2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 979-8-3503-7770-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IROS58592.2024.10801407

Proprioception Is All You Need: Terrain Classification for Boreal Forests

Damien LaRocque,[1] William Guimont-Martin,[1] David-Alexandre Duclos,[1]
Philippe Giguère,[1] François Pomerleau[1]

***Abstract*— Recent works in field robotics highlighted the importance of resiliency against different types of terrains. Boreal forests, in particular, are home to many mobility-impeding terrains that should be considered for off-road autonomous navigation. Also, being one of the largest land biomes on Earth, boreal forests are an area where autonomous vehicles are expected to become increasingly common. In this paper, we address the issue of classifying boreal terrains by introducing `BorealTC`, a publicly available dataset for proprioceptive-based terrain classification (TC). Recorded with a *Husky* A200, our dataset contains 116 min of Inertial Measurement Unit (IMU), motor current, and wheel odometry data, focusing on typical boreal forest terrains, notably snow, ice, and silty loam. Combining our dataset with another dataset from the literature, we evaluate both a Convolutional Neural Network (CNN) and the novel state space model (SSM)-based Mamba architecture on a TC task. We show that while CNN outperforms Mamba on each separate dataset, Mamba achieves greater accuracy when trained on a combination of both. In addition, we demonstrate that Mamba's learning capacity is greater than a CNN for increasing amounts of data. We show that the combination of two TC datasets yields a latent space that can be interpreted with the properties of the terrains. We also discuss the implications of merging datasets on classification. Our source code and dataset are publicly available online: `https://github.com/norlab-ulaval/BorealTC`.**

I. Introduction

With the ongoing development of field robotics, it has become common for robots to navigate through increasingly complex and challenging terrains [1]. To prevent and handle situations where an uncrewed ground vehicle (UGV) may get stuck or immobilized, vehicles must be able to accurately assess and identify the terrain they are navigating on. Such terrain awareness is often framed as a classification problem over the different terrain types a UGV might traverse [2], [3]. The problem of terrain classification (TC) has been applied in many contexts, including traversability assessment [1], terrain-aware path planning [4], and as a prior for predicting energy consumption [5].

Although being the largest land biome on Earth [6], boreal forests have received little attention for the development of autonomous navigation. Moreover, terrain awareness is essential in the context of the boreal forest, where a multitude of terrain types can significantly hinder the mobility of UGVs [7]. Subject to large seasonal variability, boreal forests are especially suitable for developing systems that are capable of multi-seasonal navigation [8]. Hence, TC in these regions is essential to the advancement of field robotics in more challenging conditions.

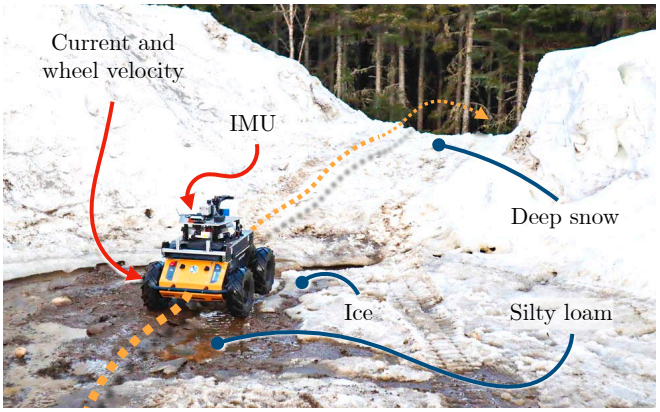

Fig. 1: An example of challenges caused by terrain in boreal forests. A small UGV with a controller relying on a friction coefficient expecting concrete will struggle to follow a path when confronted with complex ground-wheel interactions.

Although a traversed terrain can be determined with cameras [2], this approach is not universal for in-situ TC, depending on environmental conditions. For example, boreal forests challenge conventional visual-based TC, as their dense coniferous canopies obstruct the sunlight, yielding low visual feature contrast [7]. Furthermore, boreal forests are in regions with large illumination variances, having shorter days during winter. Hence, a TC method that relies on light does not work anytime in winter when light is usually scarce [9]. While lidars can be used in low lighting conditions, they provide little semantic information about the nature of the surface, given they only provide geometric information from the surroundings [10]. Both cameras and lidars are influenced by adverse environmental conditions such as snowstorms [11], heavy rain, or fog [12]. Finally, as illustrated in Figure 1, the terrain configuration in boreal forests can be quite complex, and only a thin layer of snow accumulation can hinder the capacity of visual sensors to see the ground composition. This occlusion is particularly true when a fresh layer of snow covers loam or ice, derailing controllers that assume a constant friction coefficient [13]. In this context, proprioceptive sensing is more robust for TC in harsh conditions [3], such as those found in boreal forests.

For practical reasons, many authors have gathered terrain data indoors, in urban settings, or on university campuses, where terrain usually does not impede movement [2], [5], [14]. In contrast, our work focuses primarily on off-road terrains that pose challenges for wheeled skid-steering mo-

[1] The authors are with the Northern Robotics Laboratory, Université Laval, Quebec City, Canada,
damien.larocque@norlab.ulaval.ca,
francois.pomerleau@norlab.ulaval.ca

bile robots (SSMRs). In particular, we consider deep snow, known for immobilizing wheeled SSMRs [7]; ice, a rigid slippery surface; and silty loam, a muddy and slippery soil that hampers rotations.

