



Article

# MineSegSAT: An Automated System to Evaluate Mining Disturbed Area Extents from Sentinel-2 Imagery<sup>†</sup>

Ezra MacDonald<sup>1</sup> <sup>2\*</sup>, Derek Jacoby<sup>2</sup> and Yvonne Coady<sup>3</sup>

University of Victoria; derekja@gmail.com (D.J.); ycoady@gmail.com (Y.C.)

\* Corresponding author: macdonaldezra@gmail.com

<sup>†</sup> Presented at the 5th International Electronic Conference on Remote Sensing, 721 November 2023.

**Abstract:** Assessing the environmental impact of the mineral extraction industry plays a critical role in understanding and mitigating the ecological consequences of extractive activities. This paper presents MineSegSAT, a model that presents a novel approach to predicting environmentally impacted areas of mineral extraction sites using the SegFormer deep learning segmentation architecture trained on Sentinel-2 data. The data was collected from non-overlapping regions over Western Canada in 2021 containing areas of land that have been environmentally impacted by mining activities that were identified from high-resolution satellite imagery in 2021 [1]. The SegFormer architecture, a state-of-the-art semantic segmentation framework, is employed to leverage its advanced spatial understanding capabilities for accurate land cover classification. We investigate the efficacy of loss functions including Dice, Tversky, and Lovasz loss respectively. The trained model was utilized for inference over the test region in the ensuing year to identify potential areas of expansion or contraction over these same periods. The Sentinel-2 data is made available on Amazon Web Services through a collaboration with Earth Daily Analytics which provides corrected and tiled analytics-ready data on the AWS platform. The model and ongoing API to access the data on AWS allow the creation of an automated tool to monitor the extent of disturbed areas surrounding known mining sites to ensure compliance with their environmental impact goals.

**Keywords:** remote sensing; semantic segmentation; environmental monitoring; mineral extraction

## 1. Introduction

This paper introduces MineSegSAT, a deep learning image segmentation model using the SegFormer architecture that is designed to segment areas of land that have been environmentally impacted by mineral extraction operations. The mining industry has seen a considerable expansion in recent years [2] driven by growing demand for raw materials [3] and demand trends indicate this growth will continue [4]. While this sector is important for the industrialization of the global economy, mining sites can have adverse impacts on the immediate and near environment during mining operations and after closure. Identifying specifically environmentally impacted areas of land can benefit regulators and mining operations internally to ensure environmentally conscientious practices are upheld.

This project extracts Sentinel-2 tiles in Western Canadian provinces including British Columbia and Alberta that overlap with features of mining sites like waste rock dumps, pits, water ponds, tailing dams, heap leach pads and processing/milling infrastructure [1]. The training, validation, and test datasets use 12 of the 13 bands captured from Sentinel-2, omitting Band 10 as it is typically used for atmospheric correction. 134 tiles were extracted from Sentinel-2 tiles, each spanning 7,680m<sup>2</sup> and their mining types were manually annotated based on overlapping OpenStreetMap views and the Canadian Minerals and Mining Map [5]. More than 85 percent of the tiles that are included in the dataset contain at least one mining pixel, though only roughly 4 percent of the total pixels included in the dataset contain environmentally impacted sections of land. This paper proposes a lightweight



**Citation:** MacDonald, E.; Jacoby, D.; Coady, Y. . *Environ. Sci. Proc.* . . .

Published:



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

implementation of the SegFormer model architecture, a vision transformer-based deep learning model, which was introduced in 2021, by Xie, E. et. al [6] with the minor modification that the most performant model replaces the cross entropy loss function used in the original work with the Tversky Loss function [7].

## 2. Related Work

Deep learning has broad and significant applications in the context of remote sensing and environmental monitoring while Sentinel-2 data has proven to be useful for land cover and land use mapping and for improving the automation of environmental monitoring [8]. In the context of monitoring mineral extraction operations, deep learning models and remote sensing data have been used to assess the significance of environmentally impacted areas [9] and to identify unregistered and illegal mining operations [10] [11].

