts in segmentation tasks. By developing an efficient Segformer environment for identifying landslides in high-resolution UAV and satellite images, we demonstrated superior accuracy in landslide region detection compared to existing systems. This success underscores the importance of further exploring and refining such architectures for landslide detection. In the future, research ought to be extended to assess diversified datasets and enhance data modalities. The role of domain-specific refinements of the Segformer architecture should be investigated to further improve the accuracy of landslide detection.

## References

1. Bragagnolo, L., et al.: Convolutional neural networks applied to semantic segmentation of landslide scars. CATENA **201** (2021). https://doi.org/10.1016/j.catena.2021.105189
2. Chen, L.C., et al.: Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv (2018)
3. Cheng, L., Li, J., Duan, P., Wang, M.: A small attentional YOLO model for landslide detection from satellite remote sensing images. Landslides **18**(8), 2751–2765 (2021). https://doi.org/10.1007/s10346-021-01694-6
4. Cordts, M., et al.: The cityscapes dataset for semantic urban scene understanding. arXiv (2016)
5. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv (2021)
6. Froude, M.J., Petley, D.N.: Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci. **18**(8) (2018). https://doi.org/10.5194/nhess-18-2161-2018
7. Lei, T., et al.: Landslide inventory mapping from bitemporal images using deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. **16**(6) (2019). https://doi.org/10.1109/LGRS.2018.2889307
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. arXiv (2015)
9. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. arXiv (2019)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. arXiv (2015)
11. Tang, X., et al.: Automatic detection of coseismic landslides using a new transformer method. Remote Sens. **14**(12) (2022). https://doi.org/10.3390/rs14122884
12. Xie, E., et al.: Segformer: Simple and efficient design for semantic segmentation with transformers. arXiv (2021)

13. Xu, Y., et al.: Cas landslide dataset: a large-scale and multisensor dataset for deep learning-based landslide detection. Sci. Data **11**(1) (2024). https://doi.org/10.1038/s41597-023-02847-z
14. Zhou, B., et al.: Scene parsing through ade20k dataset. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2017). https://doi.org/10.1109/CVPR.2017.544
15. Zhou, W., et al.: Mffenet: multiscale feature fusion and enhancement network for RGB-thermal urban road scene parsing. IEEE Trans. Multimed. **24** (2022). https://doi.org/10.1109/TMM.2021.3086618

# Author Queries

**Chapter 3**

| **Query Refs.** | **Details Required** | **Author's response** |
|---|---|---|
| AQ1 | Please check and confirm if the authors given and family names have been correctly identified. | |
| AQ2 | As per Springer style, both city and country names must be present in the affiliations. Accordingly, we have inserted the country name in the affiliation. Please check and confirm if the inserted country name is correct. If not, please provide us with the correct country name. | |