VitSegh24: Illegal Mining Footprints Surveillance with GeoSpatial Imagery of Ghana

Andrews Ata Kangah[1] , **Armstrong Francis Tumawu**[2]
[1]Armtos
[2]Geomappers Engineering Ltd.
andrewsatakangah@armtos.com, farmstrong03@gmail.com

Abstract

Extractive operations along the mineral-rich coasts of Ghana over so many years have begun costing the lives of indigenous people living near the affected areas. Assessing the environmental impact of these activities over the 67 years of its perpetuation, underpins all local and global climate efforts, sustainability, and prevention of its adverse effects. This paper presents VitSegh24, an image segmentation model based on the Segformer pretrained model. The model is fine-tuned on high-resolution publicly available satellite masks data covering features like waste rock dumps, pits, water ponds, tailings dams and heap leach pads all identifiable within 238,533 square kilometers land cover of the country. We propose this model and future automation of the inference pipeline as a key tool that can be scaled over the entire continent of Africa, towards climate action and responsible land use. The code and dataset is provided at https://github.com/armtos-np/vitsegh

1 Introduction

Good health and well-being, clean water and sanitation, climate action, and responsible consumption and production are a few of the United Nations Sustainable Goals. https://sdgs.un.org/goals. While Industry, Innovation and Infrastructure is also another of the sustainable development goals, its unbounded implementation may conflict with efforts against the other goals listed above, i.e climate action and responsible production and consumption of resources.

This paper presents VitSegh24, a deep learning image segmentation model based on Segformer, a novel version of the original Visual Transformer architecture, that is capable of segmenting and mapping out the footprints of mineral extraction activities within the mineral rich regions of Ghana.

The project extracts masks of Sentinel-2 imagery in affected areas in Ghana, specifically Southwestern forests of the country in Tarkwa, Obuasi, Prestea, and other neighbouring villages where mining activities have destroyed the major river bodies like the Ofin river and its tributaries. These rivers originally served as sources of drinking water to the people of those communities.

188 tiles of labeled Sentinel-2 imagery were collected and prepared from global mining footprint dataset provided by [Tang and Werner, 2023]. The dataset contains polygons mapped from high-resolution satellite imagery. For monitoring over only Ghana, the dataset is cropped to cover the affected areas in Ghana covering 1003.8 km2 of the country. This dataset is applied in a fine-tuning task using the pytorch framework to build a segmentation model on Geotiff images.

2 Previous Work

Deep Learning generally underpins all current advancements in machine learning. It has been applied in domains like healthcare, agriculture, transportation and engineering. Efforts have also been applied in the field of mining monitoring using satellite imagery to assess the footprints of mineral extractive activities in Western Canada [MacDonald *et al.*, 2023]. These applications including that of Vitsegh24 are possible due to curated labeled dataset of the effects of mining activities using Sentinel-2 satellite imagery by [Tang and Werner, 2023].

2.1 Sentinel-2 Global Mine Polygons

Tang & Werner [Tang and Werner, 2023] provide a publicly available global mining footprint dataset of 74,548 mine area polygons covering 65,585.4 km2 of mapped mining areas indicating hot and cold spots of global mining activity. Features like pits, waste rock dumps, water ponds, ore stockpiles, processing infrastructure and tailings dams are represented and labeled in the dataset.

2.2 MineSegSat

MineSegSat [MacDonald *et al.*, 2023] utilizes the above dataset 2.1. They propose an automated system for evaluating mining disturbed extents in Western Canada. The project builds off SegFormer [Xie *et al.*, 2021], a state-of-the-art semantic segmentation framework. They also propose a novel approach by replacing the original cross-entropy loss function with Tversky Loss function for improved efficiency.

2.3 Flood Detection

Sen1floods11 [Bonafilia *et al.*, 2020] curates a global scale flood dataset for permanent and flood water segmentation.

The project further discusses that training Fully Convolutional Neural Networks (FCNN) models on plentiful automatically generated labels from remote sensing algorithms performs better than models trained on scarce hand labeled data (Bonafilia et. al). As a result, future research into automating computer vision approaches to flood detection could build on the Sen1floods11 dataset.

2.4 Prithvi and HLS Foundation

The IBM and NASA team provide Prithvi - a 100 million parameter foundation model. Prithvi is a first-of-its-kind temporal Vision transformer (Jakubik et. al). It is a 1TB data-sized model, pre-trained on multispectral satellite imagery from the Harmonized Landsat-Sentinel 2 (HLS) dataset. The model is a self-supervised pre-trained model aimed at enabling few-shot training with labeled data for downstream tasks.

