

## Deep Learning Point Cloud Odometry: Existing approaches and Open Challenges

Bernardo Teixeira<sup>1</sup>, Hugo Silva<sup>2</sup>

<sup>1</sup>INESC TEC-Institute for Systems and Computer Engineering, Technology and Science - Associate Laboratory – INESC TEC, Faculty of Engineering campus, Rua Dr. Roberto Frias, Building I, 4200-465 Porto, Portugal; Doctoral Program in Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal ([bernardo.g.teixeira@inesctec.pt](mailto:bernardo.g.teixeira@inesctec.pt)) ORCID [0000-0002-3377-0023](https://orcid.org/0000-0002-3377-0023); <sup>2</sup>INESC TEC-Institute for Systems and Computer Engineering, Technology and Science - Associate Laboratory – INESC TEC, Faculty of Engineering campus, Rua Dr. Roberto Frias, Building I, 4200-465 Porto, Portugal ([hugo.m.silva@inesctec.pt](mailto:hugo.m.silva@inesctec.pt)) ORCID [0000-0002-8133-7216](https://orcid.org/0000-0002-8133-7216)

### Abstract

Achieving persistent and reliable autonomy for mobile robots in challenging field mission scenarios is a long-time quest for the Robotics research community. Deep learning-based LIDAR odometry is attracting increasing research interest as a technological solution for the robot navigation problem and showing great potential for the task.




In this work, an examination of the benefits of leveraging learning-based encoding representations of real-world data is provided. In addition, a broad perspective of emergent Deep Learning robust techniques to track motion and estimate scene structure for real-world applications is the focus of a deeper analysis and comprehensive comparison.

Furthermore, existing Deep Learning approaches and techniques for point cloud odometry tasks are explored, and the main technological solutions are compared and discussed.

Open challenges are also laid out for the reader, hopefully offering guidance to future researchers in their quest to apply deep learning to complex 3D non-matrix data to tackle localization and robot navigation problems.

**Author Keywords.** Deep Learning, Odometry, Motion Estimation, Point Cloud, LIDAR.

**Type:** Research Article

 Open Access  Peer Reviewed  CC BY

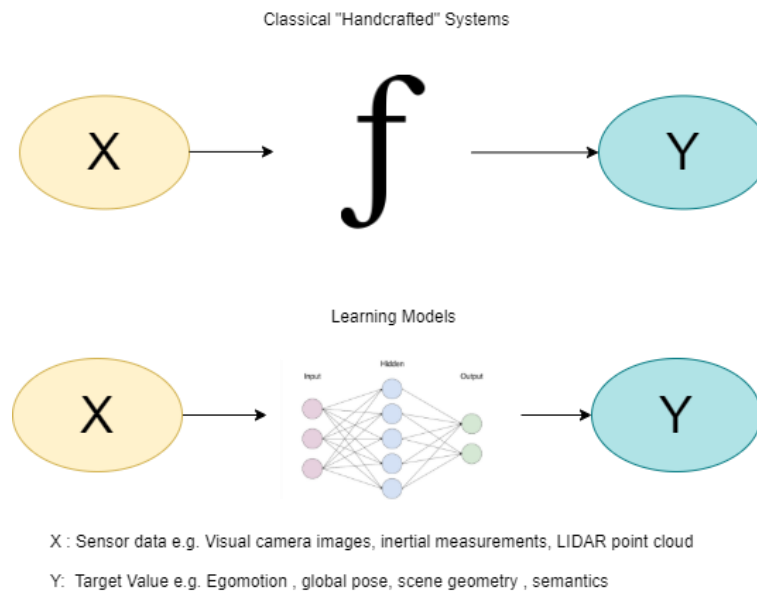
### 1. Introduction

The tasks of localization and scene mapping are a fundamental prerequisite for both human and mobile robot locomotion. As an example, human perception of their surroundings and self-motion is closely tied to their multimodal sensory perception, and that enables us to locate and navigate in complex three-dimensional (3D) scenarios. If the perception component of the human body was lacking, it would severely hinder cognition and motor control.

In similar fashion, mobile robots must always be able to perceive their environment and estimate their internal system state based on their on-board sensorization, because otherwise it would be impossible to properly develop safe and reliable robotic control systems and by association other higher-level tasks such as path planning and/or object avoidance.