Current approaches in TC generally incorporate temporal aspects, either through frequency domain representations or by employing recurrent models. Building upon these foundations, we introduce two models. Our first model improves upon previous Convolutional Neural Network (CNN) baseline architectures applied to spectrograms generated from proprioceptive sensor data. We employ recent deep learning techniques and integrate windowing functions when computing spectrograms to improve the model's performance. The second model draws inspiration from recent advancements in state space models (SSMs), specifically the Mamba architecture introduced by Gu *et al.* [15]. Mamba is a promising approach for TC, due to its ability to apply selectivity on raw sequential data while scaling linearly with sequence length.

The main contributions of this paper are (i) `BorealTC`, a novel dataset for TC with wheeled SSMRs in wintry off-road conditions; (ii) an improvement on state-of-the-art methods for data-driven TC based on CNNs; (iii) the exploration of SSM-based models for TC; and (iv) a study of the challenges of merging datasets acquired by different UGVs of the same model.

II. Related Work

A. Sensor modalities

Exteroceptive sensors have been extensively used for terrain classification (TC), as they can predict terrains at a distance. Among these sensors, cameras stand out for their ability to capture appearance-based features, providing valuable insights into the physical nature of surfaces. For example, Atha *et al.* [4] classified terrain from Martian rovers' mast camera images. Walas *et al.* [16] classified terrains from lidar data, by creating intensity-based and geometry-based feature vectors. TC can also benefit from the combination of sensors employing different modalities. For instance, Schilling *et al.* [10] leveraged both geometric and appearance-based features from lidar and camera data to assess terrain traversability. Audio signals have also been combined with camera images [2], [14] and radar scans [12] for TC. Proprioceptive data were used with camera images to classify terrain in both agricultural [17] and urban contexts [18]. While exteroceptive sensors offer many advantages, they can be significantly challenged in boreal forests. For instance, cameras will suffer from illumination variability [7], [9]. Moreover, lidars are affected by inclement weather and extreme precipitation [11], [12], typical of the same region.

To circumvent these limitations, an alternative approach is to base terrain classification primarily on proprioceptive data. Proprioceptive sensors present the advantage of directly informing about the physical characteristics of a surface through their impact on the dynamics of a UGV. Hence, they do not require an unobstructed line of sight with the surface, nor do they rely on surface illumination. Common proprioceptive sensors are Inertial Measurement Units (IMUs), wheel odometry, and motor ammeters, with the latter providing indirect torque measurements. IMUs yield accelerations that can be used to classify terrains. These measurements are especially useful for legged robots, where the body is far from the ground [19], [20]. For finer surface information, the accelerometer can be dragged on the ground [21], bypassing the damping effect of legs or wheels. In addition, actuator and haptic signals can be leveraged for TC when dealing with legged robots. For example, leg force measurements from the *Messor* walking machine [22] or leg haptic signals from an *ANYmal* walking robot [23] can provide valuable information for classification. Allred *et al.* [5] also classified terrain using leg joints data, in conjunction with images of the front-facing camera of a Spot robot. For wheeled robots with encoders and IMU, Reina *et al.* [24] have shown that proprioceptive data can be used to evaluate slip and motion resistance coefficients to predict the terrain on which a UGV was driven. By adding motor currents to wheel velocities, and IMU, Vulpi *et al.* [3] demonstrated a mean accuracy of 91.5 % for TC over four types of terrain. Given that relying on proprioceptive sensors to record terrain signatures will be strongly related to the sensor placement, the geometry of the robot, and the type of locomotion, we will use their data in combination with ours to analyze *domain shift*, which makes knowledge transfer between vehicles challenging.

B. Methods for classification

Earliest approaches involved expert systems [24] and Support Vector Machines (SVMs) for lidar-based [16] and vibration-based terrain classification (TC) [17], [20]. Yet, this family of Machine Learning (ML) techniques has the drawback of relying on features that were hand-crafted with a priori expert knowledge, which adds an inductive bias to the learning.

More recently, deep learning approaches have gained popularity due to their representation learning capabilities and their ability to process any type of sensor information. For instance, CNN architectures have been used to classify terrains based on camera images [4], [5], [18]. In the case of [18], the incorporation of proprioceptive data through a second parallel network improved classification accuracy.

Inspired by speech recognition applications, 1-D data, such as IMU-recorded vibrations and audio signals, are often transformed into frequency representations. One example of this technique is the utilization of short-time Fourier transform (STFT)-based spectrograms by Vulpi *et al.* [3], who employed a CNN to classify terrains with proprioceptive data from an IMU and the drive system of a UGV. Similarly, Zürn *et al.* [2] applied the same type of spectrogram for unsupervised acoustic feature learning. These learned acoustic features are part of a self-supervised framework for audiovisual-based TC using neural encoders. Likewise, Ishikawa *et al.* [14] employed Variational Auto-Encoders (VAEs) and a Gaussian Mixture Model (GMM) to learn terrain types from audiovisual data autonomously. In their approach, the audio signals were represented as Mel-frequency cepstral coefficients (MFCCs), again inspired by speech recognition tech-

niques. Building upon these methods, our CNN classifier uses a STFT-based spectrogram. In contrast to Vulpi *et al.* [3], we demonstrate that applying a windowing function mitigates spectral leakage.