The HLS Foundation [Jakubik *et al.*, 2023a] provide sample applications of Prithvi as a backbone for training other models. For example: It is used together with mmsegmentation framework, pytorch and openmim to train segmentation models on the Sen1floods11 [Bonafilia *et al.*, 2020] dataset

2.5 Vitsegh24

From these previous works, we explore using the pre-trained Segformer [Xie *et al.*, 2021] model to fine-tune a localized dataset of 112 masks out of 118 total data points extracted from Tang & Werner [Tang and Werner, 2023] with their corresponding band or channel files. We closely follow the approach of [MacDonald *et al.*, 2023] MineSegSat to achieve our results. We make customizations to train on satellite data that cover only areas of Ghana as it is our region of interest. We also make customizations during the data extraction and cleaning stages to remove bands without data because the available satellite data does not cover all the areas of interest.

3 Data

Masked tiles used in the training are collected from Tang & Werner [Tang and Werner, 2023] global dataset on mining footprints. The project originally provides polygon data, not raster data, of the various labeled mining sites. While the polygon data is more accurate, it extended the amount of data work needed to be done before training. Using GIS techniques, the shape file provided in the dataset was loaded in QGIS, an open-source Geographic Information System (GIS) software. A separate polygon data file (.shp file) containing the boundaries of Ghana was also loaded as a separate layer. Using feature selection and cropping, the bounded areas of Ghana were extracted from the global dataset resulting in labeled polygons for identified mining sites in Ghana.

3.1 Data Collection

Labeled tiles were already provided by [Tang and Werner, 2023] 3. Therefore, only input data for validation and testing were missing for the training step.

In previous experiments, the corresponding Google Maps tiles for the Ghanaian regions were extracted as input data. This was motivated by the final goal of Vitsegh24; to build a real-time mining site detection platform based on Google Maps tiles.

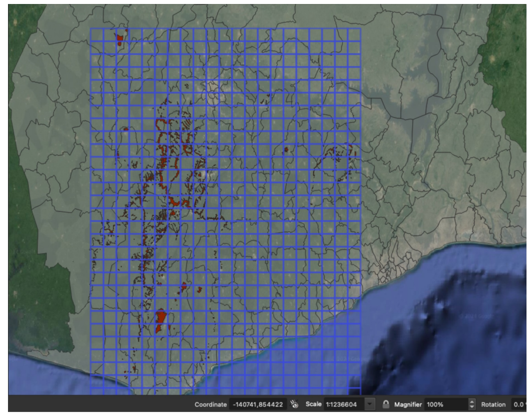

Figure 1: Grid localized to Ghana mining footprints

This approach did not provide the desired results during the training. Therefore, attention switched to using the sentinel-2 satellite dataset as input data.

Following the automated code provided in [MacDonald *et al.*, 2023], satellite images are downloaded from Amazon Web Services, which is provided by a collaboration with Earth Daily Analytics. The images are cloud corrected and analytics-ready.

The data extraction process is automated so that a duration can be provided for which available images within that period can be downloaded. This is adapted closely from the work that has been done by [MacDonald *et al.*, 2023].

For Vitsegh24, sentinel-2 images are collected between the dates 2020-01-01 and 2021-01-01 since the mask dataset provided by [Tang and Werner, 2023] was prepared with images within the same period.

3.2 Data Validation, Cleaning, Transformation

As any model is as good as its data, the created tiles needed wrangling to be suitable for training. The download script discussed in 3.1 was therefore modified from [MacDonald *et al.*, 2023] to perform the following operations to prepare the final data for training:

1. Due to the regions of Ghana and the period within which satellite images are collected, some tiles had blank data after the tiling process was finished. As a result, those tiles needed to be removed from the input dataset.
2. Organize input images and masks into the correct directory structure for training. The training script required input images and labels to be organized in the structure: *masks/mask.tif* and *image-grid-location/B0.tif* etc.
3. Split tile images into train, validation and test datasets. This step was crucial in preparing the data within the correct distributions of 60% for train, 20% for validation and 20% for testing.

Aside from these cleaning steps, additional steps were added to compute a co-occurrence matrix over the dataset to compute the distribution of mask pixels in the input images used in training.

