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

Robot-Dependent Traversability Estimation for Outdoor Environments using Deep Multimodal Variational Autoencoders

Matthias Eder[1] and Gerald Steinbauer-Wagner[1]

Abstract— **Efficient and reliable navigation in off-road environments poses a significant challenge for robotics, especially when factoring in the varying capabilities of robots across different terrains. To achieve this, the robot system's traversability is usually estimated to plan traversable routes through an environment. This paper presents a new approach that utilizes Deep Multimodal Variational Autoencoders (DMVAEs) for estimating the traversability of different robots in complex off-road terrains. Our method utilizes DMVAEs to capture essential environmental information and robot properties, effectively modeling factors that influence robotic traversability. The key contribution of this research is a two-stage traversability estimation framework for various robots in diverse off-road conditions that integrates robot properties in addition to environmental information to predict the traversability for various robots in a single model. We validate our method through real-world experiments involving four ground robots navigating an alpine environment. Comparative evaluations against state-of-the-art traversability estimation methods demonstrate the superior accuracy and robustness of our approach. Additionally, we investigate the transfer of trained models to new robots, enhancing their traversability estimation and extending the applicability of our framework.**

I. INTRODUCTION

The need for reliable navigation in robotic systems in applications ranging from search and rescue to environmental monitoring has underscored the importance of developing robust and adaptive traversability estimation methods in outdoor environments [1]. For successful navigation, robots must not only perceive their environment, but also assess the feasibility of traversing it given their capabilities [2], [3]. Existing approaches, primarily based on machine learning (e.g. [4], [5]), often focus on specific robots and limited environment scopes, failing to provide a comprehensive solution that is able to deal with heterogeneous robots and varying terrains. In response to the growing need for comprehensive traversability estimators, the importance of developing adaptable methods becomes apparent, especially as the robotic landscape expands to a wide range of applications. As technological advances drive the introduction of new robots, each with different physical attributes and sensory configurations, the need to consider a robot's capabilities in traversability analysis becomes necessary [4], [6], as retraining traversability models from scratch for each new robot becomes time-consuming and impractical. The ability to transfer knowledge and insights from previously trained models to new robots accelerates the deployment of ground robots with different capabilities in diverse environments [7]. It also maximizes the utilization of existing data sets and expertise, minimizing the data collection effort required for each new robot.

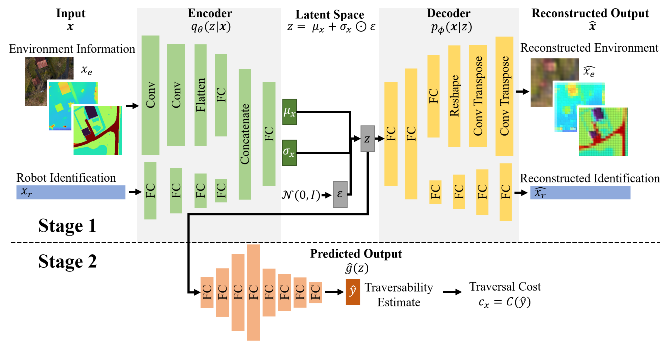

Fig. 1. Overview of the proposed two-stage DMVAE framework. In stage one, a variational autoencoder is trained on environmental data and on a robot identifier to represent the input in a probabilistic latent space. In stage two, the latent space is used to generate training samples to learn a neural network that predicts the traversability for a given robot.

This paper presents a new data-driven framework for traversability estimation for heterogeneous robots in diverse environments that can be used for global path planning. Using deep multimodal variational autoencoders (DMVAEs) [8], [9], our framework combines robot-specific properties to represent a specific robot and environmental information from aerial imagery to improve traversability predictions, with the goal of increasing the applicability of the framework to different robotic systems. At the core of this research fosters the development of a holistic system capable of estimating the traversability for different types of robots in different environments. Therefore, the DMVAE-based framework works in two stages, as can be seen in Figure 1. In the first stage, it captures probabilistic latent representation of the input data, including both environmental information and robot-specific properties, which have already been shown to have an impact on traversability estimation [6]. By design and the used input, this stage of the framework has the advantage that it can be trained on data that is relatively easy to obtain, since there is no need to directly deploy a robot in the field. In the second stage, multiple robot systems are deployed in the field to collect data that can be used to estimate robot-dependent traversability in a given environment. This data is then used as ground truth to learn the traversability for a robot in a given environment. Furthermore, the probabilistic nature of latent representations is used to generate multiple training samples for a ground truth data point in order to incorporate the latent distribution

This work was partially funded by the Austrian Research Promotion Agency (FFG) with the project FutureWoodTrans.

