A Survey of Traversability Estimation for Mobile Robots

CHRISTOS SEVASTOPOULOS AND STASINOS KONSTANTOPOULOS

1Department of Computer Science and Computer Engineering, University of Texas at Arlington, Arlington, TX 76019, USA
2Institute of Informatics and Telecommunications, NCSR “Demokritos,” 15341 Agia Paraskevi, Greece

Corresponding author: Christos Sevastopoulos (christos.sevastopoulos@mavs.uta.edu)

ABSTRACT Traversability illustrates the difficulty of driving through a specific region and encompasses the suitability of the terrain for traverse based on its physical properties, such as slope and roughness, surface condition, etc. In this survey we highlight the merits and limitations of all the major steps in the evolution of traversability estimation techniques, covering both non-trainable and machine-learning methods, leading up to the recent proliferation of deep learning literature. We discuss how the nascence of Deep Learning has created an opportunity for radical improvement in traversability estimation. Finally, we discuss how self-supervised learning can help satisfy deep methods’ increased need for (challenging to acquire and label) large-scale datasets.

INDEX TERMS Mobile robots, traversability estimation, deep learning, robot perception, machine learning, data-driven.

I. INTRODUCTION

In leading-edge mobile robotics research, a wide array of outdoor navigation applications such as planetary exploration, military operations, agricultural tasks etc. entails the necessity of adapting to the conditions encountered and, in particular, to address the challenges imposed by the terrain’s contextual complexity. For every mobile robot, it is of indispensable importance to be able to identify its surroundings and translate the information perceived by its sensors to a meaningful volume of required knowledge. Subsequently, it will obtain the capacity to determine whether it can navigate in a safe while efficient manner. A vital part of autonomous navigation implies that the perceived structure of the environment has to be precisely illustrated in order to ensure whether a specific region can be traversed or not. Uneven terrains characterized by dense vegetation, foliage and potential presence of obstacles create a fundamental need to choose carefully between proprioceptive and exteroceptive sensors: Proprioceptive sensors measure values regarding the state of the robot itself (such as encoder clicks and joint angles) while exteroceptive sensors measure values from the environment (such as temperature and distance).

Traversability illustrates the difficulty of moving through a specific region and encompasses the suitability of the terrain for traverse based on its physical properties, such as slope, roughness, surface condition [1], as well as the mechanical characteristics and capabilities of the robot. Furthermore, it might also establish a bedrock for path planning algorithms since it is inevitably incorporated in terrain indices (such as terrain roughness, terrain inclination) which are of central importance when considering the optimal path [2].

Although traversability estimation was initially framed as a binary classification problem [3], currently it can be viewed through the prism of multiple classes categorization with respect to the levels of traverse facilitation. Being able to evaluate terrains’ traversability is a constitutional step towards designing a perception system [4] for such rough and rugged terrains while processing voluminous data acquired by different sensory techniques.

Deep vision expands the traversability estimation field, as it facilitates the detection of features that conventional geometry-based approaches do not have access to; that is estimating compliance from visual information. Strictly speaking, although, from a geometric point of view, the
presence of obstacles might insinuate that the paths appear to be non-traversable, a robot might still be able to penetrate a compliant obstruction such as grass, foliage etc. Therefore, it would be vital to determine whether deep learning vision techniques can enrich the environment’s perception with semantic environment information on top of geometric information. Various studies have expounded the importance of traversability analysis as a fundamental step towards motion planning. Papadakis [3] presents a wide-angle review of how the multi-sensor acquisition of input data, incorporating laser/stereo/color information etc. juxtaposed with the accurate representation of vehicle models interacting with the terrain, can lead to meaningful inference of traversability estimation for structured and unstructured environments. Kostavelis & Gasteratos [5], on the other hand, discuss various methods of extracting semantic mapping information along with their potential applications in the field of mobile robotics, including traversability assessment.

Having the opportunity to examine the traversability analysis notion from a variety of research angles, this survey paper delineates the contemporary advances in traversability estimation through the prism of deep learning techniques as well as conventional machine learning and non-trainable methods (Figure 1). Inferring terrain’s traversability from geometric information can frequently bump into limitations as a consequence of the problems’ high dimensionality while meaningful information is extracted from image data. Therefore, due to the intricacy that complex terrains portray, this article aims to illustrate how the nascence of data-driven methods such as Deep Learning, and especially Self-Supervised Learning (SSL), exhibits advances in traversability estimation scenarios. On the grounds that data-driven methods can create structured representations derived solely by the available data and thus do not require the personal expertise of a human expert when it comes to accurate labelling and careful features’ selection, it is arguable that they are favorable methods when handling large-scale data. However, both human-engineered and data-driven methods exhibit certain advantages and limitations (Figure 2).

Summarizing, we aim to present the traversability estimation problem through the following viewpoints:

- highlight important elements of certain geometric and conventional machine learning techniques
- present an overview of state of the art deep learning techniques
- discuss the challenges that arise in different learning scenarios
- point out why SSL has the qualities to be the most promising direction

II. NON-TRAINABLE METHODS

A. GRID-BASED REPRESENTATION

In order to construct an appropriate discretization of the environment that encloses the required knowledge about the places that have been traversed, sensory information is transformed using techniques such as occupancy grids, digital elevation maps, and traversability maps. Principally, the elevation map consists of a two-dimensional regular grid, where each cell stores a height value and variance. While the robot explores the environment and collects new information, this map is continuously updated and traditionally, terrain traversability would refer to grid cell traversability computation. With regards to elevation maps, they can be built by using on-board sensing such as lidars and IMUs and by exploiting the geometry of the adjacency grid [6], traversability can be determined and thereupon, initiate the ground for successful motion planning.

Similarly to the elevation map, the traversability map uses the regular grid representation. A 2D grid-based approach, using fused stereo and visual data that segments the environment into equally sized spatial cells [7], while the principal difference against the occupancy grid-map is that, in this case, each cell illustrates the traversability rather than the occupancy of the space examined. A local traversability value is assigned to each cell of the elevation map and by computing the traversability map, the traversability of a certain robot pose can be interpreted. Traversability maps’ direct and intuitive nature relies heavily on the fact that they can be built with respect to each sensors acquired findings. For instance, Wermelinger et al. [8] construct the traversability map by incorporating three fundamental terrain characteristics: slope, terrain roughness and step height. Fan et al. [9] represent the map as a collection of terrain properties (e.g., height, risk) over a uniform grid; The estimation of traversability shall encircle a number of potential threats, such as collisions, step size, tip-over, contact loss, slippage and uncertainty in sensors, along with localization errors towards creating a planning framework by addressing the problem as an Model Predictive Control (MPC) problem.

Properly identifying and calculating the values of these aforementioned topographical characteristics, together with the robot’s mechanical capacity [10], an enriched traversability elevation map can be constructed. Consequently, it facilitates dynamic exploration tasks since it is going to dictate the exact location on which e.g. a walking robot can successfully land a valid step by adjusting its position and orientation accordingly [11], [12].
Alternatively, obtaining the traversability map can be an intermediate step towards calculating robot’s control commands (steering, velocity). In particular, Xie et al. [13] build a terrain map from laser sensor findings. This map is then converted to a traversability map by assigning a Traversability Index (TI) value to each cell in the terrain map. Ultimately, through the utilization of a one-dimensional histogram (Traversability Field Histogram), the robot can efficiently navigate to its target location. Martin et al. [14] highlight through their experiments how, by exploiting the use of onboard sensors (GPS, accelerometers, gyroscopes etc.) four traversability metrics (power consumption, longitudinal slip, lateral slip and vehicle orientation) can efficiently generate traversability costmaps.

B. CONVENTIONAL COMPUTER VISION

Early Computer Vision (CV) techniques on traversability estimation relied on making predictions based on the output that obstacle detection algorithms yielded. The system presented by Thrun [15], utilizes monocular color vision data. It focuses on the appearance of individual pixels and local visual attributes, such as intensity, color, edges, and texture. The quintessence of the method presented lies upon the detection of pixels different in appearance than the ground and classifying them as obstacles i.e. meaning that the detection of any discrepancy between a single pixel that differs in appearance from the ground is classified as an obstacle. Similarly, Huertas et al. [16] elaborate on the differentiation of the boundaries that tree trunks can have against the background as decided by an edge detection algorithm.