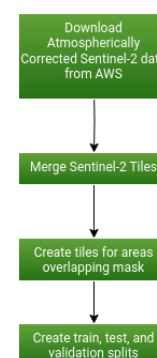
The original transformer architecture proposed an encoder-decoder architecture using multi-head attention and feed-forward layers respectively to globally attend to information from a pair of input sequences [12]. The self-attention component in the architecture updates the value for each element in a sequence by aggregating information from the entire input. Variations of the original architecture that typically do not include cross-attention components detailed in the original architecture have since been applied to computer vision problems such as image classification and segmentation with considerable success. Both image classification and segmentation transformer-based models respectively have demonstrated state-of-the-art performance on benchmark data sets [13] [14] [15] [16]. Transformer-based model architectures have similarly been applied to remote sensing segmentation and classification benchmarks with performance that rivals and in many cases improves on the performance of state-of-the-art model architectures [17].

In this paper, we explore using an implementation of the SegFormer model, a simple semantic segmentation model that uses a UNet architecture - unifying Transformer encoders with lightweight multilayer perceptron (MLP) decoders. This model is of particular interest since it uses lightweight multi-layer perceptron (MLP) layers in the decoder and outputs multi-scale features from the Transformer encoder layers. Since this model is outputting multi-scale features, it does not need to train positional embedding layers which can lead to decreased performance when the training and testing resolutions differ [6]. In the context of remote sensing, this model has been used to extract information for water bodies [18], detect buildings using optical remote sensing images [19], segment coastal wetlands from Sentinel-2 data [20], and for performing road segmentation [21].

## 3. Data

This model was trained using Sentinel-2 tiles, extracted from Amazon Web Services through a collaboration with Earth Daily Analytics which provides atmospherically corrected analytics-ready data on Amazon Web Services. The observation dates for the tiles used for training were from April 1st to September 1st in 2021 for the training period and then a comparison of tiles overlapping the test dataset in 2021 were compared from 2022 for inference and inspection. The tiles that were chosen have less than 1 percent cloud coverage and the tiles within the same coordinate reference system (CRS) were merged using the reverse painter's algorithm. The intersecting ground truth masked data from [1] was reprojected to the CRS of the merged Sentinel-2 tiles and then converted to a raster with 10m resolution to match the resolution of the input tiles.

Given that mining sites in Canada are subject to federal, provincial, and territorial environmental laws and are identified by organizations like Natural Resources Canada [5], the dataset overwhelmingly includes tiles that include environmentally impacted areas of mineral extraction sites identified by Tang and Werner [1]. From the mineral extraction sites that were identified in the original paper, Figure 1 shows the types of



mineral extraction operations that could be confidently intersected with the ground truth masks used to train this model based on visual inspection with overlapping mapped areas using OpenStreetMap and [5].

Each mask and corresponding band file for each period were then split into tiles with dimensions 768x768, 384x384, and 128x128 pixels respective to resolutions of 10m, 20m, and 60m. Note that 4 percent of the pixels included in the dataset used for training, validation, and testing include environmentally impacted areas of mining sites and 86 percent of the 134 tiles include environmentally impacted sites. The dataset can be downloaded from Google Drive using [this link](#).

Mining Operation	# of Sites
Coal	40
Aggregate	28
Precious Metals	31
Mercury	4
Uranium	3

Figure 1. Types of Mineral Extraction Operations Identified

#### 4. Model

The SegFormer model was introduced in 2021 as a lightweight transformer-based image segmentation model with an encoder-decoder or UNet architecture. Differing from the originally proposed Transformer architecture which requires encoding positional information for inputs, the SegFormer encoder layers use feed-forward layers to output multi-scale features, which avoids the need to interpolate positional codes which can lead to decreased performance when training and testing input resolutions differ [6]. The model architecture avoids complex decoders, instead using multilayer perceptron (MLP) layers that aggregate information from each feature map generated by encoder layers. The model architecture used for this experiment is very similar to the B3 implementation-sized model proposed in the original SegFormer paper [6] and differs in that the number of transformer layers used in the second of four layers in the encoder is 4 instead of the 3 used in the original model. This distinction was made as including the additional encoder layer proved to be beneficial in early model verification training iterations.

