Towards Robust Terrain Classification Using Hybrid CNN Models

Spring 2024: EE/CS 5841 Machine Learning Project

Group 3: Ian Q. Mattson, Anders Smitterberg, Satyanarayana Velamala

Overview

Successful locomotion of quadruped and bipedal robots is reliant on a good understanding of the ground surface. While most work in the field is completed on flat indoor or outdoor surfaces, such as tiles and concrete, there are many more terrains a rugged quadruped may be expected to traverse. Different terrains however, have different surface conditions, and as the terrain changes, the robot must adapt it's gait and controls to account for loose, sticky, or slippery surfaces, such as sand, dirt, snow, rocks, and grass. Neural networks can be trained on images of various terrain types to inform the robot when to adjust gait based on terrain. Furthermore, we propose that use of point cloud information in addition to standard images from a RGB depth camera can improve terrain classification accuracy.

Background

The design of a robot is typically a tradeoff between optimizing locomotion in some environments at the expense of locomotion in others. Quadrupedal robots are uniquely positioned to overcome this challenge as they are able to adopt many different gaits, or styles of walking, similarly to the way four legged animals can traverse a variety of terrains. If terrain information is passed to the robot, it could adapt it's gait accordingly.

Dataset Generation

Table 1: Data quantities listed by class for images and pointcloud, notable totalling almost 80k images and 18.3k point clouds representing a combined total of roughly 180 gigabytes

Class	Reference	Total Images	Total Point Clouds	Location
Asphalt	1	15548	4228	MTU Parking Lots
Bricks	2	1226	214	Husky Statue Pavers
Dirt	3	9726	1151	Tech Trails, Walker Lawn
Grass	4	13060	1582	Walker Lawn
Gravel	5	2689	401	Tech Trails, Husky Statue
Plants	6	3217	241	Dillman Rock Garden
Sidewalk	7	12911	3482	MTU Campus
Snow	8	15942	5452	Tech Trails
Steel Grate	9	1440	476	Walker Lawn
Tile Floors	10	3718	1153	Tech Trails Waxing Center
Total	-	**79477 images**	**18380 point clouds**	**10 Locations**
Useful Size	-	**37.2 GB**	**141.0 GB**	**178.2 GB Total Dataset Size**

Due to the unique nature of our terrain available around MTU's campus in early Spring, we collected ROS bags of both the standard visual image from the Intel Realsense camera, as well as the RGB-D point cloud from the infrared depth stereo camera. Individual ROS bags were captured by terrain type, so as to assist in labelling clusters of images and point clouds with their respective classes.

