

# Chapter 1

## Results

In this section, we evaluate five convolutional neural networks: the model proposed by Chavez-Garcia et all ? and the four MicroResnet presented in the last chapter. We will select the best performing one based on the classification score on the test set. We will show the results of the same architecture trained on regression instead of classification and prove the latter yields better results. Finally, we qualitatively evaluate the best model by predicting traversability in real-world terrains.

### 1.1 Hardware

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

### 1.2 Experimental validation

We selected as *validation* set ten percent of the training data. Since we stored each run of Krock as a *.csv* file, validation and train set do not overlap.

#### 1.2.1 Metrics

**Classification:** To evaluate the model's classification performance we used two metrics: *accuracy* and *AUC-ROC Curve*. Accuracy scores the number of correct predictions made by the network while the 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 selected the model with the higher AUC-ROC Curve during training to be evaluated on the test set.

**Regression:** We used the Mean Square Error to evaluate the model's performance.

### 1.3 Training setup

Initially, to train the models we first used Standard Gradient Descent with momentum set to 0.95, weight decay to  $1e - 4$  and an initial learning rate of  $1e - 3$  as was originally proposed in He et al.

? . However, we later switched Leslie Smith’s 1cycle policy ? that allowed us to train the network faster and with higher accuracy. The average training time is  $\approx 10$  minutes. We minimized the Binary Cross Entropy (BCE) for the classifier and the Mean Square Error (MSE) for the regressor.

## 1.4 Quantitative results

### 1.4.1 Model selection

We compared the four different MicroResnet architectures ?? and the Vanilla CNN ?? proposed in the original work ? on the test set using as metric the ROC AUC. Ideally, the second MicroResnet model should be able to extrapolate more features from the patch thanks to the big kernel size. We evaluated those models using a time window of two seconds, a threshold of 20cm and we applied data augmentation as we described in section ?? . We run five experiments per architecture and we selected the best performing network. Table 1.1 shows the results for each architecture. The best performing model was MicroResNet7x7-SE that obtained a ROC AUC of 0.896 on the test set. Moreover, the Squeeze and Excitation operator improved the performance of both MicroResNet3x3 and MicroResNet7x7 but only latter gained a huge boost of 0.021.

Model	ROC AUC		Params
	Mean	Top	
Vanilla	0.892	0.890	974,351
MicroResnet3x3	0.881	0.883	302,610
MicroResnet3x3-SE	0.883	0.888	313,64
MicroResnet7x7	0.867	0.875	303,250
MicroResnet7x7-SE	0.888	<b>0.896</b>	314,28

Table 1.1. Comparison of the four MicroResnet architectures ?? and the original convolutional neural network (Vanilla CNN) proposed in the original work ? on the test set using as metrics the ROC AUC. The best performing model is MicroResNet7x7-SE.

## 1.5 Final results

### 1.5.1 Classification

Table 1.2 table shows in deep the score of the best network for each dataset. We obtained 95% accuracy and 0.96 AUC on the validation set and 88% accuracy and 0.89 AUC on the test set.

### 1.5.2 Regression

We regressed on the advancement using the best model selected previously, MicroResNet7x7-SE. Table 1.3 table shows in deep the score of the best network for each dataset. The regressor can be used to classify patches using different thresold by just see if the predicted advancement is greater or lower than a threshold. However, we always need to fix a threshold to estimate traversabilit. Once the value is fixed, the same model trained on the same dataset to classify patches drastically

Dataset			MicroResNet		Size(m)	Resolution(cm/px)
Type	Name	Samples	ACC	AUC		
Synthetic	Training	429312	-	-	10 × 10	2
	Validation	44032	95.2 %	0.961	10 × 10	2
	Arc Rocks	37273	85.5 %	0.888	10 × 10	2
Real evaluation	Quarry	36224	88.2 %	0.896	32 × 32	2

Table 1.2. Classification results on different datasets.

