

# Semantic Segmentation Of Aerial Images Of Illegal Construction Waste

Case Study of Misgav Regional Council, Israel

Aryeh Gorun  
328634373

Yoni Tsur  
204617963

## Abstract

Aerial imagery has emerged as a powerful tool for environmental analysis and decision-making, enabling us to gain valuable insights. In this article, we present a comprehensive approach for performing semantic segmentation on aerial images of illegally dumped construction waste. We focus on the detection and analysis of the waste content. Leveraging the Segment Anything Model (SAM) developed by Meta, we produced highly accurate masks from aerial images. We created a dataset of over 22,000 manually labeled masks, which serve as ground truth for training and evaluation. Then we fine-tuned the ResNet-50 classification model together with image processing and augmentation techniques. Our methodology combines the predictions of the classification model with these detailed masks to produce the final segmentation map. The resultant segmentation offers a comprehensive understanding of the area under scrutiny, allowing for an informed decision-making process concerning the profitability and environmental necessity of cleanup operations. The analysis also opens avenues for the potential treatment of different waste streams, further aiding in waste management strategies. We achieved 0.7 test accuracy on our full dataset. We also present a fast and easy pipeline for annotating images using masks produced from the SAM model.

**Keywords:** illegal waste site, construction and demolition waste, SAM, semantic segmentation, computer vision, machine learning.

## 1 Introduction

### 1.1 Problem Definition

According to the Israeli State Comptroller, 6.2 Million Tons of C&D (Construction and Demolition) waste are generated in the country annually. Of which 2.19 Million Tons, or approximately 38%, are dumped illegally in open areas and unauthorized sites [1]. The waste is often dumped in remote areas, making it difficult to detect and clean up. Due to the high cost of cleaning up the waste, and the uncertainty about the profitability of the operation, many sites are left unattended for years.

The local authorities find it difficult to deal with the scope of the criminal activity and are forced to remove the environmental hazard themselves. This is a complex process that includes first the collection of the waste and its transportation to the authorized site, and later the cleaning and restoration of the contaminated land. This process is expensive and involves significant budget expenditures by the municipalities, which usually place the burden on the taxpayers. In addition, construction waste in open areas takes up valuable real estate space and quick and efficient mapping will help clear the land and prepare it for other uses.

Unmanaged waste contributes to air and water pollution, fire, flooding, and transmitting diseases while harming animals that unknowingly consume waste and negatively affecting the economy.

A pilot research by the Porter School of Environmental and Earth Science at Tel Aviv University suggested that "it is feasible to identify valuable materials left on the ground in the form of unattended, illegally disposed waste. Our initial national estimates for the illegal waste cleanup based on the pilot results suggest that the treatment cost in Israel can be reduced by 58 million USD and even reach zero, with the potential to generate up to 82.8 million USD profits" [2].

Indeed, many sites contain construction waste that includes recyclable materials such as concrete, ceramics, gypsum, polymers (rigid and flexible), metal, cardboard, and paper.

The research used drones to capture aerial images of the sites, and we continue this approach in our project. Using drones allows us to capture high-resolution images of the sites, and to cover large areas in a short time.

## 1.2 Project Goal

Our goal is to perform semantic segmentation on aerial images of waste. We aim to develop a tool that will allow us to analyze the images and identify the waste content. We hope that authorities can better understand the quantities and qualities of the waste and what the costs are for the required cleaning. This will allow them to make informed decisions about the profitability of the operation and the environmental necessity of the cleanup.

## 2 Dataset

### 2.1 Categories

We divided the waste into 8 categories:









Metal	Wood	Carton + Paper	Polymer	Mineral	Rubber	Background	Other
							

Table 1: Images of different classes

Polymer consists of both rigid polymers such as plastic buckets, and flexible polymers such as nylon. Mineral includes concrete, gypsum, and ceramics. Other includes styrofoam, textile, glass, and other.

### 2.2 Data Collection

We Started with  $\sim 85$  high-resolution drone images ( $5280 \times 3956$ ), from 3 sites in Misgav Regional Council. We split each image into 16 smaller images ( $1320 \times 989$ ) in order to enhance visibility. Additionally, it helps with the computational efficiency of the SAM model, since running it 16 times on a 1.3M pixel image is lighter than running SAM on a 20.8M pixel image [3]. From each split image, we produced  $\sim 70$  masks using SAM. When running SAM with its default parameters, we found that most of the objects are very small, and many masks are not very informative. By choosing the parameters for SAM carefully, we were able to produce masks that captured most areas of interest in the image, while ignoring less informative areas.

### 2.3 Annotation process

We first created an annotation pipeline in Google Colab, but it was very slow and not user-friendly. We then created a web-based annotation tool, which allowed us to annotate the images and save the results in a Google Drive folder. The average time per split improved to 5-10 minutes. We dedicated to the annotation process 1.5 months. We uploaded all the data to a university server.

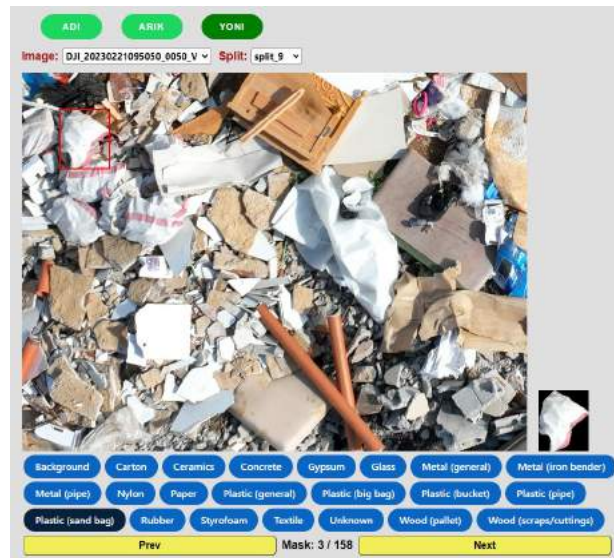


Figure 1: Annotation app

## 2.4 Data Preprocessing

Compared to other segmentation datasets, our dataset is quite hard to work with. We found a large class imbalance in the data which resulted in a skewed distribution. We have many examples of background, but very little rubber and metal, which can affect the model's performance.