[1] Matthias Eder and Gerald Steinbauer-Wagner are with the Institute of Software Technology, Graz University of Technology, Graz, Austria. {matthias.eder, steinbauer}@ist.tugraz.at

into the traversability estimation. This comprehensive fusion of data sources enables our model to recognize intrinsic features from the input data that govern the traversability of a specific robot, resulting in more accurate and personalized predictions. To validate our approach, we conducted real-world experiments in an outdoor environment using data collected with four different robots. The results show the ability of our DMVAE framework to consistently improve the performance of traversability estimation, and demonstrate its robustness across a range of scenarios and terrains. Furthermore, our evaluation investigates the transferability of the proposed framework to new robots, showing its adaptability within a broader set of applications. Extensive evaluations are performed against a baseline method to assess the effectiveness of our approach. The results highlight the superiority of our approach in terms of performance and robustness, thus confirming the suitability of deep DMVAEs for traversability estimation.

The main contributions of this work are: 1) the integration of a robot identifier as additional parameter in a deep learning model for a multi-robot traversability estimator, 2) the proposal of a deep multimodal variational autoencoder (DMVAE) framework for improved traversability estimation, 3) the use of the generative capabilities of variational auto encoders to improve results and reduce the required amount of training data, and 4) the investigation of the ability to transfer traversability estimation models between heterogeneous robots and new environments.

The rest of this paper is organized as follows. Section II discusses related work in the area of traversability estimation. Section III formulates the problem and describes the learning process. Section IV presents a performance evaluation of the proposed framework and a comparison with a baseline approach. Section V concludes the paper.

II. Related Work

The task of traversing an environment can be optimized for a variety of objectives, including energy consumption, traversal time, smoothness, and collision risk [10], [11]. An alternative approach is to identify unstable motion patterns that should be avoided in order to model costs based on a robot's motion stability [12]. Some studies (e.g. [13]) focus on estimating a single objective, while others (e.g. [14], [15]) define traversability as a combination of multiple objectives.

One of the main objectives used in traversability estimation is to predict the traversal time through an environment. This objective is studied in a variety of contexts, with a primary focus on the dynamics of the robot in a given environment [16], [17]. The traversability for a robotic system has been studied using different sensors and data, such as proprioceptive measurements [18], [19], geometry-based representations [20], or vision sensors [21]. However, these traversability estimates do not take into account the effect of the environment on the robot's movement, but they can be used in conjunction with the robot's theoretical dynamics to compute a time-optimal path. To account for environmental factors, the authors of [22] propose a data-driven approach that also considers the effect of the environment on the traversal time estimate.

Another popular objective for quantifying the traversability of a terrain is energy consumption. For example, Quann et al. [13] propose a physics-based approach to predict the energy consumption for a robot by first acquiring friction coefficients from given areas and then applying Gaussian process regression to estimate the coefficients for new areas. The method also suggests using a grayscale satellite image to gain more information about unseen environments. While this method works well on terrain with available energy measurements, it is also shown that it does not generalize well to new environments where no data is available. Wei and Isler proposed a data-driven approach to predict energy consumption based on height information around the robot using deep learning [5]. It is shown to work well on the same type of terrain, but does not generalize well to other types of terrain without retraining the network with new data. Building on this work, the authors of [22] adapted this approach and integrated semantic terrain information in addition to the geometric information retrieved from an elevation map to improve generalization on different types of terrain. Other works further adopt the use of environmental information in data-driven approaches for traversability estimation based on other objectives such as traversal time or smoothness [23], [24]. While the aforementioned approaches mostly assume environmental information as a prerequisite (e.g. provided by aerial data as in [22]), other work shows the applicability of a data-driven pipeline using on-board perception [25]. More recently, work has been done on estimating probabilistic traversability models to enable risk-aware motion planning in off-road environments [26], [27]. In [28], the authors propose the use of autoencoders to predict the traversability of an environment by aiming to reconstruct safe terrain. By using the reconstruction error, safe traversable areas could be identified. Another work uses the reconstruction error of autoencoders to provide a confidence rating for traversal in a given environment [23]. The authors of [29] construct a latent space representation of the environment from deep encoders to improve the dynamic models of a robot used for optimizing trajectory planning. Similarly, conditional variational autoencoders are used to predict the trajectory of robots [30]. Furthermore, the authors of [27] propose the use of latent spaces for feature extraction from environment representations to perform traversability analysis. In [31], Generative Adversarial Networks (GAN) are used to classify motions as traversable or not, using visual data as input.

Most proposed works so far focus on a specific robot at a time. However, the use of transfer learning to transfer the traversability estimates to a new robot system or a new environment is already being explored [7], [32]. It has been shown to have advantages for the application of traversability estimation by providing faster training, convergence, and generally requiring fewer training samples [33], [34]. While faster training and convergence have already been investigated for traversability analysis in [23], an investigation of the amount of data required to transfer a traversability estimation to a new robot system is still missing.

In this work, we estimate the traversability for a robot through an environment with the objective of minimizing the traversal time, using both geometric and semantic environmental information derived from aerial imagery, by presenting a data-driven traversability estimator that can be applied to different (heterogeneous) types of robots. Furthermore, we investigate the generalizability of the proposed approach to test how well traversability estimation for new robots can be done using a pre-trained architecture. Then, an investigation of the required amount of training data to transfer a pre-trained model to a new robot system is conducted.

