2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Contrastive Learning-Based Attribute Extraction Method for Enhanced Terrain Classification

Xiao Liu* , Hongjin Chen*, and Haoyao Chen

***Abstract*— The outdoor environment has many uneven surfaces that put the robot at risk of sinking or tipping over. Recognizing the type of terrain can help robot avoid risks and choose an appropriate gait. One of the critical problems is how to extract the terrain-related knowledge from sensor data collected as the robot traversed the ground. Many existing vision-based approaches are limited in directly perceiving the intrinsic properties of various terrains. The intuitive approach entails directly analyzing data recorded by the robot's proprioceptive sensors. However, it faces challenges in being specific to certain robot leg configurations or in the lack of interpretability of the extracted features. In this paper, a terrain attribute extraction algorithm is proposed based on contrastive learning. It leverages the haptic data generated from the interaction between the robot's legs and terrain to automatically extract terrain attributes. The results demonstrate that the attributes extracted using this method strongly correlate with the actual softness of the terrain. Furthermore, these attributes played an important role in achieving high accuracy in terrain classification tasks.**

I. INTRODUCTION

The classification of terrains is important for autonomous robots working in unstructured scenarios, such as rescue robots [1] or planetary exploration robots [2], [3], [4]. Different terrains possess distinctive properties, such as softness or friction, which greatly impact the robot's locomotion. For example, NASA's Mars rover, Spirit, failed to climb out of a sand pit until the mission was aborted [5]. Thus, we envision that accurate terrain recognition will enable robots to adapt their gait for optimized movement [6], conserve energy [7], and avoid suffering non-geometric hazards [8].

The task of categorizing terrain has been a hot research topic in robotics. Previous related works can be divided into two sensing categories: exteroceptive and proprioceptive.

For exteroceptive sensing, the main visual sensor is usually a camera [9]. This type of terrain classification task is often solved by computer vision methods like feature-based methods [10] as well as deep neural network-based methods [11], [12]. In addition, some studies have attempted to predict the physical properties of terrain by utilizing semantic segmentation output [8], [13]. Despite the high accuracy achieved by these studies, these visual approaches have several limitations. First, when applied to uncontrolled field environments, the performance of these methods is susceptible to weather and lighting conditions. Second, these methods typically require large amounts of labeled, high-quality data, making them difficult to adapt to diverse application scenarios. Lastly, these methods face challenges in precisely perceiving the intrinsic physical properties of the terrain.

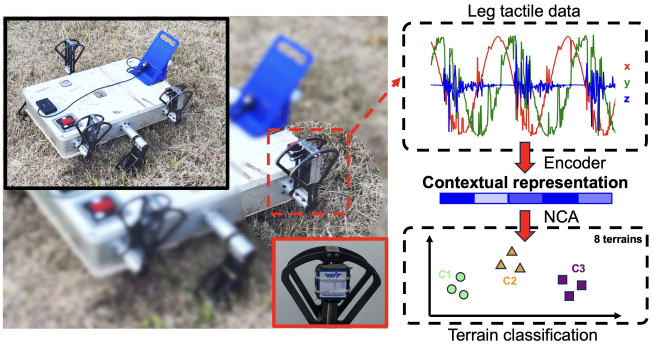


Fig. 1. An overview of our proposed method in the work. The haptic data is encoded into a contextual representation through an encoder that is trained using contrastive learning techniques. Utilizing these representations, we subsequently extract terrain attributes and complete the classification task.

Proprioceptive sensors directly record signals from the robot's interaction with the terrain during locomotion. These signals are strongly correlated with terrain attributes such as softness and friction. Researchers have utilized data recorded by various conventional sensors to achieve robust terrain recognition. Examples include acoustic signals [14], haptic data [15], [16], and IMU vibration data [17]. Some researchers have also attempted to develop their custom proprioceptive sensors for terrain classification [6], [18], [19]. These approaches generally utilize neural networks to process the collected sensor data. Supervision is conducted through a classification loss function to improve the accuracy of terrain recognition; this necessitates supervised training of the model using precise and unambiguous labels for terrain categories. As a result, the model focuses on encoding data into representations optimized for classification, neglecting the capture of intricate similarities or distance relationships between attributes across different types of terrain; this inevitably compromises the interpretability of the model. Alternatively, some other researchers have accomplished terrain classification by modeling sensor-ground interactions and making inferences about the terrain's mechanical properties based on the sensor data [20]. While this approach requires expertise, the extracted features are intricately tied to the robot's dynamic model and the acquisition method of sensors' data. Consequently, the approach exhibits a limited scope of generality.

* Equal contribution.

This work was supported in part by the National Natural Science Foundation of China (Grant No.U21A20119 and No.U1713206) and in part by the Shenzhen Science and Innovation Committee (Grant No.JCYJ20200109113412326 and No.JCYJ2021032412040003). (Corresponding author: Haoyao Chen.)

All authors are with Harbin Institute of Technology Shenzhen, P.R. China, e-mail:hychen5@hit.edu.cn.

