

Unsupervised Learning of Terrain Representations for Haptic Monte Carlo Localization

Mikołaj Łysakowski¹, Michał R. Nowicki¹, Russell Buchanan², Marco Camurri², Maurice Fallon² and Krzysztof Walas¹

Abstract—Recent studies have shown that haptic sensing can be used effectively for legged robot localization in extreme scenarios where vision sensors might fail, such as mines and sewers. However, existing methods use supervised classification, with training and evaluation executed over explicit terrain classes. This is a significant limitation in real-world applications, where prior labeling and handcrafted classes are often impractical. In this paper, we propose a novel haptic localization system based on a fully unsupervised terrain representation learned solely from the force/torque sensors located at the quadruped robot’s feet. Instead of using the detected terrain class for localization, we propose an improved autoencoder architecture to generate a sparse map on the first run and to localize against the sparse map during subsequent runs. We compare our approach to a haptic localization system based on supervised terrain classification, showing that the unsupervised method has comparable or better performance than the supervised one for the same trajectories while clearly outperforming the proprioceptive odometry estimator available on the robot. The proposed approach is therefore well-suited for a routine maintenance application, increasing the robustness of the platform.

I. INTRODUCTION

The increased maturity of legged robots has made them suitable for a number of industrial applications. In some cases, the ability to operate despite sensory failure is paramount, especially for long routine tasks in dangerous environments, such as mines (Fig. 1). In this work, we consider an autonomous quadruped robot performing an inspection task. The goal is to reliably localize with only proprioceptive sensors (IMU, joint states, tactile sensors) to complete the task at hand or return to safety even when visual sensors are unavailable.

It is possible to determine the class of terrain from tactile measurements [1], [2], [3], which in turn can be used to localize a legged robot in a prior map [4]. However, existing approaches are limited as they require a dense and terrain-annotated map of the environment which often necessitates the use of external, ground truth sensors. Moreover, the terrain classes are often differentiated by human perception, rather than the haptic signals from the robot.

¹ Institute of Robotics and Machine Intelligence, Faculty of Control, Robotics and Electrical Engineering, Poznan University of Technology, Poznan, Poland michal.nowicki@put.poznan.pl,
² Oxford Robotics Institute, University of Oxford, UK
[{russell, mcamurri, mfallon}@robots.ox.ac.uk}](mailto:{russell, mcamurri, mfallon}@robots.ox.ac.uk)

*This work is supported from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 780883, THING. M. R. Nowicki is supported by the Foundation for Polish Science (FNP).



Fig. 1: An ANYmal B300 quadruped robot performing an inspection of conveyors 800 m underground. The failure of vision or LiDAR-based localization can happen due to hot, humid, and dusty working conditions. Our system can increase the robustness of the robot in these repetitive tasks.

Our goal is to achieve a localization solution that does not require prior human input while utilizing a sparse map built only with the robot’s internal sensors. Therefore, the proposed unsupervised method learns a compact, internal representation of the traversed terrains. The terrain representation is provided by an improved autoencoder network, which compresses the force/torque signals from the robot’s sensorized feet into a low-dimensional latent vector representing the terrain. The representation is firstly used to generate a sparse, tactile map of the environment when vision or LiDAR-based localization is available. Then, when the robot blindly re-visits the same area, it uses the compressed tactile sensing to localize against the sparse map making it analogous to visual *teach & repeat* algorithms [5].

The contributions of our work can be summarized as:

- An Improved AutoEncoder (AIE) network for unsupervised, compressed haptic terrain representations for walking robots.
- A novel Haptic Monte Carlo localization method based on unsupervised, compressed haptic signals with a sparse map of the environment.
- Experimental evaluation of the localization on the ANYmal B300 quadruped robot and its comparison with the state of the art.

II. RELATED WORK

In this section, we review the relevant literature on haptic terrain classification, unsupervised encoding algorithms, and localization of legged robots.

A. Haptic terrain classification

Supervised haptic terrain classification for walking robots is a well-researched topic [1], [2], [3], [6] with methods achieving accuracy up to 90%. However, these methods require a significant number of supervised training samples while offering poor generalization to unseen terrain classes.