III. Concept

The task of the proposed framework is to predict the traversability along a path segment p_x of unit length for different types of robots. This is done by training a deep neural network on multimodal input consisting of an environment patch x_e along this path segment (as shown in [5]) and a robot's identifier x_r that is assigned to a robot. Typically, the traversability along a path is described by a cost-based representation that can be derived from a traversability objective. The entire process of predicting the traversability costs c_x for a given traversability estimate y (e.g. speed) can be described by a cost function C on y, which in turn is derived from a (learned) function f that predicts the traversability for a given robot r in environment e along path segment p_x using x_e and x_r as input:

$$c_x = C(y), \text{with } y = f(x_e, x_r). \tag{1}$$

In this work, we focus on approximating f using machine learning (ML) in two stages: a sample generator (stage one) and a predictor (stage two). An overview of the concept is shown in Figure 1. There, the concept of variational autoencoders (VAE) [35] is used to obtain a probabilistic encoder that maps the high-dimensional input $\mathbf{x} = (x_e, x_r)$ to some latent variables z. Then, the latent variables z are used as input to an ML-based predictor to estimate traversability.

More precisely, in the first stage, we aim to create a probabilistic encoder that maps the data $\mathbf{x}$ to the latent variables z. This mapping is usually shown as a probabilistic distribution $q_\theta(z|\mathbf{x})$, with θ representing its parameterization. It models the posterior distribution of the latent variables z in relation to the input data $\mathbf{x}$. The encoder can be depicted as a Gaussian distribution with mean μ and variance σ

$$q_\theta(z|\mathbf{x}) = \mathcal{N}(\mu_x, \sigma_x) \tag{2}$$

where μ_x and σ_x are the mean and variance of the Gaussian distribution determined by the encoder output for $\mathbf{x}$. In practice, the parameterization θ of the encoder is learned using a neural network that maps the input $\mathbf{x}$ to the mean vector μ_x and the covariance matrix σ_x. The learned parameters of the Gaussian distribution are then used to generate samples $z \sim \mathcal{N}(\mu_x, \sigma_x)$. To make the model differentiable and to facilitate the training of VAEs, the reparameterization trick is used. It allows sampling from $q_\theta(z|x)$ by sampling from a standard Gaussian $\mathcal{N}(0, I)$ with an identity matrix I as covariance and then transforming it using the mean and variance obtained from the encoder:

$$z = \mu_x + \sigma_x \odot \varepsilon \tag{3}$$

where ε is a vector randomly sampled from $\mathcal{N}(0, I)$.

To train the encoder, a decoder $p_\phi(\mathbf{x}|z)$, parameterized by ϕ, must be used to form a variational autoencoder (VAE). The decoder represents the probability of generating the input data $\mathbf{x}$ given the latent variables z. Again, ϕ is a parameterization that is usually learned using neural networks. The training objective for a VAE is to maximize the evidence lower bound (ELBO) [36], which can be represented as a combination of the reconstruction loss (computed from the original input $\mathbf{x}$ and the reconstruction $\hat{\mathbf{x}}$) and the Kullback-Leibler (KL) divergence [37]. The KL divergence is computed between the approximate posterior $q_\theta(z|\mathbf{x})$ (encoder) and the prior distribution over the latent variables $p(z)$, which is chosen to be the standard Gaussian $\mathcal{N}(0, I)$:

$$ELBO = E_{q_\theta(z|\mathbf{x})}[\log p_\phi(\mathbf{x}|z)] - KL(q_\theta(z|\mathbf{x}), p(z)) \tag{4}$$

The second part of the proposed framework is to generate samples from the encoder using input $\mathbf{x}$ and to estimate the traversability for multiple robots with different identifiers x_r in environment e from recorded ground truth data y^*. Therefore, a function $\hat{y} = \hat{g}(z)$ is learned that takes a single latent variable z as input and predicts the traversability estimate $\hat{y}$ for z (and consequently for $\mathbf{x}$). The function $\hat{g}$ can be approximated by training a neural network by minimizing the root mean squared error (RMSE) between the ground truth y^* and the prediction $\hat{y}$. As input, the ground truth data y^* must be recorded, which is linked to z via the environment and robot information in the encoder input $\mathbf{x} = (x_e, x_r)$.

To estimate the cost of traversal for a robot system on a path segment p_x, the framework can now be used to estimate f in two stages such that

$$c_x = C(y), \text{with } y = f(x_e, x_r) \approx \hat{g}(z) \text{ and } z \sim q_\theta(z|\mathbf{x}) \tag{5}$$

Here, $C(y)$ represents a predefined cost function for a given traversability estimate (e.g. for *speed* it can be its inverse $C(y) = y^{-1}$). The advantage of this two-step process is to take benefit of the generative capabilities of the VAE, which can be used to generate multiple samples based on the learned latent distribution. This, in turn, allows more training samples to be generated for the predictor, with the goal of reducing the required number of ground truth samples that must be laboriously collected and to improve overall performance.