Another way to process 1-D time series is to input them into a neural network designed for sequential data. In such cases, recurrent networks like Long Short-Term Memorys (LSTMs) are commonly used. Allred *et al.* [5] achieved high accuracy on TC by applying a LSTM on the joint data of a legged robot. A more complex variant of Recurrent Neural Networks (RNNs), a convolutional LSTM (C-LSTM), was used by Valada *et al.* [25] for audio-based TC on MFCCs. As Vulpi *et al.* [3] demonstrated, the CNN architecture is more stable and more accurate than LSTM and C-LSTM for proprioceptive-based TC on a wheeled UGV, hence these methods won't be investigated.

In addition to LSTM-based approaches, other methods inspired by natural language processing (NLP) have emerged for TC. Transformers, in particular, have recently been proposed to process long sequences and classify terrains as accurately as RNNs [23]. However, the performance of transformer-based approaches comes at the expense of quadratic scaling, in proportion to the sequence length. To address this limitation, other RNN-like approaches have been suggested. Most notably, Gu *et al.* [15] introduced Mamba, a SSM-based architecture that offers significant performance gains while scaling linearly. While Mamba aligns with the recurrent nature of RNNs, it distinguishes itself by employing selectivity, emphasizing key time steps in data sequences while leveraging its operations' parallelism. Although Mamba has been shown to beat recent architectures in various downstream tasks such as DNA sequence classification [15], its application to TC remains unexplored. Therefore, we evaluate the potential of Mamba for this task.

C. Datasets for terrain classification

For practical reasons, some studies have focused their efforts on classifying data acquired indoor [16], [22], [23] or in urban environments [5], [14], [25]. For example, wheeled UGVs were recently used for multi-modal data acquisition in urban areas, such as for the *Freiburg Terrains* and the *Jackal robot 7-class terrain* datasets [2], [18]. Terrain classification (TC) was explored in various off-road contexts, such as an experimental farm [3], [17], [24], a volcanic island [20], and Mars [4]. Given that the boreal forest remains covered in snow for at least half of the year, any dataset covering such an environment ought to include data on snow and ice. These two wintry terrains have been studied with legged robots [5], [19]. However, to the best of our knowledge, no publicly available TC datasets contain labeled snow and ice data from a wheeled UGV.

Many have publicly released their datasets to help generalize classification across different kinds of terrains [2]–[5], [18], [23]. Aligned with this endeavor, we propose `BorealTC`, a publicly available dataset containing annotated data from a wheeled UGV for various mobility-impeding terrain types typical of the boreal forest. Our data were acquired on deep snow and silty loam, two uncommon terrains in an urban setting, both in winter and spring. Additionally, to encompass a variety of wintry terrains, we recorded data while driving our UGV on an ice rink. Our dataset also includes experiments on asphalt and flooring, two prevalent terrains in recent datasets. These types of terrain facilitate the comparison of the learned representations of our models with those obtained from other datasets.

III. Methodology

We propose an approach to classify terrains using proprioceptive data from an IMU, and the drive system of a UGV. Figure 2 shows the general overview of the pipeline used for the evaluation. Following Vulpi *et al.* [3], we divided our sensor signals into 5 s partitions. These partitions were then split into train and test subsets, to evaluate the performance of our models with a *k*-fold cross-validation strategy. To overcome the class imbalance in the data, the partitions in both subsets were oversampled, such that all classes have the same number of samples. Subsequently, two models were applied to each sample: a CNN and a Mamba classifier.

A. Convolutional Neural Network

Following the method of Vulpi *et al.* [3], we apply a CNN to spectrograms generated from proprioceptive data. The spectrograms were computed on each 1.7 s sample by applying STFTs with a window length of 0.4 s and an overlap of 0.2 s. The original implementation [3] applied the STFT directly on these small windows, which is equivalent to applying boxcar filters. This filter results in spectra with artifacts, called *spectral leakage*, that must be avoided [26]. Instead, we use a Hamming window $w[n]$, such that

$$w[n] = a_0 - (1 - a_0) \cdot \cos\left(\frac{2\pi n}{N_{win}}\right), \tag{1}$$

for $0 \leq n \leq N_{win}$, with N_{win} being the number of samples per window in the STFT and $a_0 = 0.54$ [27]. As in Vulpi *et al.* [3], the resulting spectrograms of each channel were then padded, concatenated, and fed as an input to the CNN. The convolution operations of the CNN were performed across all channels of the spectrograms, by moving the kernel through the frequency-time planes.

B. Mamba

In light of state-of-the-art results on multiple tasks involving sequential data, we suggest using the Mamba architecture [15] for terrain classification (TC). Based on recent work with discrete SSMs, Mamba introduces attention-like selectivity and recurrent-like parallel associative scanning for linear scaling in sequence length. Notably, Mamba obviates the necessity for domain-shifting the samples to spectrograms, and consequently the need for preprocessing steps, like padding or downsampling, required by CNNs for data uniformity. Mamba thus directly processes the proprioceptive data in its sequential form, making it a promising solution for proprioceptive-based TC.