The proliferation of robotic agents in our current society, whether it would be in the form of self-driving vehicles, delivery drones or home service robots, is highly dependent on the

evolution of reliable sensing data processing that can lead to safe autonomous decision-making. When it comes to mobile robots, enabling a high level of autonomy requires a high degree of precision and robustness in localization, as well as incrementally building and maintaining a world model (i.e., the travelled trajectory or a full-fledged map of their surroundings), with the capability to continuously process new information and adapt to various scenarios.



**Figure 1:** Comparison of classical system vs novel learning models

The problem of localization has been widely studied and researched. However, typically the best performing solutions rely heavily on intricate hand-crafted and manually tailored systems. Across a wide variety of necessary subtasks for achieving robust robot navigation (i.e., odometry estimation, place recognition or global localization) the better performing solutions are usually the product of closed-form mathematical solutions that are then finetuned to tackle some typical issues that may occur.

The robotics research field presents different challenges to deep learning approaches when compared to computer vision. A robot is an agent that acts in and interacts with others within a physical real-world environment. It perceives the world with its different sensors, builds a coherent representation of the environment, and updates this model over time. Ultimately, a robot must make decisions, plan actions, and execute these actions to fulfill a task anchored on this perception it constructs of the environment.

For robotic vision, perception is only one part of a more complex and goal-driven system. The immediate outputs of Robotic vision (whether it is object detection, segmentation, depth and/or pose estimation, etc.), will ultimately result in actions in the real world. In a simplified comparison, whereas computer vision takes images or point-cloud data and translates them into useful information, robotic vision translates images into actions or interactions with the environment.

### **1.1. Learning-based approach vs classical system**

As reported in [Figure 1](#), classical system design focus on hand-designing algorithms that given a certain input X can calculate the output Y to a high degree of accuracy. However, unforeseen environment conditions or imperfect sensor measurements may negatively impact these handcrafted models and prevent them to function as intended. In addition, it is not realistic to expect a hand-crafted system to be able to handle all types of complex environmental

dynamics without working under some assumptions that may impose unrealistic constraints that impact both the accuracy and reliability of such hand-crafted systems.

Data-centric methods, on the other hand, instead place the onus of accuracy and reliability on the data itself. Learning approaches allow the algorithm to learn to construct a function that maps the inputs  $X$  (e.g., visual, inertial, LIDAR data or other sensors) to the outputs  $Y$  (e.g., displacement, orientation, scene geometry or semantics).

Learning-based methods leverage the automatic feature detection and powerful feature space representation offered by convolutional neural networks, that allows for finding a lot of task-relevant cues in the scene. As an added bonus, neural network feature representations are usually way more robust to environment conditions such as lightning changes, motion blur, imperfect camera calibration or visual artifacts, which are extremely hard to model by hand and there are no straight known mathematical formulations to mitigate their negative effect. However, this great potential of learning-based approaches can only be unlocked if certain limitations can be overcome (Sünderhauf et al. 2018).

	Classical systems	Learning-based systems
Representation	Explicit: physics-based model of the environment, mathematical formulation	Implicit: neural architectures, feature space representation
Generality	Broadly applicable to different scenarios, though sometimes unable to work under different operational conditions	Risk of overfitting to training scenario settings
Robustness	Small basin of convergence: requires good models and estimates thereof	Large basin of convergence: highly robust in trained regime
Data	Only needed for system identification	Requires abundance of data to develop a working model
Computational efficiency	Overall dependent of task complexity	After having a trained model, powerful and efficient representation

**Table 1:** Classical systems vs learning-based systems properties

### 1.2. Survey relevance

There are several survey papers that have extensively discussed model-based localization and mapping approaches in the context of robotic navigation. Most notably, Cadena et al. (2016) is one of the most comprehensive compilations of the existing approaches of robot navigation and Simultaneous Localization and Mapping (SLAM). However, and despite a brief mention of deep learning models, it does provide much when it comes to overviewing deep learning research, especially because it predates the explosion of research in this field in the last 3- 5 years. Other renown works are more focused in specific parts of the problem such as the probabilistic formulation (Thrun 2000) or visual odometry (Scaramuzza and Fraundorfer 2011). When it comes to leveraging Deep Learning for Robotic tasks, there are two important works to mention: Li, Wang, and Gu (2018) explore the transition from geometric model-based to data-driven approaches by providing a comprehensive technical review of the underlying benefits and motivations and, more recently, Chen et al. (2020) provided a more extensive compilation of the use of learned models for all components of spatial navigation, inclusively coining the term “Spatial machine Intelligence”. A discussion of the limits and the incredible potential of Deep Learning in the broad context of robotics is provided in Sünderhauf et al. (2018), underlining some of the important research directions to overcome the current technological limitations and challenges.