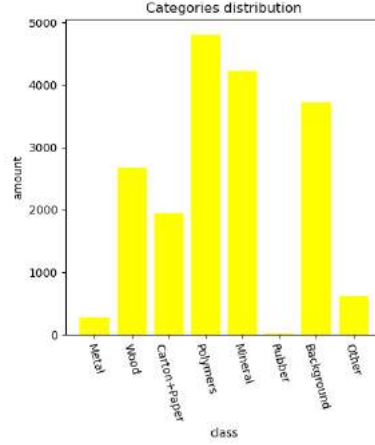


Figure 2: Class distribution

In addition, the waste is often piled up and sometimes hidden by grass and trees, making it hard to distinguish one object from another and to extract a full mask of an object. Additionally, there are challenges with outdoor light conditions such as shaded and high-saturation areas, and the fact that the images are taken from a drone, which is constantly moving.

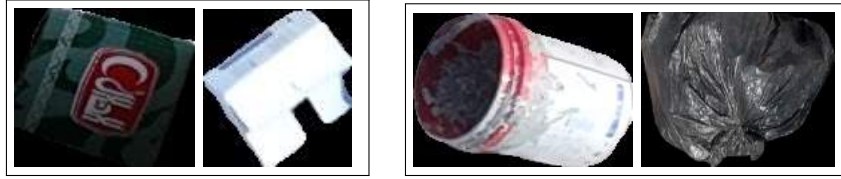


Figure 3: From left to right: Two pictures of Carton (one is colorful, with distinct edges and in the shade, the other is white, blurry, and very saturated); Two pictures of Polymers (one is a paint bucket which is a rigid polymer, the other is a black nylon garbage bag which is a flexible polymer)

We tried to overcome these challenges by performing preprocessing to the dataset prior to training. We first discarded all the masks with an area of less than 2K pixels because they were not informative. Then, in order to produce more data, we used augmentations of 90, 180, and 270-degree rotations, horizontal and vertical flips, as well as contrast, brightness and saturation augmentations. This improved the model training performance, but also resulted in over-fitting. The performance on the validation set was inconclusive, so we decided to abandon this option.

We analyzed the results of the model on 3 different mask features:

Size	
Category	Pixel Range
very small	$\leq 5K$
small	$\leq 10K$
medium	$\leq 40K$
large	$\leq 100K$
very large	$> 100K$

Coverage	
Category	Coverage Range
good	$> 70\%$
medium	$> 40\%$
poor	$\leq 40\%$

Ratio	
Category	Ratio Range
regular	$< 1.5$
big	$< 2.5$
very big	$\geq 2.5$

We found that the model performs badly on masks with poor coverage (less than 40% of the bounding box is covered by the mask).

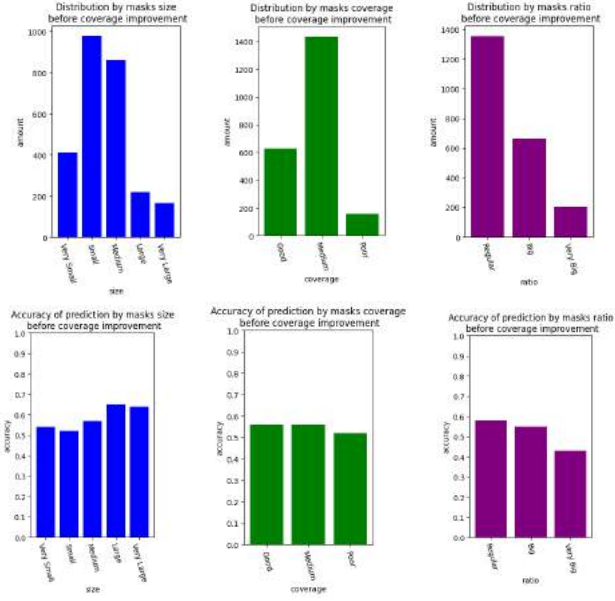


Figure 4: Top: Data features distribution before heuristic. Bottom: Accuracy per feature before heuristic

To improve this, we created a heuristic that rotated the mask by 15-degree steps, and then chose the rotation that produced the best bounding box coverage. This helped us to improve the model performance from 0.56 to 0.59. This was done early in our project, and it still helps in our final model.

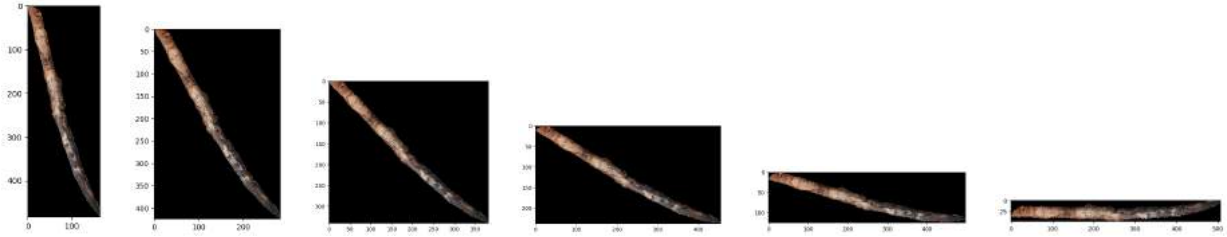


Figure 5: Our heuristic to improve coverage - ordered from left to right, are the different considered images. original mask (16% coverage), 15° (10% coverage), 30° (9% coverage), 45° (11% coverage), 60° (20% coverage), 75° (49% coverage)

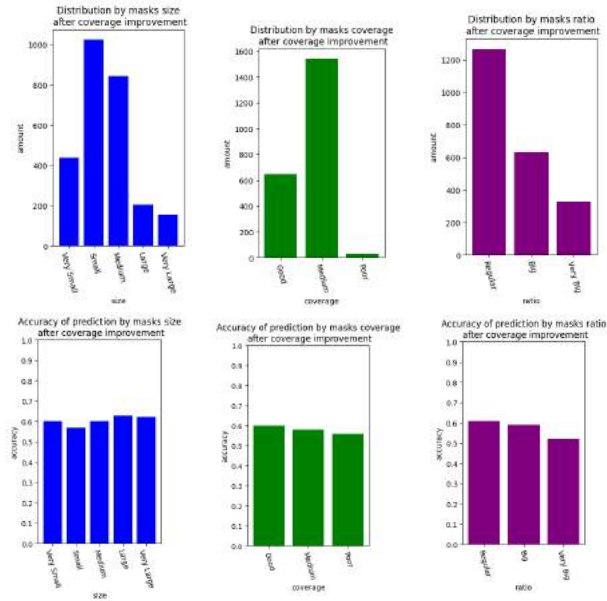


Figure 6: Top: Data features distribution after heuristic. Bottom: Accuracy per feature after heuristic

As seen from the graphs, the heuristic improved the model's performance on all features, and effectively eliminated the poor coverage masks.

### 3 Baseline

As a baseline, we chose a Naïve classification method: color histogram.

In image processing, a color histogram is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges, that span the image’s color space, the set of all possible colors.