Dataset			MicroResNet		Size	Resolution(cm/px)
Type	Name	Samples	MSE			
Synthetic	Training	-	-	-	10 × 10	2
	Validation	44032	0.006	0.006	10 × 10	2
	Arc Rocks	37273	0.020	0.020	10 × 10	2
Real evaluation	Quarry	36224	0.022	0.022	32 × 32	2

Table 1.3. Classification results on different datasets.

outperforms the regressor. Table 1.4 shows the accuracy of two MicroResNet7x7-SE on classification and regression after a threshold of twenty centimeters is selected. For this reason, we choose the MicroResNet7x7-SE trained on the classification task to be evaluated.

	Regression	Classification
	ACC	
Top	72.8%	<b>88.2%</b>
Mean	73.6%	<b>87.8%</b>

Table 1.4. Regression and classification accuracy for the same model, MicroResNet7x7-SE, once a threshold of 20cm is fixed. The classifier outperforms the regressor by 15%.

## 1.6 Qualitative results

We qualitatively evaluated the classifier predictions by plotting the traversability probability on different maps in 3D. We used a sliding window to extract the patches from the tested maps to collect the model’s predictions. Then, we created a texture based on the traversable probability, we have colored it using a colormap and finally applied on the 3D model of each map. We evaluated four rotations, from bottom to top, top to bottom, left to right and right to left since those are the most human understandable. By using those angles, we can optimize the cropping process since to avoid rotate each patch based on the Krock’s head position. Instead, we can just rotate the entire map beforehand.

### 1.6.1 Quarry

The first map we evaluate is Quarry, this map is  $32 \times 32$ m long and has a maximum height of 10m. We can use some of the terrain interesting features, such as are three bis slopes and the rocky ground on the top, to evaluate the model. For instance, we expect the trail on the slopes to be traversable at almost any rotation, especially from left to right and vice-versa. The top part should be hard to traverse in almost any case. Figure 1.1 shows the traversability probability directly on the map.

some where place the colormap bar

Correctly, the lower part of the map, composed by flat regions, was labeled with high confidence as traversable in all rotations. On the other hand, the traversability of the slopes and the bumps on the top region depends on the robot orientation.

### 1.6.2 Bars

Bars is a map composed of different heights wall, thus we expected similar probabilities with different orientation. Figure ?? shows the traversability probabilities on the terrain.

Definitely, this was a hard map for the robot due to the high number of not traversable walls. Interesting, we can identify a tunnel near the bottom center of the maps that shows how the model correctly label those patches depending on the orientation. Figure 1.3 highlights this detail.