Stereo imagery, along with color, hue, texture information can be practical in efforts of building a multi-algorithm approach [17] that is independently detecting numerous obstacles of governing characteristics such as tree trunks, water, excessive slope etc. by taking into account the terrain’s attributes. Using stereo modelling and outliers detection, Bajracharya et al. [18] build a uniform terrain mapping system that incorporates information on elevation, slope, roughness along with categorization of positive and negative obstacles. In specific, they focus on distinguishing thin structures with respect to the large amount of depth singularities and rich textural information involved, such as grass or sparse bushes, that cannot pose a genuine threat to the robot’s safety.

Based on terrain features, such as slope and roughness, Castejon et al. [19] estimate the traversability characteristics, by exploiting the power of Voronoi Diagrams to model the XY dimensions of the outdoor environment as well as creating a qualitative representation that defines the traversability model. The latter provides useful geometrical information for constructing the Digital Elevation Map that eventually contributes in discretizing the workspace and isolate the cells that offer traversability information. As a means to execute a traversability analysis in complex catacomb-like environments, Bogoslavskyi et al. [20] perform experiments on a mobile robot solely collecting input from depth images drawn by a Kinect-style sensor. They use a sequential way of extracting the traversability interpretation starting from local traversability as a result of a single depth image and afterwards, proceed with integrating all the single-image traversability estimates, into a local traversability map. The way to ensure the efficacy of their method is to perform a pixel by pixel comparison between the traversability estimates and the custom-made structures with known 3D geometry. During certain evaluation trials, it is found that the traversable regions were dependent on a specific steering of the robot and thus their method is facing serious limitations.

III. CONVENTIONAL MACHINE LEARNING

Early endeavors in incorporating sensory input with machine learning techniques determining the presence of an obstacle or taking an action based on the environment’s structure. Specifically, Dima et al. [21] present a framework governed by the use of multiple sensors such as lasers, camera and infrared imagery that contribute towards detecting humans, negative obstacles and terrain’s traversability. This is achieved through merging the strengths of three different classifiers (AdaBoost, stacked generalization, experts) on the manually collected data. Pomerleau [22] explores how the performance of a neural network-based autonomous driving improves via human demonstration. An Artificial Neural Network (ANN) receiving camera input is shown to produce more accurate levels of the output i.e steering of an autonomous vehicle by employing domain-specific knowledge during its training process.

Given the fact that unstructured environments often embody large levels of uncertainty, regression methods are engaged in the endeavor of traversability estimation. Ho et al. [23] employ the Gaussian Process (GP) Regression technique on the way to predict a planetary rover’s attitude and configuration angles by learning the vehicle response on unstructured terrain from experience. Hence, the estimation of traversability relies heavily on the formulation of the GP regression problem. Using exteroceptive data as the training input, they aim towards a direct calculation of traversability by adopting an architecture for estimating the
kernel function in order to monitor the evolution of vehicle states and propagation of uncertainty. Subsequently, GP provides a continuous representation of the terrain, and accurate estimation of traversability in areas with little or no exteroceptive data. They conclude that, combining exteroceptive with proprioceptive learning can yield more complete and more accurate traversability maps. Extending the Gaussian Process regression model, Oliveira et al. [24] implement an Uncertain-Inputs GP model that provides the usefulness of scrutinizing localization and execution noise as an effort of modeling terrain’s roughness using vibration data as an input.

Support Vector Machine (SVM) [25], a kernel-based method has been in established usage for tasks involving classification, regression, novelty detection etc. due to its ability to make classification decisions for new input vectors rather than yield probabilistic outputs [26]. Predominantly, it provides general solutions for maximizing the margin between two particular classes. In the direction of detecting road traversability, Bellone et al. [27] carefully select a feature set generated through normal vector analysis. The hypothesis made, implies that by using a normalized descriptor, that is enriched with both geometric and color data, augments the generalization of the space descriptor. Veritably, their proposed descriptor, through the use of an SVM with 4 different kernels, manages to portray higher levels of efficiency than certain standard descriptors and it can subsequently detect road traversability from point clouds acquired in outdoor environments. Zhou et al. [28] use the AdaBoost algorithm associated with Fuzzy SVM for feature selection on a 3D point cloud for the ground surface, in order to create a self-supervised visual learning for terrain surface detection in forest environments. As a means to train the classifier, a triangulated irregular network (TIN) is employed to model the ground plane and extract training points from the 3D point cloud dataset.

Alternatively, an attempt to exploit the features of combined color or depth descriptors and thus generating a textural descriptor is described by [29]. These textural, along with the color, features were accepted as training and testing inputs in the SVM classifier for performing terrain classification on areas covered with sand, grass, pavement, gravel and litterfall. Another research attempt in which textural features take a leading role on the way to derive terrain’s classification is described by Kingry et al. [30]. In particular, key features are extracted from captured visual-spectrum images and the use of an Artificial Neural Network (ANN) facilitates the way to identify terrain types of grass, concrete, asphalt, mulch, gravel, and dirt.

An instance of incrementally training an unsupervised learning scheme, where the classifier’s ultimate goal is to provide predictions about the visible terrain’s traversability is delineated by Kim et al. [31]. Autonomously collected by stereo, labeled visual features corresponding to traversable or non-traversable examples are fed to an online classifier learning algorithm that its operation aligns with the axes of 1) identifying and collecting appearance feature vectors from training examples, and 2) classifying newly collected image patches based on the learned models. The efficiency of the running mechanism articulates that labeled data gathered within a certain time window ought to be used for training as a means to fruitfully map the image patches input to the terrain map cells. Subsequently, the desired output will highlight areas classified as traversable, non-traversable or unknown. After performing experiments, it was shown that the traversability that was on-line learnt could enable the robot’s access to a goal location despite being surrounded by tall grass of high density while a conventional planner, which function counts exclusively on estimating the cost map from the elevation map computed through stereo ranging, was proven to be unsuccessful as a result of lack of representation of the feature space.

An endeavor embodying the integration of geometry and traversability en route to generating a terrain’s assessment is presented by Happold et al. [32]. In this sense, a multilayer perceptron (MLP) with one hidden layer is trained with stereo images, in a supervised manner, on eight-dimensional geometric vectors depicting features such as slope, density, height, vertical distance etc. The LAGR (Learning Applied to Ground Vehicles) platform collects stereo pairs and then the human expert labels each explored cell as low, intermediate, high, or lethal with respect to the difficulty encountered while traversing this particular cell. After gathering a total of 4000 labeled cells, it is shown that height and slope were the most important features in terms of determining terrain’s traversability. Subsequently, by integrating the classification from geometry features with color information, a cost map was generated for path planning purposes.

A. PROBABILISTIC

Thrun [15] makes a statement of fundamental gist that ’As robots are moving away from factory floors into increasingly unstructured environments, the ability to cope with uncertainty is critical for building successful robots’.

Navigation in outdoor complex environments encapsulates the need to handle the uncertainty risen as an amalgam of different factors such as sensors’ noise and error, robot’s mechanical limitations and most importantly the environment’s unpredictable nature which renders its modelling as a quite challenging task. A plethora of approaches that can represent uncertainty using probabilistic distributions and modelling has been introduced for deriving terrain traversability. In the work of Olillis et al. [33] the robot learns from human demonstration to calculate terrain costs, which indicate the probability of an obstacle’s presence. Through the combination of using Bayesian estimates and geometric information collected by stereo vision, the final terrain costs follow a certain distribution and thus it can be determined whether the path is traversable by articulating that those cells with higher values of features ought to be less traversable. Using the human interference in a similar manner but excluding any presuming correlation between the features and the traversability, in [34] training data is generated throughout the
safe journey that the human operator drives the robot throughout. Thereupon, the notion of learning is addressed with the application of the Positive Naive Bayes (NB) classifier that estimates the frequencies of observed features by finding the parameters of the probability distribution for the traversability. As being examined within the aforementioned parts of this survey, the formulation of the traversability map acts as a powerful tool in representing the path that the robot can safely pursue. Furthermore, in order to augment the accuracy in detection, Sock et al. [7] fuse, using the Bayes’ formula that combines input from independent sources to estimate a single entity, the traversability maps obtained by a visual camera and a LIDAR. The resulting map is modelled as a Markov Random Field and the cells are being independently updated.