We took the color histogram for each of our labeled masks and averaged them per class. Then for an input mask, we take its color histogram and return the label with the highest cosine similarity score. We achieved 38% accuracy on our labeled data.

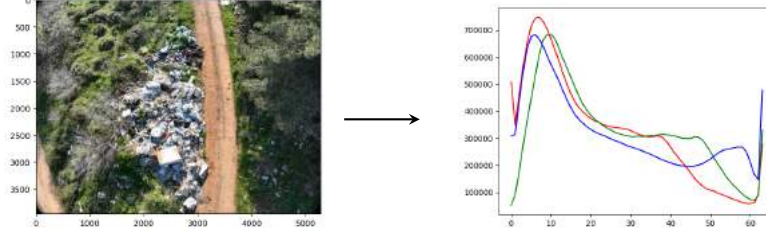


Figure 7: An example of a color histogram

Table 2: Classification baseline results

Index	Metal	Wood	Carton + Paper	Polymers	Mineral	Rubber	Background	Other	Avg	Weighted Avg
Accuracy	0.07	0.32	0.05	0.3	0.4	0.0	0.68	0.0	0.23	0.38

## 4 Methodology

### 4.1 Segmentation

Image segmentation is the process of partitioning an image into several regions. The pixels of these regions generally should share certain characteristics. In our case, we would like to group together into the same category two waste objects of the same class, i.e. a plastic sand bag and a bucket should both be classified as Polymers and thus be categorized together. Semantic segmentation has been explored in many ways in the past decade with various computer vision techniques. Historically used Fully Convolutional Networks (FCNs) [4] [5], and more recently transformer-based methods as well [6].

#### 4.1.1 SAM

We used the Segment Anything Model (SAM) by Meta, for the segmentation task. SAM is a deep learning tool for image segmentation, it has the ability to take an image and automatically create good segmentation masks around most of the objects in the image. Although this doesn’t directly equate to semantic segmentation, we can then use classification methods to classify each object and create a semantic segmentation map. We chose not to go in the direction of fine-tuning existing models because those require a lot of labeled data, which we don’t currently have. Additionally, SAM already guarantees good masks, so all we need to focus on is classification, which is an easier computer vision task. We use the following parameters for SAM:

Parameter	Value	Description
<code>predicted_iou</code>	0.92	Prediction for the quality of the mask
<code>points_per_side</code>	25	The number of points to be sampled along one side of the image.
<code>stability_score_thresh</code>	0.75	An additional measure of mask quality
<code>crop_n_layers</code>	0	The number of layers to crop from the image
<code>crop_n_points_downscale_factor</code>	1	The downscale factor of the image
<code>box_nms_thresh</code>	0.35	The box IoU cutoff used by non-maximal suppression to filter duplicate masks
<code>min_mask_region_area</code>	1000	Postprocessing to remove disconnected regions in masks with area smaller than <code>min_mask_region_area</code>

Table 3: Chosen Parameters for SAM



## 4.2 Model

### 4.2.1 Classification

We used a ResNet-50 model [7], pre-trained on ImageNet [8], and fine-tuned it on our dataset. ResNet-50 is a 50-layer residual convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer). We replaced the last layer with a linear layer of 256 neurons and added an additional linear layer with 8 neurons. We tried different structures, but in the end landed on this, because it was simple, and performed among the best.

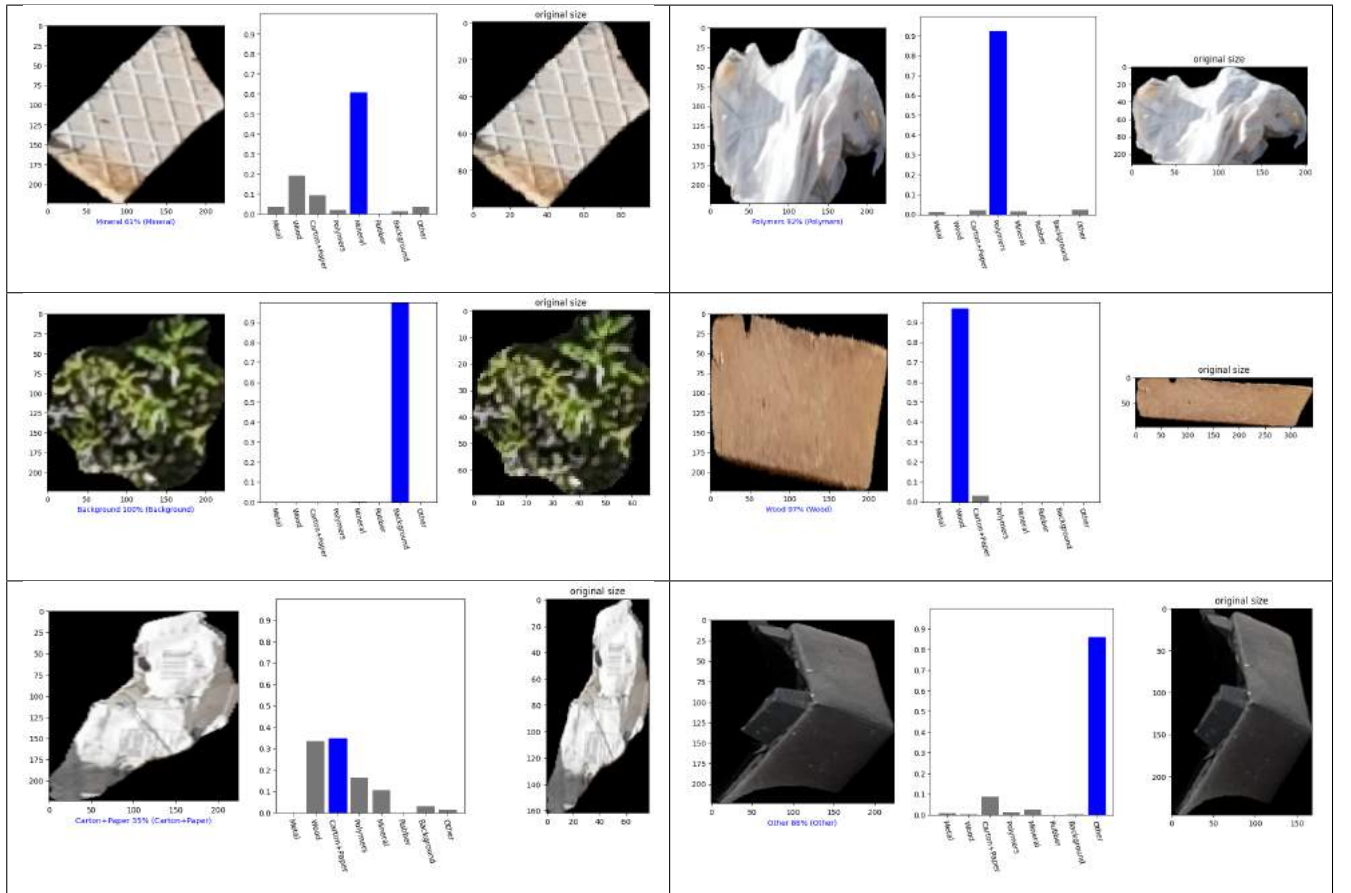
The input size of ResNet-50 is 224 x 224, and our masks are widely different and varied in size. We first tried to pad smaller masks and downsized larger masks. This approach didn't work well, we found that the model is performing worse when the image's coverage is low.