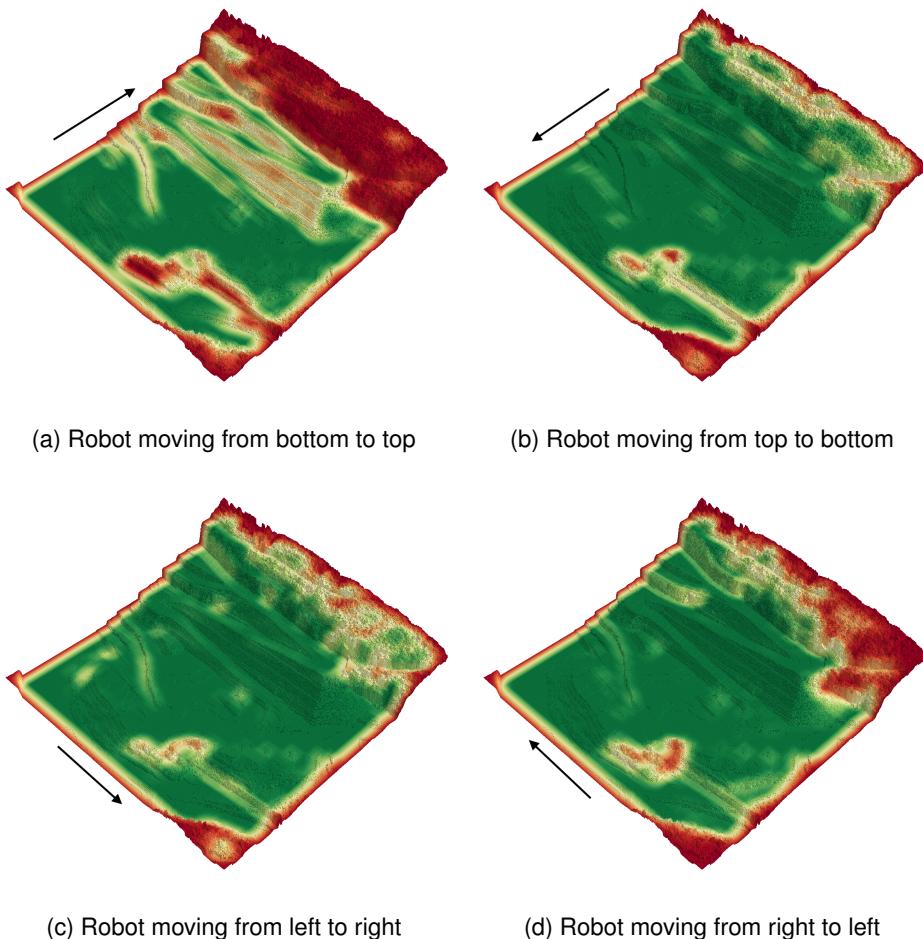


Figure 1.1. Traversability probability on the Quarry map,  $32 \times 32$ m, for different Krock's rotation. The values are obtained by sliding a window on the map to create the patches to predict their traversability.

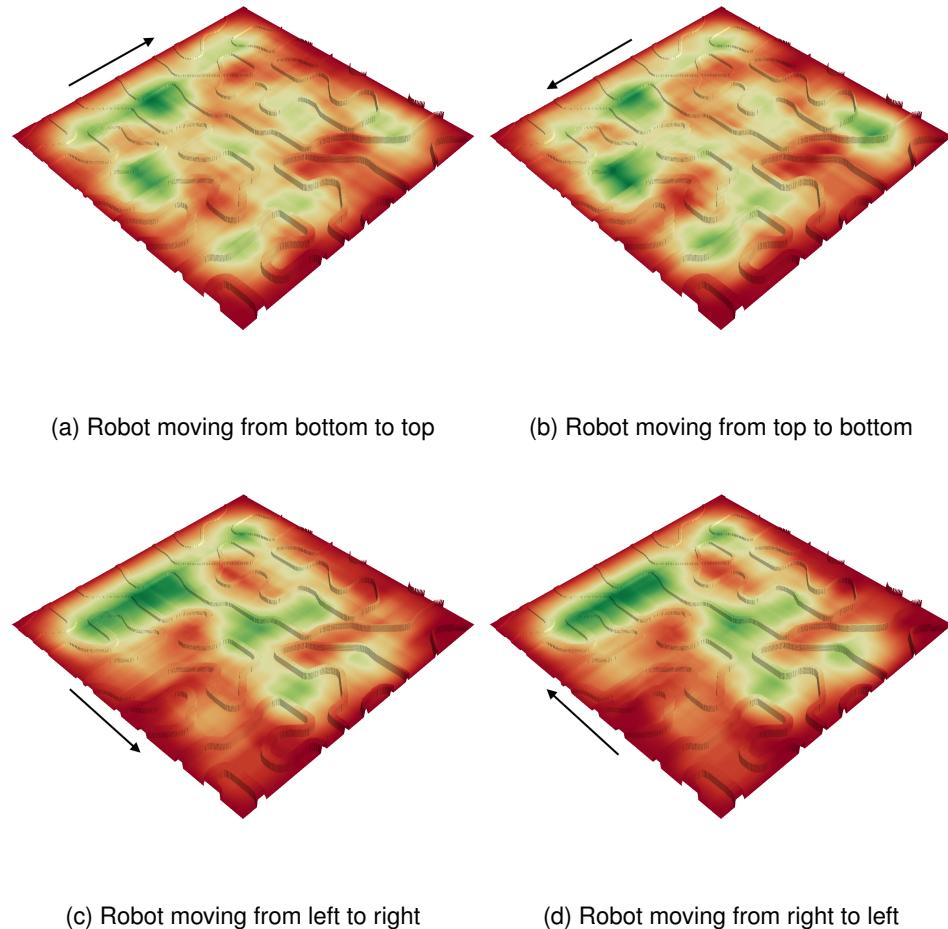


Figure 1.2. Traversability probability on the bars map, a  $10 \times 10\text{m}$ , for different Krock's rotation. The values are obtained by sliding a window on the map to create the patches to predict their traversability.