Another regularly implemented technique which aims to autonomously improve traversability estimation capabilities in unknown terrains is the use of a self-learning framework where 3D information corresponding to a densely vegetated terrain is extracted from the point cloud and is afterwards fed, through the form of geometric features, to a geometry-based classifier [35]. The main rationale en route to estimate the ground’s traversability implies that the geometry classifier supervises a second color-based classifier and hence an iterative process that the system is retrained while new labelled data enriches the representation of the ground’s model which is constructed with the use of Gaussian Mixture Models (GMM). Leveraging the advantages that a self-learning framework offers reciprocally with the use of superpixels as visual primitives, Kim et al. [36] employ vision sensing to estimate the traversability of the terrain based on its appearance instead of using the geometric stereo vision information. Their superpixel-based approach which produces higher levels of accuracy on image region classification, involving features containing color and texture information in RGB, computes traversability using Bayes’ rule along with a modified k-nearest neighborhood (k-NN) algorithm. As a way to distinguish between known and unknown regions, i.e. frontiers of the traversability map obtained by laser scans during autonomous exploration, Tang et al. [37] make use of the reachability map that reduces the traversability map’s dimension. By enforcing the k-means clustering method on the frontier candidates, along with the use of the A* Algorithm for finding the optimal paths, the cells on the map are being labelled as reachable, dangerous or unknown. Furthermore, defining the boundaries, upper and lower, of the terrain map Fankhauser et al. [38] propose a mapping approach using proprioceptive sensing (kinematic and inertial measurements) relying on the current pose of the robot that is being constantly updated as well as the noise and uncertainty of the sensor and roll, pitch angles respectively. The gist of their method is built upon creating a robust robot-centric elevation map that generates its data through the uncertainty derived from the robot’s incremental motion in the form of mathematical equations. Although their experiments using legged robotic hardware managed to apprehend and make use of the environment’s uncertainty, their implementation seems to be facing limitations that the authors manage to address for maps of larger size along with localization singularities due to their platform-specific method. Similarly, another platform-specific effort that the proposed robot-centric mapping system aiming to derive traversability using laser-based 3D SLAM is illustrated by Droeschel et al. [39]. Using proprioceptive sensing (IMU and local odometry) along with rotating laser scanner measurements that in the surface element representation are going to be interpreted as Gaussian Mixture Model observations, a pose graph is assembled by the maps of all the adjacent key poses in the direction of successfully computing the robot’s localization. By performing graph optimization, local dense 3D maps are constructed and integrated to a global one that yields information for the robot’s real-time pose and can ultimately provide traversability costs for rough terrain navigation for each map cell.

A conjunctional viewpoint of both the terrain’s topographical properties and the robot’s kinematic/dynamic configuration that impacts its motion is explored by the work of [40]. A prediction-based terrain traversability assessment method relying heavily on the RRT algorithm is presented. By creating a reference map for prediction, the algorithm determines the path the walking robot follows upon acting on a reference map created prior experiments on rough terrains. Traversability is decided through evaluating footholds, feet trajectories and other constraints among the cells of the map.

IV. DEEP LEARNING

While conventional machine learning techniques can face serious restrictions in terms of being able to process the collected data in their initial state, Deep Learning methods offer the prospect of creating better representations and thus leading to better understanding without onerous engineering struggles. A Deep Learning asset that can go beyond conventional machine learning methods in traversability estimation scenarios is that it offers the prospect of creating implicit relationships among data. Traversability estimation has been examined from the various perspectives of unsupervised, semi-supervised, supervised and self-supervised learning. Contemporary methods, are often associated with models being trained in an end-to-end supervised fashion, as a means to simplifying the training process. Equivalently, unsupervised and semi-supervised methods that use pretrained models’ features shortly before training on a supervised dataset for a specific downstream task have been a sharing a large amount of popularity too. A common approach in deep vision traversability estimation techniques pinpointed by the latest research efforts, is to process an input RGB image through a series of pre-processing techniques and convolutional layers of a self-supervised network, before the meaningful features are fed to a traversability prediction network, i.e., a classifier (Figure 3).
A. SUPERVISED

Supervised learning has been a prevailing tool in traversability computation for rough terrains due to its applicability in predicting and classifying terrains and regions that correspond to a specific class. Especially, after obtaining a training set, input sensory data needs to be mapped to a specific target value frequently concerning whether the terrain is traversable or not.

In [41] it is investigated whether a traversability classifier learnt from synthetic heightmaps performs well on real heightmaps. Specifically, the traversability estimation problem is addressed as a heightmap classification problem and the answer to whether the examined patch is traversable or not is given by the outcome of a comparison between a standard computer-vision feature extracting technique and a CNN. As far as the deep learning architecture built, it consists of adjacent convolutional and max-pooling layers followed by a Fully Convolutional (FC) layer with 2 output neurons. It is reported that the Convolutional Neural Network (CNN) estimator outperforms the feature-based approaches both on synthetic and real-world heightmaps. Besides that, it is noted that the training on simulation could successfully get transferred to traversability maps reflecting unseen real-world terrains.

In the interest of performing long-range terrain segmentation using RGB stereo images in outdoor environments, Zhang et al. [42] design an end-to-end training deep CNN architecture aiming to augment the network’s generalization efficiency. An encoder-decoder scheme using input feature and reference maps (calculated by the disparity images) is used. In particular, the encoder includes five layer ensembles consisting of convolutional layers followed by batch normalization ones right before the ReLU activation function and the pooling layers. The decoder uses upsampling layers that perform a deconvolution operation on the input features maps before the latter being processed by the aforementioned ensemble. It is reported that the introduction of the reference maps at the 1st, 3rd, and 5th ensemble layers provided a good balance between the segmentation guidance and the noise suppression. Subsequently, the output of the decoder is fed into a multi-class softmax classifier to generate predicted labels while performing pixel-wise classification. For their experiments, the used six hand-labeled image datasets containing RGB images and the disparity maps drawn from areas of dirt, foliage, natural obstacles (trees and dense shrubs), mulch etc. Subsequently, each pixel is classified into one of three terrain classes of traversable, non-traversable, and unknown regions respectively.

An assessment of terrain traversability for performing autonomous classification of Martian terrains is explored in [46] where a framework named Soil Property and Object Classifier (SPOC) that provides pixel-by-pixel image classification into one of 17 terrain classes is constructed upon a CNN. In a supervised fashion, human experts append the imagery dataset with new classes while the Mars Rover explores different sorts of terrains. A fully differentiable and trained for 6 hours end-to-end Fully Convolutional network, including multiple stages of filtering, various CNN layer dimensions (64,128,256,512) and downsampling, is used as the understructure before an upsampled penultimate layer classifies the input raw image (orbital or ground). The output of the classification acts as the input to the cost function of the optimal route planner for landing site traversability analysis and also for building a robust slip prediction model.

A human-like inspired system trained in a supervised way involving the fusion of CNNs with decision-making process was explored in the work of Tai et al. [45] where the network’s output generates the control commands for the robot to explore an indoor environment. In terms of the deep learning model used, an input depth map is fed to a three-times repeated assemblage of convolution/activation/pooling plus a fully connected layer of five nodes each one corresponding to robot’s the possible states (going forward, 2x turning left, 2x turning right). Using a Turtlebot and a Kinect sensor, the training set comprised of indoor depth information while the ground-truth was defined by a human operator and the robot’s decision-making actions demonstrated similarities to human-inspired intelligence.

Pfeiffer et al. [43] use a global motion planner, through which the robot learns a navigation policy in a supervised manner. They feed a CNN with fused pre-processed collected laser data with the relative goal position. Their CNN
architecture encompasses two residual building blocks including shortcut connections that can address training complexity. Moreover, they train and test two CNN models in simulation (with a small differentiation in their dimensions), before expanding to a real platform that traverses an area with obstacles such as tables, chairs etc. Their results indicate that their model was not only able to learn the desired navigation strategies but also to transfer the knowledge among different unseen environments. However, some impediments occur with the rise of the environment’s complexity since it is stated that the CNN is not able to act as global path planner.

In contrast with methods that exhibit pure adherence to frames’ binary classification as traversable or not, Palazzo et al. [44] design a supervised model that can analyze multiple traversability routes through the medium of the encoder-decoder architecture. Notably, while the problem is examined as a regression one, their aim is to estimate and predict the traversability costs of various routes even on scenarios that no labels are provided. Using collected RGB images as inputs, the utilized architecture consists of a fully convolutional network module for feature extraction, followed by two layers; a convolutional and a fully-connected respectively. The bottom line of their method lies on training a model to predict correct traversability scores on the source dataset, while carrying out unsupervised domain adaptation on the target data.

Table 1 provides an overview of the papers using Supervised Learning along with the input(modality), the total number of data in the experiment (for training and validation), the learning architecture and also describes whether the proposed method is only suited to the specific platform present in the experiments. Table 3 is organized in the same fashion. Tables 2 and 4 include an additional column that describes whether the labelling process is supervised, unsupervised, etc.