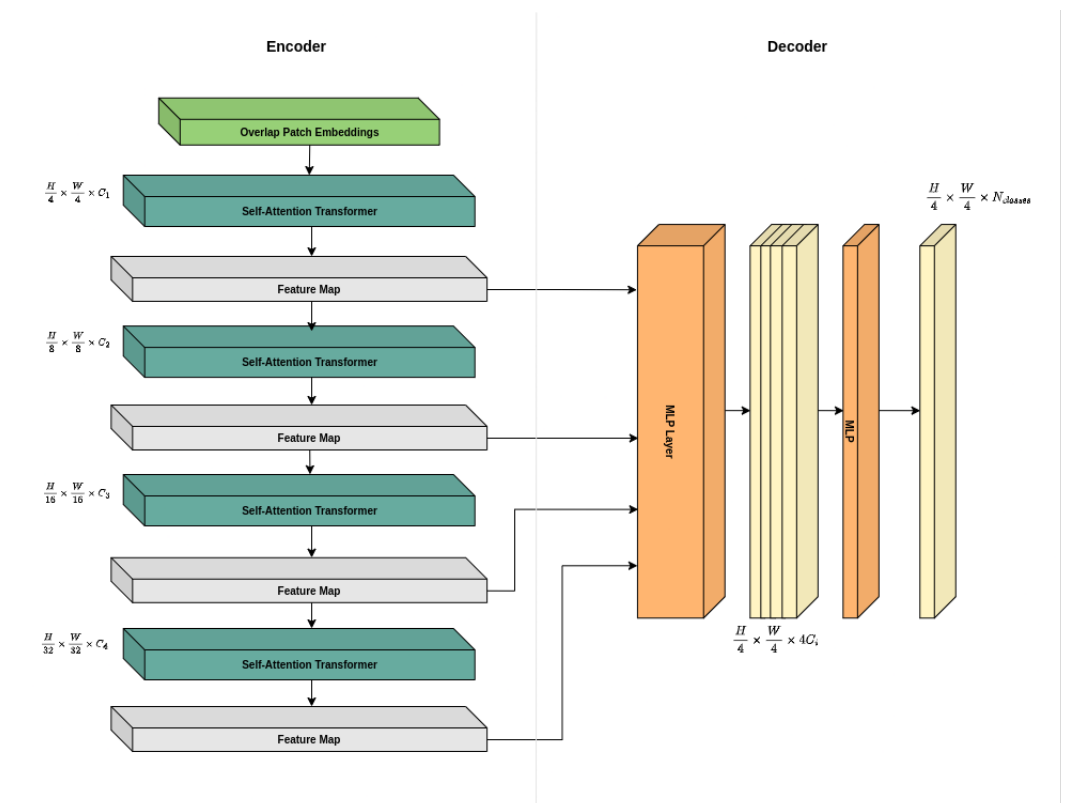


Figure 2. SegFormer Model Architecture

### 5. Experiments

Training of this model was performed using loss functions including Tversky Loss, Dice Loss, and Lovasz-Hinge Loss. For brevity, we focus on the results of the Tversky Loss function since it outperformed the other loss functions. Consider an image as a tensor  $x \in [0, 1]^{L \times H \times W}$  with a corresponding segmentation mask  $y \in \{0, 1\}^{C \times H \times W}$ , and let  $\hat{y} \in \{0, 1\}^{C \times H \times W}$ , where  $C = 1$  since we are using this loss function for binary segmentation. Now let  $\alpha, \beta > 0$ . Then the Tversky Loss denoted by  $T$  corresponding to  $y$  and  $\hat{y}$  is given by

$$T_{\alpha, \beta}(y, \hat{y}) = \frac{y\hat{y}}{y\hat{y} + \alpha(1 - y)\hat{y} + \beta y(1 - \hat{y}) + \delta}$$

where  $\delta > 0$  is a smoothing coefficient. Observe that setting  $\alpha = \beta = 0.5$  gives us the Dice Loss and setting  $\alpha = \beta = 1$  produces the Jaccard Loss.