Ahmadi et al. [7] explored the idea of using a semi-supervised approach with fewer labeled terrains while achieving similar results to supervised approaches with their Recurrent Neural Network (RNN) based on Gated Recurrent Units (GRU). Similarly, Wellhausen et al. [8] have shown that the correspondence between the feet and vision can be learned to predict the traversal cost based solely on images. This allows to reduce the number of required training samples and allows to plan the motion of the robot considering the cost of a terrain. In our most recent work [9], we evaluated the problem of supervised terrain recognition based on the practical aspects grading the solutions based on their accuracies, model sizes, and inference times.

The classes for terrain classification are typically based on visual clues. As such, the classes might not reflect the physical properties of their corresponding terrains. In [10], the authors presented an end-to-end deep learning approach to gait adaptation based on haptic signals avoiding the definition of classes. In this approach, the internal representation of the haptic signals was used to make decisions about the walking gait. Inspired by [10], we are motivated to use the internal terrain representation from force/torque signals, but we avoid relying on end-to-end learning design.

B. Encoding tactile signals

In order to use the unsupervised force/torques signals for localization, we need a method to compute the similarity of encoded signals to achieve a scalable solution with generalization capabilities to unseen terrains. While this idea has not yet been explored for the walking robots, Dynamic Time Warping (DTW) is a golden standard algorithm used for comparison of uncompressed, speed-varying signals in the time-domain [11]. In the typical application of finding the best match, the distances between the query signal and each signal from the database have to be computed to generate a list of the most similar matches in a method called DTW K-Nearest Neighbors (DTW-KNN).

The most common deep learning technique used for encoding compressed latent representations is the AutoEncoder [12] which has been adopted for many applications [13]. The AutoEncoder consists of an encoder part that reduces the dimensionality of an input signal and a decoder part that reconstructs the original signal based on this reduced latent representation. The AutoEncoder is trained without supervision and the resulting latent representations can be compared to determine the similarity of the original signals.

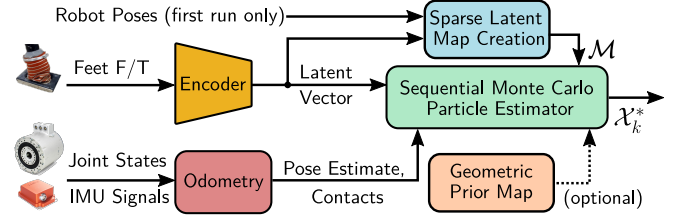


Fig. 2: A system overview of our proposed Haptic Monte Carlo localization. The odometry from the robot is fused with the latent vector from the encoder using a sparse latent map as prior. Optionally, a geometric map can be also used for increased accuracy.

C. Haptic Localization of Legged Robots

The earliest work on haptic localization applied to legged robots was presented by Chitta et al. [14]. They used a particle filtering-based algorithm to proprioceptively localize a quadrupedal robot on a known terrain. While walking, their method approximated the probability distribution of the pose of the robot's base with a set of particles that were sampled from the uncertainty of the odometry. The particle weights were determined by the L_2 norm of the height error between the prior map of the terrain and the feet' contact location. They demonstrated experiments with a Boston Dynamics' LittleDog robot using a Vicon system to control the robot.

Schwendner et al. [15] demonstrated a similar method on a wheeled robot with protruding spikes. In addition to geometric information, their method also fused information from explicit terrain classifications using a camera, which we do not require in our proposed work.

In [16], we presented a geometric only method based on Sequential Monte Carlo (SMC) Localization that estimated the 6-Degrees of Freedom (DoF) pose trajectory of the robot. Building on this work in [4], we added proprioceptive terrain classification to improve localization on flat grounds without distinct geometric features. However, this method required explicit classification of different terrains which may not always be practical in real scenarios. In this work, we tackle this problem by introducing a new method of unsupervised terrain representation.

III. PROBLEM STATEMENT

Given a quadruped robot with 12 active DoF, we wish to blindly localize the robot against a prior map, given a source of odometry and the force/torque (F/T) signals measured at the feet. Additionally, it is assumed that a low-drift or drift-free pose estimate of the robot is available to build the prior map. For our experiments, we use the ANYmal B300 [17] equipped with the sensorized feet from [18] and the odometry estimator by Bloesch et al. [19]. Because the estimate depends on the shape of map, the probability distribution of the estimated robot pose may not be Gaussian. Thus, we will use a SMC method to find the optimal pose estimate of the robot in the map.

