Semantic Front-End Filter Going to the Jungle

Mike Zhang, Dawid Jeczmionek
Supervisors: Jonas Frey, David Hoeller

1 Motivation

ANYmal [3] fails to accurately estimate the support surface when the robot moves through dense vegetation, reducing the effectiveness of locomotion and planning. We aim to use a learning-based approach to filter these corrupted measurements building on existing work in learning-based terrain estimation [2, 4].

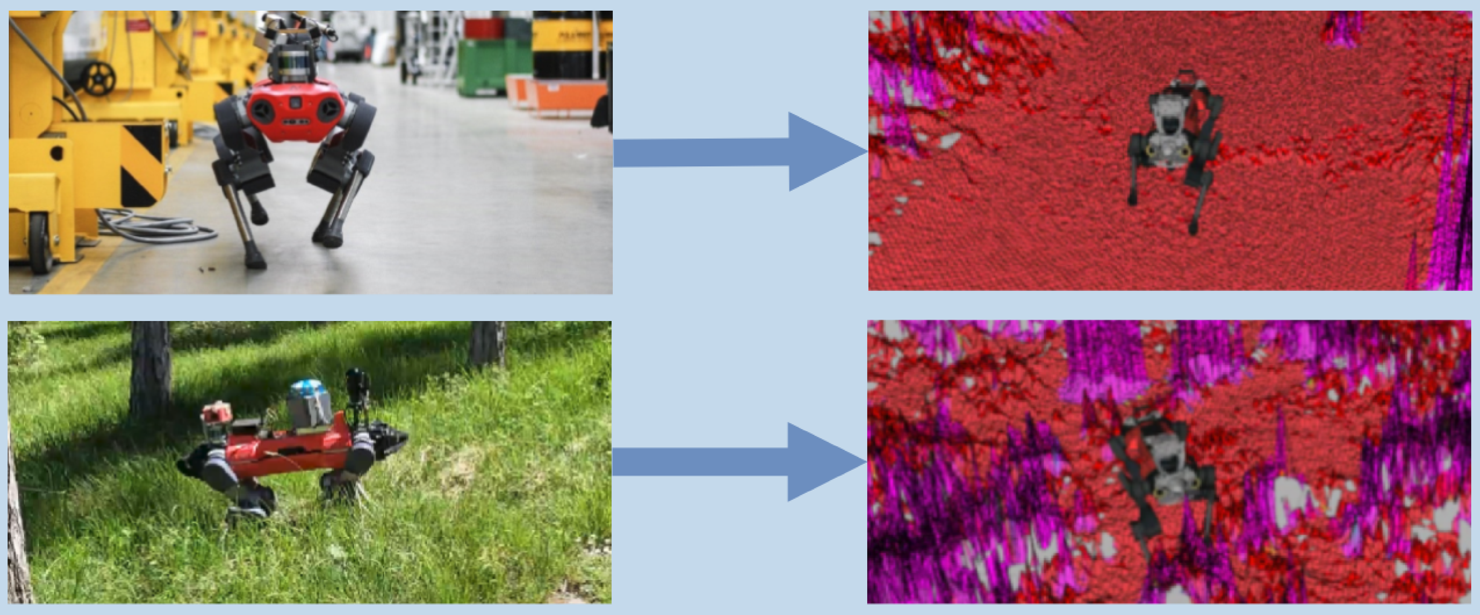

Figure 1: Degradation of the support surface estimate when traversing through vegetation.

2 Approach

We formulate support surface estimation as a supervised learning problem, where the inputs are colored point clouds. The colors in the point clouds provide the semantic cues to the estimator. The training targets are obtained by leveraging hindsight knowledge of where ANYmal walked to obtain ground-truths from the feet contacts.