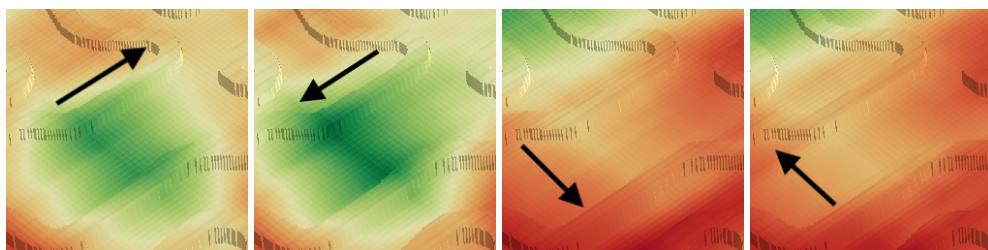


Figure 1.3. Detail of a region in the bars map where there are two walls forming a tunnel. Correctly, when the robot is travel following the trail the region is label as traversable.

### 1.6.3 Small village

We applied the same procedure to also a  $10 \times 10\text{m}$  small village map. Figure 1.4 describes its traversability for four different rotations.

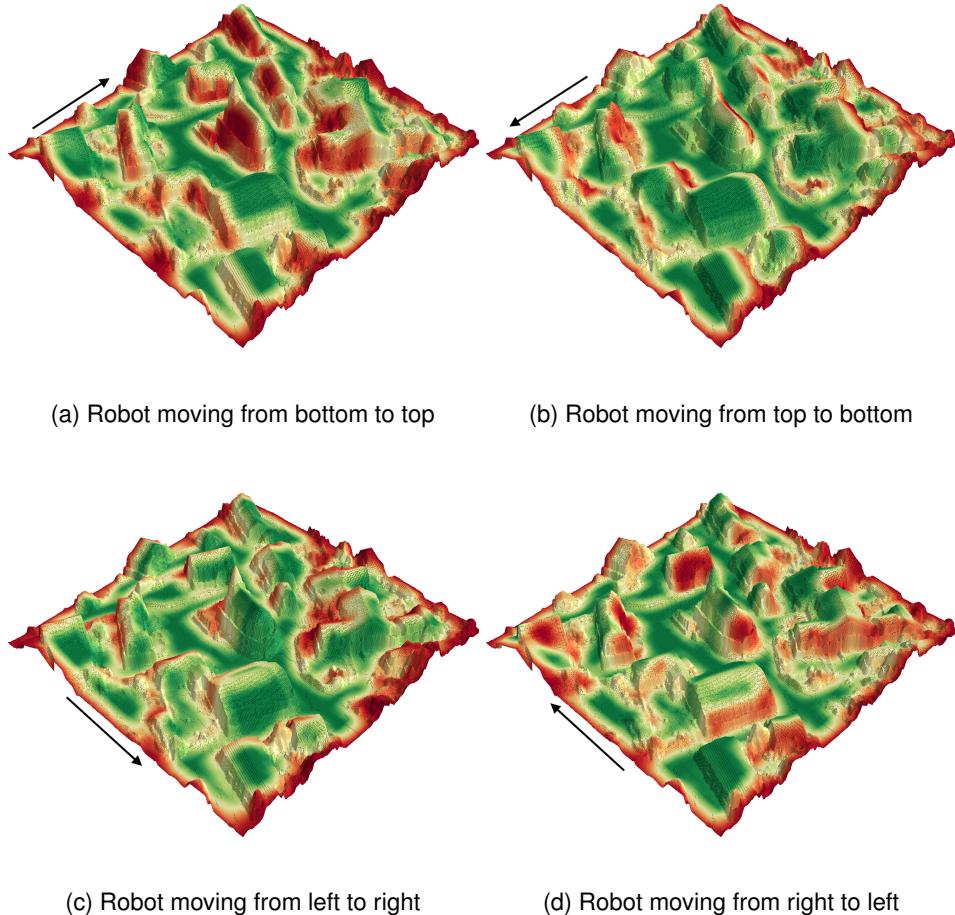


Figure 1.4. Traversability probability on the map of a small village for different Krock's rotation. The surface covers  $30 \times 30\text{m}$  and has a maximum height of 10m. The values are obtained by sliding a window on the map to create the patches to predict their traversability.