Instead, we resized all the images to 224 x 224 and normalized them using the mean and std of ImageNet (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]). We used weighted cross-entropy loss (the weights were the inverse of the class frequency in the dataset), SGD optimizer, learning rate 0.002, momentum 0.9, weight decay 0.5 every 10 epochs.

### 4.2.2 Textures

We tried to use texture features to improve the model's performance. In order to extract the texture of the masks, we used skimage.features library with the following parameters: contrast, dissimilarity, homogeneity, energy, correlation, ASM. These parameters should represent aspects of the object's texture, and help the model to distinguish between different textures. However, empirically the results were better without the texture, so we decided not to use it.

### 4.2.3 Analysis



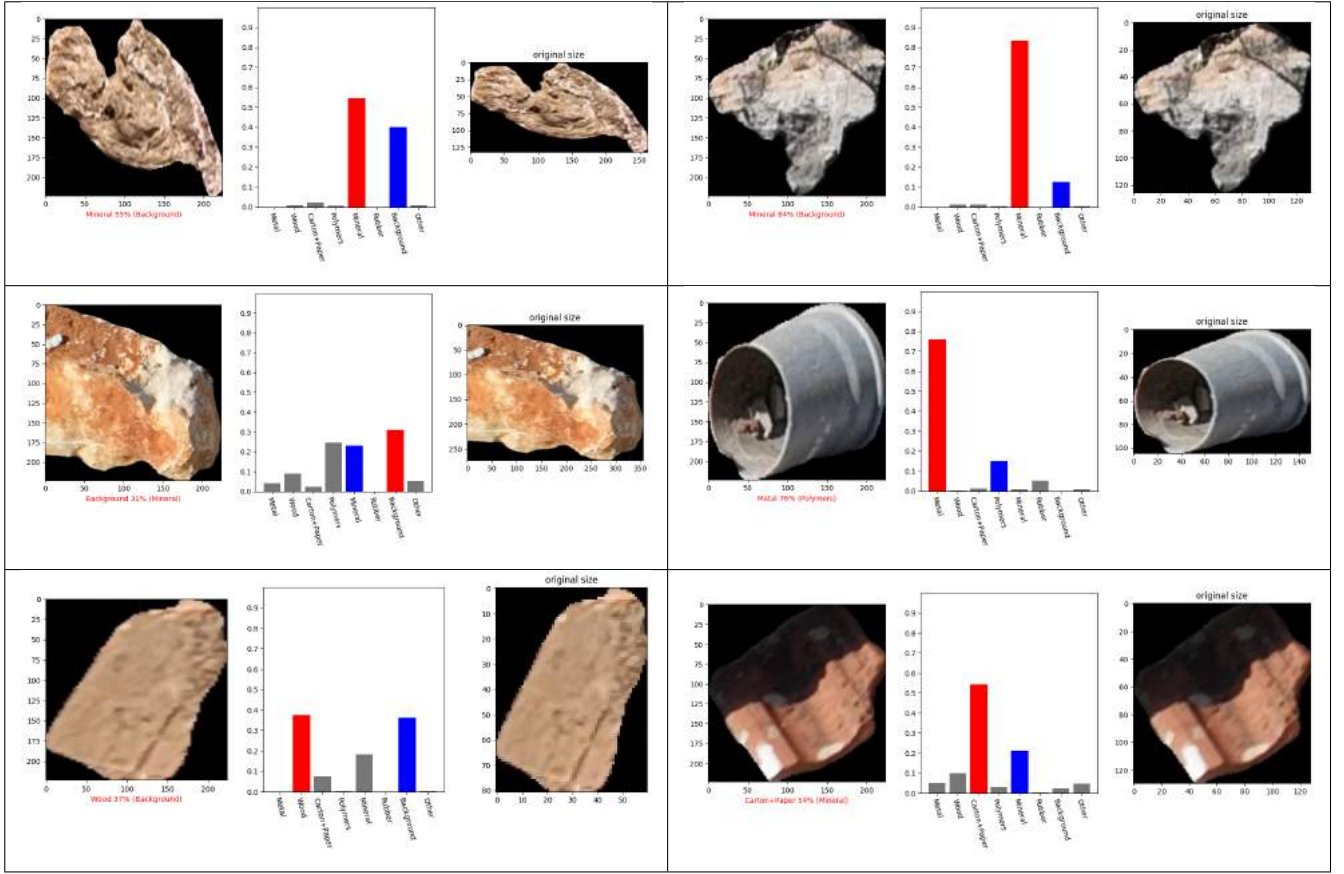


Figure 8: Classification analysis

We can observe that the model is very confident in most cases where it is clear that an object is of a certain type. Some of the errors the model makes can be attributed in our opinion, to the similarity between some classes, such as carton and wood. In other cases, we hypothesize that training the model for longer could produce a better outcome.

Additionally, one of the flaws of our annotated data is that it's sometimes not so clear whether an object belongs in the 'Background' category, or in the 'Minerals' category. Since they both contain rock-shaped objects and in some cases might be entirely dependent on the context of the image, which is missing here.

### 4.3 Final Pipeline

Combining the masks generated by SAM, and our classification model, we created a pipeline that takes an image, and produces a segmentation map. We first split the image into 16 smaller images, and run SAM on each of them. We then run the classification model to create a prediction for each mask. Finally, We merge the predictions and produce a segmentation map for each image.