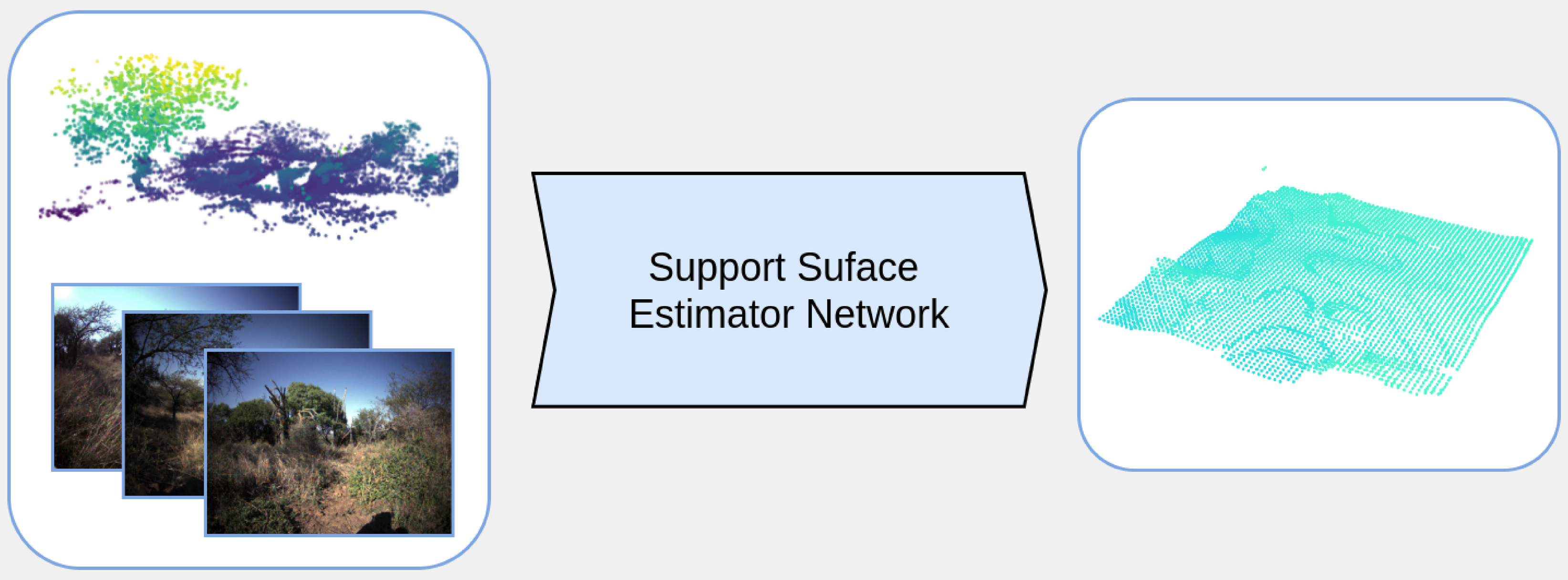

Figure 2: The proposed learing-based estimator pipeline.

3 Data Preparation

Network Inputs: Combine measurements from several lidars and color their points by projection to RGB images from onboard cameras.

Figure 3: Colored point cloud measurement inputs to the estimator.

Training Targets: Support surface measurements from feet contacts are interpolated via Gaussian Process to give denser training targets.