All the streets were labeled as traversable, while the church's roof traversability depends on the orientation. For example, most steep roofs are not traversable when walking uphill. On the other hand, if Krock walks side by side they can be traversed. Figure 1.5 shows this behavior on the church's roof.

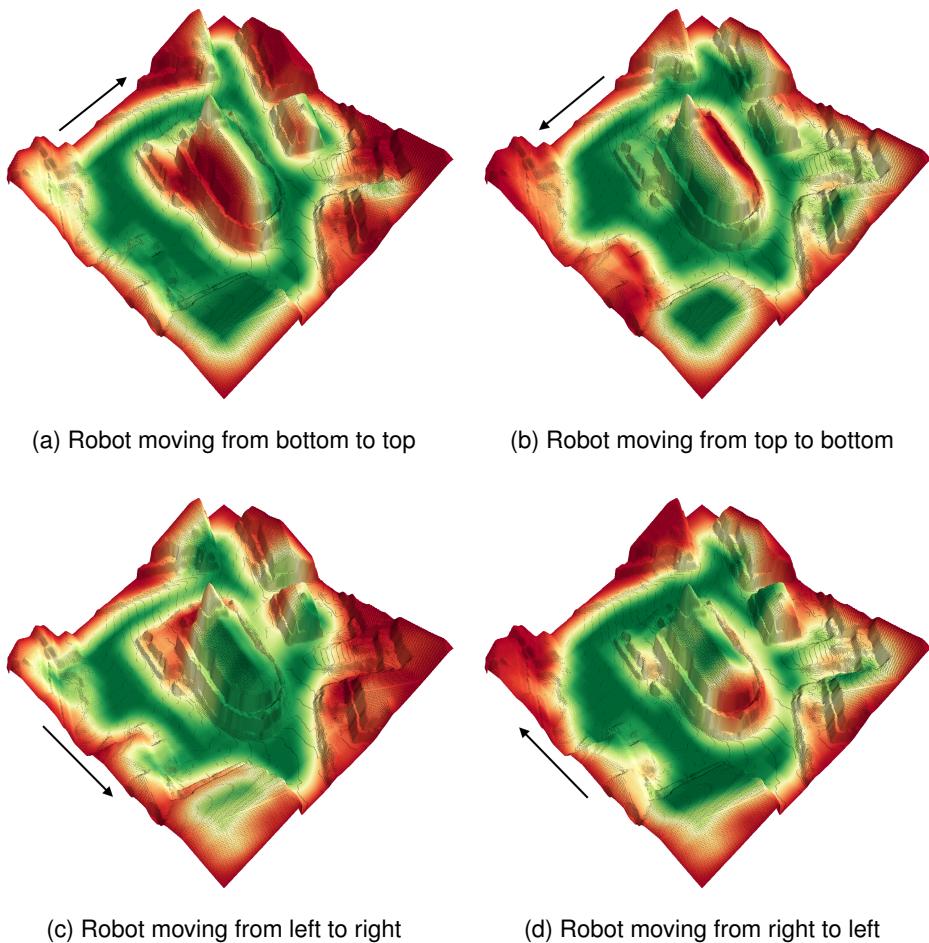


Figure 1.5. Detail of the church in the small village map for different Krock's rotation. All the fours images are correctly labeled accordingly to the robot orientation. In the first two images only downhill part of the roof is traversable. While in the last two the robot is able to travel till the end.