Until now, while some studies have utilized unsupervised learning for image-based terrain recognition tasks [21], [22], few have tried to apply contrastive learning to model robot terrain interaction data unsupervised. In this paper, we propose a terrain attribute extraction method based on contrastive learning, which has proven effective in mining the intrinsic attributes of data [23]. This method includes steps of preprocessing, feature extraction, dimension reduction, and unsupervised clustering. The haptic data is converted into a contextual representation, from which we extract the dominant terrain attribute parameters, thereby accomplishing the task of terrain classification.

In the experimental part, the amphibious robot AmphiHex-II [24], equipped with accelerometers on its legs, was used to collect haptic data across eight different terrains in an outdoor environment. This robot is not only well-suited for navigating rough terrains but also excels in perceiving ground properties through its crawling gait. Our experimental results indicate that our terrain attribute extraction method effectively distinguishes softness attributes of various terrains, achieving a terrain classification accuracy rate of 96.947% on the test set.

II. METHODOLOGY

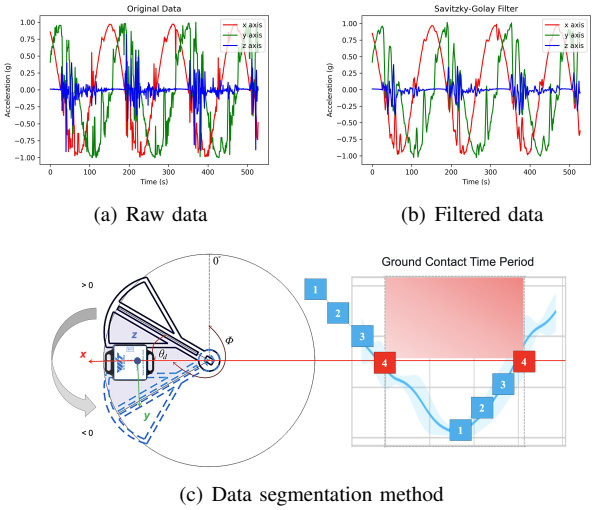
A. Overview

We propose a method for terrain attribute extraction and classification, which leverages contrastive learning techniques and utilizes haptic data from robots as input. Firstly, the filtering and time series segmentation algorithms were used to generate data samples that capture the tactile interactions between the robot and the terrain. Next, we present an unsupervised contrastive learning algorithm with a hierarchical contrastive loss function. This function operates in three dimensions—time, instance, and feature—to learn high-dimensional contextual features from the haptic samples of the terrain. Then, the contextual feature is dimensionally reduced using Neighborhood Components Analysis (NCA) [25] to eliminate redundant information. Finally, we model the terrain attribute feature distribution as a combination of multiple Gaussian distributions. Unsupervised clustering is performed by the Gaussian Mixture Model (GMM) [26] to estimate terrain attributes from the features.

B. Data Preprocessing

A significant amount of noise exists in the raw sensor data, which is unrelated to terrain interactions. Therefore, the Savitzky-Golay filter was used to perform noise reduction on the time series data. As shown in Figure 2(b), the filtering process smooths out the haptic data, effectively reducing burrs and spikes while preserving the original signal's shape details.

We introduce a sliding window-based algorithm that segments continuous haptic data into specific periods during which the robot's legs contact the ground. As shown in Figure 2(c), the acceleration sensor's x-axis data transitions from greater than 0 to less than 0 when the robot's leg starts to contact the ground. Consequently, the mean value of the



(a) Raw data (b) Filtered data

(c) Data segmentation method

Fig. 2. The haptic data before and after Savitzky-Golay filter. Fig. 2(c) shows our sliding window sampling method to segment out the haptic data when the robot's legs are in contact with the ground.

x-axis data in the four sliding windows can be calculated to ascertain whether the robot leg is in contact with the ground. If the mean value of the first three windows is greater than 0 and the mean value of the fourth window is less than 0, it is considered that the robot leg starts to contact the ground. Similarly, if the mean value of the first three windows is less than 0 and the mean value of the last window is greater than 0, it is considered that the robot leg ends in contact with the ground. Using this method, we effectively segment the measurement data of the robot's leg into discrete periods corresponding to ground interaction and provide accurate time-series data samples for subsequent modules.

C. Feature Extraction

Contrastive learning compares differences between data samples to mine intrinsic attributes within data distributions and to enhance feature representations, thereby improving performance in downstream tasks like classification and regression. Given the advantages of contrastive learning, we propose an unsupervised learning algorithm for feature extraction of the robot's haptic data. The pipeline of the method is as follows. First, two sub-sequences of equal length with overlapping segments are extracted from the input raw time series. They are then fed into the model's encoder to extract their feature representations. Finally, the model is trained using a hierarchical contrastive loss function to guide the encoder model to learn a high-quality contextual representation.