IV. Experiments and Results

In this section, we present the experimental setup and process used to investigate the performance of the proposed framework. Therefore, we describe the environment, the data generation process, and the network before presenting the experimental results. To validate the framework, several experiments were conducted. First, the impact of the generative capabilities of VAEs on the overall performance is investigated. Then, the proposed approach is compared to a baseline method to investigate the usefulness of using robot

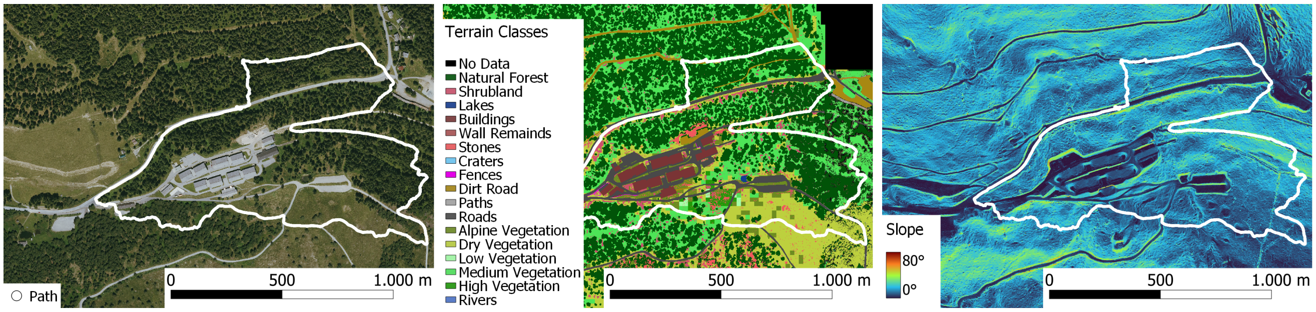

Fig. 2. Overview of the test environment. Left: orthographic overview of the experimental environment. Center: segmented environment into 17 land cover classes of which 10 are traversable. Right: elevation information on the environment represented as slope.

identification as an additional input to identify various robots and predict their traversability using a single network. Finally, the amount of additional ground truth data required to transfer a pre-trained network to a new robot is investigated.

A. Experimental Setup

The experiments were conducted on the Seetaler Alps, an alpine off-road environment and military training area in Austria. To train the encoder (stage one) of the proposed framework, environmental information retrieved from earth observation data from satellite imagery was used to generate an orthophoto O of the environment as well as an elevation map H with a resolution of $0.25m/px$. From this information, the orthophoto was manually segmented to annotate the terrain classes according to the label definitions from [38], resulting in a terrain map T with 17 terrain classes, 10 of which are typically traversable by a robot. A visualization of the environment maps is shown in Figure 2.

Furthermore, a robot identifier was added as an additional input to provide information about a robot system. For this, seven properties of a robot were chosen as identifiers, which are able to describe the robot and are known to affect the traversability of a robot [6]. The chosen properties are its dimensions (width, length, height), ground clearance, weight, maximum speed, and locomotion type. This was done instead of using a single variable as identifier in order to use information that could be potentially important when it comes to estimating a robot's traversability. It should be noted, however, that variational autocoders do not take into account correlations between data points, such as between robot properties (e.g., weight to dimensions) in the latent distribution. For this, the design of the priors would need to be adapted to allow for the training of correlations and explicit data properties, as suggested in [39], [40].

The training data for the first stage is then generated by extracting environment patches $x_e = (O_x, H_x, T_x)$ of size $2m \times 2m = 8 \times 8px$ (as described in [22]) around a random pose from the three environment maps O, H, T and assigning random robot properties x_r to them. Since extracting the environment patches and identifying the robot properties is relatively easy, this stage can be trained with a large number of training samples. This also allows learning a generalized representation that can improve performance in new environments (see Section IV-D). A total of 100.000 data samples were generated as training data and an addi-

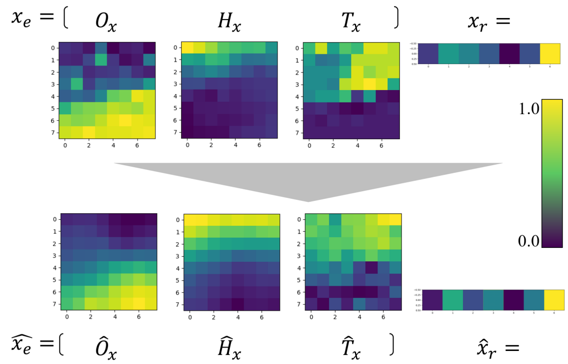

Fig. 3. Example reconstruction result from the first stage. Input data is normalized to [0,1] using the maximum values as described in [22]

tional 25.000 as test data for the first stage. The training architecture of the encoder consists of three convolutional layers for the environmental information, as used in [23], and is extended by an additional fully connected network for the robot capabilities (see Figure 1). A total of $|z| = 8$ variables were chosen for the latent representation. The network was trained with ELBO as loss function (see Section III) and managed to learn a decent reconstruction, as shown in Fig. 3.