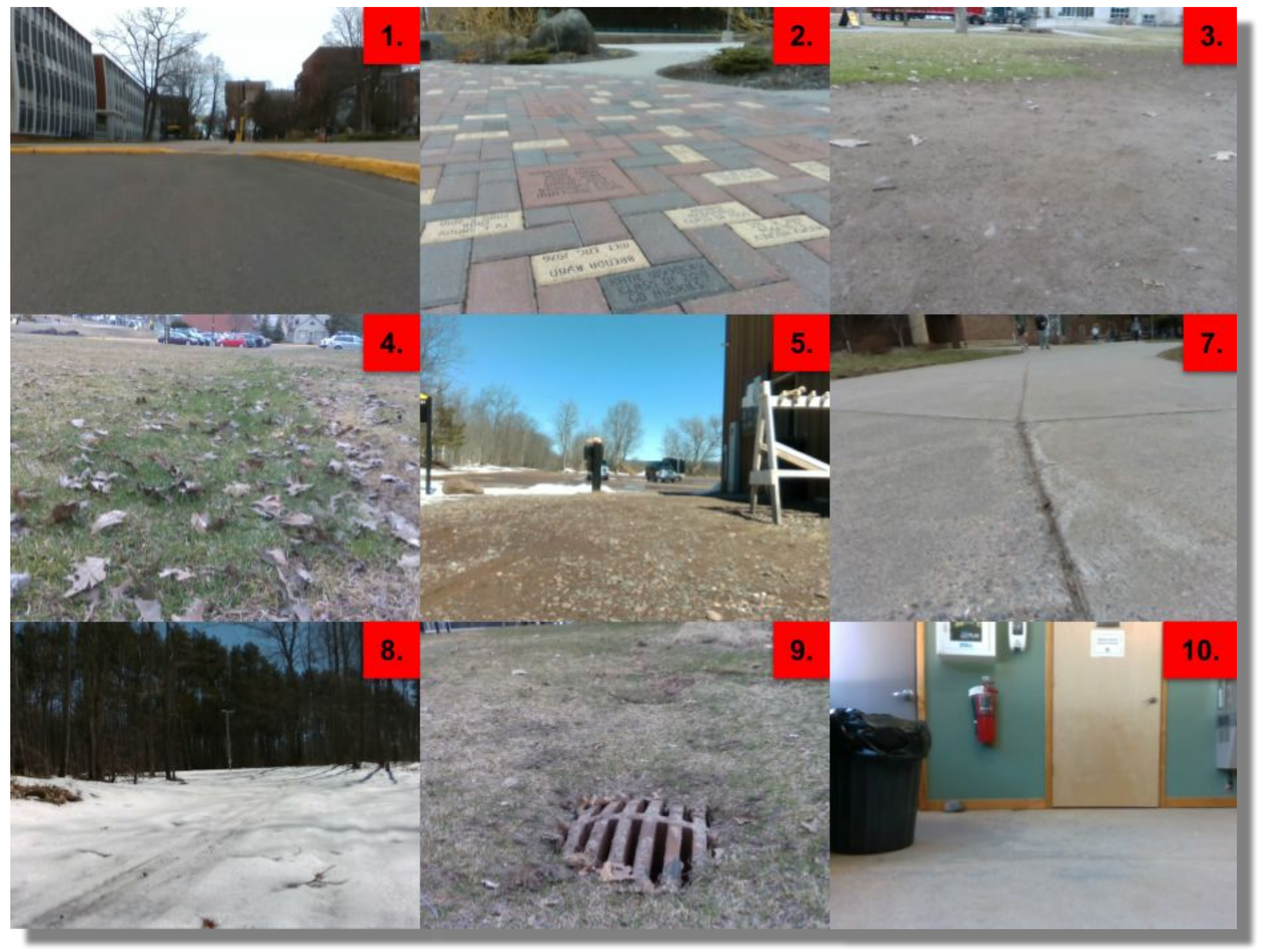


Figure 1: Samples of each terrain classified and present in the dataset. A complete description of each class, as well total representation in the dataset, is illustrated in Table 1.

Point clouds are organized as sets of XYZ coordinate points in standard euclidean space. They are typically generated from laser scanners or LiDAR, and have accurate depth perception. We used RGB point clouds to add context to aid in terrain classification.