IV. PROPOSED METHOD

An overview of our SMC system is presented in Fig. 2. Firstly, the encoder determines a terrain representation for each footstep (Section IV-A). The computed encoding of each footstep is associated with an accurate source of robot poses (e.g., from a SLAM system) to build a sparse haptic latent map (Section IV-B).

Once the map is built, the localization is performed with a predefined number of particles at regular intervals, as described in Section IV-C. At each new step the SMC estimator updates the state of its particles using the odometry from the robot, the current tactile signals, and optionally the geometry of the terrain.

A. Learned Terrain Representation

As the sparse map of terrain representation contains steps from previous walks over the same route, the map's size naturally grows linearly with the length of the trajectory. Assuming the systematic inspection over 1 km with steps taken every 10 cm, the approximated latent map size for unsupervised, uncompressed signals can exceed 200 MB. Therefore, we seek a reduced encoding that will generate a tactile description of size b called a latent vector that will decrease the sparse latent map size while preserving the properties of the original signals and increasing its generalization abilities.

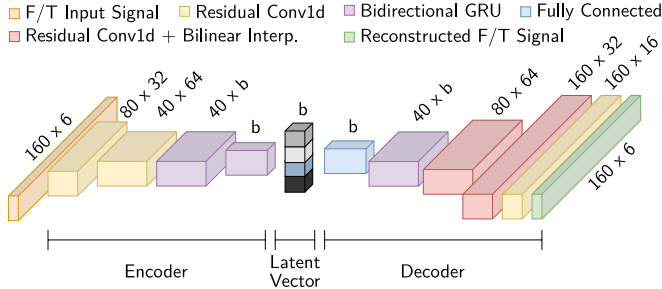


Fig. 3: The scheme of the Improved AutoEncoder (IAE) architecture was used to generate a b -dimensional latent vector. The main blocks of the network are convolutional layers (CNN), bidirectional GRU layers (Bidir GRU), and fully connected layer (FC).

The encoding is performed with the proposed architecture of Improved AutoEncoder (IAE) consisting of three types of blocks: convolutional, recurrent, and fully connected layers as presented in Fig. 3. This combination of blocks, inspired by the supervised terrain classification network from [4], extracts high-level feature representation of the time-series signals and can process variable-length data. The IAE model uses two residual layers with 1D convolutions in the CNN module and two bidirectional layers with two GRU cells in the recurrent component of an encoder. The convolutional part of the encoder is responsible for extracting the features of signals and reducing the spatial dimension of a sequence by using a stride of 2. Next, in contrast to a simple convolutional AutoEncoder, the bidirectional GRU layers are used

to capture the global and local dependencies between the features. Finally, the RNN block produces the b dimensional latent vector.

The decoder's recurrent block consists of a fully connected (FC) layer and a bidirectional GRU layer, which is followed by three residual 1D convolution blocks. The FC layer takes as an input the latent vector and produces an initial state to the bidirectional GRU layer. The RNN block is used to construct a sequence of features, which is then upsampled with bilinear interpolation in a signal's length dimension and reduced in terms of a number of features with convolutions. The model uses dropout layers with a rate of 0.4 in every block and a batch normalization with a momentum equal to 0.6 in residual blocks.

The training process of the IAE lasted for approximately 640 epochs until a convergence was observed. We used the AdamW optimizer to minimize Mean Squared Error (MSE) loss with exponential decay of learning rate and cosine decay of weight decay. The size of the mini-batch was equal to 64.

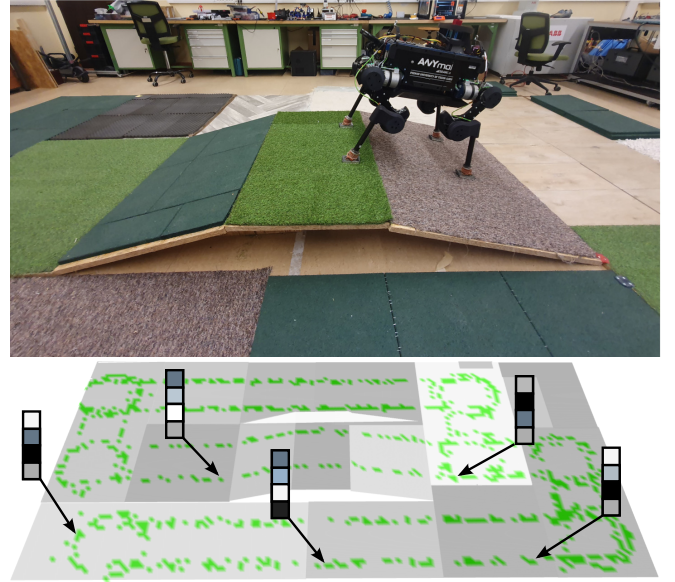


Fig. 4: The ANYmal robot walking over the prepared terrain course (top) and the visualization of the sparse latent map generated from several trials recorded prior to the localization phase (bottom). Each cell of the latent map containing the compressed tactile signal is marked by a green color.