In the second stage, the actual traversability of different robot systems is estimated based on the samples obtained from the latent space distribution. For this purpose, motion data were recorded using four different mobile robots, which differ in various ways, such as locomotion type, size, or weight. The robots used were a tracked robot (Rovo3) from Mattro[1], a wheeled skid steer robot (Husky) from Clearpath Robotics[2], a legged robot (Spot) from Boston Dynamics[3], and a wheeled robot with double Ackermann steering (Mercator) developed by TU Graz [41]. Figure 4 shows the robots used, and Table I lists some properties used as input in the proposed framework.

The data for the second stage was recorded along a predefined route (see Fig. 2) with the goal of covering realistic and diverse environment settings (different combinations of slope and land cover) that represent the diversity of the traversability in the area. The traversability objective was chosen to be the speed of a robot through the environment along a path segment p_x of length $1m$. To capture this, the robots were manually controlled by an operator who

[1] www.hawe.com/products/robot-platform/
[2] www.clearpathrobotics. com/husky-unmanned-ground-vehicle-robot/
[3] www.bostondynamics.com/products/spot

Fig. 4. The four robots used within the experiments. From left to right: Mattro Rovo3, Clearpath Robotics Husky, Boston Dynamics Spot, TU Graz Mercator.

TABLE I

SPECIFICATIONS OF THE FOUR ROBOTS USED FOR DATA RECORDING.

	Rovo3	Husky	Spot	Mercator
Dimensions ($L \times W \times H$ in m)	$1.2 \times 1.2 \times 0.5$	$1.0 \times 0.7 \times 0.4$	$1.1 \times 0.5 \times 0.7$	$1.8 \times 1.2 \times 1.2$
Weight (kg)	295	50	32	550
Ground clearance (m)	0.15	0.13	0.3	0.12
Max speed (m/s)	4.15	1.0	1.6	3.0
Locomotion Type	Skid Steering (Tracked)	Skid Steering (Wheeled)	Legged	Double Ackerman (Wheeled)

followed the robot and had the task of maneuvering the robot as fast as possible along the predefined path. The maximum speed of all four robots was limited to $2m/s$. The velocity for traversability estimation was obtained using a UKF on IMU data (Xsens MTi-G-710) and a high precision GNSS sensor with RTK (geo-konzept). To allow replication of our experiments and to encourage further research in off-road terrain analysis, the recorded dataset (with additional data recordings) is publicly available[4].

Since the robots have different outdoor capabilities, they were not able to follow the planned path exactly or were unable to access parts of the route. Thus, the exact recorded path and the number of training samples, as well as the distribution of land cover and slopes, differ slightly. For example, Spot was only able to travel a small portion of the planned route due to terrain that was too difficult for it to traverse, resulting in an abbreviated data collection. For Mercator, part of the route was blocked by a fence that could not be circumvented. In total, $10.16km$ were recorded with the robots and used to generate training samples for the predictor (stage two) by linking the recorded velocities of a path segment to its corresponding environmental sample and robot properties. This resulted in 35.218 data samples (29% Husky, 23% Rovo3, 33% Mercator, 14% Spot), which were randomly split $80/20$ into a training and a test set for each robot to train the second stage. More information about the recorded data set can be found in [24]. The test results of this training are shown in the following sections.

B. Performance Investigation of Generative Capabilities

As an initial verification of the functionality of the proposed framework, the system was trained on all four robots as described above, using one sample generated by the encoder. To evaluate the performance, the root mean square error (RMSE) and the mean absolute percentage error (MAPE) were calculated on the test results of the predictor. This resulted in an RMSE of $0.24m/s$ and a MAPE of 19.75%, confirming the overall functionality of the approach. The next

[4]https://robonav.ist.tugraz.at/data/

TABLE II

PERFORMANCE RESULTS OF THE TRAVERSABILITY PREDICTOR ON A DIFFERENT NUMBER OF GENERATED SAMPLES.

	# Generated Samples					
	1	**5**	**10**	**15**	**20**	**40**
RMSE (m/s)	0.24	0.20	0.19	0.18	0.17	**0.15**
MAPE (%)	19.75	15.55	15.18	13.73	13.36	**11.81**

TABLE III

TRAVERSABILITY PREDICTOR VS BASELINE RESULTS.