This work aims to be narrower in scope, limiting itself to study only LIDAR-based odometry estimation and choosing to focus on the approaches that leverage deep learning as a tool for performing this challenging task. It is worth keeping in mind that this is a new and emerging research field, and this cutting edge nature (most significant advances were produced in the

last 2 years) naturally means the literature resources are still scarce and lack great benchmark comparisons.

Notably, although the problem of localization and odometry falls is fundamentally part of robotics research, the incorporation of learning methods for such tasks is a cross-disciplinary effort that involves research areas such as machine learning, data science and computer vision. Thus, the relevance of compiling a brief survey is magnified, since the natural community barriers are often blurred, and the information is more dispersed.

### **1.3. Article organization**

The remainder of this article is organized as followed: Section 2 introduces background knowledge on deep learning and how it has been being used as a tool for solving perception tasks, delving deeper on deep learning (DL) itself and the huge advances on learning-based models for Computer Vision; Section 3 compares existing approaches to DL-based LIDAR odometry estimation, zooming in on the differences between these novel approaches; Section 4 offers a summary of open challenges and research opportunities pertaining to this topic; and finally, Section 5 wraps up the paper with concluding remarks.

## **2. Background Knowledge**

In recent years, deep learning methods have risen to predominance by showing good capability for cognitive and perceptual tasks in computer vision applications, whether at analyzing unknown features, capturing image depth or even perceiving egomotion between image frames. Thus, the development of learning-based applications aimed at improving visual-based robotic navigation has had a significant surge as of late, with plenty of new literature, new methods and techniques that are incrementally improving upon the accuracy and cross-scenario robustness for visual robot navigation tasks.

However, the onset and proliferation of Deep Learning methods is mostly tied to Computer Vision applications. This is because the vast majority of deep learning techniques are performed on Euclidean data, that is represented in 1-dimensional or 2-dimensional structures, typically in a grid-like topology. However, for Robotic systems, there usually is sensor equipment able to capture rich underlying information about the scene that outputs sparse 3-dimensional (3D) data and being able to adequately process point cloud data structures to capture as much useful information as possible has the potential to be a giant leap for robotic perception capability. This presents a whole bunch of different new challenges when compared to visual image data that current research is only now trying to address.

### **2.1. Challenges on 3D matrix data**

Robotic applications are progressively being equipped with more 3D perception sensors that provide point-cloud representations of the environment. Thus, it is important to take a closer look at the specific challenges non-matrix data poses to the use deep neural network architectures. Applying deep learning on 3D point cloud data comes with many different new challenges. For instance: occlusions which can happen in cluttered scenes; noise/outliers which cause the appearance of unintended points and/or points misalignment. However, there are more practical considerations that introduce more pronounced challenges when it comes to application of deep learning on point cloud data which can be categorized into the following:

- **Irregularity:** Point cloud data is highly irregular, i.e., the 3D points are not evenly sampled across the different regions of an object/scene, so some regions could have a

denser point concentration while most patches have sparsely constructed point clouds.

- **Unstructured:** Point cloud data is not displayed on a regular grid. Each point is scanned independently and its distance to the neighboring points is not always fixed or fully known. In contrast, pixels in images are represented on a 2D grid, and the spacing between two adjacent pixels is always known and fixed.
- **Unorderdness:** A point cloud of a scene, regardless of representation, is obtained by acquiring data around the objects in the scene and is usually stored as a list in a file. That means there often is no implicit order on the point set, introducing ambiguity whereas there are multiple possible point cloud forms for representing the same scene.

