

1 Results

In the section we show and evaluate the models results. We will start by presenting to the reader the networks score on each metric, then we will use the best models to predict the traversability of real world terrain. Finally, we will use handcrafted patches, for example a wall of a certain height in a specific position, to test the robustness of the network by trying to highlighting its behaviour.

1.0.1 Experiment Setup

We run all the experiment on a work station with Ubuntu 18.10 operating system. The machine is equipped with a Ryzen 2700x, a powerful CPU with 8 cores and 16 threads, and a NVIDIA 1080 GPU with 8GB of dedicated RAM.

To train the network we follow the best practice on residual network [he2015deep] using Standard Gradient Descent with momentum set to 0.95 and weight decay to $1e - 4$ with an initial learning rate of $1e - 3$. We fix the maximum number of train epochs to 30 and reduce the learning rage on platuoe by a factor of 0.1 with a patience of 4. We used early stopping to stop the training if the validation accuracy does not increase in 6 epochs.

fix this typo

We select a threshold of $0.2m$ on a time window of two seconds to label the patches, meaning that a patch with an advancement less than 20 centimeters is labeled as *no traversable* and viceversa.

1.0.2 Experimental validation

We select as *validation* ten percent of the training data. We remain to the reader that we store each run of *Krock* as a *.csv* file. So, to avoid any biases, we used completely different dataframes, meaning that train and validation sets are composed by non overlapping data from the simulations. The test set is composed by fifty simulations run on the *Quarry* map.

add more maps if we add them to the test set

1.0.3 Metrics

To evaluate the model's performance we used two metrics: *accuracy* and *AUC-ROC Curve*. Accuracy is defined simply by the number of correct predictions made by the network while AUC-ROC Curve represents degree or measure of separability, informally it tells how much model is capable of distinguishing between classes. For each experiment, we select the model with the higher AUC-ROC Curve during training to be evaluated.

1.1 Quantitative Results

The following table shows the final results on various test dataset made by using real-world heightmaps.

table with results

Moreover, we would like to also show the different steps we made to reach this result. The following table shows the metric's score without any data-augmentation. Adding dropout increases the results.

table with results

With dropout plus coarse dropout.

table with results

1.2 Qualitative results

We qualitative evaluate the models in real world scenarios by computing the traversability probability for each map with different rotation. Specifically, we used a sliding window to extract the patches from the heightmaps and colour by blue the relative region with the coresponding traversability probability. A brighter colour yields an higher probability. The following pictures show the traversability probability on the *Query* assuming *Krock* is walking from bottom to top.

We can notice that the robot has no problems traversing the slopes near the top of the map thanks to its legs.