		Husky	Rovo3	Mercator	Spot	All
[22] Seperately trained	**RMSE (m/s)**	0.15	0.20	0.42	0.17	-
	MAPE (%)	12.86	19.74	30.03	11.31	-
[22] Combined No Identifier	**RMSE (m/s)**	0.19	0.29	0.54	0.20	0.36
	MAPE (%)	19.66	24.96	34.12	18.1	25.43
[22] Combined w. Identifier	**RMSE (m/s)**	0.16	0.19	0.29	0.17	0.22
	MAPE (%)	15.47	19.4	18.43	12.78	16.99
DMVAE (ours)	**RMSE (m/s)**	**0.13**	**0.15**	**0.18**	**0.11**	**0.15**
	MAPE (%)	**11.71**	**12.50**	**12.90**	**8.33**	**11.81**

step was to investigate whether the trained latent distribution of the encoder could be used to improve the prediction results by exploiting the generative capabilities of the VAE. This is done by retraining the predictor on a larger number of training samples which are generated by running the encoder multiple times on an input sample to generate multiple latent samples (z) that represents the distribution of the latent space. Table II provides an overview of the test results using a different number of generated samples.

It can be seen that when the number of training samples generated by the encoder is increased, the overall errors decrease significantly. The optimum in our experiments is achieved for the chosen maximum of 40 generated samples with an RMSE of $0.15m/s$ and a MAPE of 11.81%. This confirms the applicability of the proposed framework, especially the use of the generative capabilities of VAEs to improve the prediction performance.

C. Comparison to Baseline

A performance comparison was performed to compare the proposed approach to another state-of-the-art approach that uses patch-based environment information from aerial data to predict the traversability of a robot. Therefore, the results were compared to the approach presented in [22], which was used as basis for this work.Since the baseline is trained using only a single robot system, the following adaptations and comparisons were made. First, the baseline is trained separately on each robot type to obtain the prediction performance of each robot. Then, the training data from all four robots are combined and used to train the baseline again. To investigate whether robot identification can improve the performance of the baseline, a robot identifier was added to the baseline as a third competitor. Table III shows the test results of the performance comparisons. When the baseline was trained separately on each robot ('[22] Seperately trained'), good overall performance results were obtained for Husky, Rovo3, and Spot. The performance of Mercator is significantly worse, although no direct reason for this could be observed. When retraining the same network with all robot data combined ('[22] Combined No Identifier'), the

Fig. 5. Excerpt of the new environment information. For annotation, the same legends are used as in Fig. 2.

performance of all robots gets worse, as expected. This is due to the fact that the training samples of all four robots were recorded in the same environment, but have different ground truth values for the traversability. To see if this problem could be solved with additional information, a robot identifier containing the robots' properties was added as an additional input to the baseline method ('[22] Combined w. Identifier'). This significantly improved the overall performance of the network, and even improved the prediction performance for Mercator, which had the worst performance. This suggests that by using an robot identifier and multiple robots for training, general features can be learned that can be applied to robot systems that lack these features. When comparing all baselines to the proposed framework, our method outperforms the baselines for every single robot, highlighting the performance of the framework. However, it should be noted that the proposed approach gives similar results as the baseline '[22] Combined w. Identifier' when trained on only one generated sample (see Section IV-B).

The results highlight two main benefits of the proposed framework. First, the approach demonstrates the utility of robot identifiers in predicting the traversability of different robots within a unified network, and can even improve the overall prediction performance of individual robots by using general learned features. Second, the use of generative functionalities inherent in the proposed framework provides a valuable avenue for performance improvement.

D. Adaptability in new Environments

To investigate the generalizability of the proposed framework to new environments, a performance comparison with the baseline '[22] w. Identifier' was performed in a completely new environment. Therefore, velocity data was recorded for a single robot (Husky) in a new environment more than 40 km away from the training environment. Figure 5 shows a section of the new environment used for this evaluation. The results show that the proposed approach (RMSE $0.21 m/s$, MAPE 18.79%) outperforms the baseline (RMSE $0.29 m/s$, MAPE 27.45%), suggesting a better generalizability of the proposed framework in new environments. This may be due to the large amount of input data generated for the first stage from environmental data and robot properties, as this may allow better learning of relevant environmental features in the latent space.

E. Transfer to new Robots

To test how well the proposed approach works on new robots, a performance comparison was made between a robot trained on a pre-trained predictor ('pre-trained') and a predictor trained from scratch ('newly trained') on different amounts of training data. Therefore, each of the four robots was trained once alone from scratch and once using a pre-trained network which was trained on the data of the other three robots. This was done several times using different percentages of training samples. Figure 6 shows the average MAPE on the test data, computed from all four robots trained on the predefined network and from scratch, over a different percentage of training data. The upper and lower limits indicate the best and worst performance of the four robots. It can be seen that the pre-trained models outperform a newly trained one when using less training data, while both converge to the optimum when using the full dataset. Furthermore, it can be observed that the performance deviation of the robots from the average is smaller when using a pre-trained model, indicating a higher training stability.

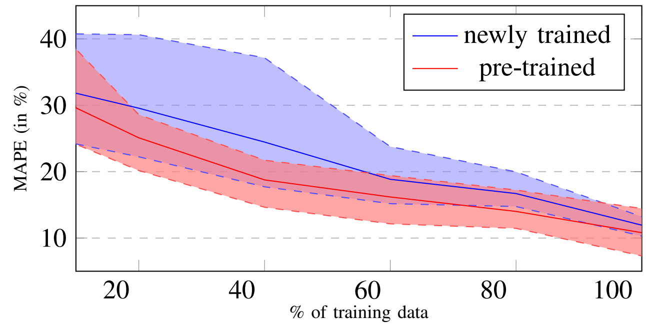

Fig. 6. Average performance results of all four robots, once trained on a pre-trained network and once trained from scratch.