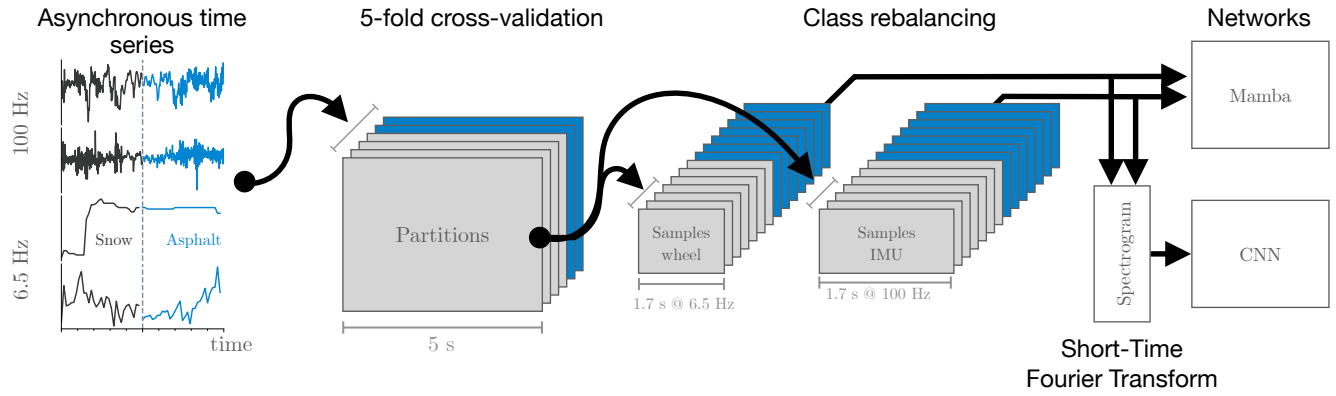

Fig. 2: Overview of the training process. From the left, data from asynchronous sensors were recorded and hand-labeled according to the terrain on which the robot was driven. To allow a 5-fold cross-validation, trajectories are split into 5 s partitions. Classes are then rebalanced through oversampling before being fed to the different networks. The CNN performed classification on spectrograms, while Mamba classified the samples directly in the time domain.

IV. Experiments

In this section, we first describe the platform used to record the `BorealTC` dataset. We then give details about the terrain and highlight differences between ours and `Vulpi` dataset. We finally give implementation details for both ML architectures used in our analyses.

A. Vehicle and sensors

Our dataset was recorded with a *Husky* A200, a wheeled UGV from *Clearpath Robotics* (Kitchener, Ontario, Canada). The *Husky* is a SSMR with a weight of 70 kg and a wheel baseline of 0.6 m. While the UGV has four wheels, the wheels are mechanically coupled with timing belts, such that a single gearmotor drives two wheels on the same side of the vehicle. The motor currents are measured by MDL-BDC24 motor drivers, while the wheel speeds are collected using optical encoders mounted at the output of the gearmotors. Both currents and wheel velocities are provided by the *Husky* at a rate of 6.5 Hz. The robot is also equipped with a Xsens MTi-30 IMU, which records three angular velocities and three linear accelerations at a frequency of 100 Hz. To ensure consistency with the data of Vulpi *et al.* [3], the IMU data was transformed to the base reference frame of the robot, using the Coriolis formula as described by Deschênes *et al.* [28]. Finally, all sensor data was recorded with Robot Operating System (ROS) 2. The robot and its sensors are shown in Figure 1.

B. Dataset description

Our `BorealTC` dataset was collected by driving the *Husky* on five different types of terrains, namely ASPHALT, FLOORING, ICE, SILTY LOAM, and SNOW, as shown in Figure 3. ASPHALT, FLOORING, and ICE data were acquired in an urban setting, on the campus of *Université Laval* (46°46′52.47″N, 71°16′27.74″W). ICE data was recorded on an ice rink on the same campus, ensuring that the ice was consistent between experiments. SILTY LOAM and SNOW data were taken at *Forêt Montmorency* (47°19′19.29″N, 71°8′50.13″W), the experimental boreal forest of *Université Laval*, 75 km north of its main campus. During all seasons, excluding winter, the soil of *Forêt Montmorency* is a podzol typical of boreal forests. More specifically, the trails on which the *Husky* was driven are dug in the silty loam layer of the podzol. As we recorded the SILTY LOAM data at the end of the winter, the silty loam was saturated in water, making it slippery. All the data were collected on relatively flat surfaces, with a pitch smaller than 5°, to avoid effects from the slope of the terrain on the classification.

To better evaluate our methods, we used our `BorealTC` dataset in conjunction with the `Vulpi` dataset [3]. As the data follows the work of Reina *et al.* [17], which was recorded in San Cassiano, Lecce, Italy (40°3′35.40″N, 18°20′50.98″E), we extrapolate that the dataset comes from the same location. As with our dataset, the `Vulpi` dataset was collected with a *Clearpath Robotics Husky*, meaning that both datasets were acquired with vehicles of similar dimensions and weights. While the acquisition rates of our IMU and wheel data are at 100 Hz and 6.5 Hz respectively, the `Vulpi` dataset has a rate of 50 Hz for their IMU and 15 Hz for their wheel data. A summary of both `BorealTC` and `Vulpi` datasets is given in Table I, where the ranges of motion commands that were sent to the UGVs are compared for each class. This comparison is done by computing the median $\tilde{\cdot}$ and the interquartile range (IQR) of the absolute values $|v_x|$ and $|\omega_z|$ of the linear and angular velocities. Notably, our `BorealTC` dataset is an order of magnitude larger and contains a total of 116 min of sensor data, while `Vulpi` contains 13 min. In addition, our dataset includes a significant amount of data with rotational motions, such as turns-on-the-spot, while the `Vulpi` dataset contains mostly forward linear motions. The inclusion of rotational motions enables a better representation of the entire input space of a UGV [13], and is thus crucial for accurate terrain modeling.