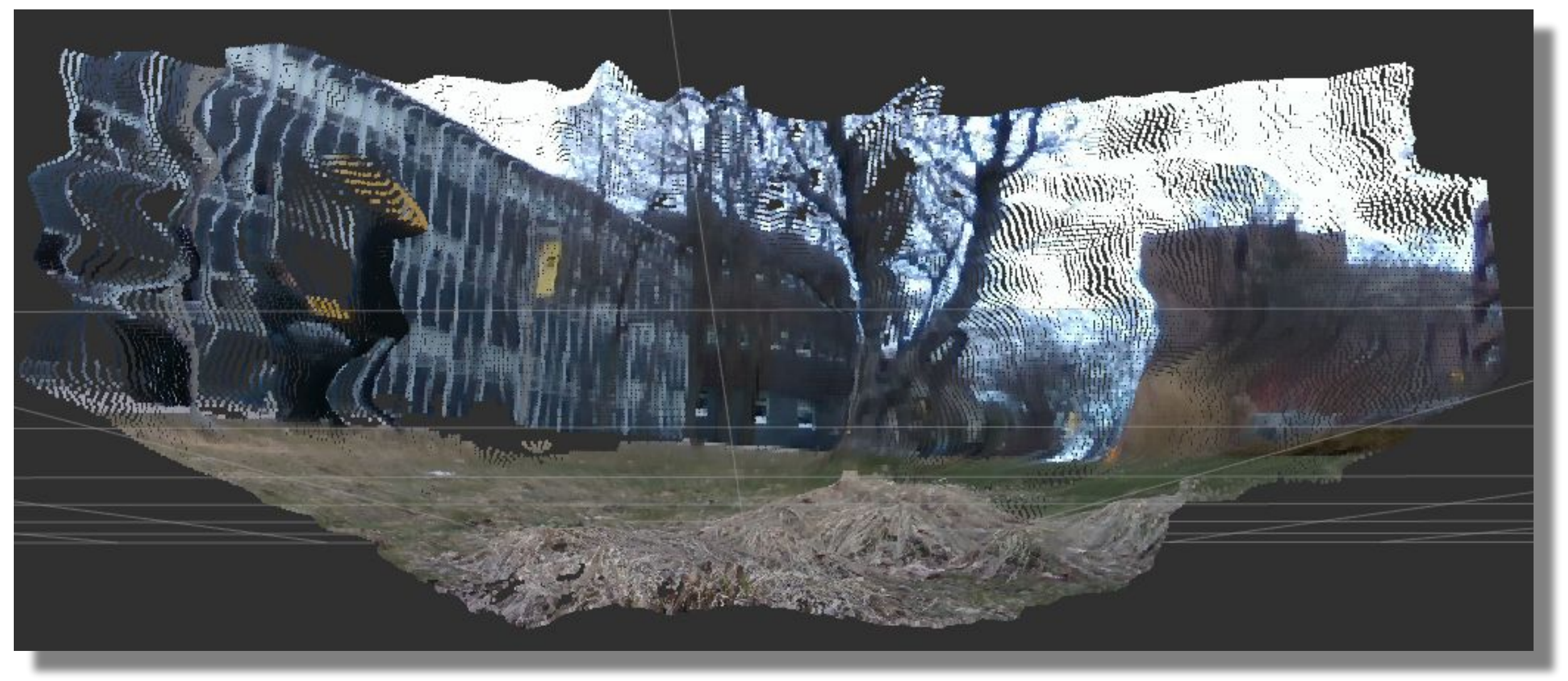


Figure 2: Representative RGB colorized-point cloud of the Dillman Rock garden illustrating terrain depth variation due to plants and vegetation

CNNs – Two of Them

A convolutional neural network to classify terrain types based on images was the obvious choice as they have been demonstrated to be extremely effective at image classification tasks and are relatively transformation invariant. ResNet50 is a well trained classifier that could be practically trained on our GPU hardware. A CNN was also chosen for the point cloud data for similar reasons. Point cloud data could be passed into the neural network as a four dimensional image, and processed by the network. A famous example of pointnet classification is pointNet, which also uses CNNs.

Our system trained two separate CNNs, for images, and point clouds, and used information from both of their outputs to determine terrain type

