Report 1

Aerial Perspective Object Detection

Prepared by:

Jukai Hu	400485702
Ray Albert Pangilinan	400065058
Luke Vanden Broek	400486889

October 12, 2022

Table of Contents

Jukai Hu	400485702
Ray Albert Pangilinan	400065058
Luke Vanden Broek	400486889

1. Problem Statement

In this project, what we have to achieve is to train a model to identify the different objects in birds' eyes view and then apply it to identify the risk level of specific areas in which pedestrians might have conflicts with vehicles and other obstacles. This is an object classification question, where input images-in this project, they are those original images beheld by birds-are fed to the algorithms which will then classify the objects in those images with different labels (eg, "paved-area", "dirt", "grass", etc.) based on different color assigned to each pixel. After that, one way to determine the risk level of a certain classified area is to estimate the distance from this area to classified cars, bicycles, obstacles, and conflicting areas, which are potential obstacles for pedestrians.

2. Approach

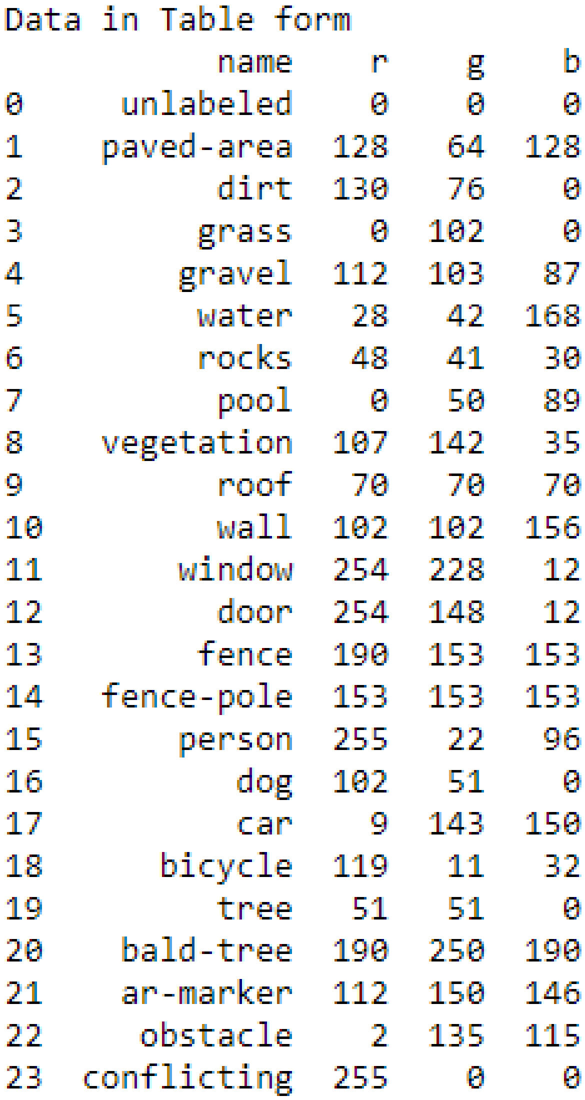
In this report, we will try to solve the segmentation problem with conventional machine learning approaches (Random Forest, KNN, and NB). Firstly, to create a training dataset for image classification, we should have annotated images with labels that will be used by the algorithm. Here in our project, we are provided with those label images semantic and RGB color images, which are already labeled for training. Then, we will feed those images to our supervised algorithms for training. After that, we could derive a classification of objects in input images based on different colors assigned to those objects. This is because we would have already trained our models with those training dataset, in which all objects are separated from each other using different pixel values.

In the next report, we would apply deep neural networks to work with the dataset training to see if it is better than Random Forest, KNN, and NB on behalf of final accuracy and computation complexity. All code for this report can be found in the linked GitHub page (Hu *et al.,* 2022).

3. Dataset Description

The dataset used for this project includes 400 different images at resolutions of 6000 pixels x 4000 pixels. Included in the files are images of the data in their original format, semantic format and with an RGB color mask to identify different objects, as well as a .csv file that indicates the RGB color values representing each classification type (*Semantic Drone Dataset*, n.d).