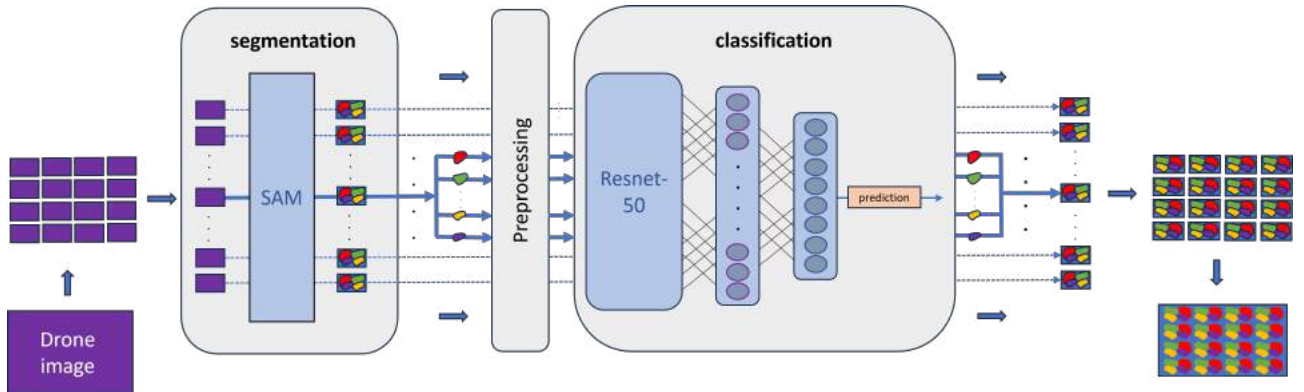


Figure 9: Pipeline Flow

## 5 Results

After analyzing the results of the model, we found that the high class imbalance severely damages the model's performance. Specifically the classes Rubber, Metal, and Other constitute less than 5% of the dataset, and the model performs poorly on them, and consequently on the other classes as well. We first show the results of our model trained and evaluated on the entire dataset and then show the results of the model trained and evaluated on a reduced dataset with just the main 5 classes.

### 5.1 Full Model

We present the final results and statistics of the classification model, as well as some sample segmentation maps. Overall the test accuracy of our full model is 0.70 on our current data.

Table 4: Classification test top-1 Results Per Class

	Metal	Wood	Carton + Paper	Polymers	Mineral	Rubber	Background	Other	Avg	Avg no Background
Amount	33	307	250	606	588	3	400	60	280.9	263.9
Precision	0.22	0.69	0.59	0.84	0.73	0.67	0.86	0.2	0.6	0.56
Recall	0.67	0.74	0.59	0.7	0.68	0.67	0.8	0.47	0.66	0.64
F1	0.33	0.71	0.59	0.76	0.71	0.67	0.83	0.28	0.61	0.58

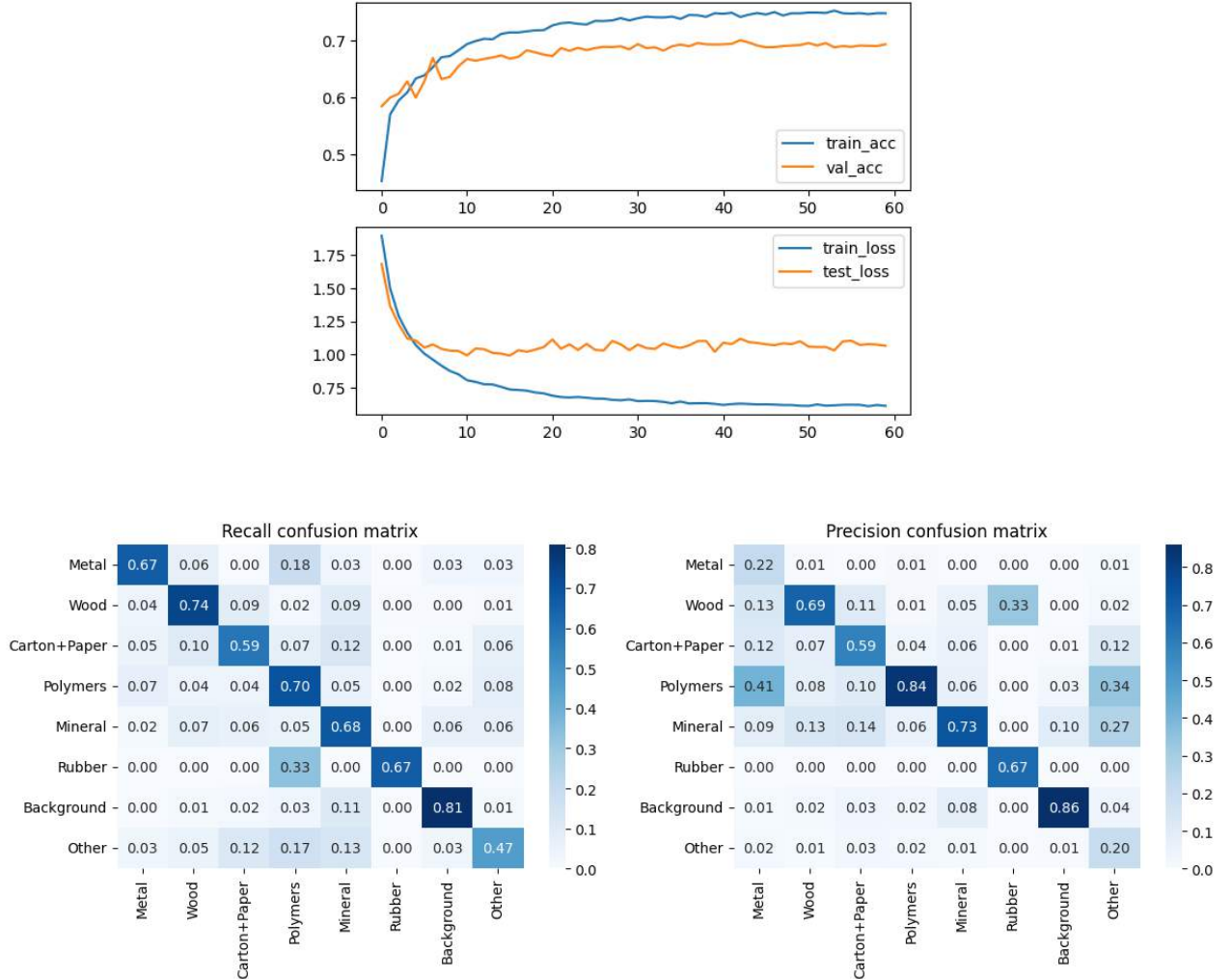
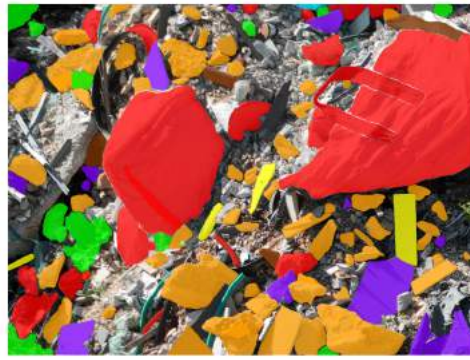