These properties of point cloud are very challenging for some deep learning techniques, especially for Convolutional Neural Networks (CNN's). This is because convolutional neural networks are more well suited to work with ordered, regular and on a structured grid data. Early approaches overcome these challenges by converting the point cloud into a structured grid format, i.e., projecting the point cloud into some sort of image structure using cylindrical projection, spherical projection or 2D panoramic view projection. In recent years, researchers have been working have been developing approaches that directly use deep learning on raw point cloud data to great success which can unlock the full potential of deep learning as a technology for scene perception and understanding from point cloud data.

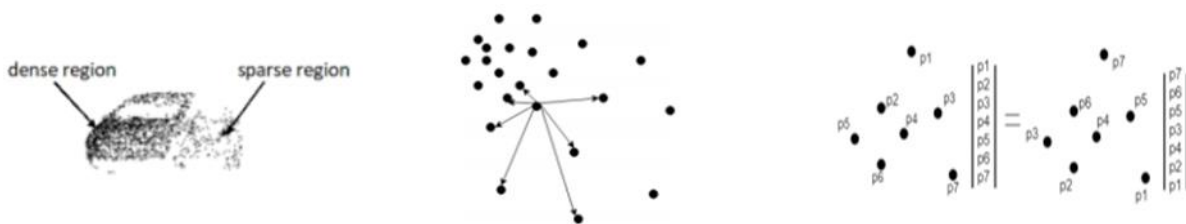


Figure 2: Illustration of the classical challenges of working with point cloud data.  
Adapted from Bello, Yu, and Wang (2020)

### 3. Comparison of Existing Approaches

LIDAR sensors are a great option for mobile agents to perceive their surroundings, with the added bonus of allowing to detect the 3D scale of the world, which visual cameras cannot directly obtain. Classical LIDAR odometry frameworks are very competitive for the motion estimation task, especially because they do not suffer as much from the inaccurate depth prediction and scale drift as visual camera information. Its performance though, is also very sensitive to point cloud registration errors caused by non-smooth motion. In addition, the data quality of LIDAR measurements is also highly affected by weather conditions such as rain or fog.

Geometry-based methods like point-to-point Iterative Closest point (ICP), point-to-plane or GICP (Segal, Haehnel, and Thrun 2010) were designed to solve the point cloud registration problem, but usually the odometry estimation task requires an extra sensitivity to tackle. LOAM (Zhang and Singh 2017) for instance, has long been considered the state-of-the-art LIDAR motion estimation framework. It works by extracting the line and plane features in LIDAR data and saving these features to the map for edge-line and plane-surface matching. LOAM achieves low-drift and good performing real-time odometry estimation by having two modules running in parallel: accurate motion estimation and mapping. The estimated motion of scan-to-scan registration is used to correct the distortion of point clouds and guarantee the real-time performance of registration while simultaneously the odometry outputs are

optimized jointly with the map. Given the success of LOAM, deep learning approaches have come out to try to emulate its success as far as accuracy is concerned, while maintaining the benefits learning-based solutions can have when compared with classical-handcrafted methods.

Since point cloud data is challenging to be directly fed to neural networks due to their sparsity and irregular sampling format, some data preprocessing and/or domain adaptation technique is usually employed. A common strategy to handle point cloud data in the scope of neural networks is to utilize a spherical or cylindrical projection to convert point cloud data information to regular matrix type data. After that, the feature extraction process can follow the typical pipeline of convolutional neural network representation. This step avoids memory inefficiency issues associated with conventional modules such as 2D convolution and upconvolution on non-grid like data typology.

For the point cloud registration problem, multiple feature-based methods aimed at detecting ever more accurate correspondences between consecutive scans have been developed in the recent past. Other methods for point cloud registration follow the working principle of iterative local methods like classical ICP variants, which learn to align two scans directly. However, when compared with registration, odometry problems are much more challenging because they require much higher accuracy under real noisy environments to prevent drifting. It is still early times for deep LIDAR odometry estimation, but multiple different approaches have been surfacing, with widely distinct techniques.

In particular, [Velas et al. \(2018\)](#) predicted pose by framing the odometry estimation problem as consecutive classification tasks and exhausting all possibilities. This approach, although it addresses the task of odometry estimation, it is not explicitly modelled as an odometry estimation system. LO-Net ([Li et al. 2019](#)) was the first instance of directly formulating the odometry estimation problem as a numerical regression of pose estimates and the first success towards true learning-based LIDAR odometry.