The metrics that were used for evaluating the efficacy of the predictions outputted by the model were the F1-Score, Precision, and Recall as stated below, where True Positives is given by  $TP$ ; False Positives is given by  $FP$ ; and False Negatives are given by  $FN$ .

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1_{score} &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned}$$

Each of these metrics are computed on a per-pixel basis for each image and the returned score is the average of the score for each image.

Given that the band dimensions vary in their resolution, each of the respective 20m and 60m bands were up-sampled to the 10m resolution bands using interpolation. Before feeding data to the model during training, validation, and testing stages, data transformations were applied to each instance. The transforms applied during training were applied with a given probability detailed in Figure 3 on each input tile, while only random cropping was used for validation data, and patch merging for the test dataset. The model size that was used on the dataset was 512x512 and the batch size used during training and validation was 8. During the test stage, individual images were fed into the model.

Train Transforms	Probability
Random Crop	100%
Vertical Flip	50%
Random Rotation	50%
Horizontal Flipping	50%
Channel Shuffle	30%

Figure 3. Transformations applied during training

The model was trained for 1,000 epochs with a learning rate of 0.0003 using a Cosine Annealing learning rate and was then fine-tuned for another 1,000 epochs with a starting learning rate of 0.0001. The dataset was split such that 70% of samples were used for training, 15% for validation, and the remaining 15% of samples for testing. An additional experiment was performed to verify the results of the test dataset on the test region in the ensuing year, which did not result in additional environmentally impacted mining areas being identified. Code that accompanies this paper can be found [here](#). The results for the most performant model across the training, test, and validation datasets were as follows:

	F1	Precision	Recall
<b>train</b>	0.5743	0.5514	0.5992
<b>val</b>	0.7261	0.7216	0.5461
<b>test</b>	0.5035	0.4890	0.5189