1) Encoder: We borrow our encoder from TS2Vec [27], a framework that has the ability to extract features of time series at multiple scales while capturing different levels of contextual information. The encoder takes in a set of time series data, denoted as $\mathcal{X}$, which consists of N elements: $\{x_1, x_2, \cdots, x_N\}$. Each sample $x_i \in \mathbb{R}^{T \times F}$ represents a time-series data segmenting the contact interaction between the robot and the terrain. The length of data is T, and the number of sensor channels is F. Our objective is to utilize

the encoder to learn a mapping function that transforms each input x_i into its corresponding feature representation $r_i = \{r_{i,1}, r_{i,2}, \cdots, r_{i,T}\}$, where $r_{i,t} \in \mathbb{R}^K$ is a timestamp-level feature representation. These K-dimensional features contain contextual knowledge related to terrain attributes.

2) Loss Function: We train the contrastive model using a hierarchical loss function in three dimensions: time, instances, and features. The temporal contrastive loss and the instance-wise contrastive loss guide the model to learn discriminative representations at different timestamps and across different instances [28]. For example, the model can learn the trends of haptic signals in the time dimension and the features between different terrains in the instance dimension.

We select the representation $r_{i,t}$ and $r'_{i,t}$ which are at the same timestamp as positive pairs. These representations are extracted from two augmented context views r_i and r'_i, generated from the same input time series x_i. In the temporal contrastive loss, the negative pairs are comprised of feature representations that originate from different timestamps, such as $r_{i,t}$ and $r_{i,t'}$, or $r_{i,t}$ and $r'_{i,t'}$. We formulate L_{time} as,

$$L_{\text{time}}^{(i,t)} = -\log \frac{\exp(r_{i,t} \cdot r'_{i,t})}{\sum_{t'=1}^{T} \left(\exp(r_{i,t} \cdot r'_{i,t'}) + \exp(r_{i,t} \cdot r_{i,t'})|_{t \neq t'}\right)}, \quad (1)$$

where i and t are the instance and timestamp index.

In the instance-wise contrastive loss, feature representations originating from different instances but sharing the same timestamp are classified as negative pairs. We formulate $L_{instance}$ as,

$$L_{\text{instance}}^{(i,t)} = -\log \frac{\exp(r_{i,t} \cdot r'_{i,t})}{\sum_{t'=1}^{B} \left(\exp(r_{i,t} \cdot r'_{j,t}) + \exp(r_{i,t} \cdot r_{j,t})|_{i \neq j}\right)}, \quad (2)$$

where B is the batch size.

Besides that, we propose to use the structural similarity index (SSIM) [29]. This loss function evaluates the structural similarity between two feature representations to provide a more accurate quantification of their differences. This approach is widely adopted in the field of image quality assessment. We contend that the positive sample pairs generated from terrain proprioceptive sensing data can be interpreted as distortions of the original signals caused by cropping and masking techniques. Therefore, if we can effectively measure the change in structural information between the distorted signal and the original signal, we can provide better learning objectives for the contrastive learning model, thus improving the model's learning effect. We formulate L_{ssim} as,

$$L_{\text{ssim}}^{(i,t)} = -\log \frac{(2\mu_{r_{i,t}} \cdot \mu_{r'_{i,t}} + c_1)(2\sigma_{r_{i,t}r'_{i,t}} + c_2)}{(\mu_{r_{i,t}}^2 + \mu_{r'_{i,t}}^2 + c_1)(\sigma_{r_{i,t}}^2 + \sigma_{r'_{i,t}}^2 + c_2)}, \quad (3)$$

where μ_r, σ_r^2, and $\sigma_{r_{i,t}r'_{i,t}}$ represent the mean, variance, and covariance, respectively, of the positive sample pairs. And c is a constant value introduced to prevent division by zero.

The computational workflow of the hierarchical loss function is detailed in Algorithm 1. When the timestamp length of the feature representation exceeds one unit, a continuous maximum pooling operation is applied to reduce the timestamp dimension of the feature representation. Next, the contrastive loss functions for the time, instance, and feature dimensions are computed separately and accumulated to the total loss value. The total loss value is averaged to obtain the hierarchical contrastive loss values.

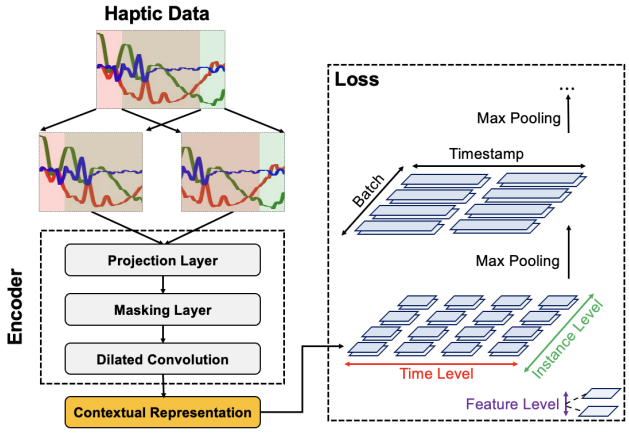


Fig. 3. The pipeline of contrastive learning. The input time-series data is segmented to yield two subsequences containing identical fragments. Then these subsequences are encoded into contextual representation. Training is performed in three dimensions: time, instance, and feature using hierarchical contrast loss function.

