Research Article

Learning-Based Visual Navigation under Extreme Conditions for Planetary Exploration

Vibhor Pal*

Independent Researcher

Abstract

Planetary visual navigation is essential for facilitating autonomous exploration of alien landscapes in harsh climatic circumstances. This research introduces a learning-based visual navigation system that combines SLAM-net with path planning and motion control to make Mars rover operation more reliable. The system uses DB-Net and Value Iteration Network (VIN) architectures that were trained on datasets of Martian topography to make sure that decisions are made quickly and paths are created. The DB-Net model, which combines global and local feature fusion, is better than VIN at navigation accuracy (95.6%) and success rate (93.3%). The addition of Vision Transformer-based SLAM improves the accuracy of localization and mapping by up to 20% in low-light and dusty circumstances. Testing the Athena rover in Mars-like environments shows that it can reliably avoid obstacles, accurately estimate its position, and change its navigation. Combining stereo vision with LiDAR sensors makes 3D perception and mapping even better. The suggested method has a 95% success rate for navigation in NASA ROAMS simulations, showing that it might be useful for durable, autonomous planetary exploration. The structure works on several sorts of terrain, such as sandy, gravel, and hard ground. This work advances the development of next-generation intelligent navigation systems for forthcoming interplanetary missions.

Keywords: Planetary Visual Navigation, Mars Rover, SLAM, DB-Net, Value Iteration Network, Sensor Fusion, 3D Mapping, Autonomous Exploration

Introduction

In order to enable robots to operate independently in a range of often unstructured situations, autonomous navigation has emerged as a critical area of robotics research. Among the various sensing modalities, vision has emerged as a primary tool due to the compactness, low cost, and high informational richness of cameras [1]. Real-time location, mapping, planning, and interaction with their environment are all possible for robots when vision-based systems are combined with developments in computer vision and ML to extract the environment's intricate geometric and semantic representations[2]. Recent advancements in hardware, such as solid-state LiDAR, event-based sensors, and high-resolution global-shutter cameras, along with advanced software, such as DL frameworks, differentiable optimization, and large-scale simultaneous localization and mapping (SLAM), have sped up the adoption of vision-driven autonomy in a variety of applications, from industrial robotics and autonomous cars to underwater inspection and planetary exploration.

These positive developments notwithstanding, there is still a major challenge of realizing a powerful navigation in harsh and unpredictable conditions. The terrains of the planets especially introduce special challenges, including extreme changes in lighting, dark, shadows or bright areas, dust storms, and severe constraints in computing power[3]. These conditions directly influence the effectiveness of the perception and control systems and tend to restrict the efficiency of the traditional navigation algorithms[4]. In the case of autonomous robots, localization would be necessary and precise interpretation of the surrounding environment is critical towards safe and efficient decision making. Mapping and localization of space using the traditional Visual SLAM techniques, which are the basis of the visual method, are based on quality images. They tend to have impaired performance in semi-dark or low-light situations, however, when feature detection and pose estimation is invalid.

To overcome these limitations, learning-based techniques have come into existence where deep neural networks (DNN) are used to achieve robustness and flexibility in the face of challenging conditions. These techniques enhance the perception and navigation in places where the classical algorithms fail

because of the manipulation of contextual and time information [5]. Nevertheless, the current literature has largely concentrated on a nominal or extreme dark environment, and semi-dark planetary environments have not been prioritized. The present research is dedicated to learning-based visual navigation in severe conditions to explore the planets, which would improve the localization, mapping, and decision-making of complicated extra-terrestrial environments and support the safe and effective navigation of autonomous rovers within the context of long-term missions.

It is essential to have autonomous robots be able to navigate safely in remote and unstructured planetary environments with the aim of exploration missions[6]. Deep learning (DL) and SLAM Vision-based navigation have high perception and decision-making features. Nevertheless, harsh environmental factors like light depletion, reflective or non-textured landscapes, dust storms, and limited computation overpower traditional algorithms and lower localization accuracy and confidence and require learning-style methods to achieve robust navigation.

Organization of the Paper

The paper provides a comprehensive overview of planetary visual navigation. Section I introduces the background and objectives. Section II reviews planetary visual navigation. Section III covers learning-based navigation techniques. Section IV discusses SLAM integration. Section V discusses the Analysis and Discussion of Learning-Based Visual Navigation Techniques. Section VI presents a literature review, and Section VII brings significant decisions and plans for the future to a close.

Planetary Visual Navigation Exploration

Planetary visual navigation is a key part of autonomous exploration because it lets rovers view, interpret, and move over difficult alien terrain without needing constant human help. It employs vision-based perception, mapping, and path planning to make sure that transportation is safe and successful in tough places like Mars or Europa[7]. Through powerful AI and machine learning (ML) models, visual navigation improves the precision of localisation, the ability to avoid obstacles, and the knowledge of topography. This supports protracted missions and broadens the field of planetary exploration.