### B. SELF-SUPERVISED

Self-supervised learning (SSL) is a form of supervised learning that human intervention, in terms of labeling, is not necessitated. In specific, the agent investigates a partition of unlabeled data, interprets it, and then, by developing a reliable representation, it is able to produce the labels missing and thus develop a sturdy perspective about the remaining part while automatically creating a labeled dataset. A key aspect of SSL that renders it as the contemporary most promising direction towards traversability estimation in unknown environments, is the ability to establish larger proportions of data efficiency in deep learning models that aim, as a consequence of reduced demand, for hand-labeled training data. Subsequently, it battles against the pure reliance on extensive amounts of data, and it is proven to be highly

---

**Table 1. Overview of papers using Supervised learning.**

| Reference | Input | Data Size | Architecture | Platform-specific |
|-----------|-------|-----------|--------------|------------------|
| Pfeiffer et al., 2017 [43] | 2D Laser,Position | 4.3M | CNN | Yes |
| Palazzo et al., 2020 [44] | RGB | 5k | CNN | No |
| Tai et al., 2017 [45] | RGB-D | 1104 | CNN | No |
| Zhang et al., 2018b [42] | RGB | 700 | CNN | No |
| Kong et al., 2018 [30] | RGB | 15k segments | ANN | No |
| Chavez-Garcia et al., 2018 [41] | Heightmap Patch | 65k | CNN | No |
| Bel lone et al., 2017 [27] | RGB Point Cloud | 5k | SVM | No |
| Rothrock et al., 2016 [46] | RGB | 700 | FCNN | Yes |
| Sock et al., 2016 [7] | 3D Laser, RGB | 215k | SVM | No |
| Narvaez et al., 2018 [29] | Depth,IR,Color | 820 | SVM | No |
| Oliss et al., 2007 [33] | RGB | Unspecified | Bayesian | No |
| Ho et al., 2013 [23] | RGB-D Point Clouds | 215k | GP Regression | Yes |
| Dima et al., 2004 [21] | IR,Color | 28k | Stacked Generalization | |
| Olivera et al., 2020 [24] | IMU | Unspecified | GP Regression | No |
| Gao et al., 2021 [47] | RGB, IMU, Laser | 12k | CNN | No |

**Table 2. Overview of papers using hybrid learning.**

| Reference | Input | NumOfData | Labels | Architecture | Platform-specific |
|-----------|-------|-----------|--------|--------------|------------------|
| Goh et al., 2022 [48] | Stereo | 17k | SSL,Super | CNN | No |
| Shah and Levine, 2022 [49] | RGB, GPS | Unspecified | SSL,Super | CNN | No |
| Zurn et al., 2020 [50] | Stereo,Audio | 25k | SSL,Unsuper | CNN Autoencoder | No |
| Sekar et al., 2020 [51] | RGB | 1000steps/episode | SSL, RL | CNN,RNN,MLP | No |
| Hirose et al., 2017 [52] | RGB | 78k | Super,Unsuper | GAN, FCL | No |
| Tai and Liu, 2016 [53] | RGB-D | 32images/iteration, 4k iterations | Super,RL | CNN, FCN | No |
| Happold et al., 2006 [32] | Stereo | 4k | Super,Unsuper | MLP | No |
beneficial especially for scenarios that involve updated data collections for different tasks as described by [51].

One of the first endeavors in determining long-range traversability using short-range data and self-supervision is described in [40]. In pursuance of training a vision classifier for a four-wheeled rover in a Martian-like rough terrain, the authors present two self-supervised approaches for local and distant terrain classification respectively. Short range data input, acquired both from vision and vibration sensors, creates a “local training” framework fusing texture, color and geometry information for all the encountered classes i.e. rock, sand, grass. Using the short range training, the second approach for “remote training” employs stereo processing to identify the distance to patches in the image and, by position estimation, to identify when the rover has driven over a particular patch of the terrain. Consequently, long-range data was classified with respect to the classes having previously been identified for the “local” scenario. Collecting data while the rover traverses the terrain and setting a threshold for the data points collected for the visual classifier, training was realized with an SVM classifier, and by fusing the class likelihoods of the input large image patches corresponding to traversable areas. With regards to the features extractors involved, a total of four approaches working in an interleaved fashion is presented, each one being trained with either labeled or unlabeled offline data. The efficiency reported in their results is produced by the use of a multilayer CNN that was initialized with deep belief net training, consisting of two convolutional and max-pooling between them layers, which was responsible for independently pretraining each layer in both unsupervised and supervised manners.

Significant amounts of research attention due to recent advances in the field of Deep Learning is placed on the novelty detection problem for indoor/outdoor robot navigation. A distinguishable work has been presented by Richter and Roy [55] in which the novelty detection scenario is addressed through the utilization of autoencoders irrespective of the extent of appropriate training that the robot has received. In specific, the robot repeatedly gathers training data, labels it in a self-supervised manner and feeds it to a conventional feedforward neural network consisting of three hidden layers, followed by sigmoid activation functions and a softmax output layer that predicts collisions or not. On the other end, the autoencoder, comprising of three hidden sigmoid layers and a sigmoid output layer, reconstructs similar inputs and determines whether the new images bare enough resemblance to those of the training data. In the case of detecting something novel in the environment, the robot decides towards a safe behavior. In any other case, it can just augment the array of familiar environment types. Nonetheless, instances of misclassification of novelty in images might occur due to inadequate training of the collision predictor. The authors address this incident by matching the size and architecture of the hidden layers of the two networks.

Another work, interpolating a long-range vision classifier trained in self-supervision is portrayed by Hadsell et al. [54]. The classifier’s output allow successful detection of trees, obstacles etc., by having the horizon as its perspective, and thus determining the traversability of the input large image patch patches. With regards to the features extractors involved, a total of four approaches working in an interleaved fashion is presented, each one being trained with either labeled or unlabeled offline data. The efficiency reported in their results is produced by the use of a multilayer CNN that was initialized with deep belief net training, consisting of two convolutional and max-pooling between them layers, which was responsible for independently pretraining each layer in both unsupervised and supervised manners.

Significant amounts of research attention due to recent advances in the field of Deep Learning is placed on the novelty detection problem for indoor/outdoor robot navigation.
A conjointment of supervision between an unsupervised acoustic proprioceptive classifier that self-supervises an exteroceptive visual one is explored by Zurn et al. [50]. The robot equipped with both a stereo camera and a microphone, traverses various complex terrains and collects visual (terrain patches) and audio (vehicle-terrain interaction) data respectively. This is achieved by associating the visual features of a ground’s patch right in front of the robot with the auditory features that correspond to the area that the robot is traversing. Projecting the camera images into a birds-eye-view perspective, they act as weakly labeled training data for the semantic segmentation network trained in an unsupervised manner. By teleoperating a rubber-wheeled robot, they collect visual (24 thousand images) and auditory data (4 hours of video) of five different terrains: asphalt, grass, cobblestone, parking lot, and gravel. In regard to the architecture used, the authors implement an encoder/decoder architecture for the audio data and the MobileNet V2 model (pretrained on the ImageNet dataset) for the visual feature extraction network. Moreover, for the semantic segmentation of the terrains, they adopt the AdapNet++ network with an EfficientNet backbone. Their self-supervised exteroceptive semantic segmentation model achieved a comparable performance to supervised learning with manually labeled data.

An automated self-supervised learning method and the corresponding prediction of navigation-relevant terrain properties is presented in [62]. Specifically, they conduct their experiments using the ANYmal robot [63] by measuring the interaction, during locomotion, between the robot’s sensorized feet and the terrain and then projecting the robot’s footholds into camera images. Using this foothold projection system, they annotate semantic classes in the images by assigning a semantic label to each time step in the sequence while human annotation is only implicated in observing and marking the possible transitions between terrain types with a concrete time stamp and the terrain’s type. The possible terrain types involved are asphalt, gravel path, grass, dirt, and sand. A learnt, while walking, terrain property, called ground reaction score, provides an alternative way to generate the footholds’ labels in a self-supervised way. For the semantic segmentation purposes of this study, the authors use a CNN which architecture is based on ERFNet [64]. By teleoperating the robot through different environments, they collect a dataset of 70000 training and 15000 validation images respectively. Ultimately, in addition to the application of their approach to a legged robot like ANYmal, it is mentioned that it could be deployed to other types of ground robots as well.