B. Sparse Haptic Map Creation

The sparse map is created from steps recorded prior to localization from the robot equipped with sensorized feet as presented in Fig. 4. Whenever a contact between the ground and a foot was detected by ANYmal's state estimator, we collected 160 samples of 3-axis force and 3-axis torque signals from a foot. Then, the signals are processed by the compression algorithm resulting in a dataset mapping global position of feet and a corresponding latent vector.

To store the compressed signals from this phase, we defined a data structure called a latent map based on the

implementation of a grid map [20]. The latent map stores the sparse information about the compressed representation of tactile signals and their Euclidean position in the 2-dimensional space.

C. Sequential Monte Carlo Localization

The core localization algorithm used in this work is the same as [4]. We provide some explanation of the method here, however, details are omitted due to space constraints.

Let $\mathbf{x}_k \in SE(3)$ be a robot's pose at time k . We use the notation $\tilde{\mathbf{x}}$ to represent a pose estimate from an external estimator, and $\mathbf{x}^*$ to represent its most likely estimate from our SMC filter. Our objective is to approximate the posterior distribution of the state $\mathbf{x}_k$ given a history of measurements $\mathbf{z}_0, \dots, \mathbf{z}_k = \mathbf{z}_{0:k}$ as follows:

$$p(\mathbf{x}_k | \mathbf{z}_{0:k}) = \sum_i w_{k-1}^i p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_{k-1}^i) \quad (1)$$

where w^i is the *importance weight* of the i -th particle; $p(\mathbf{z}_k | \mathbf{x}_k)$ is the measurement likelihood function and $p(\mathbf{x}_k | \mathbf{x}_{k-1}^i)$ is the motion model for the i -th particle state. Since $p(\mathbf{x}_k | \mathbf{z}_{0:k})$ is typically unknown, the state $\mathbf{x}_k$ is normally sampled from $p(\mathbf{x}_k | \mathbf{x}_{k-1}^i)$, yielding:

$$p(\mathbf{x}_k | \mathbf{z}_{0:k}) = \sum_i w_{k-1}^i p(\mathbf{z}_k | \mathbf{x}_{k-1}^i) \delta(\mathbf{x}_k - \mathbf{x}_{k-1}^i) \quad (2)$$

where $\delta(\cdot)$ is the Dirac delta function.

Blind locomotion requires conservative footstep placement, hence we opt for a statically stable gait, which guarantees stability at all times even when the motion is stopped mid-flight phase. Since only one leg can be moved at the time, as soon as the swing leg touches the ground the robot enters into a four-support phase; at this time a new set of particles are sampled:

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}^i) = \mathcal{N}(\mathbf{x}_k, \Delta \tilde{\mathbf{x}}_k \mathbf{x}_{k-1}^i, \Sigma_k) \quad (3)$$

where $\Delta \tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_{k-1}^{-1} \tilde{\mathbf{x}}_k$ is the pose increment measured by the onboard state estimator at times $k-1$ and k . At time k , the new particles $\mathbf{x}_k^i$ are thus sampled from a normal distribution centered at the pose estimated from the odometry with its corresponding covariance.

1) *Unsupervised Haptic Measurement Model:* During the localization phase, the F/T signals recorded from the sensorized feet are processed to generate localization likelihoods as shown in Fig. 5. Given a particle state $\mathbf{x}_k^i$, the estimated contact position in the world coordinates for each foot f is defined as the concatenation of the estimated robot base pose and the location of the end effector, in base coordinates:

$$\mathbf{d}_f^i = (d_{xf}^i, d_{yf}^i, d_{zf}^i) = \mathbf{x}_k^i \mathbf{d}_f. \quad (4)$$

The corresponding measured F/T signal for the foot f is encoded into a latent vector $\mathbf{v} \in \mathcal{R}^b$, where b is the dimension of the latent vector.