Algorithm 1: Hierarchical Contrastive Loss

Input: Feature representations r, r'
Loss Function L_{time}, L_{instance}, L_{ssim}
Output: hierarchical loss $L_{\text{hierarchical}}$

L_{total}=0, $L_{\text{hierarchical}}$=0, d=0;
while timestamp_length$(r) > 1$ **do**
  //Max pooling for time dimension.
  $r \leftarrow \text{maxpool1d}(r, \text{kernel\_size} = 2)$;
  $r' \leftarrow \text{maxpool1d}(r', \text{kernel\_size} = 2)$;
  //Calculate for all positive pairs.
  $L_{\text{total}} \leftarrow L_{\text{total}} + L_{\text{time}}(r, r') + L_{\text{instance}}(r, r') + L_{\text{ssim}}(r, r')$;
  $d \leftarrow d + 1$;
end
$L_{\text{hierarchical}} \leftarrow L_{\text{total}}/d$;
return $L_{\text{hierarchical}}$;

D. Attribute Estimation

After processing the contrast learning model, each haptic data sample x_i is extracted to obtain a high-dimensional semantic feature representation r_i. However, high-dimensional representations often grapple with issues of correlation and multicollinearity among different features, which can undermine the effectiveness and interpretability of the data. The NCA method maximizes both intra-class similarity and minimizes inter-class similarity in the data dimensionality reduction process, and thus can extract relatively more essential features in the data. For this reason, this paper employs NCA to reduce the dimension of the features extracted from the haptic data. This technique maps the complex data into a lower-dimensional space. Consequently, it yields a clearer feature distribution, facilitating a more precise estimation of terrain attribute parameters. We can also directly utilize a classifier on these dimensionally reduced features to identify the types of terrain.

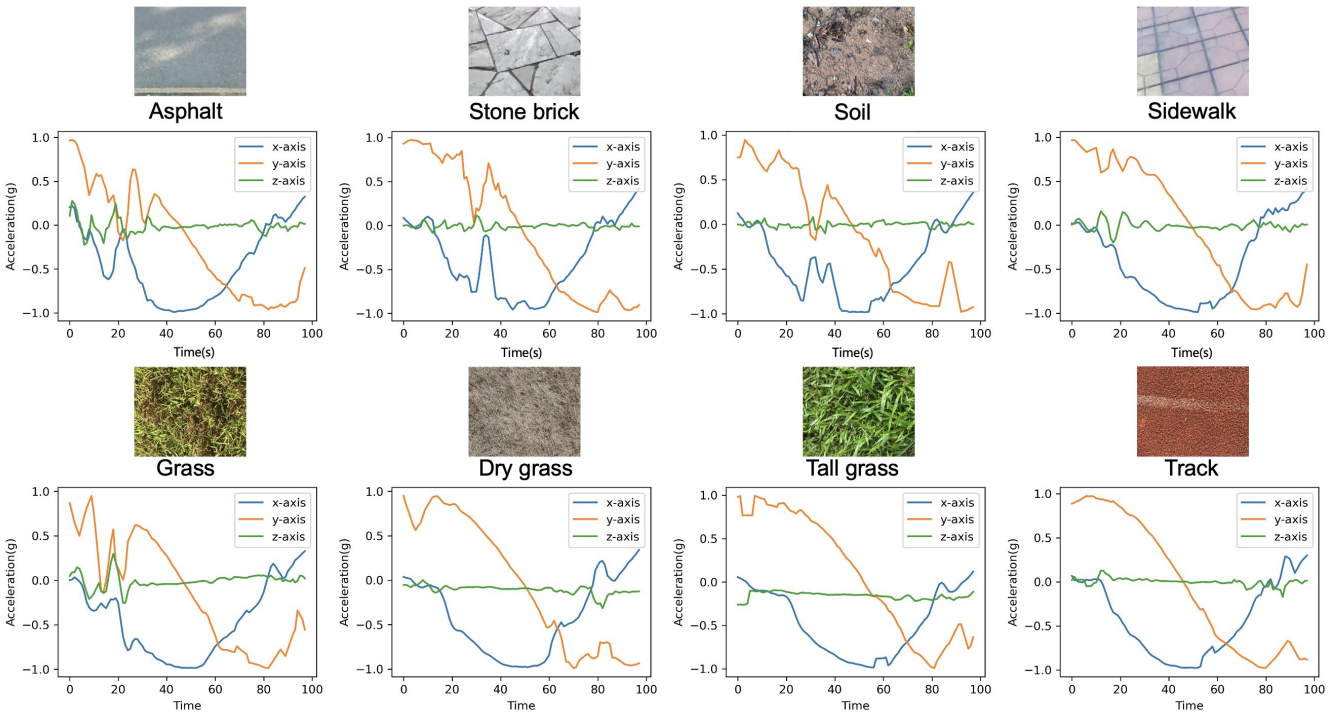


Fig. 4. The dataset contains eight types of terrain in outdoor scenes on campus.