V. Conclusion

This paper proposes a new framework for estimating the traversability in off-road environments for different robots using a two-stage process. Stage one aims to learn a latent distribution from environmental information and robot properties by learning a variational autoencoder on readily available data about the environment and the robots. Stage two uses the latent distribution to generate training samples to learn a traversability estimator for the different robots. It is shown that robot identifiers can be used to better train the traversability estimation for various robots within a single network, and that the proposed framework outperforms baselines by exploiting the generative capabilities of variational autoencoders. Furthermore, the framework is able to better generalize to new environments due to the large amount of training data available for the first stage. The proposed approach not only advances the state of the art in traversability estimation, but also contributes to the broader area of field robotics by contributing to the challenges of outdoor navigation. To further improve the traversability estimation and to exploit the similarities between robots, the integration of correlations between the robot properties used as input will be done using correlated variational autoencoders [39]. Moreover, the benefits of incorporating the robot description directly into the encoder of the VAE is to be investigated.

References

[1] Z. Zhu, N. Li, R. Sun, D. Xu, and H. Zhao, "Off-road autonomous vehicles traversability analysis and trajectory planning based on deep inverse reinforcement learning," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 971–977.

[2] P. Wolf and K. Berns, "Data-fusion for robust off-road perception considering data quality of uncertain sensors," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 6876–6883.

[3] B. Gao, S. Hu, X. Zhao, and H. Zhao, "Fine-grained off-road semantic segmentation and mapping via contrastive learning," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5950–5957, 2021.

[4] B. Suger, B. Steder, and W. Burgard, "Traversability analysis for mobile robots in outdoor environments: A semi-supervised learning approach based on 3d-lidar data," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 3941–3946.

[5] M. Wei and V. Isler, "Predicting energy consumption of ground robots on uneven terrains," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 594–601, 2022.

[6] M. Wermelinger, P. Fankhauser, R. Diethelm, P. Krüsi, R. Siegwart, and M. Hutter, "Navigation planning for legged robots in challenging terrain," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 1184–1189.

[7] J. Zelinka, M. Prágr, R. Szadkowski, J. Bayer, and J. Faigl, "Traversability transfer learning between robots with different cost assessment policies," in *Modelling and Simulation for Autonomous Systems*, J. Mazal, A. Fagiolini, P. Vasik, M. Turi, A. Bruzzone, S. Pickl, V. Neumann, and P. Stodola, Eds. Cham: Springer International Publishing, 2022, pp. 333–344.

[8] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational autoencoder for deep learning of images, labels and captions," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.

[9] M. Suzuki and Y. Matsuo, "A survey of multimodal deep generative models," *Advanced Robotics*, vol. 36, no. 5-6, pp. 261–278, 2022. [Online]. Available: https://doi.org/10.1080/01691864.2022.2035253

[10] C. Sevastopoulos and S. Konstantopoulos, "A survey of traversability estimation for mobile robots," *IEEE Access*, vol. 10, pp. 96 331–96 347, 2022.

[11] P. Borges, T. Peynot, S. Liang, B. Arain, M. Wildie, M. Minareci, S. Lichman, G. Samvedi, I. Sa, N. Hudson, M. Milford, P. Moghadam, and P. Corke, "A survey on terrain traversability analysis for autonomous ground vehicles: Methods, sensors, and challenges," *Field Robotics*, vol. 2, no. 1, pp. 1567–1627, 2022.

[12] M. Prágr, P. Čížek, and J. Faigl, "Traversal cost modeling based on motion characterization for multi-legged walking robots," in *2019 European Conference on Mobile Robots (ECMR)*, 2019, pp. 1–6.

[13] M. Quann, L. Ojeda, W. Smith, D. Rizzo, M. Castanier, and K. Barton, "Off–road ground robot path energy cost prediction through probabilistic spatial mapping," *Journal of Field Robotics*, vol. 37, no. 3, pp. 421–439, 2020.

[14] Z. Jian, Z. Lu, X. Zhou, B. Lan, A. Xiao, X. Wang, and B. Liang, "Putn: A plane-fitting based uneven terrain navigation framework," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 7160–7166.

[15] E. Wallin, V. Wiberg, F. Vesterlund, J. Holmgren, H. J. Persson, and M. Servin, "Learning multiobjective rough terrain traversability," *Journal of Terramechanics*, vol. 102, pp. 17–26, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022489822000313

[16] M. Lepetič, G. Klančar, I. Škrjanc, D. Matko, and B. Potočnik, "Time optimal path planning considering acceleration limits," *Robotics and Autonomous Systems*, vol. 45, no. 3, pp. 199–210, 2003.