C. UNSUPERVISED AND SEMI-SUPERVISED

Semi-supervised and unsupervised learning provide auspicious ground for focusing on the essential segments rather than precise pixel-wise classification that require labelling of the training images.

One preceding method is described by Shneier et al. [65] using range and color information, build unsupervised models solely derived from the geometry and appearance of the scene. These models are learnt using clusters of neighboring data, extracted from the same physical region and can enclose an estimate of the traversability cost. However, due to the fact that the features learnt within the models are self-sufficient, not requiring any range data, causes the association between the region models and the traversability estimation to be uncertain for distant regions.

In recent research, a fundamental aim of the use of unsupervised learning in traversability estimation problems is to learn a particular set of features that can be transferred to the network that will subsequently be trained on a specific downstream task (such as image classification) which determines traversable areas. Transfer learning approaches normally require a large-volume training dataset (e.g., Pascal
Voc, Kitti, Imagenet) to train, and by using the pre-trained weights of a model, a classification task can yield higher levels of accuracy and abstraction.

An instance of a deep-learning-based architecture, considering the unsupervised problem as supervised, that aims to train a generative model is served by using Generative Adversarial Networks (GANs) [66]. GANs have enjoyed wide use in computer vision research over the span of the previous decade and can automatically train a generative model while using both a generative and a discriminative model.

Exploiting the notion of transfer learning, Hirose et al. [52] train a model with positive examples automatically spawned by an on-board fisheye camera attached on a Turtlebot (and an attached laptop) within a time window of 7.2 hours. With regards to the architecture presented, it consists of the conventional ensemble of the two adversarial modules i.e a generator and a discriminator which are both designed as standard CNNs that are trained in a merely supervised fashion. It is noted that only a slight annotation provided drastic improvement on the performance. On top of that, the generator receives the latent vector $z$ as an input which has been previously generated by an additional network, the inverse generator that is being trained simultaneously along with the aforementioned generator and discriminator. In order to enhance the levels of accuracy of the unsupervised method, an additional Fully Connected (FC) layer is trained in a supervised way to classify the scenes as “GO” or “NOT GO”. This linear classifier is using the GAN knowledge extracted by three specific GAN features. In order to test the network’s effectiveness, a saliency map is used to provide the meaningful segments of the images which, as it was reported, correspond to the right and left side of the input images. These segments of the images they indicate the presence of a wall or a corridor and thus portray a fundamental role in determining indoor traversability. Ultimately, it is stated that the method presented could be an imperative tool towards the effort of building cost maps.

Extending their work in performing GAN predictions for indoor traversability scenarios, Hirose et al. [67] introduce the GONet framework that uses the same aforementioned idea of semi-supervised learning incorporating a small amount of negative training data that can be proven to be more advantageous than solely including positive data, in improving traversability estimation. Indicatively, the GONet architecture consists of two models, one responsible for extracting features from positive automatically labeled examples of traversable areas extracted by a fisheye camera mounted on a robot and the second which performs the final classification after been trained on both positive and negative examples. Additionally, in order to exploit the temporal nature of the collected data, an LSTM unit that captures the temporal dependencies in the data is added, creating a new separate model named GONet+$T$, and the output it yields is further fed to a fully connected layer responsible for subsequently predicting the traversability probability. Strengthening the performance of GONet, a second extension named is introduced. GONet+TS is trained identically as the GONet+$T$ and addresses the limitations in prediction owing to environment’s structural complexity captured in stereo images. Performing indoor experiments with the TurtleBot2 platform and using the saliency map, despite the effectiveness presented in all methods performed, GONet+$T$ and GONet+TS highlighted smoother predictions in indoors traversability estimation than GONet due to the inclusion of the LSTM layer.

Utilizing GONet’s application with VUNet, a dynamic-scene view synthesis method [68], the authors present a unified system that can single out the traversable areas in the robot’s vicinity. VUNet is the result of combining two supplementary networks SNET and DNet that have the ability to model static and dynamic transformations based on robot’s actions. SN T is responsible for predicting static (S) and DNet for predicting dynamic (D) changes in the parts of the environment due to robot motion respectively. SNET’s architecture is based on the encoder-decoder scheme, and ought to the fact that the sampling procedure reuses original pixels of the input image, sharper images are generated. On the other hand, DNet is built upon a conditional adversarial network architecture. As a consequence, due to the bifold character of that synthesized approach, both static (e.g., walls, windows, stairs) and dynamic (e.g., humans) components of the environment can be predicted from different camera poses in future time steps. In order to estimate future traversability, two applications based on assisted tele-operation are introduced i.e. early obstacle detection (moving pedestrians) and multi-path future traversability estimation. As inputs to VUNet, the last two acquired images and a virtual navigation command, i.e., a linear and angular velocity are used. During the first experiment, VUNet predicts the motion of the human in the image and informs the teleoperator using warning signals and emergency stop commands. With regards to the second application, the system is able to generate virtual velocities for five different paths around the robot, and hence by predicting the images using scene view method, it can compute the traversability for each of the paths.

Using GONet and VUNet as a solid baseline, an 8-convolutional layer architecture named PoliNet [69] is trained to learn the Model Predictive Control-policy (MPC) for performing safe visual navigation of a mobile robot with mere human supervision. Concretely, by combining VUNet-360 (a variant of VUNet that uses input from a 360 camera) with the aforementioned traversability estimation network GONet, PoliNet can produce the velocity commands necessary for the robot to successfully follow a visual path in a safe manner. The control policy tries minimize the difference between an image taken from the 360 camera at time $t$ and the next sub-goal image in the trajectory. Hence, the control policy is responsible for finding the appropriate location in a way that the current image looks similar to the one of the sub-goal’s. PoliNet is trained offline before getting transferred to the online setup. Data was collected both in simulation and in the real world. With regards to the real data, the robot was
teleoperated and gathered a total of 10 a half hours of 360-camera RGB images. Although their experiments show their method to be generally robust, there were instances in which the robot was not able to circumvent large obstacles, mainly due to the fact that traversability was only considered as a soft constraint in the optimization problem.

Training GANs occasionally suffers from an array of reasons such as catastrophic forgetting [70] as well as difficulties in convergence, mode collapse and instability, due to design-related issues such as network architecture, appropriate selection of objective function, etc. [71]. For the cases of gathering data in an unsupervised manner or with scarce labels, such as an autonomous visual data collection by a mobile robot, recent advances in self-supervised contrastive learning offer the advantage of optimizing the learning capabilities of the designed model or operating in conjunction with the semi-supervised learning approach that is tailored to the downstream task that is examined. For instance, Goh et al. [48] use the popular approach of SimCLR [72] to perform a Martian terrain segmentation analysis with limited data corresponding to classes such as oil, bedrock, sand, big rock, rover and background. Using supervised contrastive learning, Gao et al. [47] manually label a set of anchor patches in their effort to efficiently create a feature representation that is able to distinguish different traversability regions. Shah & Levine [49] combine (a) the output of a heuristic model trained on teleoperated prior data using the contrastive InfoNCE loss function [73]; with (b) the output of a local traversability model towards successful path planning.

D. DEEP REINFORCEMENT LEARNING

As described by Sutton [74], in uncharted territory-where one would expect learning to be most beneficial-an agent must be able to learn from his own experience. Reinforcement learning (RL) enables a robot to autonomously discover an optimal behavior through trial-and-error interactions [75]. What differentiates the field of robotics in terms of applying reinforcement learning to, is the amount of challenges encompassed. One major arduousness is the high dimensionality and complexity of the states involved as well as the adversity in performing a complete and noise-free observation of the true state. What is more, another considerable difficulty that RL is facing in Robotics, derives from the fact that interactions between a mechanical system and its environment can harm the platform or any humans involved. However, RL can be proven to be an effective arrow in the quiver when the robot is navigating through complex and dynamic environments [76]. On top of that, conventional RL algorithms fused with Deep Learning, can handle many practical problems, where the incorporated states of the Markov Decision Process exhibit high levels of dimensionality and thus optimal policies are easier to be learnt.