After dimensionality reduction, we obtain a more compact low-dimensional feature representation $\mathcal{Z} = \{z_1, z_2, \ldots, z_N\}$, where $z_n \in \mathbb{R}^D$. In the subsequent step, it is necessary to model the distribution of terrain attribute features within the low-dimensional space. Different terrains possess distinct physical attributes. This paper assumes that each attribute feature follows a Gaussian distribution. Consequently, the overall distribution across all terrains can be represented as a combination of multiple Gaussian distributions. We employ GMM to estimate the terrain attribute distribution parameters automatically. The GMM performs unsupervised clustering on the feature distribution. The attribute parameters characterizing each type of terrain are quantified using the coordinates of the corresponding cluster centers.

III. Experiment

A. Data Collection

The robot used in this paper is AmphiHex-II, chosen for its ability to locomote in a manner that allows for the perception of both normal and tangential forces exerted by the terrain. This capability enhances the robot's sensitivity to the terrain's properties. The CPG control algorithm is integrated into its control circuit system, which enables the robot to switch gaits smoothly during movement. The variable stiffness leg structure allows the robot to move freely over asphalt, masonry, grass, mud, forest, and other terrains, providing superior field traversal performance than other robots. To collect haptic data across these different terrains, the AmphiHex-II is equipped with a three-axis accelerometer.

In this paper, the haptic perception data of the robot in the outdoor campus scene was collected in December. As shown in Figure 4, eight kinds of field terrain were collected in our dataset, including asphalt pavement, brick stone, forest soil, sidewalk pavement, plastic runway, and three types of grass. Tall grass is the dense grass at the edge of the sidewalk. The dry grass is the withering grass in the forest. Moreover, the playground grass is relatively lower. Through the time series preprocessing method introduced in this paper, a terrain haptic dataset containing 2,223 training data samples and 954 test data samples is finally constructed.

B. Metrics

In order to quantitatively analyze the effectiveness of the terrain attribute extraction method proposed in this paper, we have chosen three evaluation metrics: classification accuracy, clustering quality, and estimation error. For evaluating classification accuracy, this paper uses the k-nearest neighbors algorithm (KNN) classification model. We formulate $Accuracy$ as,

$$\text{Accuracy} = \frac{\sum_{i=1}^{N}[f(r_i) = y_i]}{N}, \tag{4}$$

where r_i denotes the representation vector, y_i denotes the true label of the terrain class, and N is the number of samples.

For clustering quality, Silhouette Coefficient (SC) [30] is used for evaluation. This metric computes a score for each sample by considering its distance from other samples within the same cluster and the nearest neighboring cluster. By doing so, the metric captures the level of intra-cluster similarity for samples of the same terrain and the inter-cluster variability for samples of different terrains. For each sample, the Silhouette Coefficient can be formulated as,

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \tag{5}$$

where a_i denotes the average distance of sample i from other samples within the same cluster, and b_i denotes the average

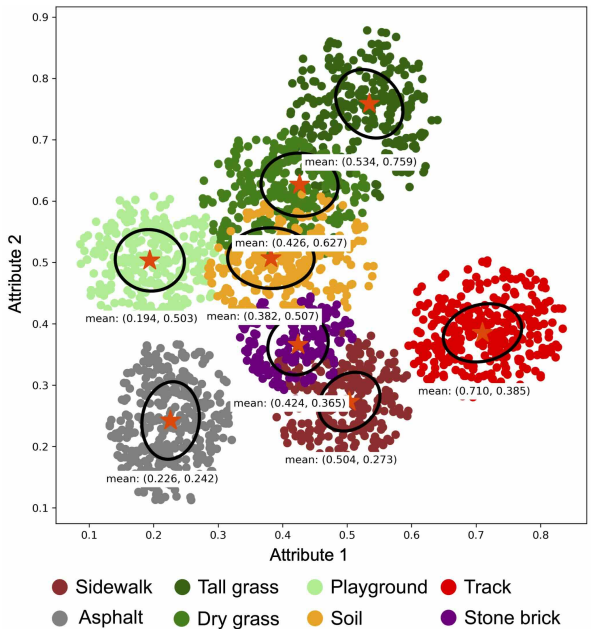


Fig. 5. Gaussian distribution of terrain attributes obtained by our method.

distance of sample i from all samples in the nearest neighbor cluster. The clustering quality can be defined as the mean of the Silhouette Coefficients for all samples.

$$\mathrm{SC} = \frac{1}{N}\sum_{i=1}^{N} s_i. \tag{6}$$

To evaluate estimation error, we first employ the Hungarian algorithm to find the optimal match between estimated and empirical values of terrain attribute distribution parameters. Subsequently, the Root Mean Square Error (RMSE) is calculated to quantify estimation accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{1}{P}\sum_{p=1}^{P}(x_i - \mu_i)^2}, \tag{7}$$

where P is the number of terrain attribute distributions and x_i is the estimated value of the ith terrain attribute parameter, and μ_i is the empirical value of the ith terrain attribute parameter.