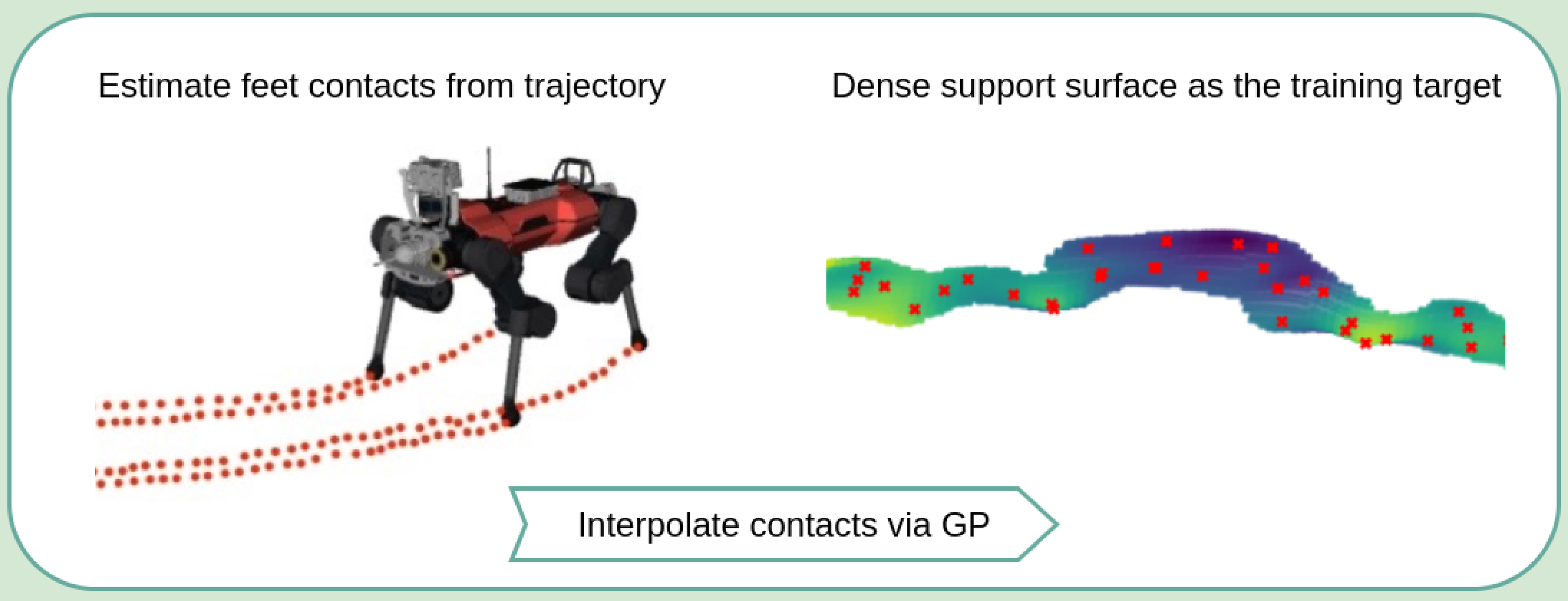

Figure 4: Training targets from feet contact locations.

4 Learning Setup

Network Architecture: The estimator is a 3D convolutional encoder-decoder based on [2]. For efficient processing of 3D data, all quantities are represented as sparse tensors and processed via the MinkowskiEngine [1]. Thus, the inputs point clouds are voxelized before being consumed by the network.
A sequence of measurements are processed in an autoregressive manner where the previous estimate is used for the current prediction.