During the localization stage, we determine the latent vector $\mathbf{m} = \mathcal{S}(d_{xf}^i, d_{yf}^i)$ from a sparse latent map $\mathcal{S}$ that is

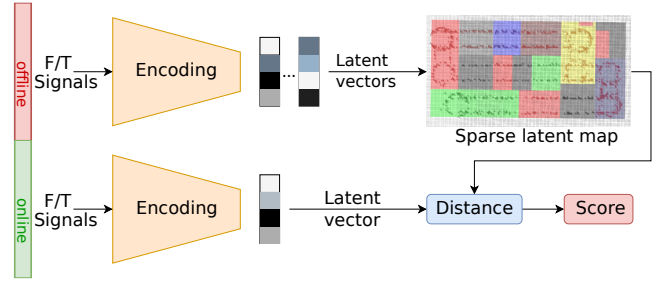


Fig. 5: Comparing encoded measurement with a prior map of latent vectors using a selected distance measure. Note that the latent map is sparse (based on previous steps) and the colored segmentation is simply for visualization

the closest to the current foot placement $\mathbf{d}_f^i$ and located in the latent map at global coordinates (s_{xf}^i, s_{yf}^i) . We compute the distance d_u between the measured latent vector $\mathbf{v}$ and the latent vector $\mathbf{m}$ from the latent map using either the Euclidean L_2 norm (f_{L_2}) or cosine similarity ($f_{\cos}$):

$$d_u = f_{L_2}(\mathbf{v}, \mathbf{m}) = \sqrt{\sum_{i=1}^n (v_i - m_i)^2}, \quad (5)$$

$$d_u = f_{\cos}(\mathbf{v}, \mathbf{m}) = \frac{\mathbf{v} \cdot \mathbf{m}}{\|\mathbf{v}\| \|\mathbf{m}\|}, \quad (6)$$

where v_i and m_i indicate the i -th components of the encoded latent representations.

To model the likelihood, we use the univariate Gaussian distribution centered at the cell matched in the sparse latent map:

$$p(z_k | \mathbf{x}_k^i) = \begin{cases} \frac{1}{\sigma_u \sigma_d 2\pi} & d_{2D} > d_t \\ \mathcal{N}(d_u, 0, \sigma_u) \mathcal{N}(d_{2D}, 0, \sigma_d) & d_{2D} \leq d_t, \end{cases} \quad (7)$$

where $d_{2D} = \sqrt{(d_{xf}^i - s_{xf}^i)^2 + (d_{yf}^i - s_{yf}^i)^2}$ is the Euclidean distance between the foot's position (d_{xf}^i, d_{yf}^i) and the matched, closest step in the sparse latent map (s_{xf}^i, s_{yf}^i) . We reject matches if the Euclidean distance exceeds the threshold of $d_t = 25$ cm, which was set to match the similar Euclidean threshold for terrain classification in [4]. The sigma values of $\sigma_u = 0.4$ determines the standard deviation of the F/T signal while sigma value of $\sigma_d = 0.4$ captures the decreased credibility of the sparse map measurement when the foot was placed further away from the known location in the map. The sigma values σ_u and σ_d were chosen empirically.

2) *Geometric Measurement Model:* The geometric measurement likelihood is the same as [4] and is modeled as a univariate Gaussian centered at the local elevation of each cell. Given the estimated contact position of the foot f in the world coordinates $\mathbf{d}_f^i$ by a particle state $\mathbf{x}_k^i$, the measurements and their relative likelihood functions for the

i -th particle are:

$$z_k = d_{zf}^i - \mathcal{M}(d_{xf}^i, d_{yf}^i) \quad (8)$$

$$p(z_k | \mathbf{x}_k^i) = \mathcal{N}(z_k, 0, \sigma_z) \quad (9)$$

where: d_{zf}^i is the vertical component of the estimated contact point location in world coordinates of foot f , according to the i -th particle; $\mathcal{M}(d_{xf}^i, d_{yf}^i)$ is the corresponding dense map elevation at the xy coordinates of d_f^i and σ_z equal to 0.01 is a measurement variance.

V. EXPERIMENTAL RESULTS

A. Dataset

The SMC algorithm was evaluated over a $3.5 \text{ m} \times 7 \text{ m}$ area containing flat terrain with two ramps covered by eight different terrain types (rubber, carpet, PVC, foam, sand, artificial grass, ceramic, and gravel). The ground truth robot movement was recorded from the external motion capture system (OptiTrack) providing poses at 100 Hz. The reference map for the geometric component was captured with a sub-centimeter accurate survey-grade laser scanner (SURPHASER 100HSX).