C. Implementation details

As our preprocessing pipeline is inspired by Vulpi *et al.* [3], we ported their publicly available MATLAB implemen-

Fig. 3: Types of terrains considered in our dataset. From left to right: silty loam, deep snow, asphalt, flooring, and ice.

TABLE I: Description of our `BorealTC` dataset and the `Vulpi` dataset [3]. For each class, we give the number N of 5 s partitions, the location (Loc.), as well as the median ($\tilde{\cdot}$) and the IQR of the absolute values of the linear speed $|v_x|$ and of the yaw rate $|\omega_z|$. Locations include San Cassiano (SC), *Forêt Montmorency* (FM), and the main campus of Université Laval (UL).

Terrain	N	Loc.	$\lvert\tilde{v}_x\rvert$ (IQR)	$\lvert\tilde{\omega}_z\rvert$ (IQR)
`Vulpi` [3]				
CONCRETE	24	SC	0.56 (0.26)	0.00 (0.00)
DIRT ROAD	16	SC	0.56 (0.25)	0.00 (0.00)
PLOUGHED	60	SC	0.56 (0.26)	0.00 (0.00)
UNPLOUGHED	56	SC	0.56 (0.25)	0.00 (0.00)
`BorealTC` (ours)				
ASPHALT	111	UL	0.46 (0.58)	0.01 (0.09)
FLOORING	423	UL	0.23 (0.05)	0.02 (0.09)
ICE	450	UL	0.24 (0.38)	0.27 (0.52)
SILTY LOAM	126	FM	0.00 (0.24)	0.10 (0.17)
SNOW	281	FM	0.00 (0.31)	0.10 (0.26)

tation[1] to Python. To ensure a fair comparison with the baseline from Vulpi *et al.* [3], we kept the same pipeline parameters, using five folds for the cross-validation, with partitions and sample durations set at 5 s and 1.7 s, respectively. Our pipeline implementation was then validated by running it on the `Vulpi` dataset. In line with our Python-based pipeline, we implemented our models with PyTorch Lightning,[2] thereby facilitating replicability and adhering to prevailing standards in deep learning.

Our CNN and Mamba classifiers were trained, validated, and tested on the `Vulpi` and the `BorealTC` datasets, as well as on the combination of both datasets. In each scenario, channel-wise normalization was performed using the minimal and maximal values derived from the training data.

For the combined dataset, downsampling each similar sensor to the smallest available frequency was necessary for the CNN to ensure that the dimensions of the spectrograms were compatible. Hence, we resampled our IMU data to 50 Hz and the wheel data from `Vulpi` to 6.5 Hz. It is important to note that this downsampling step was not needed for Mamba, as it can handle varying frequencies without requiring adjustments. We kept the original labels from both datasets, resulting in a classification on nine terrain types.

For our CNN, we first applied a convolution layer with a kernel of size of one, effectively applying a Multilayer Perceptron (MLP) individually to each frequency-time element across all channels. Subsequently, the network sequentially applied a batch normalization (BN) layer, followed by a convolution layer with a kernel of size of three, and another BN. Finally, predictions were generated by applying a fully connected layer on the flattened feature maps.

For our Mamba classifier, we used two branches to treat the IMU and wheel velocity data separately. Each branch consists of a fully connected layer used to project the input data to a high-dimensional feature space, followed by a Mamba block. This larger feature space enhances training stability, as the latent representation can be encoded on additional channels. Using two branches allows the models to handle both data types independently, without requiring preprocessing steps such as padding or downsampling for input compatibility. While multiple Mamba blocks can be stacked one after the other, we found that it did not improve the model's performance; we thus only used one Mamba block per branch. As Mamba is a causal model, we follow Gu *et al.* [15] and only keep the final hidden state of each block. To predict the terrain type, we concatenate the final hidden states of both branches and feed them to a fully connected layer.

Our classifiers were trained on an NVIDIA RTX A6000 GPU, an AMD Ryzen Threadripper 3970X 32-core CPU, and 128 GB of RAM. For the training phase, we utilized a further subdivision of 10 % dedicated to validation, allowing us to monitor the models' performances during training. Our source code and the `BorealTC` dataset are publicly available in our `BorealTC` repository.[3] In addition, all training details, including hyperparameters and model checkpoints, are given in the same repository.

V. RESULTS

This section presents the performance of our models on both evaluated datasets. For all models, we reported the following metrics over all folds: the precision, the recall, the F1 score, and the accuracy. We then analyze the influence of the train dataset size on the test error of our models. Finally, we discuss the labeling of both datasets by comparing the latent space of both datasets.