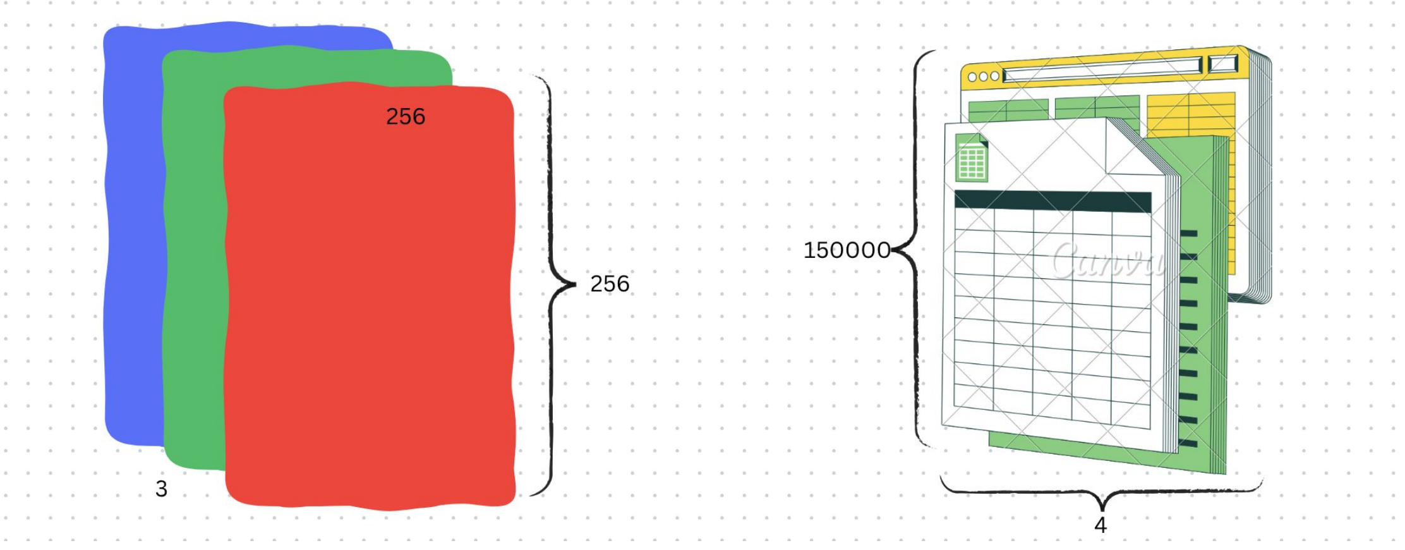
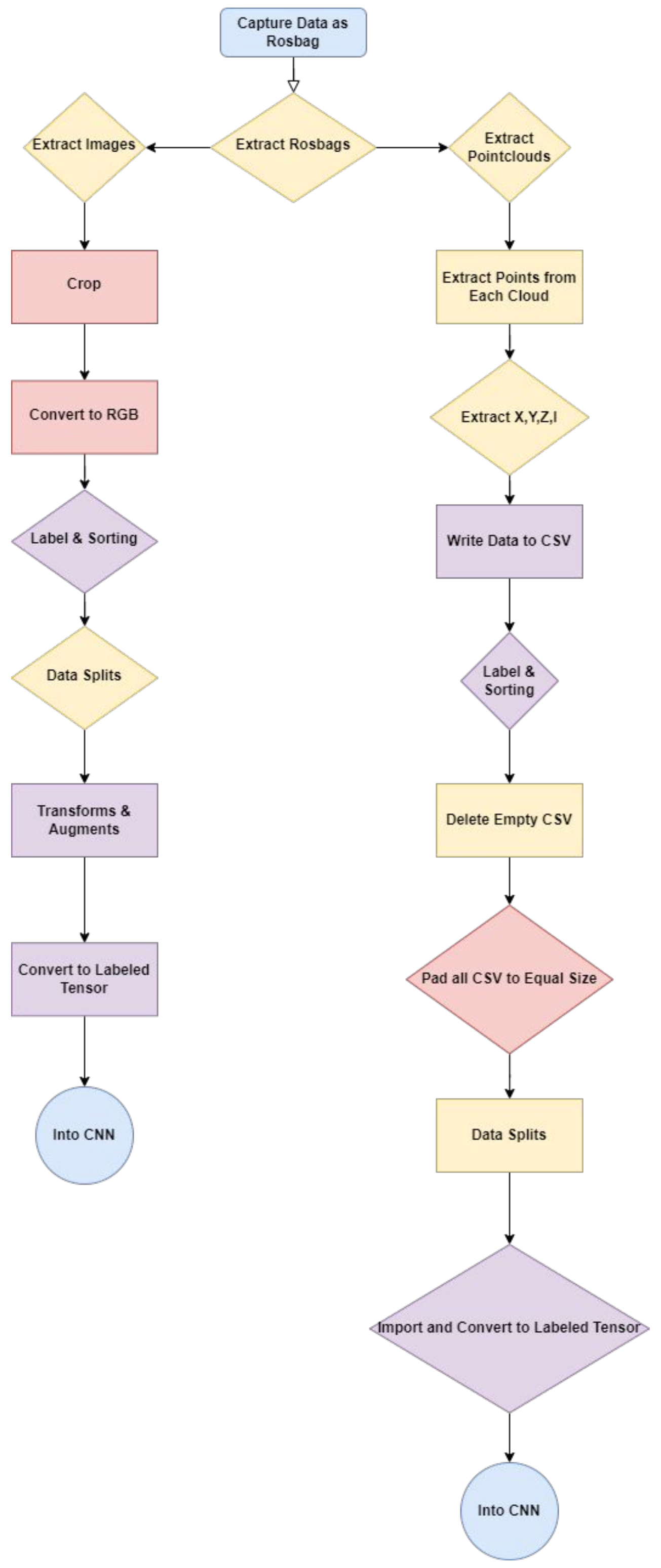


Figure 4: Input layer size from RGB images and point cloud data. 4 pointcloud layers are representative of x,y,and z coordinate and reflective intensity

Processing Pipeline

All of the data for this project was collected and processed by us. In total hundreds of gigabytes of data were collected, and had to be labeled by hand and prepared for consumption by the neural networks. The data were captured from a variety of locations with both the 3D point cloud data and images simultaneously. Both the color images and point clouds were labeled and prepared to be processed. Several scripts were used to speed up this work, however even so this was by far the most time consuming and tedious portion of the project. The image data was resized and converted to RGB, then prepared to be processed by ResNet50. The point cloud data was extracted to a four dimensional CSV, X,Y,Z, and Intensity, and then padded such that all the incoming data was the same dimension. These data could then be fed into the two separate neural networks.