This setup was used to capture the terrain classification PUTANY dataset [9]. We chose the PUTANY dataset as several supervised terrain classification algorithms based on the F/T signals were already evaluated on this dataset [9]. In our case, we measured the performance of the compressed terrain representations using accuracy (Acc) averaged over 1148 terrain samples in the testing part of PUTANY dataset (Sec. V-B).

The performance of the haptic SMC localization (Sec. V-C) was measured over three trials accumulating to the total length of 715 meters, 4054 seconds, and 6658 steps. The sparse map for these trials was created based on previous, independent runs. This haptic localization dataset is the same as in our previous work [4] and thus the localization results can be directly compared.

B. Unsupervised Compressed Terrain Representation

The goal of the unsupervised compression of the tactile signals is to find a reduced latent representation that still contains the original information. In our case, we measured the ability to store original information on the terrain classification problem arguing that greater accuracy corresponds to better discrimination and thus is preferred for the haptic SMC localization. The terrain classification results obtained by compression algorithms with varying lengths of latent vector b are presented in Fig. 6. The five best unsupervised results are compared to state-of-the-art supervised terrain classification algorithms in Table I. Classification with unsupervised methods was performed using the nearest neighbor algorithm, performed on the similarity measure between latent representations.

While the supervised methods outperform the unsupervised approaches, the latter still provide reliable accuracy (above 70%). The highest accuracy among the unsupervised methods was achieved with IAE-512- L_2 equal to 74.2%,

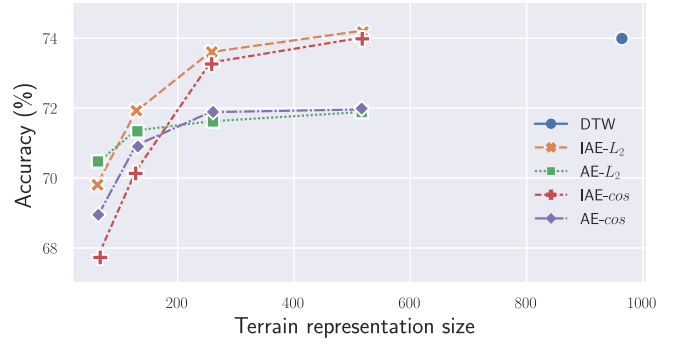


Fig. 6: The terrain recognition accuracy for the unsupervised, compressed tactile signals depending on the length of the latent vector (representation) used to capture the original properties of the F/T signals

which compresses the original signal to half of its length. The second-best result was achieved with DTW-KNN [9], which directly compares the signals in the time domain making it unsuitable for larger terrain maps.

TABLE I: Accuracy comparison of supervised and unsupervised methods.

Method	Supervised	b	d_u	Acc [%]
CNN-RNN [4]	Yes	-	-	93.0
HAPTR [9]	Yes	-	-	91.7
IAE	No	512	L_2	74.2
DTW-KNN [11], [9]	No	960	DTW	74.0
IAE	No	512	cos	74.0
IAE	No	256	L_2	73.6
IAE	No	256	cos	73.3

This analysis of the different configurations reveals that the L_2 measure outperforms the cosine distance $\cos$ for AE or IAE regardless of the chosen length of the latent vector. A shorter length of the compressed signal degrades accuracy but this effect is reduced for AE and IAE as the value nears 256. From Fig. 6, it is evident that IAE better captures the original signal characteristic than AE as it can achieve similar performance with shorter latent vectors. Among the evaluated solutions, the compression with IAE-256- L_2 was chosen as a good trade-off: it uses approximately four times less memory than DTW-KNN and two times less than IAE-512- L_2 while achieving an accuracy that is only worse by 0.6 percentage points than the best reported value.

C. Haptic localization with geometric and latent map

In this section, we directly compare our method to the approaches presented in [4]: Haptic Localization (HL) performed with the geometry of the environment utilizing a dense, prior geometric map (HL-C), with the supervised terrain classifications (HL-C), or with both of these sources of information (HL-GC). We also show the non-localized results from the robot's proprioceptive odometry TSIF [19].