LO-Net architecture is consisted of three parts: a normal estimation sub-network, a mask prediction sub-network, and a Siamese pose regression main network. Cylindrical projection is utilized to encode point cloud data, which is then fed to the normal estimation network. Odometry is treated as regression problem, decoupled into position and orientation learning with two learnable parameters to observe the scale between translational and rotational motions. A mask strategy is employed to compensate for dynamic objects in the scene, perceiving the image patches where geometric consistency can be correctly enforced and thus introducing robustness to the pose regression network. LO-Net managed to achieve competitive results to the classical baseline for LIDAR odometry which is LOAM.

DeepPCO ([Wang et al. 2019](#)) is an end-to-end LIDAR odometry framework that is composed of two sub-networks: a translation estimation network and a flow orientation network. The sub-networks form a deep parallel framework that regresses 6-DoF pose. LIDAR Data is encoded as a 2D panoramic-view projection images and feed to the network as two consecutive images stacked together.

DeepLO ([Cho, Kim, and Kim 2019](#)) takes a different approach, feeding a rendered vertex map and a normal map into a network and regress a 6-DoF relative pose between two consecutive frames constrained by both an ICP-inspired loss and a field-of-view loss. Leveraging these representations and having designed two different loss functions, the authors can train the framework in a supervised or unsupervised manner.

DMLO (Li and Wang 2020) utilizes cylinder encoding to represent LIDAR points in a grid-like topology first and proceeds to extract feature vectors using convolutional neural networks and comparing the similarities in a local region to get correspondences between different scans. After that, the problem is converted to a rigid transformation estimation between matched pairs in 3D space which can be simply solved by Singular Value Decomposition (SVD). It is worth noting that all these methods for LIDAR odometry make use of projections of 3D data into 2D space, while encapsulating some of the key properties of the original 3D shape. Projecting 3D data into the spherical and cylindrical domains is a common practice for representing 3D data in a fast and efficient way and have the added benefit of easing the processing of 3D data due to the Euclidean grid structure of the resulting projections. However, such representations can be not optimal for complicated 3D computer vision tasks due to the possible information loss incurred while performing this projection.

#### **4. Open Challenges and Research Opportunities**

The relative novelty of learning-based techniques for solving robotic perception and control tasks in open-world scenarios means that several research questions are far from being resolved. This section will briefly mention some of the bigger open questions that should be the focus of multiple research efforts by the research communities and if solved, could be key for the wider proliferation of Deep Learning as a core technology in Robotic Systems, towards practical applications that can grant them greater autonomy in their deployment environments. Most issues are transversal to the odometry estimation task, independently of the specific sensor modality, i.e., vision-based motion estimation frameworks face the very same challenges as point-cloud based odometry.

##### **Unified evaluation benchmark and performance metrics**

Ever since the introduction of SLAM systems, the argument persists about how to evaluate a given method's performance and benchmark it according to sensible metrics that can provide an intuitive notion of how the method is performance versus the world scenario being perceived. This discussion becomes even more relevant in the case of Artificial Neural Networks (ANN's) and systems that use them as core technologies. This stems for the data-centric nature of such methods, of which the learning mechanism is affected by the intrinsic characteristics of the training data and how it is prepared. For instance, even though the KITTI dataset is widely regarded as a great choice for visual odometry (VO) estimation evaluation, it is frequent to see different authors in the literature utilizing different data splits and different metrics to evaluate their methods. Therefore, it makes it harder to fairly draw a direct comparison between methods, in the absence of a truly universally accepted benchmark dataset with both a strictly followed data split scheme and evaluation metrics. Moreover, the lingering question of scenario variability still subsides, with serious questions being raised about whether the performance in the KITTI dataset only terrestrial urban scenarios with fairly simple dynamics (almost a 2D trajectory) is enough evaluation data for accessing a method's true capability and generalization ability. To this point, authors usually complement this benchmark with a real-world scenario test, but again, it is usually almost impossible to directly test another method in a fair and clearway. The sheer volume of data itself is also important, because even though the KITTI dataset possesses more than enough data for testing and benchmarking classical feature-based methods, deep learning architectures could perhaps benefit from a larger scale VO dataset to improve their performance and generalization ability. In the past, deep learning architectures have achieved prominence in classification and object detection tasks in part due to the availability of such large-scale task-specific datasets such as

ImageNet or Pascal VOC. With this in mind, there is a growing need for creating a complete VO benchmark datasets package that is able to cover multiple environments, as well as contain a bigger variability of motions and different scene dynamics.