Figure 2: Example of mask tile.

Distribution

From 1 it is visually clear that more unlabeled regions are covered in the identified grid than the labeled ones (marked in *red*).

In previous experiments, the aim was to achieve a balance between the labeled and unlabeled regions in our training data after the tiles were created covering the entire region of Ghana. This led to the deletion of a large number of input tiles that had zero corresponding labels or masks.

This data distribution task is now achieved directly during data extraction. An algorithm extracted from https://www.matecdev.com/posts/shapely-polygon-gridding.html is applied to split the geo-dataframe of Ghana according to a specified delta. The delta determines the size of the grids that are created. The grids are then checked against the geo-dataframe of the labeled or masks polygon to retrieve only grids that intersect. See 3.

With this approach, only grids intersecting the identified mining affected or target areas are downloaded. Tiled input images are then generated from these grids. This ensures that only mining affected regions are included in the resulting input dataset and provides a good distribution of the target class (1) against the background data classes (0).

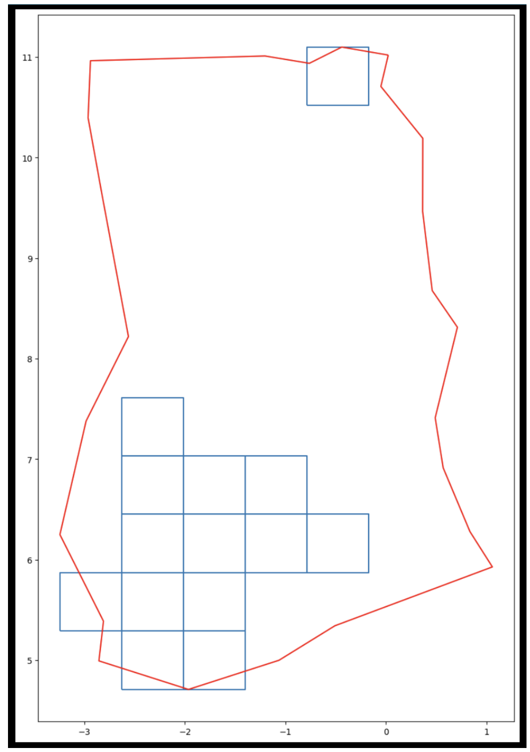

Figure 3: Intersecting grids over Ghana

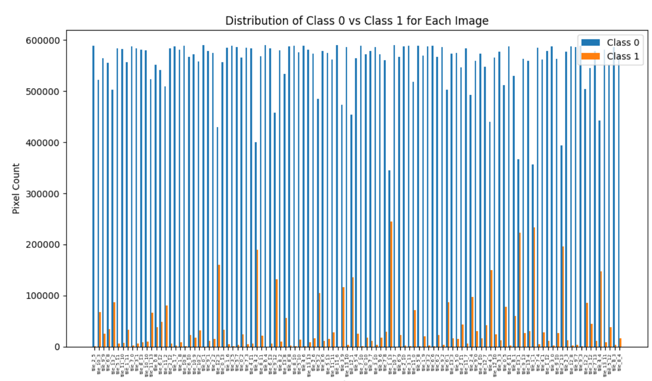

Figure 4: Distribution of classes in training data tiles

4 Model

Vitsegh24 is trained on the pre-trained weights of Segformer [Xie *et al.*, 2021] model which is a lightweight image segmentation model based on the vision transformer architecture. MineSegSat provides a good framework for fine-tuning Segformer for mining sites detection which we adopt.

4.1 Architecture

The architecture of Vitsegh24 does not differ in any way from that of [MacDonald *et al.*, 2023]. We maintain the same architecture with MineSegSat because in previous experiments, we adopted different architectures like that of [Jakubik *et al.*, 2023b] with our own customizations using HLS Foundation's framework [Jakubik *et al.*, 2023a] with MMSegmentation. This yielded less performant results than expected. Prithvi being a 100 million parameter model was also probably too complex for the simple task of binary image segmentation. Prithvi is originally a real-time *video* and image segmentation model.

We estimate that after achieving efficient results with [Xie *et al.*, 2021], the next steps will involve real-time video segmentation bringing [Jakubik *et al.*, 2023c] and [Jakubik *et al.*, 2023a] back into the purview.