We propose haptic localization with the unsupervised, compressed F/T signals working solely on the sparse map (HL-U) as well as combining this approach with the geom-

Absolute Pose Error (APE) $\mu[m]$				
Trial	TSIF [19], [4]	HL-G [4]	HL-GC [4]	HL-GU
1	0.64	0.23	0.14	0.15
2	1.28	0.25	0.11	0.18
3	0.72	0.21	0.18	0.13

TABLE II: Comparison of haptic localization performance when prior geometry of the map is used. The usage of unsupervised F/T signals (HL-GU) results in comparable gains when supervised terrain classification is used (HL-G).

etry of the terrain requiring the dense, geometric map (HL-GU). The localization accuracies obtained over three trials used in [4] are presented in Tab. II.

These results confirm that the when using unsupervised tactile signals with geometric information (HL-GU) performance is improved over methods using only geometry (HL-G). Compared to the supervised terrain classification method with geometry (HL-GC), the usage of unsupervised F/T signals (HL-GU) yields comparable results achieving the best result for trial 3. This demonstrates that the unsupervised F/T signals are able to capture the differences between signals obtained for different classes as well as different tactile responses for steps taken over the same terrain class.

D. Haptic localization with latent map only

In some cases it may not be feasible to create a prior, dense geometric map. To replicate this case, we use the same trails as in the previous section but only the sparse latent map is available for localization purposes. Results obtained without F/T signals (TSIF), with supervised terrain classification (HL-C) and unsupervised, compressed tactile signals (HL-U) are presented in Tab. III.

Absolute Pose Error (APE) $\mu[m]$			
Trial	TSIF [19], [4]	HL-C [4]	HL-U
1	0.64	0.63	0.47
2	1.28	0.73	0.57
3	0.72	0.61	0.5

TABLE III: The comparison between the localization accuracy obtained with a haptic SMC localization utilizing the internal ANYmal sensors (TSIF) combined with terrain classification (HL-C) or proposed, unsupervised terrain representation (HL-U).

The method based on the unsupervised haptic signals (HL-U) outperforms the method utilizing the supervised terrain classification (HL-C) in all considered trials while also not requiring a prior, dense map of the environment. The obtained localization accuracy is also clearly better when compared to the default state estimator TSIF, which drifts continuously without localization.

The trajectory obtained with the HL-U for trial 2 is presented in Fig. 7. The visual analysis reveals that the method is effective at decreasing the accumulated drift of the internal odometry of the walking robot. This is true both

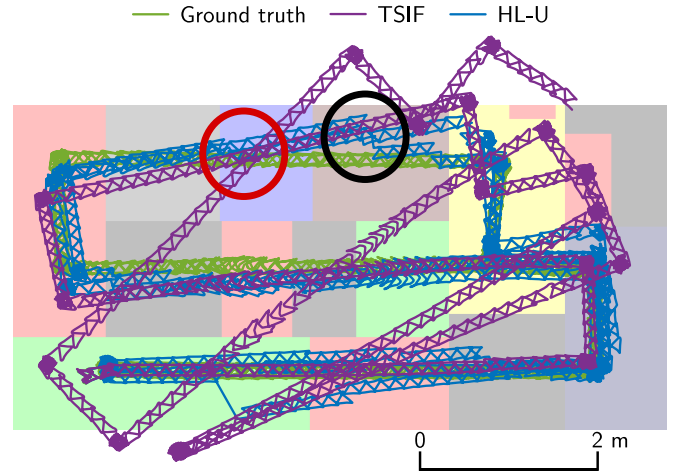


Fig. 7: The trajectory comparison between the ground truth (GT), internal ANYmal state estimator (TSIF), and haptic SMC with unsupervised, compressed tactile signals (HL-U) for trial 2. The corrections for robot localization for HL-U can be observed once a robot crosses to a different terrain class (red circle) and during localization over the same terrain type (black circle). The patches of different terrain classes are marked by different colors.

when crossing boundaries between terrain classes (red circle) and also observed while walking over long stretches of the same terrain class (black circle). The latter benefit is a direct result of using unsupervised terrain representations instead of explicit terrain class.