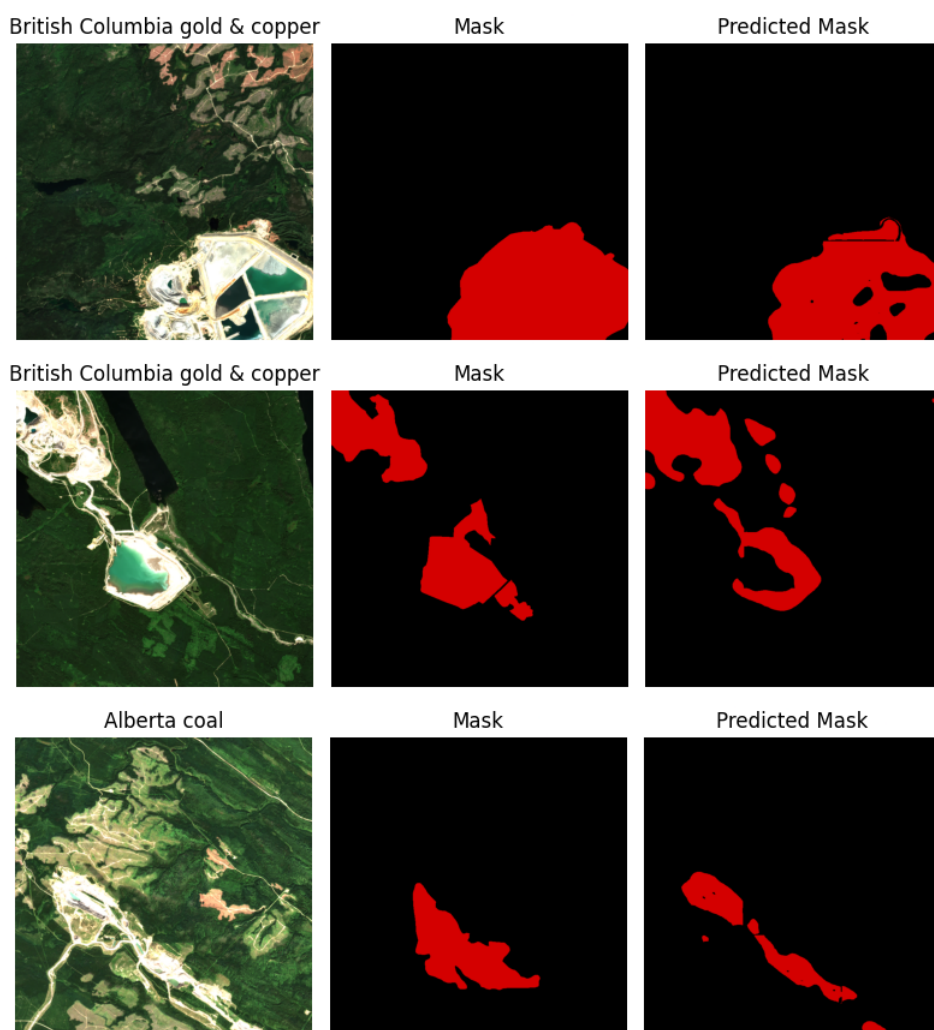


Figure 4. Model predictions on test dataset

## 6. Conclusion & Future Work

The ground truth data provided and models trained on this data are demonstrably capable of identifying the features identified in the original paper, though more work is required to assess the quality of the ground truth dataset and improve the performance of the model such that it might be useful for performing accurate environmental monitoring that can be reliably used in real-time. Correlating the identified mining areas with existing publicly identified above-ground or visibly environmentally impacted mining areas could be beneficial for better understanding the efficacy of the data and areas where the model is challenged in making accurate predictions.

From this sample dataset, the model struggles to identify characteristics like tailing ponds as seen in Figure 4 as well as built-up rock piles. Waterways like rivers, lakes, and potentially non-impacted water sources were included in the dataset including alpine and mountainous regions which may explain the model’s challenge in identifying the respective features. Further training could be done that uses a larger dataset with additional mining sites and more generalized model pre-training in an effort to improve the robustness of the model. Data sources like Sentinel-1 data that include pertinent information about ground features and more data take periods over mining areas could be used in the training and evaluation datasets to provide a more enriched analysis of a given area. Further, other deep learning architectures should be trained on the dataset to establish a performance benchmark on the dataset and improve the understanding of areas of strength and weakness in predicting the masked regions.

A model trained based on these findings could present value for identifying and assessing multiple risks posed by mineral extraction operations by identifying anomalies in the size variation of tailing ponds, rock pile formations, open-pit mines, and processing/milling infrastructure which were identified in the ground truth masks. This model would thrive in a circumstance where it is monitoring known active or previously active mineral extraction sites that pose a potential threat to the environment. Given that this data is made available by government agencies [5], implementing a meaningful deep learning-based monitoring service of this nature is not only possible but could significantly benefit the environmental monitoring of mineral extraction sites.