Capture Data as Rosbag
Extract Rosbags
Extract Images
Extract Pointclouds
Crop
Convert to RGB
Label & Sorting
Data Splits
Transforms & Augments
Convert to Labeled Tensor
Into CNN
Extract Points from Each Cloud
Extract X,Y,Z,I
Write Data to CSV
Label & Sorting
Delete Empty CSV
Pad all CSV to Equal Size
Data Splits
Import and Convert to Labeled Tensor
Into CNN

Before training the CNN, images were randomly manipulated to generate noise and variations to add robustness, especially against unintended robot motion, or varying lighting conditions.

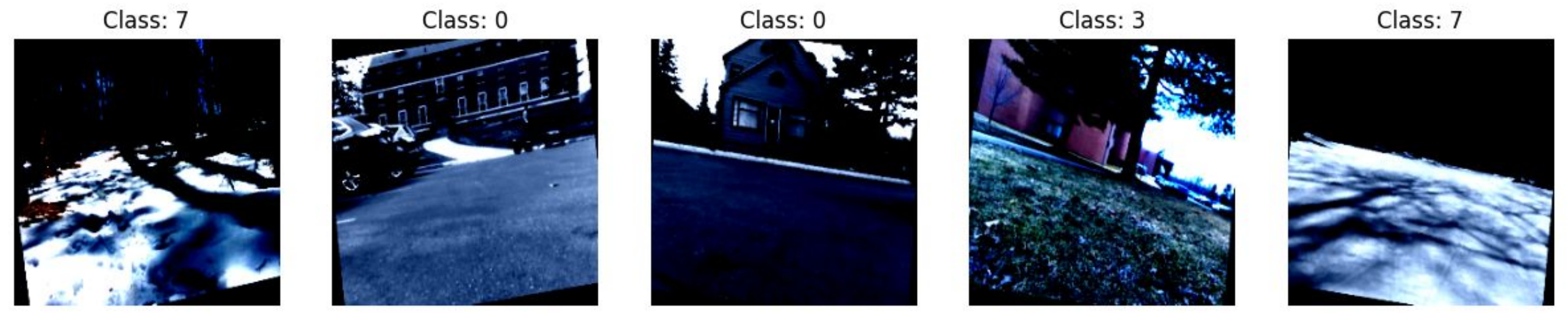
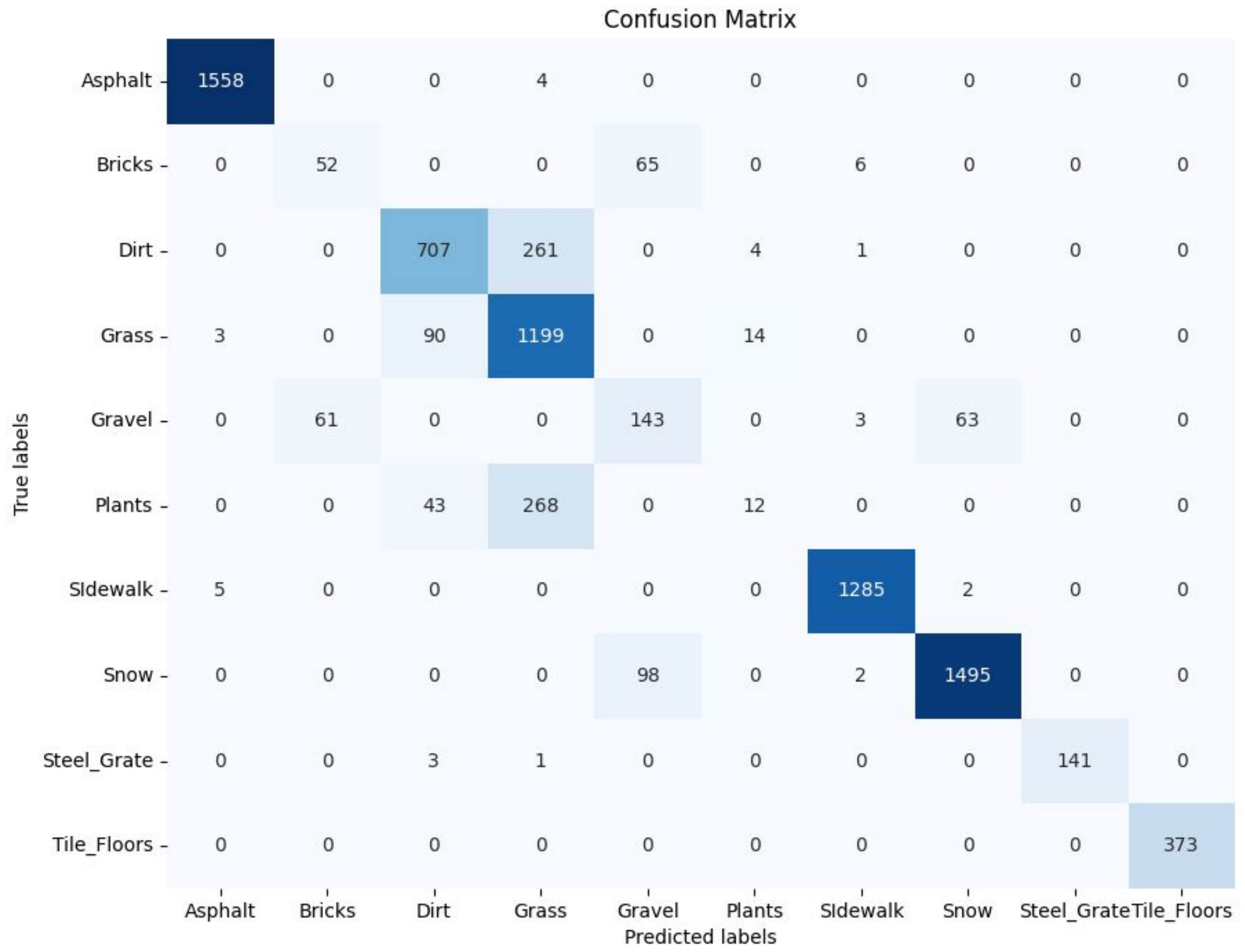