VI. CONCLUSIONS

In this work we propose to utilize the unsupervised, compressed force/torque signals directly in a haptic localization framework for routine inspection tasks. We evaluate different approaches to unsupervised terrain representation revealing that the best trade-off between the compressed signal length and performance can be obtained with the proposed Improved AutoEncoder. The chosen compression method is used with haptic SMC localization over 3 trials containing external ground truth reference from the OptiTrack systems. The localization with unsupervised tactile signals performs similarly or even outperforms the localization with the supervised terrain classification both with and without geometric information. We believe the proposed method is a valuable improvement over previous haptic localization methods. In our future work, we plan on verifying our approach outdoors while also extending the geometric component to work with a sparse map created in *teach & repeat* scenarios.

REFERENCES

- [1] M. H. Hoepflinger, C. D. Remy, M. Hutter, and R. Siegwart, "Haptic Terrain Classification on Natural Terrains for Legged Robots," in *International Conference on Climbing and Walking Robots (CLAWAR)*, 2010, pp. 785–792.
- [2] J. Bednarek, M. Bednarek, P. Kicki, and K. Walas, "Robotic Touch: Classification of Materials for Manipulation and Walking," in *IEEE International Conference on Soft Robotics (RoboSoft)*, 2019, pp. 527–533.

- [3] H. Kolvenbach, C. Bärtschi, L. Wellhausen, R. Grandia, and M. Hutter, "Haptic Inspection of Planetary Soils With Legged Robots," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1626–1632, 2019.
- [4] R. Buchanan, J. Bednarek, M. Camurri, M. R. Nowicki, K. Walas, and M. Fallon, "Navigating by touch: haptic Monte Carlo localization via geometric sensing and terrain classification," *Autonomous Robots*, 2021.
- [5] P. Furgale and T. D. Barfoot, "Visual teach and repeat for long-range rover autonomy," *Journal of Field Robotics*, vol. 27, no. 5, pp. 534–560, 2010.
- [6] X. A. Wu, T. M. Huh, R. Mukherjee, and M. Cutkosky, "Integrated Ground Reaction Force Sensing and Terrain Classification for Small Legged Robots," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 1125–1132, 2016.
- [7] A. Ahmadi, T. Nygaard, N. Kottege, D. Howard, and N. Hudson, "Semi-Supervised Gated Recurrent Neural Networks for Robotic Terrain Classification," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1848–1855, 2021.
- [8] L. Wellhausen, A. Dosovitskiy, R. Ranftl, K. Walas, C. Cadena, and M. Hutter, "Where Should I Walk? Predicting Terrain Properties From Images Via Self-Supervised Learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1509–1516, 2019.
- [9] M. Bednarek, M. Łysakowski, M. Bednarek, M. R. Nowicki, and K. Walas, "Fast haptic terrain classification for legged robots using transformer," in *2021 European Conference on Mobile Robots (ECMR)*, 2021.
- [10] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, vol. 5, no. 47, 2020.
- [11] J. Lines and A. Bagnall, "Time series classification with ensembles of elastic distance measures," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, p. 565–592, May 2015.
- [12] M. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *Aiche Journal*, vol. 37, pp. 233–243, 1991.
- [13] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," arXiv:2003.05991, 2020.
- [14] S. Chitta, P. Vernaza, R. Geykhman, and D. Lee, "Proprioceptive localization for a quadrupedal robot on known terrain," in *IEEE International Conference on Robotics and Automation (ICRA)*, April 2007, pp. 4582–4587.
- [15] J. Schwendner, S. Joyeux, and F. Kirchner, "Using embodied data for localization and mapping," *Journal of Field Robotics*, vol. 31, no. 2, pp. 263–295, 2014.
- [16] R. Buchanan, M. Camurri, and M. Fallon, "Haptic Sequential Monte Carlo Localization for Quadrupedal Locomotion in Vision-Denied Scenarios," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2020.
- [17] M. Hutter *et al.*, "Anymal - a highly mobile and dynamic quadrupedal robot," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 38–44.
- [18] G. Valsecchi, R. Grandia, and M. Hutter, "Quadrupedal locomotion on uneven terrain with sensorized feet," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1548–1555, 2020.
- [19] M. Bloesch, M. Burri, H. Sommer, R. Siegwart, and M. Hutter, "The Two-State Implicit Filter Recursive Estimation for Mobile Robots," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 573–580, Jan 2018.
- [20] P. Fankhauser and M. Hutter, "A Universal Grid Map Library: Implementation and Use Case for Rough Terrain Navigation," in *Robot Operating System (ROS) – The Complete Reference (Volume 1)*, A. Koubaa, Ed. Springer, 2016, ch. 5.