An end-to-end deep reinforcement learning approach for a mobile robot navigating an unknown environment is portrayed by Tai & Liu [53] in which the inputs are raw depth images and the control commands serve as the outputs. As a means to create a feature representation, they use a CNN architecture with three convolutional layers which weights are initialized by a pre-trained model and three fully-connected layers for exploration policy learning. Since the robot is navigating in an indoor environment filled with obstacles, the feedback in terms of negative reward is obtained by the potential collision between the robot and obstacles. After training the model for many thousands of iterations, simultaneous and real-world indoor experiments demonstrated efficacy in obstacle avoidance along with improvement in the traversable areas detection. A dueling architecture (Figure 4) named D3QN was initiated in the work of Xie et al. [13] in the pursuance of obstacle avoidance using the concepts of convolution and deep Q Learning as its main foundations. It consists of a fully convolutional neural network, that outputs depth information from an RGB image which has previously been blended with additional noise and blur to adapt to real-world scenarios, followed by a deep Q network [77] that encloses a convolutional and a dueling network [78] while the main hypothesis implies that the training, taking place in simulation (Gazebo), will provide adequate knowledge to be transferred to real-world implementations. In terms of training speed, it was proven that the D3QN architecture is almost twice faster than DQN, highlighting its efficiency on obstacle avoidance scenarios and, consequently in efforts of determining traversable obstacle-free regions. Zhang et al. [79] present an architecture that accepts depth images along with the environment’s obtained elevation map and the robot’s 3D orientation as the inputs, which are fused and fed into an Advantage Actor-Critic (A3C) model [80]. Before their merging, depth and elevation map information is each passed through a four layered convolutional structure followed by a pooling layer whereas the 3D orientation is passed directly to a fully connected layer and then merged with the elevation information. All input sources are then concatenated and fed to an LSTM that can improve capturing the underlying states of a partially observed environment. Eventually, the actor and critic components consist of a fully connected layer each with the difference that in the actor part the output vector’s values are normalized by the Softmax function. Although both training and testing phases took place in a simulative 3D environment with varying levels of terrain’s traversability, their results showed that the agent sufficiently learnt to traverse different terrains towards a predefined goal location, and occasionally around non-traversable objects, with an average Deep RL decision-making time of 0.074 seconds.

An off-policy algorithm [60] acting as a powerful tool in improving the learning capabilities of an end-to-end RL approach, named BADGR (Berkeley Autonomous Driving Ground Robot), uses a self-supervised data labelling mechanism that is not built upon any human supervision or SLAM techniques in simulation. By collecting data using a random control policy, collisions are detected either by LIDAR or IMU as the robot stores the sensory observations along
with the corresponding actions taken. Events are labeled in a self-supervised manner by the collected dataset and then appended to it. Input RGB images act as the current observation that along with a future sequence of actions such as control commands, future events can be predicted. As far as the model’s architecture is concerned, the input images are fed into three convolutional and 4 fully connected layers each one followed (except for the last one) by a ReLU activation function, in order to form the initial hidden state of a recurrent LSTM unit that will handle each future action and yield the corresponding predicted future event. After deploying the BADGR system in real-world environments, it was shown that, by using only 42 hours of autonomously collected data, it could successfully traverse areas of tall vegetation and bumpy terrains.

By taking the history of proprioceptive states into account, Lee et al. [81] undertake the rough terrain traversability estimation as a temporal problem that requires a robust controller to produce the appropriate actuation. In this context, a sequential Temporal Convolutional Network (TCN), comprised of convolutional layers each followed by a ReLU activation function, uses input from joint encoders and IMU and, accordingly, implicitly learns to analyze contact and slippage events while a four-legged robot is navigating in complex terrains including those of mud, sand, snow etc. Towards this goal, they claim that direct RL techniques might not be fruitful due to large time processing and thus a teacher-student policy is selected instead. First, the teacher policy based on ground-truth knowledge concerning the interaction between the robot and the terrain, is trained on simulation. Afterwards, it supervises a student learning and the eventual student policy acts on the real robot. An additional concept introduced during the training stage, that enhances the robustness of the method is encapsulated on the adaptive nature of synthesized terrains in order for the controller to traverse them. Finally, as a means to integrate the neural network to regulate the controller, the Policies Modulating Trajectory Generators (PMTG) is employed.

Other methods include the implementation of deep inverse reinforcement learning [82] for determining off-road traversability [83]. Towards this direction, the authors propose a two-CNN structure that encompasses the vehicle kinematics in the states (2D poses) which unavoidably leads to an increase of the state-space complexity. In their experiments, data is collected from a laser scanner which is transformed to input features for a five-layered fully convolutional network structure that by recurrently applying a convolution layer for 120 and 150 times per network, the value iteration is completed with noticeable reduction of the computational burden. Their results showed improvements on safe trajectory prediction.

V. RESEARCH CHALLENGES

Natural outdoor terrains impose challenges due to the unique structure that each environment exhibits. For this reason the challenges enforced can be observed directly (uneven surfaces, terrain singularities etc.) while others can be substantially treacherous for the robot’s safety particularly in the cases of non-compliant objects that appear to be traversable, muddy terrains and unquestionably, erratic shifts in lighting and weather conditions. On the other hand, machine learning and especially deep learning techniques require an extensive amount of data, a fact that frequently poses stringent questions to the available computational power that handles this data. Despite the fact that deep learning exhibits certain advances in features learning for traversability estimation, there are certain challenges that potentially arise and align with the notion of transferability. Generally speaking, the field of traversability estimation is unavoidably affected by uncertainty in deep learning since it relies on the appropriateness of training data collected especially in stochastic and complex environments [43], [55].

A potential solution identifying the right pre-training method and the right dataset to pre-train on, frequently requires a solid perception of the scenario upon which traversability is going to be inferred. Correctly choosing, the learning method, the dataset and the model architecture to be used for pre-training, has an immediate effect on capturing the required amount of knowledge to be transferred to the downstream task.

What is more, dataset construction along with meticulously deploying the software (simulation/algorithms) to the real-world platform are two additional topics that require accuracy and robustness. Hence, manual labelling would
TABLE 4. Overview of papers that do not rely on supervision (labeled data).

| Paper          | Input                        | Data            | Labels | Architecture          | Platform-specific |
|----------------|------------------------------|-----------------|--------|-----------------------|-------------------|
| Lee et al., 2020 [81] | Joint encoders, IMU         | 400 steps/episode | RL     | CNN, MLP              | Yes               |
| Tang et al., 2019 [37]   | Map Cells                   | Unspec          | Unsuper| Encoder-Decoder       | No                |
| Hirose et al., 2019a [68]| RGB                         | 47k             | Senti  | Encoder-Decoder, MPC  | No                |
| Hirose et al., 2019b [69]| RGB                         | 10.30 hours     | Senti  | Encoder-Decoder       | No                |
| Zhu et al., 2019 [83]    | Ladar feature map           | 2.4k scene maps | RL     | CNN                  | No                |
| Hirose et al., 2018 [67] | RGB                         | 78k             | Senti  | GAN                  | No                |
| Zhang et al., 2018a [79]| Depth Orientation, Elevation maps | 208 steps/episode | RL     | CNN, LSTM            | No                |
| Xie et al., 2017 [13]    | RGB, Depth                  | 300 steps/episode | RL     | CNN                  | No                |
| Suger et al., 2015 (ML) [34]| 3D- Laser                    | Unspec          | Senti  | Naive-Bayes          | No                |
| Shneier et al., 2008 [65]| Stereo                      | Unspec          | Unsuper| Occupancy Maps, Geometric Vision | Yes             |
| Kim et al., 2006 (CV) [31]| Stereo                      | 220k patches    | Non-trainable | Geometric Vision   | No                |

TABLE 5. Dataset/Software/Hardware-related Challenges encountered in ML/DL references.

| Challenges                        | Task                                                                 | References |
|-----------------------------------|----------------------------------------------------------------------|------------|
| Computational Complexity          | Semantic Terrain Segmentation                                       | [47] [48] |
|                                  | Terrain Exploration                                                 | [42]       |
|                                  | Terrain Classification                                              | [48]       |
|                                  | Scene Traversability                                                | [52]       |
| Dataset Construction              | Terrain Exploration                                                 | [51]       |
|                                  | Terrain Classification                                              | [42]       |
|                                  | Semantic Terrain Segmentation                                       | [29] [40] |
|                                  | Navigation                                                          | [50]       |
|                                  | Collision prediction model                                           | [60]       |
| Obstacle Avoidance               | Terrain Traversability Analysis                                     | [23] [34] |
|                                  | Indoor Exploration                                                  | [45] [37] |
|                                  | Generalization                                                      | [21] [44] |
|                                  | Motion Planning                                                     | [49]       |
|                                  | Route Traversability Prediction                                     | [44]       |
|                                  | Terrain Cost Calculation                                            | [33]       |
| Localization Uncertainty          | Mapping Algorithm                                                   | [38]       |
|                                  | Terrain Traversability Analysis                                     | [24]       |
| Robust Slip Prediction model      | Terrain Classification                                              | [46]       |
| Sensory Information Extraction    | Motion Planning                                                     | [43]       |
| Simulation to Real world deployment| Collision Avoidance                                               | [13] [79] |
|                                  | Motion Planning                                                     | [43]       |

It eventually require a skilled human expert who can guide the agent in areas that terrain information is nebulous.