Figure 3: Five random samples of transformed training data, where valid transformations are random horizontal flip, rotation, and grayscale

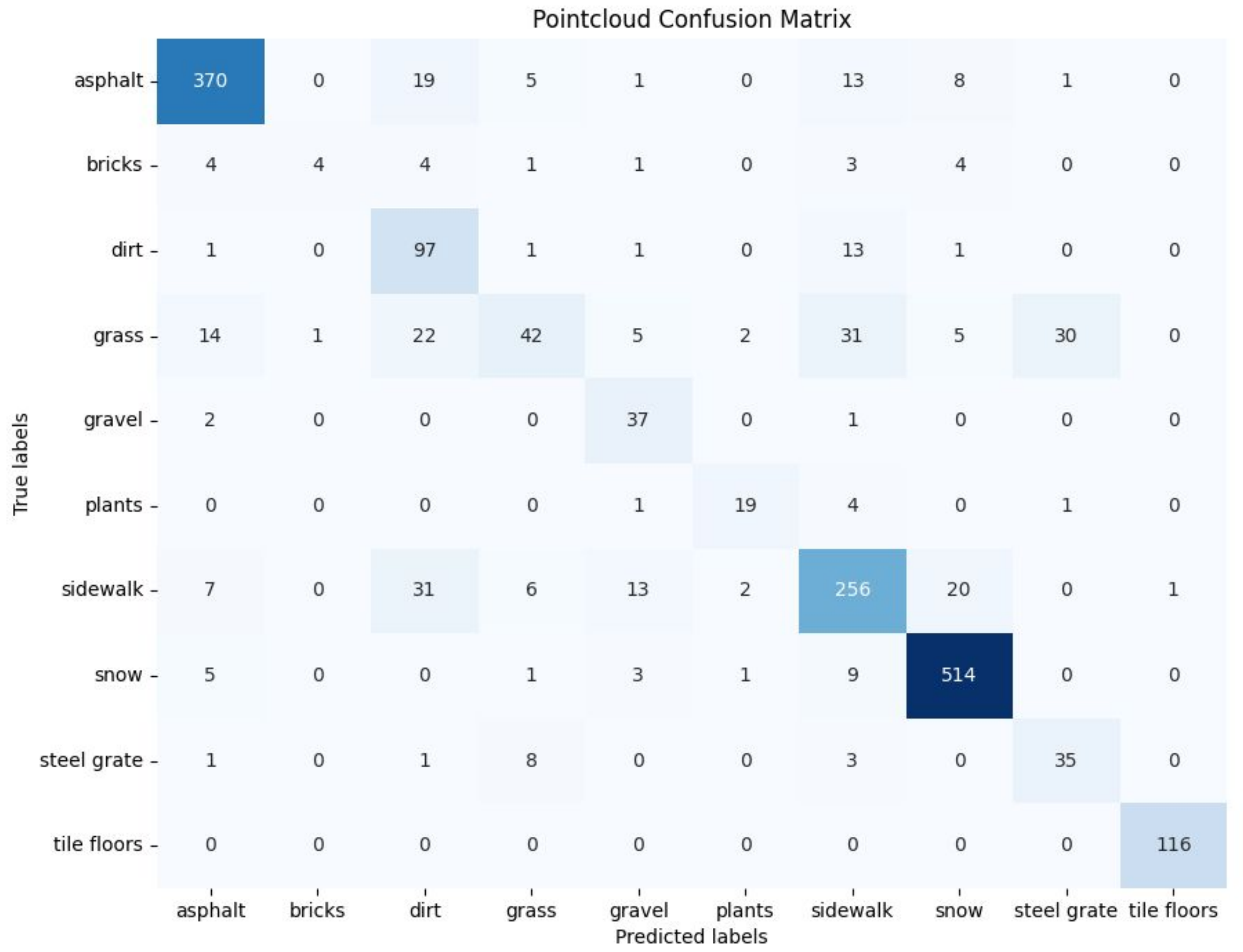
Results

Depicted below are the separate confusion matrices for both the RGB image classifier and point cloud classifier trained on the test split of both of the full datasets. Similar patterns can be observed in both of the confusion matrices. Misclassifications are generally logical.



Confusion Matrix

True labels \ Predicted labels	Asphalt	Bricks	Dirt	Grass	Gravel	Plants	Sidewalk	Snow	Steel_Grate	Tile_Floors
Asphalt	1558	0	0	4	0	0	0	0	0	0
Bricks	0	52	0	0	65	0	6	0	0	0
Dirt	0	0	707	261	0	4	1	0	0	0
Grass	3	0	90	1199	0	14	0	0	0	0
Gravel	0	61	0	0	143	0	3	63	0	0
Plants	0	0	43	268	0	12	0	0	0	0
Sidewalk	5	0	0	0	0	0	1285	2	0	0
Snow	0	0	0	0	98	0	2	1495	0	0
Steel_Grate	0	0	3	1	0	0	0	0	141	0
Tile_Floors	0	0	0	0	0	0	0	0	0	373



Pointcloud Confusion Matrix

True labels \ Predicted labels	asphalt	bricks	dirt	grass	gravel	plants	sidewalk	snow	steel grate	tile floors
asphalt	370	0	19	5	1	0	13	8	1	0
bricks	4	4	4	1	1	0	3	4	0	0
dirt	1	0	97	1	1	0	13	1	0	0
grass	14	1	22	42	5	2	31	5	30	0
gravel	2	0	0	0	37	0	1	0	0	0
plants	0	0	0	0	1	19	4	0	1	0
sidewalk	7	0	31	6	13	2	256	20	0	1
snow	5	0	0	1	3	1	9	514	0	0
steel grate	1	0	1	8	0	0	3	0	35	0
tile floors	0	0	0	0	0	0	0	0	0	116

Conclusions & Recommendations

Confusion matrix results are reasonable and expected given multiple data types being present under some labels, as well as noise and a comparatively small amount of samples for some classes.

Going forward, it would be valuable to combine the outputs of both CNNs with a large, fully connected layer to combine data types. We would also like to test this system with nighttime data, where we anticipate Point cloud classification may be stronger.

Model	Epochs	Validation Loss	Validation Accuracy
Images Only	50	0.2138	87.71%
Point Clouds Only	15	0.5373	82.0%