### **End-to-end DL vs Hybrid Model**

The scope to which DL methods should be used in robotic systems is also subject of discussion amongst the scientific community. End-to-end fully data-driven learning models that are able to predict a given task solely from raw data have proven massively successful in achieving increasing performance in accuracy, efficiency and robustness. On the other hand, there are those who believe that the secret for persistent autonomy is to integrate deep learning modules into the pre-built physical/geometry-inspired algorithms so as to fully leverage both the intrinsic nature of the data as well our prior hand-tailored empirical knowledge of the physical world. Hybrid methods have also achieved state-of-the-art performance for some tasks, such as visual odometry or global localization and benefit from being less data hungry than pure learning methods. Thus, this is a critical task-specific fundamental question that every roboticist should pose when developing a DL system. Data availability and variability could swing the answer to more of a hybrid model but often the general overall power of data-centric learning and the ease of integration with other high-level learning task such as path planning, or robotic control can be the deciding factors for pure end-to-end learning models.

### **Real-world deployment: Practical considerations, generalization ability and scalability**

Real world deployment and real-world performance of DL systems is still a systematic unanswered question in most cases that tends to be somewhat overlooked. For once, the computational and energy resource consumption of such systems must be taken into account, especially in the case of small mobile robots that sometimes demand a lot of optimization to properly function. Every parallelization opportunity in the inference process should be taken advantage, even though these robots usually do not have any type of GPU hardware, which limits the potential parallel implementations. Deployment of learning-based algorithms efficiently is somewhat tied to the available hardware. Nvidia Jetson TX1/2 or AGX Xavier embedded modules can function as cost-efficient GPU hardware for UAVs or mobile robots, operating under relative low power demand and with a lightweight form-factor. However, it often does not meet the requirements of more complex learning-based methods. Therefore, the trade-off between performance and model size also must be considered.

### **New sensor sources**

The rise in predominance of Deep Learning as a tool for solving perception tasks was mostly tied to Computer Vision techniques, thus relying on visual cameras as the main data source. Over the years a lot of other sensors have also been used for feeding data to neural networks. Most prominently, inertial and LIDAR data has been utilized to great success in a myriad of different tasks, especially when it is possible to integrate more than one sensor modality in a given system so as to increase its accuracy and robustness. Other sensors such as event-cameras, thermal cameras, radio signals or magnetic sensors have also been utilized in the literature, but most of these are severely underexplored compared to the more mainstream visual camera sensors.

Previous experience with underwater robotics suggests that multibeam acoustics are sometimes the most effective sensor in perceiving 3D structure in confined underwater environments and are often the best the most suitable sensing equipment for reconstructing complex scenarios. For this reason, one pertinent research question for future studies, is the



inclusion of this data topology as a sensor modality for data-centric learning-based egomotion perception.

### **Safety, reliability and interpretability**

The biggest criticism of Deep Learning systems is that it is often perceived as a "black-box", depriving the system designer of the ability of easily correcting some immediate problems. This characteristic makes it extremely hard to employ DL systems in safety-critical tasks, since even the smallest error in pose or scene estimation can ultimately cause a localization drift that may severely harm the entire system. Therefore, it is also really important to develop mechanisms to mitigate the underlying uncertainty with the DL system decision-making. There are two significantly relevant approaches: interpretability and uncertainty estimation. Interpretability in AI is the degree to which a human can consistently predict the model's result. The higher the interpretability of a machine learning model, the easier it is for someone to understand why certain decisions or predictions have been made. In a more practical sense, it refers to the system having a module that produces explanations for its predictions and what is inducing errors (e.g., a faulty sensor). Uncertainty estimation on the other hand is the idea of estimating a belief metric, i.e., a representation of the extent to which we trust our predictions. There are multiple ways of estimating a method uncertainty or bias, with perhaps the most common being to rely on the supervision of an external signal. In this way, the unreliable predictions (with low uncertainty) are avoided or smoothed in order to ensure the systems to stay within safe and reliable behavior.