[1] https://github.com/Ph0bi0/T_DEEP
[2] https://github.com/Lightning-AI/lightning
[3] https://github.com/norlab-ulaval/BorealTC

A. Models performance

To quantitatively assess the performance of our classifiers, we tested them on the `Vulpi` and the `BorealTC` datasets, as well as on the combination of both datasets. All reported metrics are averaged over 5-fold cross-validation. Table II gives the performance of our CNN and Mamba classifiers on `Vulpi`. Our CNN is 2.62 %pt more accurate than the implementation of Vulpi *et al.* [3], reported at 91.5 %. An ablation study determined that the use of a Hamming window increased the accuracy by 0.6 %pt, while the rest of the improvement is due to better hyperparameter optimization. It can be seen from the metrics that the CNN outperforms Mamba on `Vulpi`. Since the dataset of Vulpi *et al.* [3] is small, we conjecture that the CNN has an innate advantage due to its stronger inductive bias and the direct utilization of spectrograms.

TABLE II: Results on the `Vulpi` dataset.

Terrain	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
CNN				
CONCRETE	**99.21**	**95.27**	**97.20**	**94.12**
DIRT ROAD	**92.40**	**92.05**	**92.22**	
PLOUGHED	**96.94**	**98.96**	**97.94**	
UNPLOUGHED	**88.20**	**90.48**	**89.32**	
Mamba				
CONCRETE	87.13	83.33	85.19	86.76
DIRT ROAD	91.34	83.90	87.46	
PLOUGHED	93.93	96.67	95.28	
UNPLOUGHED	76.08	83.93	79.81	

Table III compares the performance of both classifiers on our larger dataset. Although the CNN demonstrates higher accuracy, the difference with Mamba is not as pronounced as in Table II. We surmise that Mamba's performance catches up to the CNN due to the significantly larger size of `BorealTC`, approximately nine times that of `Vulpi`. On the other hand, the CNN achieves comparable performance on both datasets, albeit being less accurate on `BorealTC`. We attribute this lower accuracy to the higher complexity of our dataset, in terms of terrain types and input commands.

B. Impact of train dataset size on model performance

In light of Mamba's close brush with CNN in Section V-A, we performed an ablation study to examine the impact of train dataset size on test error. To increase the amount of available data for the ablation, we used the combined dataset detailed. We generated several decimated training sets, with ratios of $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and $\frac{1}{16}$ with respect to the complete dataset. For each subset, we applied a stratified fold strategy to maintain the same class distribution. The test error percentage was then obtained over a 5-fold cross-validation. Figure 4 illustrates the impact of the train dataset size on the test error for both classifiers. While CNN outperforms Mamba on smaller

TABLE III: Results on the `BorealTC` dataset.

Terrain	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
CNN				
ASPHALT	**92.98**	83.89	88.20	**93.96**
FLOORING	**97.29**	**98.70**	**97.99**	
ICE	**97.25**	**98.11**	**97.68**	
SILTY LOAM	**96.00**	**97.24**	**96.61**	
SNOW	86.84	**92.31**	89.49	
Mamba				
ASPHALT	91.90	**85.50**	**88.59**	93.68
FLOORING	95.46	98.17	96.79	
ICE	97.12	97.36	97.24	
SILTY LOAM	95.39	96.20	95.79	
SNOW	**88.68**	91.57	**90.10**	

datasets, Mamba seems to be more accurate and could possibly surpass CNN on larger datasets, which aligns with our observation in Section V-A. Mamba's trend line closely follows a linear trend in log-log space, whereas the CNN trend line is potentially sublinear. For both models, the trend follows a typical power-law curve. Yet, further studies may be needed to determine whether these trends hold true for larger datasets. Overall, the results suggest that performance is predominantly limited by the amount of training data, and not by the quantity of sensor information or the learning capacity of a classifier. Finally, we observed a worse performance compared to the separate datasets, especially for CNN. This could be due to overlaps between the classes in both datasets. We investigate this behavior in the following section.

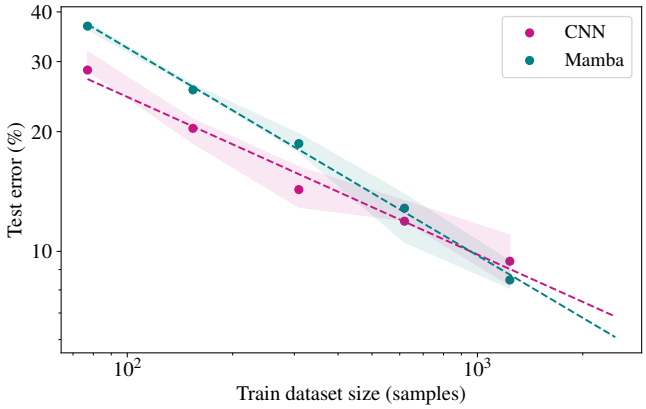

Fig. 4: Influence of train dataset size on the test error in log-log scale. Performance was assessed by combining `Vulpi`–`BorealTC`. Bands show the IQR over 5 folds.