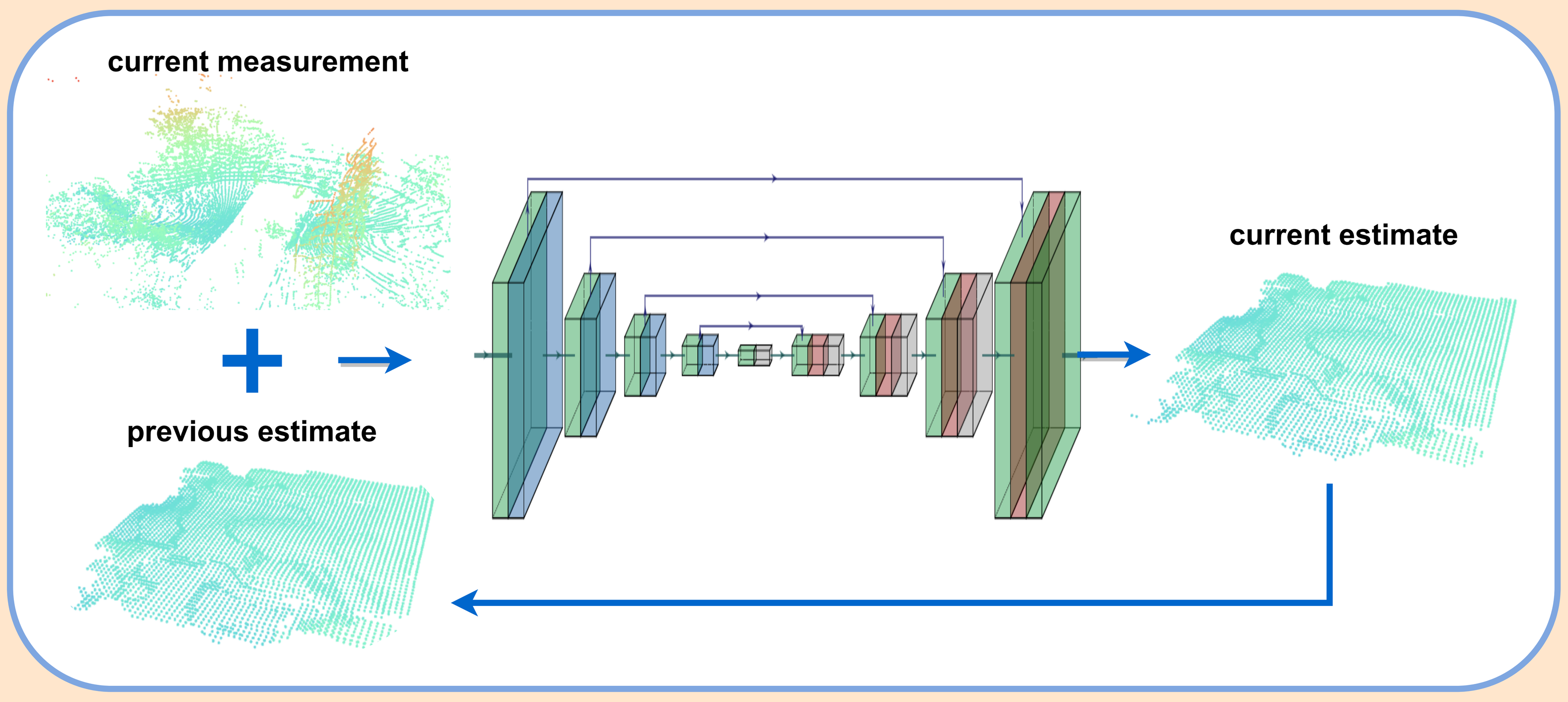

Figure 5: Network architecture.

Loss Masking: The loss must be masked such that the network can only learn from the data where the support surface targets exist.
For this purpose, we purpose a novel masking function for sparse tensors based on kernel maps.

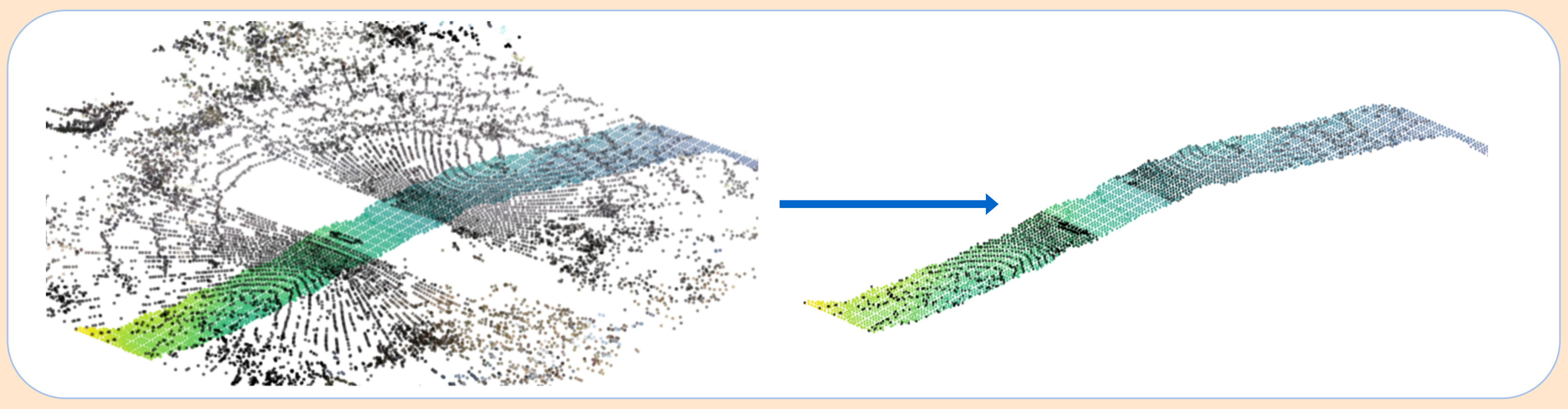

Figure 6: Loss masking visualization.