C. Result Analysis

As shown in Figure 5, we ultimately obtain a Gaussian mixture distribution of terrain attributes after implementing the algorithm in this paper on the haptic dataset. Terrain attribute parameters are estimated based on the coordinates of each distribution's clustering centers. For the classification task, we adopt the KNN classifier and achieve an accuracy of 96.947% on the test set.

To assess the validity of the terrain property parameters derived using our method, we compare them with empirical measurements of terrain softness. These empirical measurements utilized a Shore hardness tester, and the softness attribute values were averaged from multiple measurements. As shown in Table I, the comparison results reveal that the estimated terrain attribute parameters closely align with the empirical softness measurements, evidenced by a Pearson correlation coefficient of 0.966421. This result demonstrates that our algorithm not only effectively classifies various types of terrain but also accurately estimates terrain attributes, exhibiting strong interpretability.

TABLE I
THE TERRAIN ATTRIBUTE

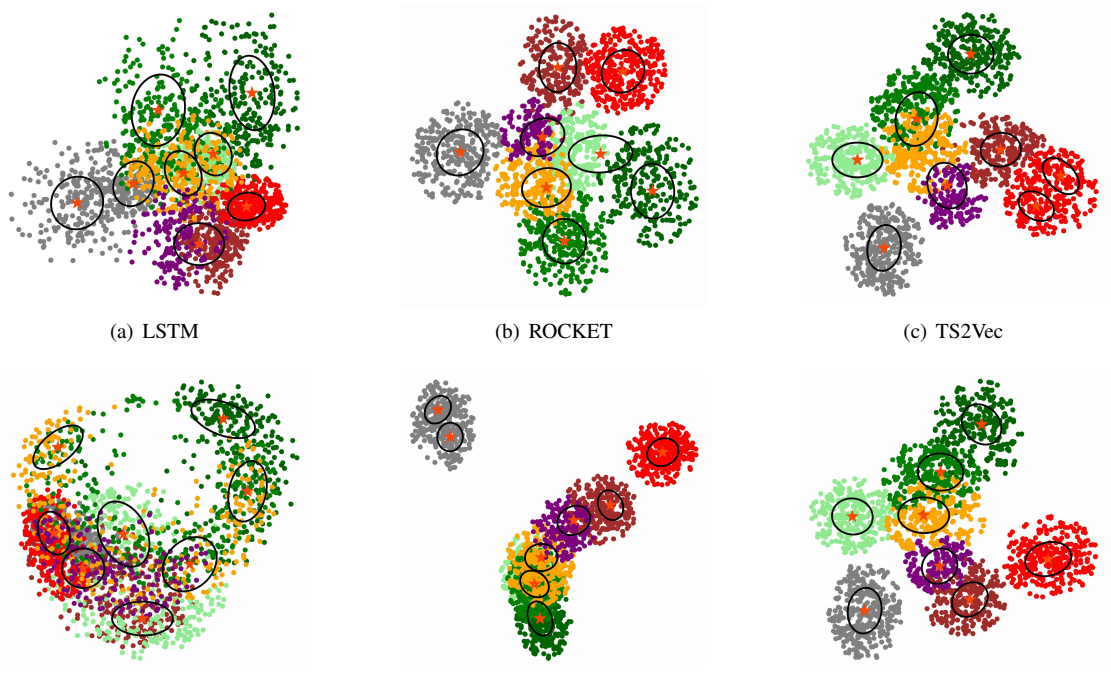
Terrain	Estimated Value (1-attribute2)	softness parameters
Asphalt	0.758	87.5
Sidewalk	0.727	94.5
Tall grass	0.241	17.5
playground grass	0.497	43.0
plastic runway	0.615	67.5
dry glass	0.373	32.0
soil	0.493	52.5
stone brick	0.635	87.5

D. Ablation Study

1) Feature Extraction Method: To analyze the effectiveness of the feature extraction method proposed in this paper, we conducted comparative experiments with three mainstream automatic feature extraction methods: LSTM-AutoEncoder [31], ROCKET [32], and Ts2Vec [27]. All methods employ a feature representation dimension of 320 and utilize the same techniques for feature reduction and parameter estimation.

The figure 6 illustrates the Gaussian distributions estimated by the Gaussian Mixture Model. In this visualization, the mean of each Gaussian distribution is denoted by a five-pointed star. At the same time, the ellipse signifies the contour corresponding to the first standard deviation of each distribution. Correspondingly, Table II presents the evaluation of the quantitative metrics associated with the four encoding methods.