C. Latent space coherence between both datasets

When combining two TC datasets with distinct terrain classes, it is important to ensure that each terrain is well delineated, since the data from two terrain classes may share common features, albeit being from two different datasets. To

assess this issue, we took advantage of having two different datasets, both acquired by two similar vehicles, to determine whether the labels in both datasets were properly delimited. We used the t-distributed stochastic neighbor embedding (t-SNE) technique [29] to visualize the latent space created by our models, as seen in Figure 5. Specifically, t-SNE was applied to project the features, extracted right before the fully connected layer of our CNN, into 2-D space. This projection allows us to visualize the proximity between labels of all classes. The result of the t-SNE is interpreted with our field observations and terrain descriptions from previous works on the `Vulpi` dataset [17]. In most cases, classes are grouped in separate clusters, which means they are well-defined. Meanwhile, CONCRETE and ICE are each spread in three different clusters, whereas SILTY LOAM is dispersed between CONCRETE, ICE, and FLOORING. This dispersion can be explained by different weather conditions or commanded body velocities. The spread of the embeddings affects the performance when both datasets are merged, as described in the last section. Next, we can see that the data for ASPHALT, CONCRETE, and UNPLOUGHED are in the same region, meaning that their labels coincide, as they are all hard rigid terrains. Similarly, the embeddings for DIRT ROAD, PLOUGHED, and SNOW are in the same zone, clustering as soft grounds. Moreover, we noted an adjacency between ICE and FLOORING data, two hard grounds. This result has previously been observed by Giguere *et al.* [19] with a legged robot. Furthermore, the ICE embeddings are close to SILTY LOAM, which were both slippery during our experiments. Finally, even if the terrains are accordingly grouped with their properties, both datasets are still separated in the t-SNE visualization. Indeed, apart from ASPHALT, all the classes from the `Vulpi` dataset (with lighter labels) are in the lower region of the t-SNE, while the classes from our dataset (with darker labels) are in the upper region. We believe that this separation indicates that both datasets are distinguishable, as they were recorded with different vehicles and different experimental procedures. As such, our classifiers have a harder time consolidating both datasets' features, which agrees with the performance hit noted in Section V-B. This observation could be verified by applying a consistent recording procedure to record various datasets on a standardized fleet of vehicles.

VI. Conclusion

In this paper, we introduced our `BorealTC` dataset for proprioceptive-based terrain classification (TC), which is one order of magnitude larger than its alternative. Our publicly available dataset contains IMU, motor current, and wheel odometry signals recorded with a *Husky* A200 over five types of terrains, with a particular focus on boreal forests. In particular, `BorealTC` contains annotated data on three wintry terrain types, SNOW, ICE, and SILTY LOAM, all of which are omnipresent in boreal forests. We confirmed the capacity of a CNN and a Mamba classifier to classify terrains on the `Vulpi` dataset [3] and our dataset. Moreover, we showed that a spectrogram-based CNN excels on smaller TC datasets, while Mamba performs well on increasingly larger datasets. Additionally, a t-SNE applied on a combined TC dataset showed how embeddings of a type of terrain cluster with embeddings of terrains with similar properties. However, we determined that merging two datasets does not yield a homogeneous mix of the terrain labels. Such division could be caused by differing vehicles, sensors, and methodologies, meaning that the specificities of each dataset could have guided the classification. Future research should aim at standardizing data acquisition procedures for TC. We believe gathering various datasets by applying the same experimental procedure on similar vehicles enables proper datasets for TC without providing dataset-specific hints to classifiers. We surmise that this need for standardized TC experiments is aligned with the requirement for standardized terrain-aware vehicle characterization [13], as well as models that can classify terrains in any given biome. Finally, while we suggest that *proprioception is all you need* for TC, further research is needed to compare the performance of proprioceptive-based TC in boreal contexts with other architectures and modalities.

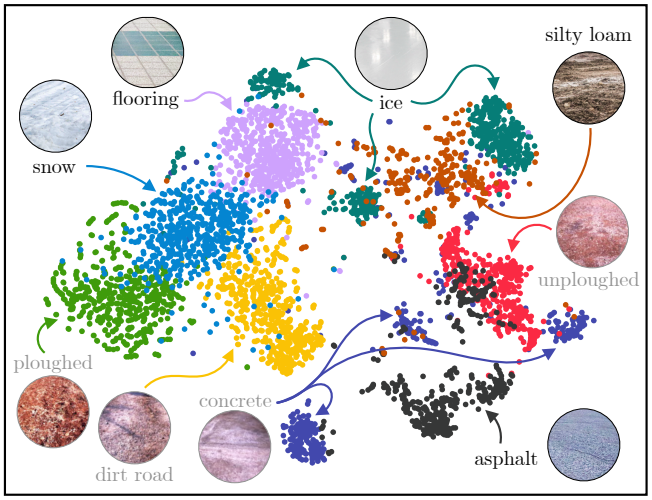

Fig. 5: An illustration of class proximity using t-SNE analysis from our CNN classifier trained on both datasets. Each colored dot represents an embedding for a given class. Each inset illustrates a terrain class in a dataset. The terrains of the `BorealTC` dataset are indicated with black labels and photos from our experiments, while the terrains of the `Vulpi` dataset are indicated with gray labels and insets from the figures of Vulpi *et al.* [3].

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the grant CRDPJ 527642-18 SNOW (Self-driving Navigation Optimized for Winter).

References

[1] P. Borges, T. Peynot, S. Liang, *et al.*, "A Survey on Terrain Traversability Analysis for Autonomous Ground Vehicles: Methods, Sensors, and Challenges," *Field Robotics*, vol. 2, no. 1, 1567–1627, Mar. 2022.

[2] 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, 466–481, Apr. 2021.

[3] F. Vulpi, A. Milella, R. Marani, and G. Reina, "Recurrent and convolutional neural networks for deep terrain classification by autonomous robots," *Journal of Terramechanics*, vol. 96, pp. 119–131, Aug. 2021.