Vitsegh24 consists of 4 encoder layers and 4 All-MLP (Multilayer Perceptron Layer) decoder layers. Based on Segformer [Xie *et al.*, 2021], the encoder layers extract the features from the input image using patches of size 4 unlike in the original Vision Transformer (ViT) architecture. The smaller patch sizes favour the dense prediction task. According to [Xie *et al.*, 2021], a Mix-FFN (Mixed Feed Forward Network) encoder is introduced. This encoder ignores paddings to leak positional information for features extracted from input images. Positional Encodings (PE)as used in ViT caused the resulting model to be unable to make predictions on images of different resolution from the training images' since the PE is finite. The Mix-FFN uses a 3x3 conovolution to represent features extracted.

Since the features are without Positional Encodings, the All-MLP decoder layers which by design have each output in the encoders fully connected with the input of each decoder, enable features to be well represented. This enables the third All-MLP layer to fuse concatenated features in the previous layers which the final layer takes to predict the segmentation mask of $\frac{H}{4}$ x $\frac{W}{4}$ x Ncls resolution.

This additional improvement within Segformer favours VitSegh24. The flexibility of inference any resolution input image makes the design of a software using the Vitsegh24 easier in the future. 5.1

The model is trained for 1,000 epochs with a batch of 8 images loaded in each training step.

4.2 Evaluation

The model is evaluated against three metrics: F1 Score, Precision and Recall. Since the F1 Score is a balancing metric, it is used during training to determine the best model throughout the 1,000 epochs of training.

The F1 Score is calculated from the Precision and Recall values which demonstrates its averaging effect over the performance of the model.

It is also the best metric for this project because inherently, the classes in consideration are generally skewed. In particular, the background pixels per each input tile is more than the mask pixels. See the distribution 4

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} = \frac{\text{2 x Precision x Recall}}{Precision + Recall}$$

4.3 Results

After training for 1,000 epochs, the best F1 Score achieved is 0.467 over the training dataset. Calculating the metrics of the model over the validation dataset, the model has a F1 score of 0.143 with a higher Recall value of 0.999 over Precision of 0.077. The same higher Recall metric of 0.997 is observed for the test dataset over 0.055 Precision value.

The higher Recall values observed over the test dataset demonstrates that the model performs well at predicting the actual relevant cases i.e actually true positive cases. 1. It also therefore deduces that the model performs well on new data that is not part of the training set.

Dataset	F1	Precision	Recall
train	0.467	0.400	0.560
validation	0.143	0.077	0.999
test	0.105	0.055	0.997

Table 1: Metrics Results

With the distribution of background pixels being higher per input image than the mask pixels, even after avoiding regions with no mining activities, the higher Recall value of the model is very satisfactory. This is because it means that the model is still able to accurately predict positive pixels in input images despite the skewed pixels in the dataset.

5 Conclusion

In this paper, we demonstrate how VitSegh24 is trained for segmentation of mining activities on land cover using satellite imagery.

The dataset used in the project was curated by an automated process. This demonstrates the feasibility of the future goals of the project. The automated data extraction process, accompanied by other software tools that will be available in the future, will serve as a real-time framework for monitoring illegal mining operations in Ghana.

5.1 Future Work

As already described in the abstract of this paper, the final goal is to create a full Machine Learning Operations (MLOps) pipeline for data collection, cleaning, processing, and inference satellite imagery. This pipeline would serve as a real-time alert system for new mining sites that spring up in Ghana advising authorities (barring all political arguments), the general public (if made publicly available) and law enforcement