Figure 10: Precision and Recall normalized Confusion Matrices

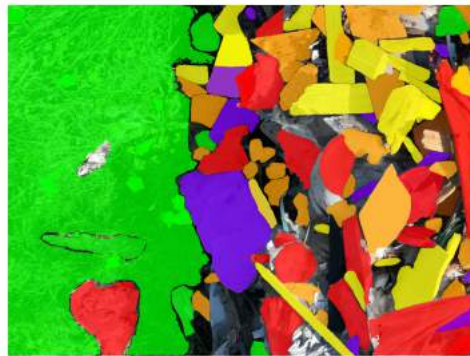




- Category Label
- Rubber
  - Other
  - Background
  - Mineral
  - Polymers
  - Carton+Paper
  - Wood
  - Metal



- Category Label
- Rubber
  - Other
  - Background
  - Mineral
  - Polymers
  - Carton+Paper
  - Wood
  - Metal



- Category Label
- Rubber
  - Other
  - Background
  - Mineral
  - Polymers
  - Carton+Paper
  - Wood
  - Metal

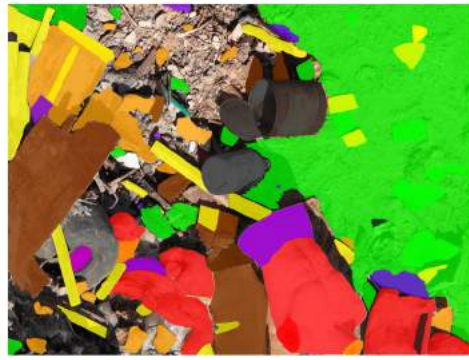


- Category Label
- Rubber
  - Other
  - Background
  - Mineral
  - Polymers
  - Carton+Paper
  - Wood
  - Metal



- Category Label
- Rubber
  - Other
  - Background
  - Mineral
  - Polymers
  - Carton+Paper
  - Wood
  - Metal





Category Label

- Rubber
- Other
- Background
- Mineral
- Polymers
- Carton+Paper
- Wood
- Metal



Category Label

- Rubber
- Other
- Background
- Mineral
- Polymers
- Carton+Paper
- Wood
- Metal



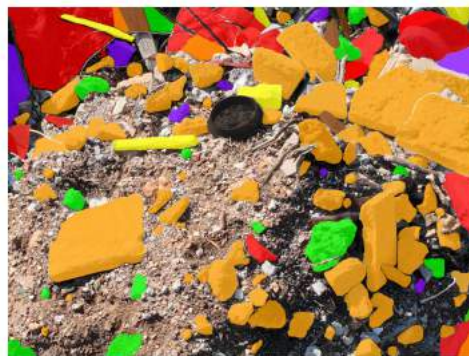
Category Label

- Rubber
- Other
- Background
- Mineral
- Polymers
- Carton+Paper
- Wood
- Metal



Category Label

- Rubber
- Other
- Background
- Mineral
- Polymers
- Carton+Paper
- Wood
- Metal



Category Label

- Rubber
- Other
- Background
- Mineral
- Polymers
- Carton+Paper
- Wood
- Metal

## 5.2 Reduced Model

The reduced model is trained and evaluated on a dataset with just the main 5 classes: Wood, Carton + Paper, Polymers, Mineral, and Background. This model also suggests that if in the future we will have more data, we will be able to train a model that will perform well on all classes.

Overall the test accuracy of our reduced model is 0.77 on our current data.

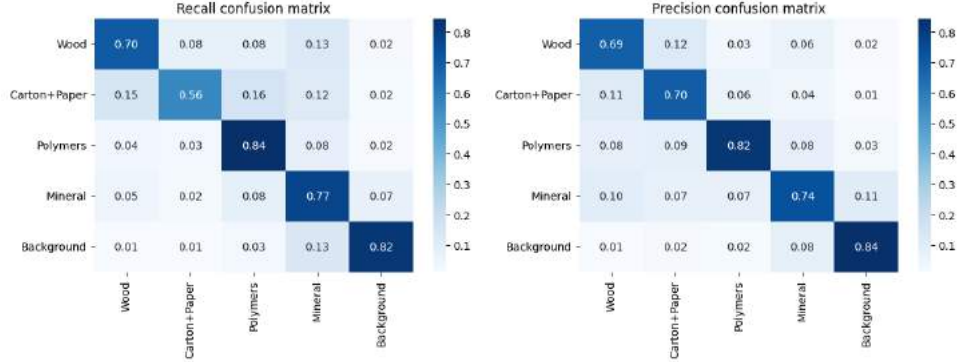


Figure 11: Precision and Recall normalized Confusion Matrices

## 6 Discussion

### 6.1 Limitations

- SAM’s overall performance is not real-time when using a heavy image encoder, which creates a problem for use in real-time applications because of the heavy computational resources it requires.
- Even though we tried to fine-tune SAM’s parameters, we still got many masks that were not very informative, as well as missing important objects. Sam is also not great at capturing masks of sparse and thin objects, such as metal tiles or a group of small pebbles. There is a trade-off between the number of masks and the quality of the masks, and there is no one configuration that will produce the perfect result for all images.
- We tried to create accurate annotations, however, labeling errors are unavoidable in some cases. In addition, we don’t necessarily have the knowledge to always distinguish between the different classes, and even though we consulted with our project leader from the School of Environmental and Earth Science at Tel Aviv University, there might still be some annotation errors.
- The heavy class imbalance, the lack of data, and the fact most of the images the model was trained on were from the same area and were taken at the same day, somewhat limit the ability to use our result as a reliable tool for analyzing never seen images. Our method might not generalize well to other areas with different climates and environments. More work is needed to ensure the robustness of the model and to improve its overall accuracy.

### 6.2 Future work

In the future, we would like the project to expand, including collecting more data from more areas. We had no over-fitting problems, so we believe that the model can be improved by adding more data. We used only drone images, but we believe that the model can be improved by adding images from other sources like images from the ground and maybe even satellite images. Furthermore, Since our classification objects are segmented masks (isolated from their environment), we might add images from a different distribution for the under-represented classes, such as metal and rubber. Metal object are relatively rare in our dataset, probably because metal collectors collect them before the drone arrives, and rubber is rare in the sites that our data came from.

When creating the dataset, we annotated the images with more specific 22 categories which we unified:

Metal (general)	Metal (iron bender)	Metal (pipe)	Wood (pallet)	Wood (scraps/cuttings)	Wood (general)
Carton	Nylon	Plastic (big bag)	Plastic (bucket)	Plastic (general)	Plastic (pipe)
Plastic (sand bag)	Gypsum	Concrete	Ceramics	Rubber	Paper
Styrofoam	Textile	Glass	Background		

Those categories can be used in the future to create a more sensitive model and more detailed segmentation map. f.e. we plan to separate the Polymer category into Rigid Polymers (such as plastic buckets, big sandbags,



etc.) and Flexible Polymers (such as nylon bags).