[4] D. Atha, R. M. Swan, A. Didier, Z. Hasnain, and M. Ono, "Multi-mission Terrain Classifier for Safe Rover Navigation and Automated Science," in *2022 IEEE Aerospace Conference (AERO)*, IEEE, Mar. 2022.

[5] C. Allred, H. Kocabas, M. Harper, and J. Pusey, "Terrain Dependent Power Estimation for Legged Robots in Unstructured Environments," in *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*, IEEE, Dec. 2022.

[6] D. J. Hayes, D. E. Butman, G. M. Domke, J. B. Fisher, C. S. Neigh, and L. R. Welp, "Boreal forests," in *Balancing Greenhouse Gas Budgets*. Elsevier, 2022, 203–236.

[7] D. Baril, S.-P. Deschênes, O. Gamache, *et al.*, "Kilometer-scale autonomous navigation in subarctic forests: challenges and lessons learned," *Field Robotics*, vol. 2, no. 1, 1628–1660, Mar. 2022.

[8] I. Ali, A. Durmush, O. Suominen, *et al.*, "FinnForest dataset: A forest landscape for visual SLAM," *Robotics and Autonomous Systems*, vol. 132, p. 103 610, Oct. 2020.

[9] M. Paton, F. Pomerleau, and T. D. Barfoot, "In the dead of winter: Challenging vision-based path following in extreme conditions," in *Field and Service Robotics*. Springer International Publishing, 2016, 563–576.

[10] F. Schilling, X. Chen, J. Folkesson, and P. Jensfelt, "Geometric and visual terrain classification for autonomous mobile navigation," in *2017 IEEE/RSJ IROS*, IEEE, Sep. 2017.

[11] C. Courcelle, D. Baril, F. Pomerleau, and J. Laconte, "18 On the Importance of Quantifying Visibility for Autonomous Vehicles Under Extreme Precipitation," in *Towards Human-Vehicle Harmonization*. De Gruyter, Mar. 2023, 239–250.

[12] D. Williams, D. De Martini, M. Gadd, L. Marchegiani, and P. Newman, "Keep off the Grass: Permissible Driving Routes from Radar with Weak Audio Supervision," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Sep. 2020.

[13] D. Baril, S.-P. Deschênes, L. Coupal, *et al.*, "DRIVE: Data-driven Robot Input Vector Exploration," in *2024 IEEE ICRA*, IEEE, May 2024, 5829–5836.

[14] R. Ishikawa, R. Hachiuma, A. Kurobe, and H. Saito, "Single-modal Incremental Terrain Clustering from Self-Supervised Audio-Visual Feature Learning," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, Jan. 2021.

[15] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[16] K. Walas and M. Nowicki, "Terrain classification using Laser Range Finder," in *2014 IEEE/RSJ IROS*, IEEE, Sep. 2014.

[17] G. Reina, A. Milella, and R. Galati, "Terrain assessment for precision agriculture using vehicle dynamic modelling," *Biosystems Engineering*, vol. 162, 124–139, Oct. 2017.

[18] Y. Chen, C. Rastogi, and W. R. Norris, "A CNN Based Vision-Proprioception Fusion Method for Robust UGV Terrain Classification," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7965–7972, Oct. 2021.

[19] P. Giguere and G. Dudek, "Clustering sensor data for autonomous terrain identification using time-dependency," *Autonomous Robots*, vol. 26, no. 2–3, 171–186, Mar. 2009.

[20] K. Otsu and T. Kubota, "Energy-Aware Terrain Analysis for Mobile Robot Exploration," in *Field and Service Robotics*. Springer International Publishing, 2016, 373–388.

[21] P. Giguere and G. Dudek, "A Simple Tactile Probe for Surface Identification by Mobile Robots," *IEEE Transactions on Robotics*, vol. 27, no. 3, 534–544, Jun. 2011.

[22] P. Dallaire, K. Walas, P. Giguere, and B. Chaib-draa, "Learning terrain types with the Pitman-Yor process mixtures of Gaussians for a legged robot," in *2015 IEEE/RSJ IROS*, IEEE, Sep. 2015.

[23] M. Bednarek, M. R. Nowicki, and K. Walas, "HAPTR2: Improved Haptic Transformer for legged robots' terrain classification," *Robotics and Autonomous Systems*, vol. 158, p. 104 236, Dec. 2022.

[24] G. Reina and R. Galati, "Slip-based terrain estimation with a skid-steer vehicle," *Vehicle System Dynamics*, vol. 54, no. 10, pp. 1384–1404, Jun. 2016.

[25] A. Valada and W. Burgard, "Deep spatiotemporal models for robust proprioceptive terrain classification," *The International Journal of Robotics Research*, vol. 36, no. 13–14, 1521–1539, Aug. 2017.

[26] F. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, 51–83, 1978.

[27] R. B. Blackman and J. W. Tukey, "The Measurement of Power Spectra from the Point of View of Communications Engineering - Part II," *Bell System Technical Journal*, vol. 37, no. 2, 485–569, Mar. 1958.

[28] S.-P. Deschênes, D. Baril, M. Boxan, J. Laconte, P. Giguère, and F. Pomerleau, "Saturation-Aware Angular Velocity Estimation: Extending the Robustness of SLAM to Aggressive Motions*," in *2024 IEEE ICRA*, IEEE, May 2024, 10711–10718.

[29] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.