[17] A. Wolek, E. M. Cliff, and C. A. Woolsey, "Time-optimal path planning for a kinematic car with variable speed," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 10, pp. 2374–2390, 2016.

[18] G. G. Waibel, T. Löw, M. Nass, D. Howard, T. Bandyopadhyay, and P. V. K. Borges, "How rough is the path? Terrain traversability estimation for local and global path planning," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2022.

[19] A. Jacoff, E. Messina, and J. Evans, "Performance evaluation of autonomous mobile robots," *Industrial Robot: An International Journal*, vol. 29, no. 3, pp. 259–267, 2002.

[20] Z. Jian, Z. Lu, X. Zhou, B. Lan, A. Xiao, X. Wang, and B. Liang, "Putn: A plane-fitting based uneven terrain navigation framework," 2022. [Online]. Available: https://arxiv.org/abs/2203.04541

[21] A. Broggi, E. Cardarelli, S. Cattani, and M. Sabbatelli, "Terrain mapping for off-road autonomous ground vehicles using rational b-spline surfaces and stereo vision," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 648–653.

[22] M. Eder and G. Steinbauer-Wagner, "Predicting energy consumption and traversal time of ground robots for outdoor navigation on multiple types of terrain," 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023. [Online]. Available: https://arxiv.org/abs/2308.15978

[23] J. Zelinka, M. Prágr, R. Szadkowski, J. Bayer, and J. Faigl, "Deep transfer learning of traversability assessment for heterogeneous robots," in *Proceedings of the 22nd Conference Information Technologies – Applications and Theory (ITAT 2022)*, L. Ciencialová, M. Holeňa, R. Jajcay, T. Jajcayová, F. Mráz, D. Pardubská, and M. Plátek, Eds., vol. 3226. Slovakia: CEUR-WS, 2022, pp. 12–20.

[24] M. Eder, R. Prinz, F. Schöggl, and G. Steinbauer-Wagner, "Traversability analysis for off-road environments using locomotion experiments and earth observation data," *Robotics and Autonomous Systems*, vol. 168, p. 13, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889023001331

[25] M. G. Castro, S. Triest, W. Wang, J. M. Gregory, F. Sanchez, J. G. Rogers, and S. Scherer, "How does it feel? self-supervised costmap learning for off-road vehicle traversability," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 931–938.

[26] H. Lee, J. Kwon, and C. Kwon, "Learning-based uncertainty-aware navigation in 3d off-road terrains," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 10 061–10 068.

[27] X. Cai, M. Everett, L. Sharma, P. R. Osteen, and J. P. How, "Probabilistic traversability model for risk-aware motion planning in off-road environments," 2023.

[28] R. Schmid, D. Atha, F. Scholler, S. Dey, S. Fakoorian, K. Otsu, B. Ridge, M. Bjelonic, L. Wellhausen, M. Hutter, and A.-a. Aghamohammadi, "Self-supervised traversability prediction by learning to reconstruct safe terrain," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 12 419–12 425.

[29] J. Vertens, N. Dorka, T. Welschehold, M. Thompson, and W. Burgard, "Improving deep dynamics models for autonomous vehicles with multimodal latent mapping of surfaces," 2023.

[30] B. Ivanovic, K. Leung, E. Schmerling, and M. Pavone, "Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 295–302, 2021.

[31] N. Hirose, A. Sadeghian, P. Goebel, and S. Savarese, "To go or not to go? a near unsupervised learning approach for robot navigation," 2017.

[32] S. Hosseinpoor, J. Torresen, M. Mantelli, D. Pitto, M. Kolberg, R. Maffei, and E. Prestes, "Traversability analysis by semantic terrain segmentation for mobile robots," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, 2021, pp. 1407–1413.

[33] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[34] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, 2016.

[35] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Foundations and Trends in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019. [Online]. Available: http://dx.doi.org/10.1561/2200000056

[36] ——, "Auto-encoding variational bayes," 2013. [Online]. Available: https://arxiv.org/abs/1312.6114

[37] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79 – 86, 1951. [Online]. Available: https://doi.org/10.1214/aoms/1177729694

[38] A. Valada, G. Oliveira, T.Brox, and W. Burgard, "Deep multispectral semantic scene understanding of forested environments using multimodal fusion," in *International Symposium on Experimental Robotics (ISER 2016)*, 2016. [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2016/OB16c

[39] D. Tang, D. Liang, T. Jebara, and N. Ruozzi, "Correlated variational auto-encoders," 2019. [Online]. Available: https://arxiv.org/abs/1905.05335

[40] X. Guo, Y. Du, and L. Zhao, "Property controllable variational autoencoder via invertible mutual dependence," in *International Conference on Learning Representations (ICLR)*, 2021.

[41] R. Halatschek, K. Ramanna, W. Url, and G. Steinbauer-Wagner, "Universal offroad robot platform for disaster response," in *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 11 2020, p. 6.