4. Data Visualization



```
Data in Table form
          name    r    g    b
0    unlabeled    0    0    0
1   paved-area  128   64  128
2         dirt  130   76    0
3        grass    0  102    0
4       gravel  112  103   87
5        water   28   42  168
6        rocks   48   41   30
7         pool    0   50   89
8   vegetation  107  142   35
9         roof   70   70   70
10        wall  102  102  156
11      window  254  228   12
12        door  254  148   12
13       fence  190  153  153
14  fence-pole  153  153  153
15      person  255   22   96
16         dog  102   51    0
17         car    9  143  150
18     bicycle  119   11   32
19        tree   51   51    0
20   bald-tree  190  250  190
21   ar-marker  112  150  146
22    obstacle    2  135  115
23 conflicting  255    0    0
```

Figure 1: RGB colour values in tabular form

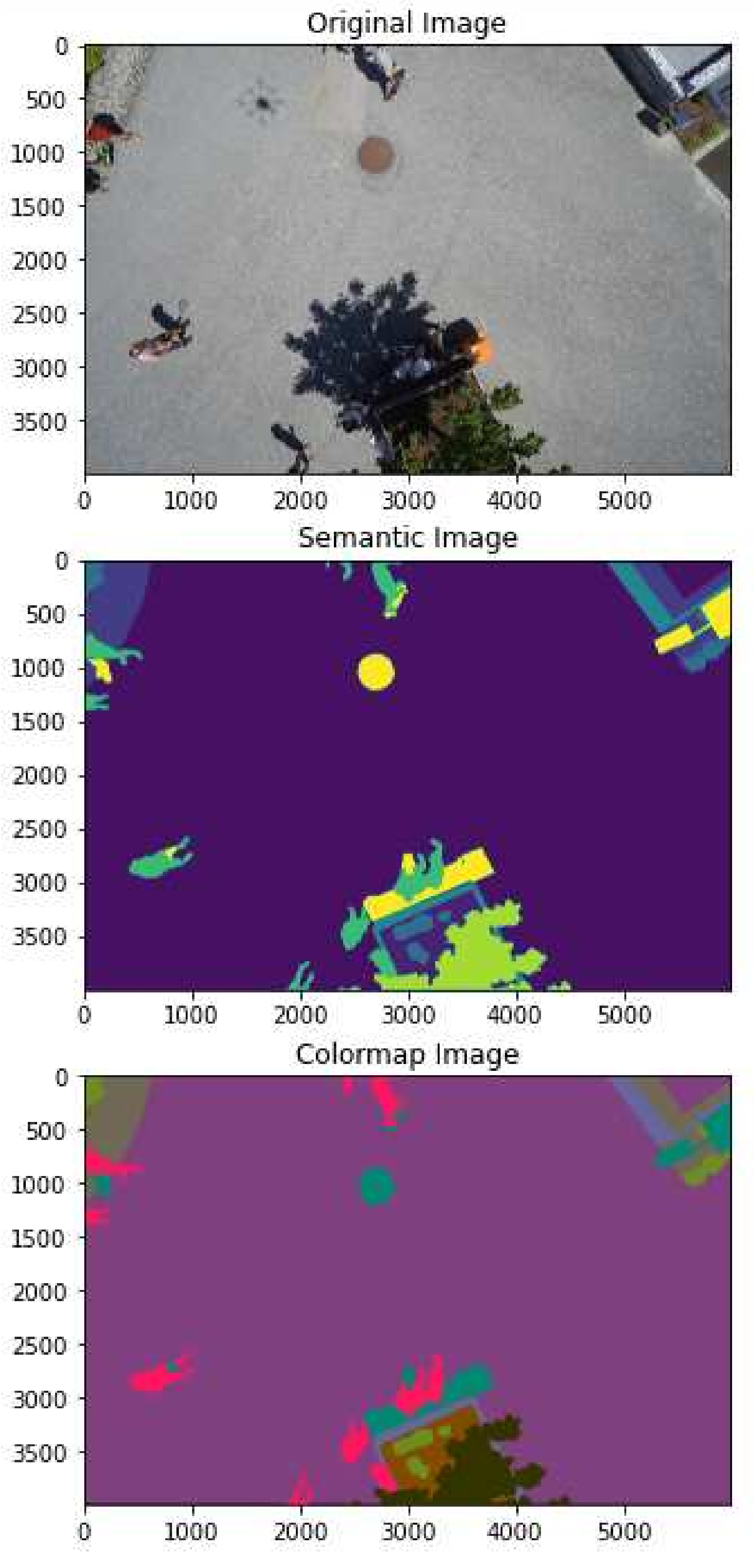


Figure 2: Plot of image 000 in original, semantic and colormap formats

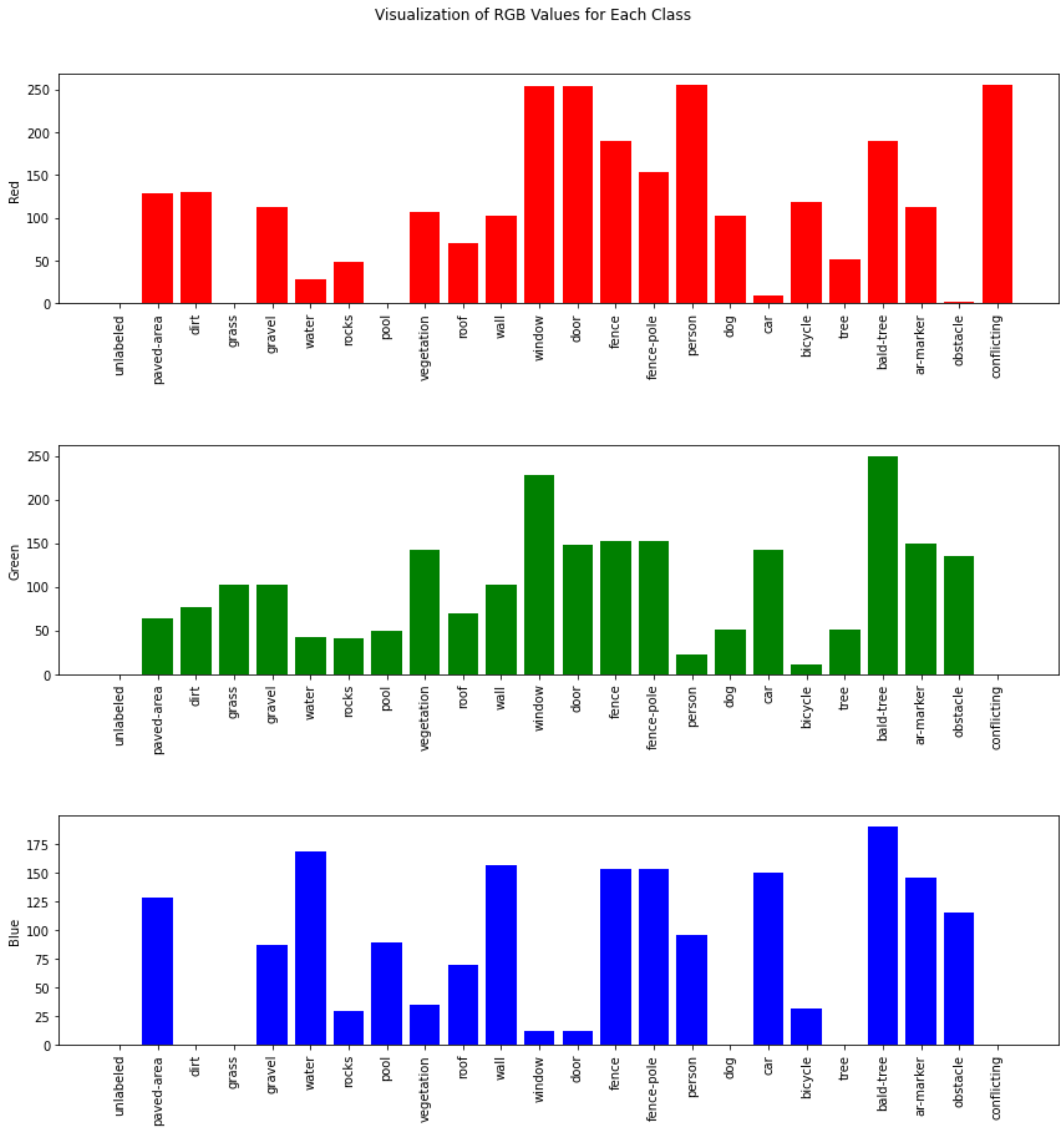


Figure 3: Visualization of RGB color values in histogram form

5. Task Description

Based on the classification of image pixels to different classes, as well as our project outcome to identify objects such as cars, people, and roads in drone imagery, we can determine that the main scope of our project will be to perform a classification task. Since the classes are provided by the dataset, we can also infer that our problem falls under supervised learning. Furthermore, this task can be extended to a multi-class classification given the quantity of twenty-four labels outlined by the dataset.