## References

1. Tang, L.; Werner, T.T. Global mining footprint mapped from high-resolution satellite imagery. *Commun Earth Environ* **2023**, *4*, 1–12. <https://doi.org/10.1038/s43247-023-00805-6>.
2. Maus, V.; Giljum, S.; da Silva, D.M.; Gutschlhofer, J.; da Rosa, R.P.; Luckeneder, S.; Gass, S.L.B.; Lieber, M.; McCallum, I. An update on global mining land use. *Sci Data* **2022**, *9*, 433. <https://doi.org/10.1038/s41597-022-01547-4>.
3. Lenzen, M.; Geschke, A.; West, J.; Fry, J.; Malik, A.; Giljum, S.; Milà i Canals, L.; Piñero, P.; Lutter, S.; Wiedmann, T.; et al. Implementing the material footprint to measure progress towards Sustainable Development Goals 8 and 12. *Nat Sustain* **2022**, *5*, 157–166. <https://doi.org/10.1038/s41893-021-00811-6>.
4. IRP, U. Global resources outlook 2019: Natural resources for the future we want. Technical report, United Nations Environment Programme, 2019.
5. Government of Canada, N.R.C. Natural Resources Canada. The Atlas of Canada. Minerals and Mining in Canada, 2021.
6. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. 2021.
7. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In Proceedings of the Machine Learning in Medical Imaging; Wang, Q.; Shi, Y.; Suk, H.I.; Suzuki, K., Eds., Cham, 2017; Lecture Notes in Computer Science, pp. 379–387. [https://doi.org/10.1007/978-3-319-67389-9\\_44](https://doi.org/10.1007/978-3-319-67389-9_44).
8. Phiri, D.; Simwanda, M.; Salekin, S.; Nyirenda, V.R.; Murayama, Y.; Ranagalage, M. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing* **2020**, *12*, 2291. <https://doi.org/10.3390/rs12142291>.
9. Gerassis, S.; Giráldez, E.; Pazo-Rodríguez, M.; Saavedra, ; Taboada, J. AI Approaches to Environmental Impact Assessments (EIAs) in the Mining and Metals Sector Using AutoML and Bayesian Modeling. *Applied Sciences* **2021**, *11*, 7914. <https://doi.org/10.3390/app11177914>.
10. Balaniuk, R.; Isupova, O.; Reece, S. Mining and Tailings Dam Detection in Satellite Imagery Using Deep Learning. *Sensors (Basel)* **2020**, *20*, 6936. <https://doi.org/10.3390/s20236936>.
11. Rangnekar, A.; Hoffman, M. Learning representations to predict landslide occurrences and detect illegal mining across multiple domains. In Proceedings of the Climate Change AI. Climate Change AI, 2019.
12. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, ; Polosukhin, I. Attention is all you need. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., NIPS'17, pp. 6000–6010.
13. 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. *ArXiv* **2020**.
14. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.S.; et al. Rethinking Semantic Segmentation From a Sequence-to-Sequence Perspective With Transformers. pp. 6881–6890.
15. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling Up Capacity and Resolution. pp. 12009–12019.
16. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention Mask Transformer for Universal Image Segmentation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1280–1289. ISSN: 2575-7075, <https://doi.org/10.1109/CVPR52688.2022.00135>.
17. Li, Z.; Wang, Y.; Zhang, N.; Zhang, Y.; Zhao, Z.; Xu, D.; Ben, G.; Gao, Y. Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey. *Remote Sensing* **2022**, *14*, 2385. <https://doi.org/10.3390/rs14102385>.
18. Yang, X.; Chen, M.; Yu, C.; Huang, H.; Yue, X.; Zhou, B.; Ni, M. WaterSegformer: A lightweight model for water body information extraction from remote sensing images. *IET Image Processing* **2023**, *17*, 862–871. <https://doi.org/10.1049/ipr2.12678>.
19. Li, M.; Rui, J.; Yang, S.; Liu, Z.; Ren, L.; Ma, L.; Li, Q.; Su, X.; Zuo, X. Method of Building Detection in Optical Remote Sensing Images Based on SegFormer. *Sensors* **2023**, *23*, 1258. <https://doi.org/10.3390/s23031258>.
20. Lin, X.; Cheng, Y.; Chen, G.; Chen, W.; Chen, R.; Gao, D.; Zhang, Y.; Wu, Y. Semantic Segmentation of China's Coastal Wetlands Based on Sentinel-2 and Segformer. *Remote Sensing* **2023**, *15*, 3714. <https://doi.org/10.3390/rs15153714>.
21. Ma, T.; Zhou, X.; Xi, R.; Yang, J.; Zhang, J.; Li, F. A Semi-supervised Road Segmentation Method for Remote Sensing Image Based on SegFormer. In Proceedings of the Artificial Intelligence and Robotics; Yang, S.; Lu, H., Eds., Singapore, 2022; Communications in Computer and Information Science, pp. 189–201. [https://doi.org/10.1007/978-981-19-7943-9\\_16](https://doi.org/10.1007/978-981-19-7943-9_16).