Based on comparative experimental results, the feature extraction method introduced in this paper demonstrates superior performance over the three alternative methods across all metrics. Specifically, our method produces a more discriminative distribution of different terrain attributes. Within this distribution, the inter-cluster distance between samples is markedly increased, while the intra-cluster cohesion among samples is strengthened. Furthermore, the method presented in this paper performs well in estimating terrain attributes with unsupervised method. In contrast, alternative approaches such as LSTM-AutoEncoder and ROCKET show greater deviations in their estimated distribution parameters from the empirical values. Additionally, the results from TS2Vec representation exhibit an error in which two sets of parameters are estimated for a single distribution. This enhanced performance is attributable to the contrastive learning method taken in this paper, which employs a structured similarity measure to formulate a comparative learning loss function tailored to accommodate the unique characteristics of terrain haptic data. This strategy effectively reduces the distance between positive sample pairs while increasing the distance between negative sample pairs in the high-



(a) LSTM (b) ROCKET (c) TS2Vec

(d) PCA (e) LDA (f) **Ours**

Fig. 6. In the experiment, we tried to replace the modules in our method using other popular approaches. LSTM(a), ROCKET(b), TS2Vec(c) were used to replace our feature extraction module. PCA(d) and LDA(e) were used to replace the feature dimensionality reduction module.

TABLE II
COMPARISON OF FEATURE EXTRACTION METHODS

Methods	Accuracy(%)	Silhouette Cofficient	RMSE
LSTM-AutoEncoder	90.085	0.054	0.032
ROCKET	96.349	0.194	0.030
TS2Vec	96.601	0.244	0.097
Ours	**96.947**	**0.256**	**0.015**

TABLE III
COMPARISON OF DIMENSION REDUCTION METHODS

Methods	Accuracy(%)	Silhouette Cofficient	RMSE
PCA	81.869	-0.111	0.174
LDA	88.259	0.194	0.165
Ours	**96.947**	**0.256**	**0.015**

dimensional space. As a result, it provides a more focused learning target for comparative learning models. Compared to conventional methods based on Convolutional Neural Networks and Long Short-Term Memory networks, the proposed feature extraction method is more effective in extracting the intrinsic attributes that differentiate various terrains.

2) Dimensional Reduction Method: To assess the efficacy of the feature dimensionality reduction techniques implemented in this study, we conduct comparative experiments with two well-established methods: Principal Component Analysis (PCA) [33] and Linear Discriminant Analysis (LDA) [34]. It should be noted that all these approaches employ identical feature extraction and attribution estimation methodologies for a fair comparison. The attribute distribution outcomes for each of these dimensionality reduction methods are presented in Figure 6(d) 6(e), while the associated quantitative metrics are detailed in Table III.

Based on the comparative experimental results, the NCA Method used in this study outperforms both PCA and LDA across key metrics. Furthermore, NCA yields a more discriminative distribution of varying terrain attributes. In contrast, PCA exhibits the least effectiveness in feature reduction. This ineffectiveness results in a high overlap among the characterized feature distributions, consequently leading to the largest attribute estimation errors. LDA performs modestly, only successfully distinguishing the hardest terrain (asphalt pavement) from the softest (plastic track) while failing to differentiate other terrain categories accurately. This superior performance of NCA considers both similarity and distance relationships between the data.

IV. CONCLUSION

In this study, we propose a novel method that extracts terrain attributes from haptic data, achieving remarkable accuracy in classifying different types of terrain. The method comprises four main components: preprocessing of haptic data, feature extraction based on contrastive learning, dimensionality reduction via NCA, and parameter estimation through GMM. Our experiments were conducted on a haptic dataset that encompasses eight distinct terrains. The results indicate that the contextual representations obtained through contrastive learning encompass terrain-related knowledge. The terrain attributes extracted from the representation exhibit a strong correlation with the terrain's softness properties. Moreover, by employing dimensionality reduction on contextual representations, we can achieve high accuracy in terrain classification with a simple KNN classifier.

REFERENCES

[1] B. M. Pillai and J. Suthakorn, "Challenges for novice developers in rough terrain rescue robots: A survey on motion control systems," *Journal of Control Science and Engineering*, vol. 2019, 2019.

[2] P. Arm, G. Waibel, J. Preisig, T. Tuna, R. Zhou, V. Bickel, G. Ligeza, T. Miki, F. Kehl, H. Kolvenbach *et al.*, "Scientific exploration of challenging planetary analog environments with a team of legged robots," *Science robotics*, vol. 8, no. 80, p. eade9548, 2023.

[3] L. Ding, R. Zhou, Y. Yuan, H. Yang, J. Li, T. Yu, C. Liu, J. Wang, S. Li, H. Gao *et al.*, "A 2-year locomotive exploration and scientific investigation of the lunar farside by the yutu-2 rover," *Science Robotics*, vol. 7, no. 62, p. eabj6660, 2022.