6. Data Distribution

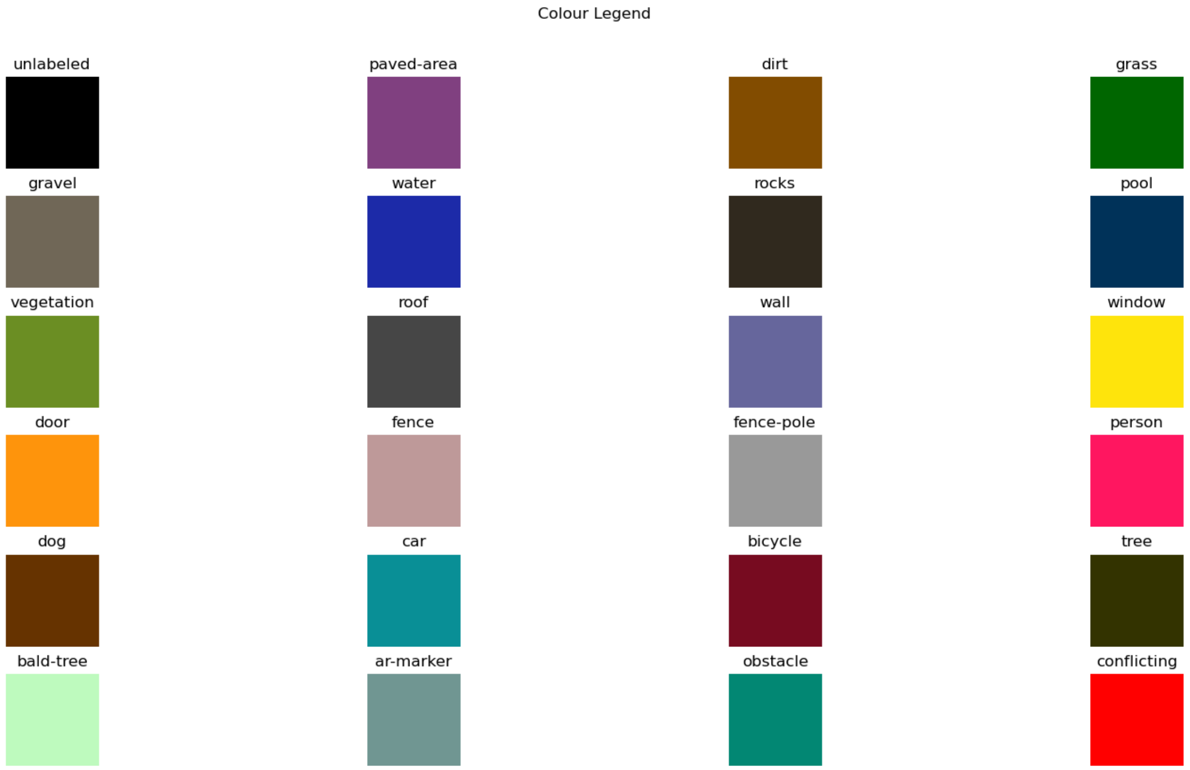


Figure 4: Legend for colormap values

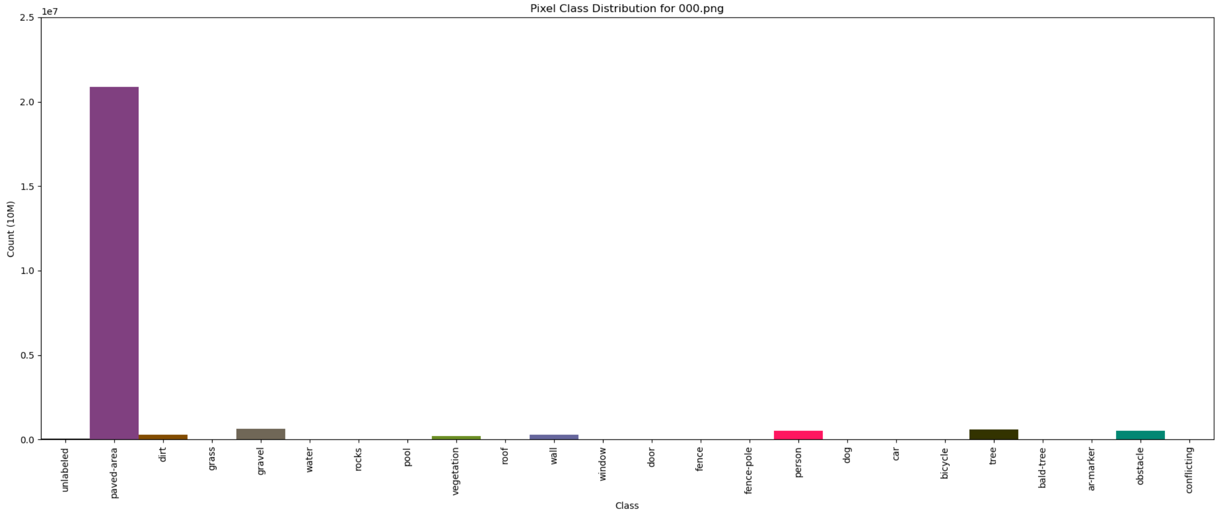


Figure 5: Pixel quantities for image 000

7. Problem Solution

We began work on our project task by first determining how we might structure the data for input into machine learning models. Given that our raw data is in the form of images, we imported them into a Python script as three-dimensional numpy arrays of dimensions 4000 x 6000 x 3, corresponding to the pixel width and height of the images as well as the RGB values of each pixel. We then flattened the initial array, resulting in a two-dimensional array with dimensions 24,000,000 x 3, with each pixel representing a data point and the RGB values representing features. Performing the same data manipulation steps on the semantic images gave us our pixel labels in a matching format With each image in a format sufficient for training conventional machine learning algorithms, we fitted nine images in the dataset to Random Forest, K-Nearest Neighbours, and Naive Bayes models, and predicted the class labels for a tenth image. While we would have liked to expand our training and test sets to include more of the dataset, as well as introduce cross-validation, the extremely high image resolution resulted in training times of about an hour. We also attempted to train a support vector machine model alongside the three aforementioned algorithms but did not see the algorithm proceeding past training the first image.