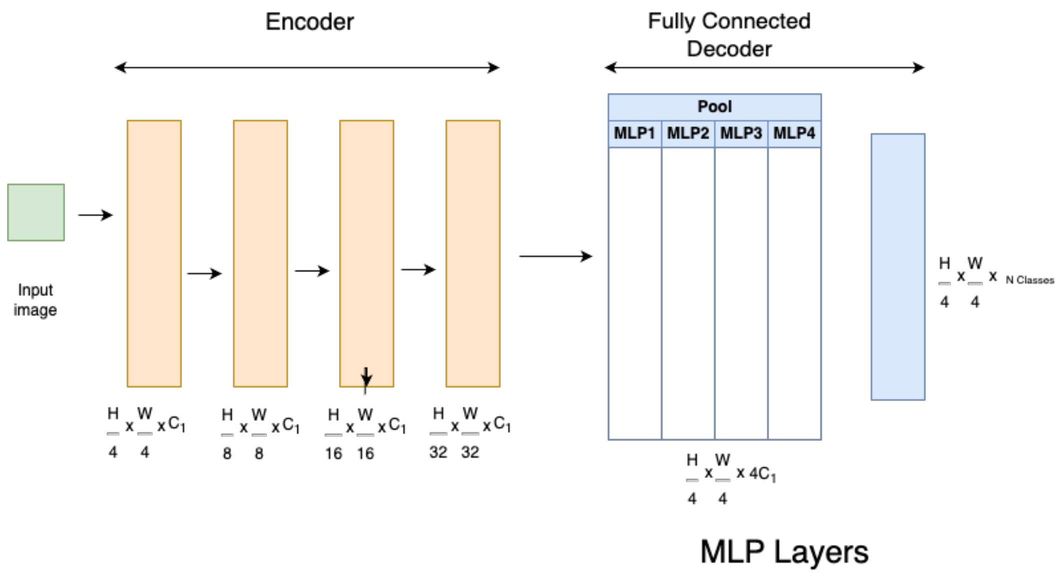

Figure 5: Model Architecture

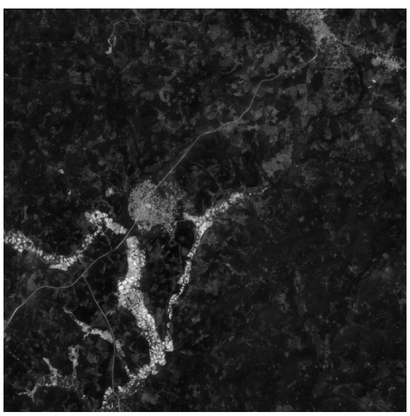

Figure 6: Input image with stacked bands

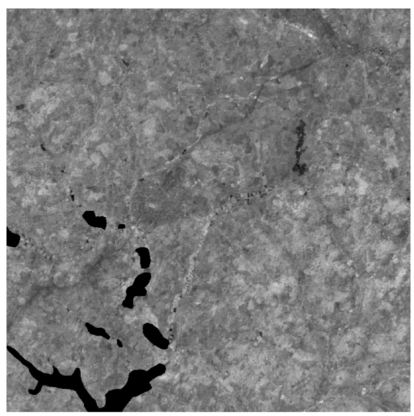

Figure 7: Input image with mask

Figure 8: Predicted segmentation mask

agencies to take action and shut down such activities if they are indeed illegal and unauthorized.

Some parts of the automated pipeline have already been created. These parts include data collection or download, cleaning, and processing.

Automated inference and direct overlay onto a 3D map over the web are a few of the requirements still missing in the current state of the project. These will be added in future iterations to improve the effectiveness of Vitsegh24 as a real-time illegal mining site dection tool in Ghana.

A. Appendices

The metrics discussed in the Evaluation section are not the only metrics that were used in the evaluation of this model. In both former and final experiments, Intersection over Union (IoU) was used.

1. The IoU is calculated as follows

$$IoU = \frac{\text{True Positive}}{\text{True Positive + False-Positive + False Negative}} \tag{1}$$

2. Co-occurrence matrix calculation

$$C(i,j) = \sum_{x=1}^{N} \sum_{y=1}^{M} \begin{cases} 1, & \text{if } I(x,y) = i \text{ and} \\ & I(x+\Delta x, y+\Delta y) = j \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Computing the co-occurrence matrix proves very essential in segmentation tasks for analysing the distribution of pixel data. We are able to remove tiles that have higher percentage of 0-class pixels by computing co-occurrence matrices. The equation above explains the components of the co-occurrence matrix.

Ethical Statement

With the advent of Large Language Models, chatbots like ChatGPT, Claude and Gemini which are able to perform generalised tasks like writing essays and scripts are made possible. While originality is desired in high quality scientific research, these chatbots ease some of the workload involved in getting from ground zero research to state-of-the-art.

For the ethical reasons of originality in research, we wish to state clearly that ChatGPT and Claude AI were used during this research to generate some of the python scripts for data cleaning and processing. These scripts were reviewed, scrutinized and modified by the authors for correctness. Where the generated code was already of high quality, no such modifications were made, only reviews were conducted to ensure correctness.

All other work not involving the above stated, for example: data collection, training and documentation of the processes and methods used within this research paper are all original works of the authors.