We would also like to investigate more of SAM’s capabilities and play around with its hyperparameters, including giving SAM input prompts such as points or boxes. This can potentially be used to generate masks for all objects in an image, including masks that were missed during the automatic segmentation.

Preferably the drone images are taken such that each split is no more than 3 meters wide and tall. Otherwise, the masks tend to be too small and of low quality. So the full image shouldn’t be more than 12 x 12 meters (The exact elevation to take the pictures from might vary from drone to drone). To capture better masks and avoid shaded areas, we think the best time to take the pictures is around noon when the sun is at its highest point.

Current results might not be good enough to use this model as a fully autonomous tool. However, under human supervision, this tool can be used to significantly reduce the time and effort required to analyze the images and produce a segmentation map. (surprisingly the model predicted rubber well, probably because ResNet-50 was trained on car wheels).

In order to produce financial outcomes and pragmatic insights, an algorithm for estimating the volume of recyclable material from the final segmentation mask still needs to be developed.

In the future, we could look into more advanced models such as ViT [9] for the classification task, as well as other methods like an ensemble of different models.

## 7 Credits

Adi Magar for her guidance and support throughout the project and in the annotation process. Dr. Moni Shahar and Ido Cohen for their help with the technological and technical aspects. Prof. Vered Blass for help with the environmental considerations and other suggestions.

## References

1. *The Israeli Ministry of Environmental Protection* 2022. [https://www.gov.il/en/departments/ministry\\_of\\_environmental\\_protection/govil-landing-page](https://www.gov.il/en/departments/ministry_of_environmental_protection/govil-landing-page).
2. Mager, A. & Blass, V. From Illegal Waste Dumps to Beneficial Resources Using Drone Technology and Advanced Data Analysis Tools: A Feasibility Study. *Remote Sensing* **14**. ISSN: 2072-4292. <https://www.mdpi.com/2072-4292/14/16/3923> (2022).
3. Kirillov, A. *et al.* Segment anything. *arXiv preprint arXiv:2304.02643* (2023).
4. Long, J., Shelhamer, E. & Darrell, T. *Fully convolutional networks for semantic segmentation* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), 3431–3440.
5. Girshick, R. *Fast r-cnn* in *Proceedings of the IEEE international conference on computer vision* (2015), 1440–1448.
6. Strudel, R., Garcia, R., Laptev, I. & Schmid, C. *Segmenter: Transformer for semantic segmentation* in *Proceedings of the IEEE/CVF international conference on computer vision* (2021), 7262–7272.
7. He, K., Zhang, X., Ren, S. & Sun, J. *Deep residual learning for image recognition* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), 770–778.
8. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), 248–255.
9. Dosovitskiy, A. *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021. arXiv: [2010.11929](https://arxiv.org/abs/2010.11929) [cs.CV].