5 Results

Qualitative results for the estimated support surface in an example of environment with dense vegetation.

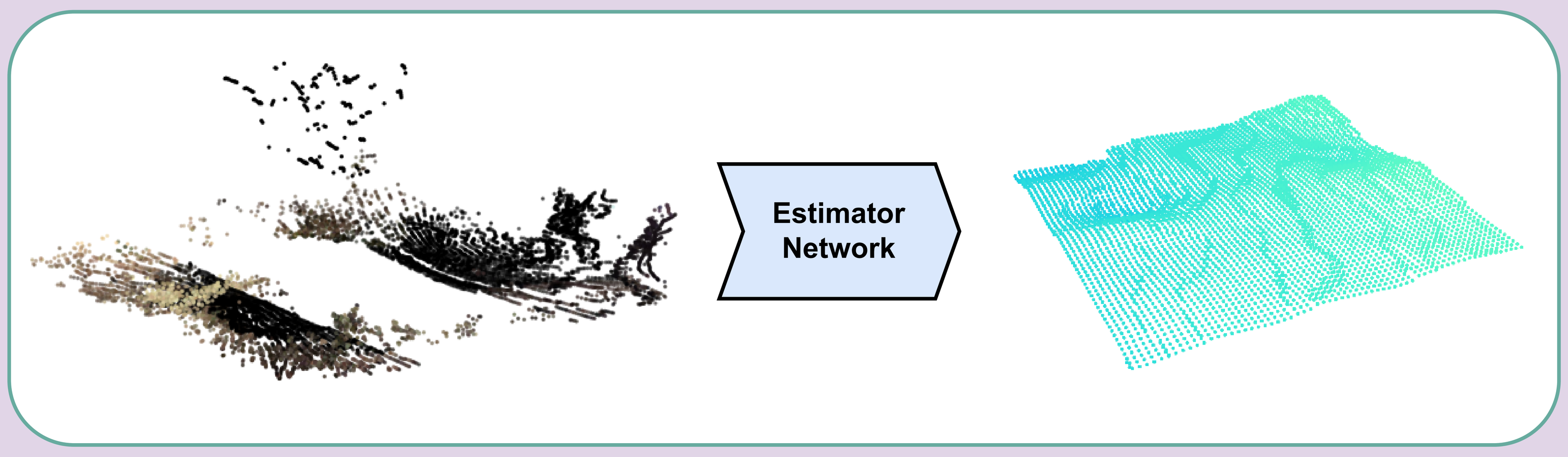

Figure 7: Example support surface estimate result from colored point cloud input.

6 Conclusion

The method is a promising solution and it should be further validated in terms of generalization on higher number of trajectories. Further, the algorithm must be tested in the real environment.

References

[1] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019.

[2] David Hoeller, Nikita Rudin, Christopher Choy, Animashree Anandkumar, and Marco Hutter. Neural scene representation for locomotion on structured terrain.

[3] Marco Hutter, Christian Gehring, Dominic Jud, Andreas Lauber, C Dario Bellicoso, Vassilios Tsounis, Jemin Hwangbo, Karen Bodie, Peter Fankhauser, Michael Bloesch, et al. Anymal-a highly mobile and dynamic quadrupedal robot. In *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 38–44. IEEE, 2016.

[4] Shehryar Khattak Marco Hutter Jonas Frey, David Hoeller. Locomotion policy guided traversability learning using volumetric representations of complex environments.