Acknowledgements

Armtos is the non-profit organization whose entities carry out this work of science. The researchers in the team consists of highly qualified and talented Computer Scientists, Geological Engineers, and Land Economists. Apart from the individuals listed as authors of the document, we want to acknowledge the valuable insights and contributions of the following people in the team, including one from the [MacDonald *et al.*, 2023] MineSegSat team. Without their advice and support, this work would not be possible.

1. Ezra MacDonald (MineSegSat)
2. Issah Abubakari Samori
3. Maame Pokua Debrah
4. Anthony Attakumah Kangah

Computing resources for training machine learning models are costly as a result of the amount of processing power required. This work did not face much of that computational cost due to support from Amazon Web Services (AWS).

We acknowledge the support of Amazon Web Services Rapid Ramp Program, and Activate Founders Program who have supported the team with credits on the AWS Cloud Platform to experiment and train Vitsegh24.

References

[Bonafilia *et al.*, 2020] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[Jakubik *et al.*, 2023a] Johannes Jakubik, Linsong Chu, Paolo Fraccaro, Ranjini Bangalore, Devyani Lambhate, Kamal Das, Dario Oliveira Borges, Daiki Kimura, Naomi Simumba, Daniela Szwarcman, Michal Muszynski, Kommy Weldemariam, Bianca Zadrozny, Raghu Ganti, Carlos Costa, Campbell Watson, Karthik Mukkavilli, Sujit Roy, Christopher Phillips, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Wei Ji Leong, Ryan Avery, Rahul Ramachandran, Manil Maskey, Pontus Olofossen, Elizabeth Fancher, Tsengdar Lee, Kevin Murphy, Dan Duffy, Mike Little, Hamed Alemohammad, Michael Cecil, Steve Li, Sam Khallaghi, Denys Godwin, Maryam Ahmadi, Fatemeh Kordi, Bertrand Saux, Neal Pastick, Peter Doucette, Rylie Fleckenstein, Dalton Luanga, Alex Corvin, and Erwan Granger. HLS Foundation, August 2023.

[Jakubik *et al.*, 2023b] Johannes Jakubik, Linsong Chu, Paolo Fraccaro, Carlos Gomes, Gabby Nyirjesy, Ranjini Bangalore, Devyani Lambhate, Kamal Das, Dario Oliveira Borges, Daiki Kimura, Naomi Simumba, Daniela Szwarcman, Michal Muszynski, Kommy Weldemariam, Bianca Zadrozny, Raghu Ganti, Carlos Costa, Campbell Edwards, Blair & Watson, Karthik Mukkavilli, Hendrik Schmude, Johannes & Hamann, Parkin Robert, Sujit Roy, Christopher Phillips, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Wei Ji Leong, Ryan Avery, Rahul Ramachandran, Manil Maskey, Pontus Olofossen, Elizabeth Fancher, Tsengdar Lee, Kevin Murphy, Dan Duffy, Mike Little, Hamed Alemohammad, Michael Cecil, Steve Li, Sam Khallaghi, Denys Godwin, Maryam Ahmadi, Fatemeh Kordi, Bertrand Saux, Neal Pastick, Peter Doucette, Rylie Fleckenstein, Dalton Luanga, Alex Corvin, and Erwan Granger. Prithvi-100M, August 2023.

[Jakubik *et al.*, 2023c] Johannes Jakubik, Sujit Roy, C. E. Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, Daiki Kimura, Naomi Simumba, Linsong Chu, S. Karthik Mukkavilli, Devyani Lambhate, Kamal Das, Ranjini Bangalore, Dario Oliveira, Michal Muszynski, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Sam Khallaghi, Hanxi (Steve) Li, Michael Cecil, Maryam Ahmadi, Fatemeh Kordi, Hamed Alemohammad, Manil Maskey, Raghu Ganti, Kommy Weldemariam, and Rahul Ramachandran. Foundation Models for Generalist Geospatial Artificial Intelligence. *Preprint Available on arxiv:2310.18660*, October 2023.

[MacDonald *et al.*, 2023] Ezra MacDonald, Derek Jacoby, and Yvonne Coady. Minesegsat: An automated system to evaluate mining disturbed area extents from sentinel-2 imagery, 2023.

[Tang and Werner, 2023] Liang Tang and Tim T. Werner. Global mining footprint mapped from high-resolution satellite imagery. *Communications Earth & Environment*, 4(1):134, 2023.

[Xie *et al.*, 2021] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers, 2021.