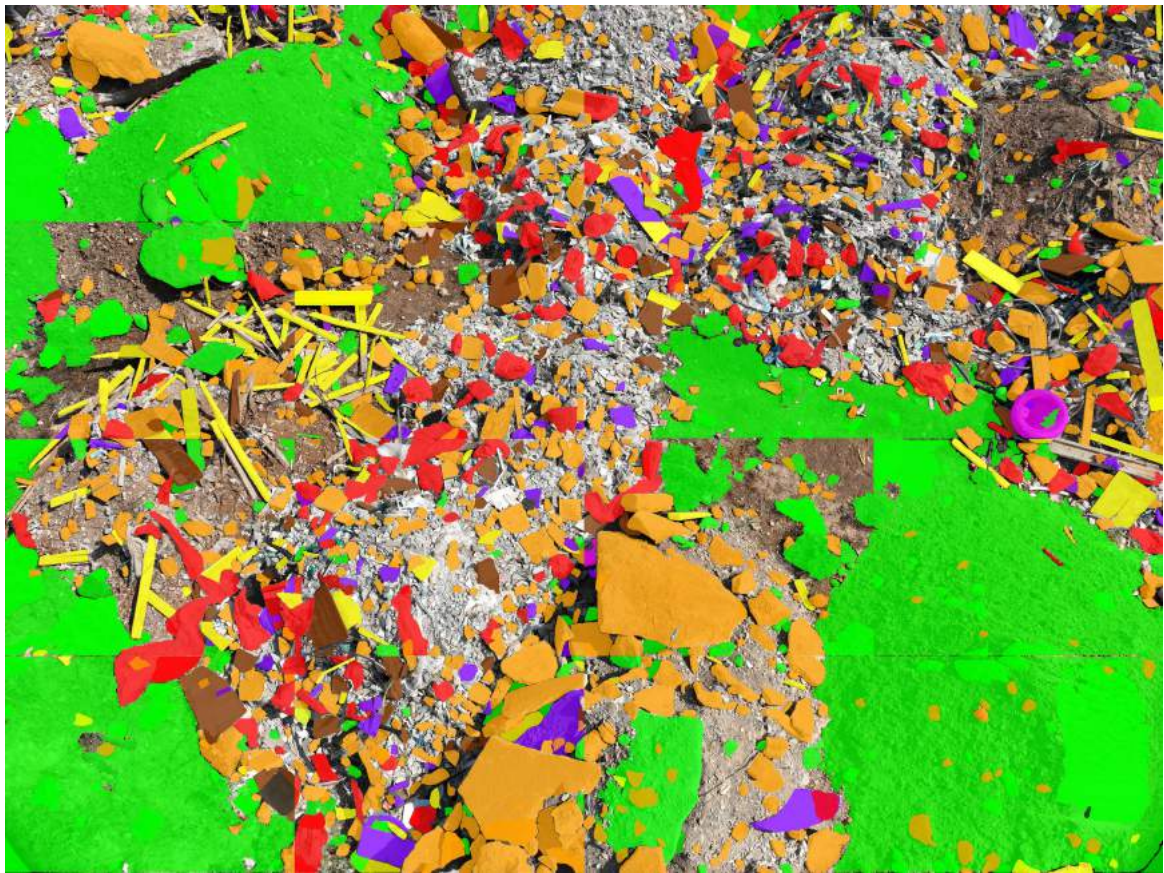


Figure 12: Full image example 1



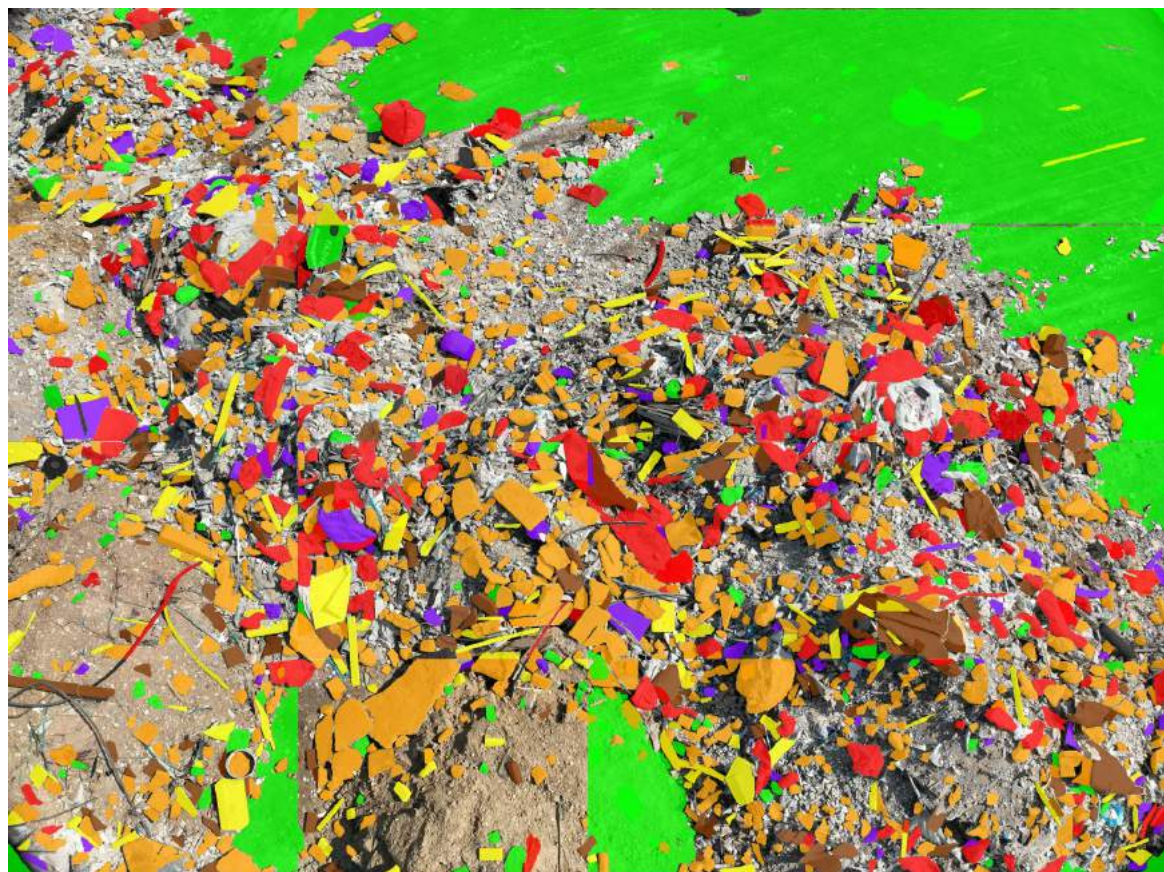


Figure 13: Full image example 2



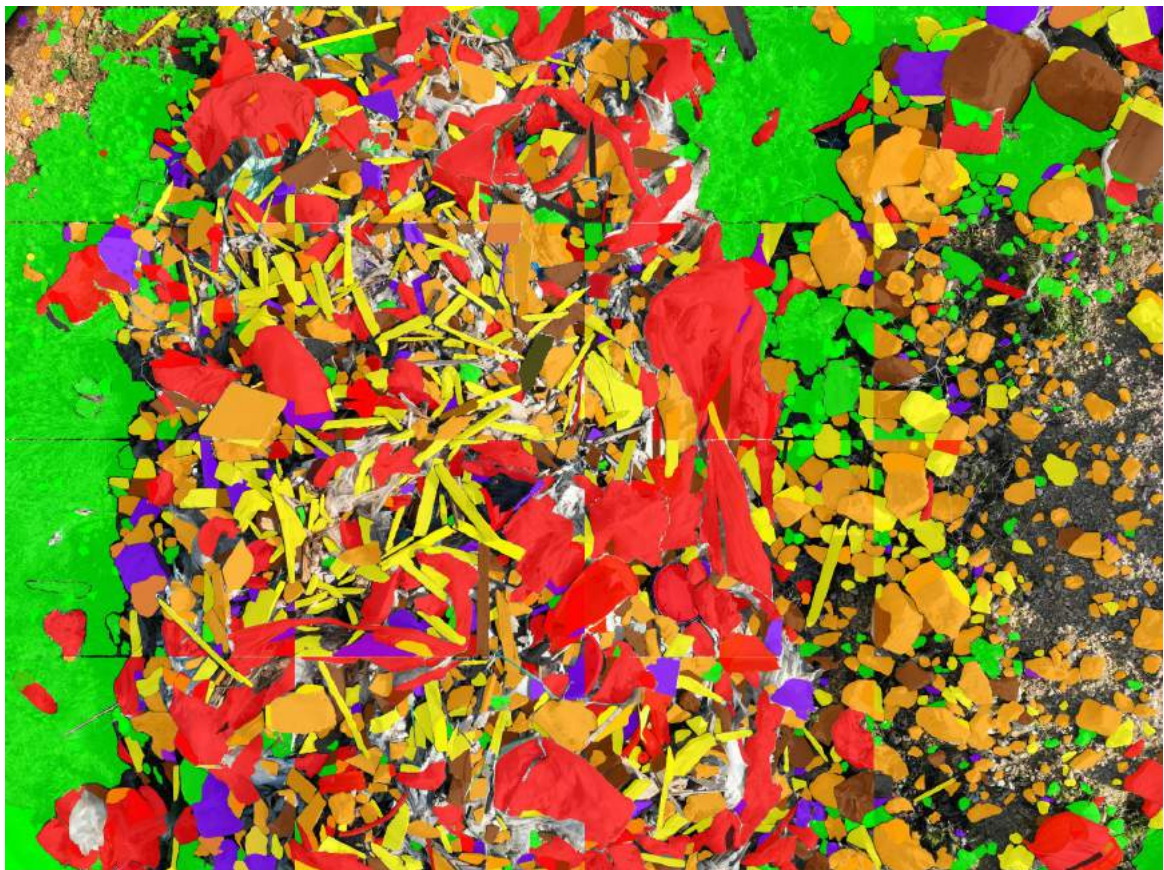


Figure 14: Full image example 3