Therefore, we can state that the challenges encountered in traversability estimation can be a result of environment’s complexity, dataset construction as well as software/hardware-related issues. It becomes apparent that these aforementioned challenges are interconnected as, for instance, a complex and unpredictable environment directly affects the dataset collection as well as the accuracy of the software representation.

Irrespective of the technique used to infer traversability, these challenges will be a perpetual factor to consider and address in traversability estimation problems as a way to reinforce and secure the platform’s safety as well as the experiment’s fruitful outcome. Collecting the references that correspond to machine learning, deep learning and non-trainable method, we cluster the challenges that each paper’s method is facing with respect to the challenge’s type (environment-related and the Dataset/Software/Hardware-related ones) Tables 5, 6, and 7 present an overview of the challenges that researchers need to address while applying different learning methods in outdoor (and merely indoor) traversability applications. To enrich the presented information, we include the specific task that each of the articles is referring to.

After quantifying some information extracted from Tables 5, 6, and 7, the following pie charts depicted in Figures 5, 6 and 7 are drawn. For Figures 6 and 7
we demonstrate the frequency of specific occurring challenges for both categories (Environment & dataset/software/hardware).

Additionally, in Figure 5 we notice that challenges related to the environment’s complexity as well as the inadequacy of computational resources have been predominantly faced...
in non-trainable methods. This fact is in consensus with the frequencies of occurrence of such challenges in trainable methods as well.

VI. CONCLUSION
An abundant number of robotic applications in both outdoor and indoor environments constitutes the concept of traversability estimation a central task for safe and successful navigation through motion planning. As different environments are governed by various levels of stochasticity, the efforts to collect and interpret data from various sensor modalities often reveal further challenges due to the type and the volume of the data acquired. In this survey, we examine how the acquisition of sensory data, in various trainable and non-trainable methods, has been interpreted towards deriving useful conclusions for outdoor and indoor traversability estimation experiments.

Traditional geometric methods representing the environment in 3D have been used along with bespoke algorithms to identify obstacle and relevant pixels in the scenes encountered. Frequently witnessed, early traversability estimation computer vision techniques used certain assumptions in order to make predictions based on the outputs of obstacle detection algorithms. This led to limitations regarding long-distance perception, deformable ground traversability and generally speaking, scientific approaches were more conservative in their implementations. Afterwards, conventional machine learning methods examined the use of merging multiple sensor modalities along with probabilistic approaches that can model environment’s uncertainty. However, features’ engineering is difficult and time-consuming for the human expert. Due to the progression of the algorithmic learning techniques, different types of learning are associated with different types of encountered challenges.

Extracting meaningful conclusions in relation to the traversability of the scene has recently been associated, on the grounds that deep learning techniques are widely popular nowadays, with the size of collected data. Powerful deep learning algorithms often require enormous datasets in order to adequately train the weights of the model making headway for multiple downstream tasks, such as image classification, semantic segmentation etc. By the same token, due to the difficulty to create models that demonstrate high-level representations, the need to generate accurate labels about the terrain or the surrounding obstacles, while addressing the various challenges arising during data collection, is closely correlated to the type of the learning technique that is implemented. This survey documents an assortment of preceding and present-day learning techniques whilst it distinguishes the impact of employing Self-Supervised Learning techniques, in current traversability estimation tasks. In proximity to the Self-Supervised Learning techniques, contemporary and future research endeavors target the use of transfer learning or representation-learning methods as a propitious way to build data-efficient and robust traversability learning frameworks while truncating the highly cumbersome and vigilant tasks of manual labelling by a human expert.

REFERENCES
[1] H. Seraji, “Traversability index: A new concept for planetary rovers,” in Proc. IEEE Int. Conf. Robot. Autom., May 1999, pp. 2006–2013.
[2] G. Ishigami, K. Nagatani, and K. Yoshida, “Path planning and evaluation for planetary rovers based on dynamic mobility index,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Sep. 2011, pp. 601–606.
[3] P. Papadakis, “Terrain traversability analysis methods for unmanned ground vehicles: A survey,” Eng. Appl. Artif. Intell., vol. 26, no. 4, pp. 1373–1385, 2013.
[4] D. Langer, J. K. Rosenblatt, and M. Hebert, “A behavior-based system for off-road navigation,” IEEE Trans. Robot. Autom., vol. 10, no. 6, pp. 776–783, Dec. 1994.
[5] I. Kostavelis and A. Gasteratos, “Semantic mapping for mobile robotics tasks: A survey,” Robot. Autom. Syst., vol. 66, pp. 86–103, Apr. 2015.
[6] Y. Pan, X. Xu, Y. Wang, X. Ding, and R. Xiong, “GPU accelerated real-time traversability mapping,” in Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO), Dec. 2019, pp. 734–740.
[7] J. Sock, J. Kim, J. Min, and K. Kwak, “Probabilistic traversability map generation using 3D-LIDAR and camera,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2016, pp. 5631–5637.
[8] L. Wellhausen, R. Ranftl, and M. Hutter, “Safe robot navigation via multi-modal anomaly detection,” IEEE Robot. Autom. Lett., vol. 5, no. 2, pp. 1326–1333, Apr. 2020.
[9] D. D. Fan, K. Otsu, Y. Kubo, A. Dixit, J. Burdick, and A.-A. Agha-Mohammadi, “STEP: Stochastic traversability evaluation and planning for risk-aware off-road navigation,” in Robotics: Science and Systems XVII. RSS Foundation, 2021.
[10] S. Li, R. Song, Y. Zheng, H. Zhao, and Y. Li, “Rugged-terrain traversability analyzing for quadroper robots,” in Proc. 2nd Int. Conf. Intell. Robotic Control Eng. (ICRC), Aug. 2019, pp. 1–6.
[11] D. Kim, D. Carballo, F. D. Carlo, B. Katz, G. Bleidt, B. Lim, and S. Kim, “Vision aided dynamic exploration of unstructured terrain with a small-scale quadruped robot,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2020, pp. 2464–2470.
[12] V.-G. Loc, I. M. Koo, D. T. Tran, S. Park, H. Moon, and H. R. Choi, “Improving traversability of quadruped walking robots using body movement in 3D rough terrains,” Robot. Auto. Syst., vol. 59, no. 12, pp. 1036–1048, Dec. 2011.
[13] L. Xie, S. Wang, A. Markham, and N. Trigoni, “Towards monocular vision based obstacle avoidance through deep reinforcement learning,” in Proc. Robot., Sci. Syst. Workshop, New Frontiers Deep Learn. Robot., Boston, MA, USA, Jul. 2017, pp. 1–6.
[14] S. Martin, L. Murphy, and P. Corke, “Building large scale traversability maps using vehicle experience,” in Experimental Robotics (Springer Tracts in Advanced Robotics), vol. 88, Quebec, QC, Canada: Springer, 2013.
[15] P. Thrun, “Probabilistic robotics,” Commun. ACM, vol. 45, no. 3, pp. 52–57, 2002.
[16] A. Huertas, L. Matthies, and A. Rankin, “Stereo-based tree traversability analysis for autonomous off-road navigation,” in Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION), Jan. 2005, pp. 210–217.
[17] A. Rankin, A. Huertas, and L. Matthies, “Evaluation of stereo vision obstacle detection algorithms for off-road autonomous navigation,” JPL, Pasadena, CA, USA, Tech. Rep. 05-1610, Jun. 2005. [Online]. Available: http://hdl.handle.net/2014/38362
[18] M. Bajracharya, J. Ma, M. Malchano, A. Perkins, A. A. Rizzi, and L. Matthies, “High fidelity day/night stereo mapping with vegetation and negative obstacle detection for vision-in-the-loop walking,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Nov. 2013, pp. 3663–3670.

[19] C. Castejon, B. L. Boada, D. Blanco, and L. Moreno, “Traversable region modeling for outdoor navigation,” J. Intell. Robotic Syst., vol. 43, nos. 2–4, pp. 175–216, Dec. 2005.

[20] I. Bogoslavskiy, O. Vysotska, J. Serafin, G. Grisetti, and C. Stachniss, “Efficient traversability analysis for mobile robots using the Kinect sensor,” in Proc. Eur. Conf. Mobile Robots, Sep. 2013, pp. 158–163.