## **5. Conclusion**

This article proposed to compile an in-depth study on Deep LiDAR odometry techniques for helping solve the robot navigation task, with the ultimate aim of providing a roadmap for future research. The relative novelty of the research field means that the literature is still not abundant and lacking wider comparisons and benchmarks. However, the general approaches and different data handling strategies detailed in this article show multiple different paths forward for tackling robot localization tasks using 3D non-matrix data. Although the potential of learning-based LIDAR techniques is clearly showcased in recent literature, the technical maturity of such approaches is still on early stages, and its increase will be key to future real-world deployments within mobile robot solutions.

Open challenges and research opportunities are also laid out, encompassing several different issues troubling deep learning/robotics researchers, that if addressed in a meaningful manner, have the potential to catapult these approaches to the foreground of robot software development.

## **References**

- Bello, S. A., S. Yu, and C. Wang. 2020. "Review: Deep learning on 3D point clouds". Preprint, submitted January 17, 2020. <https://arxiv.org/abs/2001.06280>.
- Cadena, C., L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. 2016. "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age". *IEEE Transactions on Robotics* 32, no. 6 (december): 1309-32. <https://doi.org/10.1109/tro.2016.2624754>.
- Chen, C., B. Wang, C. X. Lu, N. Trigoni, and A. Markham. 2020. "A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence". Preprint, submitted June 22, 2020. <https://arxiv.org/abs/2006.12567>.

- Cho, Y., G. Kim, and A. Kim. 2019. "Deeplo: Geometry-aware deep lidar odometry". Preprint, submitted February 27, 2019. <https://arxiv.org/abs/1902.10562>.
- Li, Q., S. Chen, C. Wang, X. Li, C. Wen, M. Cheng, and J. Li. 2019. "LO-Net: Deep real-time lidar odometry". In *Proceedings - 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8465-74. IEEE Computer Society. <https://doi.org/10.1109/CVPR.2019.00867>.
- Li, R., S. Wang, and D. Gu. 2018. "Ongoing evolution of visual SLAM from geometry to deep learning: Challenges and opportunities". *Cognitive Computation* 10, no. 6 (december): 875-89. <https://doi.org/10.1007/s12559-018-9591-8>.
- Li, Z., and N. Wang. 2020. "Dmlo: Deep matching lidar odometry". Preprint, submitted April 8, 2020. <https://arxiv.org/abs/2004.03796>.
- Scaramuzza, D., and F. Fraundorfer. 2011. "Tutorial: Visual odometry". *IEEE Robotics and Automation Magazine* 18, no. 4: 80-92. <https://doi.org/10.1109/mra.2011.943233>.
- Segal, A. V., D. Haehnel, and S. Thrun. 2010. "Generalized-ICP". Paper presented at the International Conference on Robotics Science and Systems, RSS 2009, Seattle, United States, June 28 through July 1, 2009. <https://doi.org/10.15607/rss.2009.v.021>.
- Sünderhauf, N., O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, et al. 2018. "The limits and potentials of deep learning for robotics". *The International Journal of Robotics Research* 37, no. 4-5 (april): 405-20. <https://doi.org/10.1177/0278364918770733>.
- Thrun, S. 2000. "Probabilistic algorithms in robotics". *AI Magazine* 21, no. 4 (december): 93-109. <https://doi.org/10.1609/aimag.v21i4.1534>.
- Velas, M., M. Spanel, M. Hradis, and A. Herout. 2018. "CNN for IMU assisted odometry estimation using velodyne LiDAR". In *18th IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 71-77. IEEE. <https://doi.org/10.1109/ICARSC.2018.8374163>.
- Wang, W., M. R. U. Saputra, P. Zhao, P. Gusmao, B. Yang, C. Chen, A. Markham, and N. Trigoni. 2019. "DeepPCO: End-to-end point cloud odometry through deep parallel neural network". In *IEEE International Conference on Intelligent Robots and Systems*, 3248-54. <https://doi.org/10.1109/IROS40897.2019.8967756>.
- Zhang, J., and S. Singh. 2017. "Low-drift and real-time lidar odometry and mapping". *Autonomous Robots* 41, no. 2 (february): 401-16. <https://doi.org/10.1007/s10514-016-9548-2>.

## Acknowledgements

The authors wish to acknowledge the support of the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, through PhD research grant [2020.05052.BD](https://doi.org/10.1007/s10514-016-9548-2).