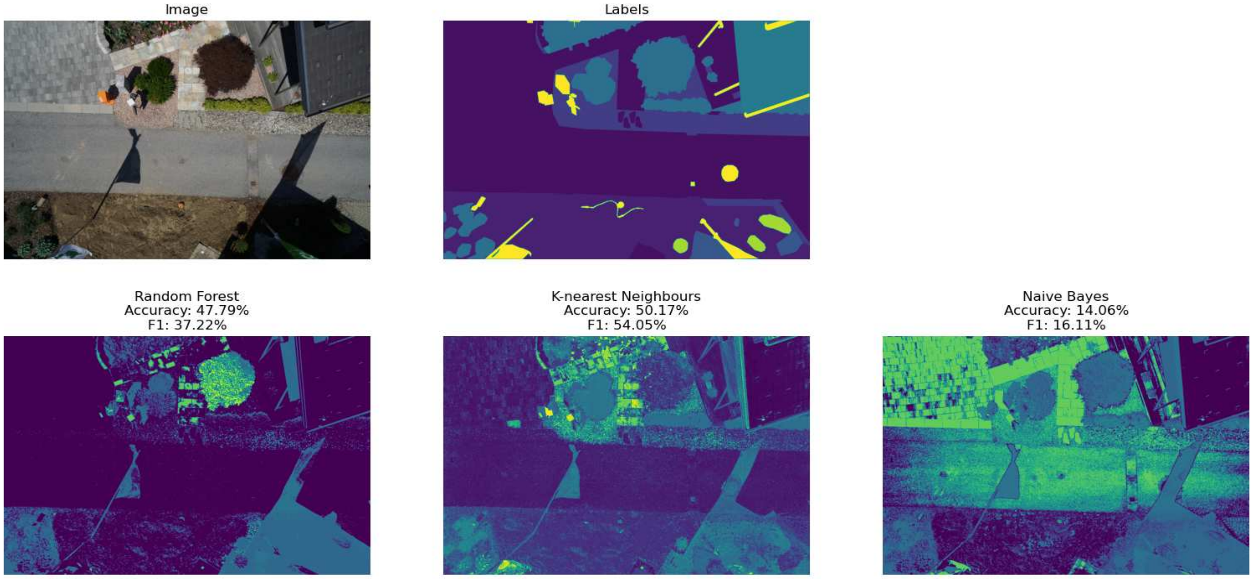


Figure 6: Results of machine learning techniques

Analyzing the prediction results of the three chosen machine algorithms, we noted their performance to be underwhelming, with both Random Forest and K-Nearest Neighbours posting accuracy ratings and F1 scores around 40-50%. Additionally, naive Bayes did not perform well at all, with about 14% accuracy and 16% F1 score. Switching to a visual perspective of the results, we observed that the algorithms tended to incorrectly classify areas covered by shadows, as well as show lackluster results in terms of segmenting regions of the image. We theorize that this was due to our algorithms performing classification on a pixel-by-pixel basis, as well as using the RGB values of the pixels as the only features. Furthermore, we believe that more complex approaches such as feature selection, normalization, hyperparameters, and the use of convolutional neural networks will help alleviate these issues as well as the long processing times noted above.

8. References

Hu, J., Pangilinan, R. A., & Vanden Broek, L. (2022). *Albertpangilinan/ SEP740*. GitHub. Retrieved October 11, 2022, from https://github.com/AlbertPangilinan/SEP740

Semantic Drone Dataset. Institute of Computer Graphics and Vision. (n.d.). Retrieved October 11, 2022, from https://www.tugraz.at/index.php?id=22387