[21] C. S. Dima, N. Vardapet, and M. Hebert, “Classifier fusion for outdoor obstacle detection,” in Proc. IEEE Int. Conf. Robot. Autom., (ICRA), Apr. 2004, pp. 665–671.

[22] D. A. Pomerleau, “Progress in neural network-based vision for autonomous robot driving,” in Proc. Intell. Vehicles Symp., 1992, pp. 391–396.

[23] K. Ho, T. Peynot, and S. Sukkarieh, “Traversability estimation for a planetary rover via experimental kernel learning in a Gaussian process framework,” in Proc. IEEE Int. Conf. Robot. Autom., May 2013, pp. 3475–3482.

[24] R. Oliveira, L. Ott, V. Guzzilini, and F. Ramos, “Bayesian optimisation for safe navigation under localisation uncertainty,” in Robotics Research (Springer Proceedings in Advanced Robotics), vol. 10, Cham, Switzerland: Springer, 2020.

[25] M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning. Cambridge, MA, USA: MIT Press, 2018.

[26] C. M. Bishop, Pattern Recognition and Machine Learning. Singapore: Springer, 2006.

[27] M. Bellone, G. Reina, L. Caltagirone, and M. Wahde, “Learning traversability from point clouds in challenging scenarios,” IEEE Trans. Intell. Transp. Syst., vol. 25, no. 1, pp. 296–305, Jan. 2018.

[28] S. Zhou, J. Xi, M. W. McDaniel, T. Nishihata, P. Slaesser, and K. Iagnemma, “Self-supervised learning to visually detect terrain surfaces for autonomous robots operating in forested terrain,” J. Field Robot., vol. 29, no. 2, pp. 277–297, Mar. 2012.

[29] F. Yandun Narváez, E. Gregorio, A. Escolá, J. R. Rosell-Polo, M. Torres-Torriti, and F. Aaust Cheein, “Terrain classification usingToF sensors for the enhancement of agricultural machinery traversability,” J. Terramech., vol. 76, pp. 1–13, Apr. 2018.

[30] N. Kiniry, M. Jung, E. Derse, and R. Dai, “Vision-based terrain classification and solar irradiance mapping for solar-powered robotics,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Oct. 2018, pp. 5834–5840.

[31] D. Kim, J. Sun, S. Min Oh, J. M. Rehg, and A. F. Bobick, “Traversability classification using unsupervised on-line visual learning for outdoor robot navigation,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2006, pp. 518–525.

[32] M. Hapold, M. Ollis, and N. Johnson, “Enhancing supervised terrain classification with predictive unsupervised learning,” in Robotics: Science and Systems VI. Cambridge, MA, USA: MIT Press, Aug. 2006.

[33] M. Ollis, W. H. Huang, and M. Hapold, “A Bayesian approach to imitaiton learning for robot navigation,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Oct. 2007, pp. 709–714.

[34] B. Sugér, B. Steder, and W. Burgard, “Traversability analysis for mobile robots in outdoor environments: A semi-supervised learning approach based on 3D-lidar data,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2015, pp. 3941–3946.

[35] G. Reina and A. Milella, “Towards autonomous agriculture: Automatic ground detection using trinocular stereovision,” Sensors, vol. 12, no. 9, pp. 12405–12423, Sep. 2012.

[36] D. Kim, S. M. Oh, and J. M. Rehg, “Traversability classification for UGV navigation: A comparison of patch and superpixel representations,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Oct. 2007, pp. 3166–3173.

[37] Y. Tang, J. Cai, M. Chen, X. Yan, and Y. Xie, “An autonomous exploration algorithm using environment-robot interacted traversability analysis,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Nov. 2019, pp. 4885–4890.

[38] F. Funkhauser, M. Blesch, and M. Hutter, “Probabilistic terrain mapping for mobile robots with uncertain localization,” IEEE Robot. Autom. Lett., vol. 3, no. 4, pp. 3019–3026, Oct. 2018.

[39] D. Droeschel, M. Schwarz, and S. Behnke, “Continuous mapping and localization for autonomous navigation in rough terrain using a 3D laser scanner,” Robot. Auto. Syst., vol. 88, pp. 104–115, Feb. 2017.

[40] C. A. Brooks and K. D. Iagnemma, “Self-supervised classification for planetary rover terrain sensing,” in Proc. IEEE AeroS. Conf., Mar. 2007, pp. 1–9.
[65] M. Shneier, T. Chang, T. Hong, W. Shackleford, R. Bostelman, and J. S. Albus, “Learning traversability models for autonomous mobile vehicles,” Auto. Robots, vol. 24, no. 1, pp. 69–86, Jan. 2008.

[66] I. Goodfellow, P. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Commun. ACM, vol. 63, no. 11, pp. 139–144, 2020.

[67] N. Hirose, A. Sadeghian, M. Vázquez, P. Goebel, and S. Savarese, “GONet: A semi-supervised deep learning approach for traversability estimation,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Oct. 2018, pp. 3044–3051.

[68] N. Hirose, A. Sadeghian, F. Xia, R. Martín-Martín, and S. Savarese, “VUNet: Dynamic scene view synthesis for traversability estimation using an RGB camera,” IEEE Robot. Autom. Lett., vol. 4, no. 2, pp. 2062–2069, Apr. 2019.

[69] N. Hirose, F. Xia, R. Martín-Martín, A. Sadeghian, and S. Savarese, “Deep visual MPC-policy learning for navigation,” IEEE Robot. Autom. Lett., vol. 4, no. 4, pp. 3184–3191, Oct. 2019.

[70] H. Thanh-Tung and T. Tran, “Catastrophic forgetting and mode collapse in GANs,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–10.

[71] D. Saxena and J. Cao, “Generative adversarial networks (GANs) challenges, solutions, and future directions,” ACM Comput. Surv., vol. 54, no. 3, pp. 1–42, 2021.

[72] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 1597–1607.

[73] A. Van Den Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” 2018, arXiv:1807.03748.

[74] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, Cambridge, MA, USA: MIT Press, 2012.

[75] J. Kober, J. A. Bagnell, and J. Peters, “Reinforcement learning in robotics: A survey,” Int. J. Robot. Res., vol. 32, no. 11, pp. 1238–1274, 2013.

[76] A. S. Polydoros and L. Nalpanidis, “Survey of model-based reinforcement learning: Applications on robotics,” Int. J. Intell. Robot. Syst. Theory Appl., vol. 86, no. 2, pp. 153–173, May 2017.

[77] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing Atari with deep reinforcement learning,” in Proc. NIPS Deep Learn. Workshop, 2013, pp. 1–9.

[78] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, “Dueling network architectures for deep reinforcement learning,” in Proc. Int. Conf. Mach. Learn. (ICML), 2016, pp. 1995–2003.

[79] K. Zhang, F. Niroui, M. Ficocelli, and G. Nejat, “Robot navigation of environments with unknown rough terrain using deep reinforcement learning,” in Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR), Aug. 2018, pp. 1–7.

[80] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, “Asynchronous methods for deep reinforcement learning,” in Proc. Int. Conf. Mach. Learn. (ICML), 2016, pp. 1928–1937.

[81] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, “Learning quadrupedal locomotion over challenging terrain,” Sci. Robot., vol. 5, no. 47, Oct. 2020, Art. no. eabc5986.

[82] M. Wulfmeier, P. Ondruska, and I. Posner, “Maximum entropy deep reinforcement learning,” 2015, arXiv:1507.04888.

[83] 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 Proc. IEEE Intell. Vehicles Symp. (IV), Oct. 2020, pp. 971–977, doi: 10.1109/IV47402.2020.9304721.

[84] D. Belter, P. Łabecki, and P. Skrzypczyński, “Adaptive motion planning for autonomous rough terrain traversal with a walking robot,” J. Field Robot., vol. 33, no. 3, pp. 337–370, 2016.

[85] C. Ye, “Navigating a mobile robot by a traversability field histogram,” IEEE Trans. Syst., Man, Cybern., B (Cybern.), vol. 37, no. 2, pp. 361–372, Apr. 2007.

[86] I. Ulrich and I. Nourbakhsh, “Appearance-based obstacle detection with monocular color vision,” in Proc. 17th Nat. Conf. Artif. Intell., 12th Conf. Innov. Appl. Artif. Intell. (AAAI/IAAI), Jul. 2000, pp. 866–871.

* * *