[4] Z. Wang, H. Chen, and M. Fu, "Whole-body motion planning and tracking of a mobile robot with a gimbal rgb-d camera for outdoor 3d exploration," *Journal of Field Robotics*, 2024.

[5] J. Matson, "Unfree spirit: Nasa's mars rover appears stuck for good." *Scientific American*, vol. 302, no. 4, pp. 16–16, 2010.

[6] H. Chen, X. Zhu, S. Zhu, H. Chen, S. Zhang, and Y. Lou, "Terrain attribute recognition system for cpg-based legged robot," in *Intelligent Robotics and Applications: 14th International Conference, ICIRA 2021, Yantai, China, October 22–25, 2021, Proceedings, Part III 14*. Springer, 2021, pp. 556–567.

[7] S. Zenker, E. E. Aksoy, D. Goldschmidt, F. Wörgötter, and P. Manoonpong, "Visual terrain classification for selecting energy efficient gaits of a hexapod robot," in *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2013, pp. 577–584.

[8] R. Zhou, L. Ding, H. Gao, W. Feng, Z. Deng, and N. Li, "Mapping for planetary rovers from terramechanics perspective," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1869–1874.

[9] Z. Wang, H. Chen, S. Zhang, and Y. Lou, "Active view planning for visual slam in outdoor environments based on continuous information modeling," *IEEE/ASME Transactions on Mechatronics*, 2023.

[10] P. Filitchkin and K. Byl, "Feature-based terrain classification for littledog," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 1387–1392.

[11] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathyamoorthy, K. Weerakoon, and D. Manocha, "Ga-nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8138–8145, 2022.

[12] V. Suryamurthy, V. S. Raghavan, A. Laurenzi, N. G. Tsagarakis, and D. Kanoulas, "Terrain segmentation and roughness estimation using rgb data: Path planning application on the centauro robot," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 1–8.

[13] P. Ewen, A. Li, Y. Chen, S. Hong, and R. Vasudevan, "These maps are made for walking: Real-time terrain property estimation for mobile robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7083–7090, 2022.

[14] J. Zürn, W. Burgard, and A. Valada, "Self-supervised visual terrain classification from unsupervised acoustic feature learning," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 466–481, 2020.

[15] 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.

[16] J. Bednarek, M. Bednarek, L. Wellhausen, M. Hutter, and K. Walas, "What am i touching? learning to classify terrain via haptic sensing," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 7187–7193.

[17] 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.

[18] X. A. Wu, T. M. Huh, A. Sabin, S. A. Suresh, and M. R. Cutkosky, "Tactile sensing and terrain-based gait control for small legged robots," *IEEE Transactions on Robotics*, vol. 36, no. 1, pp. 15–27, 2019.

[19] Z. Yu, S. Perera, H. Hauser, P. R. Childs, and T. Nanayakkara, "A tapered whisker-based physical reservoir computing system for mobile robot terrain identification in unstructured environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3608–3615, 2022.

[20] L. Ding, P. Xu, Z. Li, R. Zhou, H. Gao, Z. Deng, and G. Liu, "Pressing and rubbing: physics-informed features facilitate haptic terrain classification for legged robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 5990–5997, 2022.

[21] T. Panambur, D. Chakraborty, M. Meyer, R. Milliken, E. Learned-Miller, and M. Parente, "Self-supervised learning to guide scientifically relevant categorization of martian terrain images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1322–1332.

[22] F. Zhang, H. Bian, and M. Wei, "Contrastive learning ideas in underwater terrain image matching," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[23] P. H. Le-Khac, G. Healy, and A. F. Smeaton, "Contrastive representation learning: A framework and review," *Ieee Access*, vol. 8, pp. 193 907–193 934, 2020.

[24] B. Zhong, S. Zhang, M. Xu, Y. Zhou, T. Fang, and W. Li, "On a cpg-based hexapod robot: Amphihex-ii with variable stiffness legs," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 2, pp. 542–551, 2018.

[25] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," *Advances in neural information processing systems*, vol. 17, 2004.

[26] D. A. Reynolds *et al.*, "Gaussian mixture models." *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.

[27] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "Ts2vec: Towards universal representation of time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8980–8987.

[28] E. Eldele, M. Ragab, Z. Chen, M. Wu, C. K. Kwoh, X. Li, and C. Guan, "Time-series representation learning via temporal and contextual contrasting," *arXiv preprint arXiv:2106.14112*, 2021.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] H. Řezanková, "Different approaches to the silhouette coefficient calculation in cluster evaluation," in *21st International Scientific Conference AMSE Applications of Mathematics and Statistics in Economics*, 2018, pp. 1–10.

[31] M. Said Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Network anomaly detection using lstm based autoencoder," in *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, 2020, pp. 37–45.

[32] A. Dempster, F. Petitjean, and G. I. Webb, "Rocket: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1454–1495, 2020.

[33] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[34] P. Xanthopoulos, P. M. Pardalos, T. B. Trafalis, P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," *Robust data mining*, pp. 27–33, 2013.