Visual Navigation with SLAM-net

It is recommended that SLAM-net be combined with modules for motion control and path planning to make a visual navigation pipeline. This pipeline makes a global occupancy grid map and utilizes SLAM-net to guess the robot's position on a regular basis. After receiving the map and posture, a 2D path planner calculates a route to the destination. A local controller that generates distinct robot actions tracks this course.

Task Specification

The task is exploiting noisy RGB(D) input to navigate to goals in situations that have never been seen before. Discrete robot operations result in noisy motion, and using coordinates to specify the goal in relation to the starting posture makes it difficult to determine the robot's location. A successful navigation attempt is made if the robot stops within 0.36 meters of its successful. An accurate posture estimate is crucial for this success criterion.

Path Planner

Path planning uses a weighted algorithm on an 8-connected occupancy grid, with higher costs near obstacles and paths re-planned after each SLAM-net update[8]. Collisions are detected when forward motion is blocked, triggering obstacle registration, a turn-around, and a step back. An initial rotation policy ensures full environment coverage at the start of each episode.

Local Controller

A basic controller tracks the intended course and chooses actions to turn or advance in the direction of the path's farthest traversable point. Discrete actions are output by the controller.

Perception and Planning for Planetary Rovers

The two feature-based SLAM approaches are bundle adjustment-based and filter-based. These frameworks make operations more efficient by letting the rover execute low-level navigation chores on its own while giving high-level commands. Most modern navigation systems rely on vision to see the world around them. They use stereo cameras to do this. To design a path, construct a travers ability map that breaks the landscape up into a grid, with each cell indicating an area the rover can travel through[9]. Cells are given values that describe topographical parameters including height, slope, and roughness. Then, the travers ability map is used to plan safe ways to get to the goal. In relation to the created map, stereo visual odometry (VO) adjusts the rover's position as it moves. This involves detecting and tracking features between consecutive frames, triangulating their positions, rejecting outliers, and performing nonlinear refinement to estimate motion accurately.

Hardware Platform of the Mars Environment

Experiments were conducted using the Athena rover, a platform similar in size to Mars rovers that may be used to evaluate planetary rover capabilities on

Earth[10]. An Nvidia TX1 handles planning and navigation, while an Nvidia TK1 serves as a backup computer for the hardware interface[11]. The system uses a pair of stereo cameras installed on a pan-tilt pole at a height of 1.4 meters with a baseline of 0.30 meters, and it is mostly vision-based. Encoders measure the mast posture in relation to the rover body while images are taken at 640 x 480 resolution with an 82 x 66 degree field of view[12]. The rover does tests at a speed of around 0.042 meters per second. Experiments were conducted at the JPL Mars Yard, a 21×22m outdoor facility designed to replicate Martian terrain for testing robotic platforms (as shown in Figure 1).
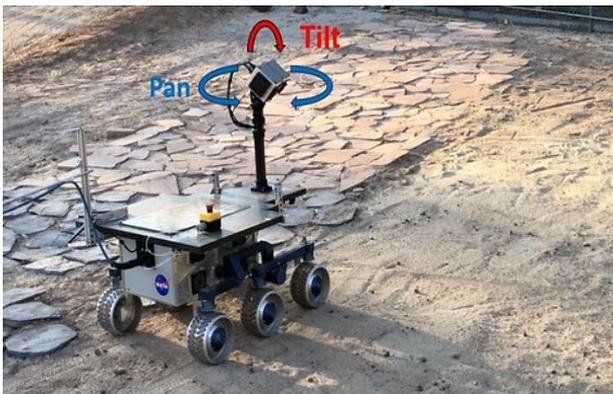
Figure 1 Athena Rover in the Mars Yard

There are a variety of feature-rich and feature-poor areas due to the terrain's sand and rocks of different sizes and forms. To simulate the texture-limited surfaces of Mars, smooth sand areas were raked prior to experiments, ensuring realistic conditions for validating vision-based navigation methods.

Navigation and Mapping Challenges.

Here's the "Navigation and Mapping Challenges" are in key points given below:
Pose estimation errors: can build up over time, which can affect both the accuracy of the navigation and the success of the mission overall.
Variability of Terrain: Any rocks, slopes, craters, and non-uniform terrains make mapping and the planning of paths challenging and require on the fly changes in mapping algorithms.
Lack of Texture: Environments with no texture, such as sandy or smooth surfaces, remove visual features from the vehicle's perception of the environment and can cause difficulty with the vision-based perception of the environment[13], causing increased reliance on non-vision cues for navigational measures.
Collisions with Obstacles and Occlusions: Any impediments not visualized on the map result in a collision and therefore it is worthwhile to have recovery and dynamic re-planning available in such a way that the autonomy may alter the plans in case of an obstacle.

Computing Resources: The computing resources on board are limited and the SLAM and path-planning algorithms are only limited to lightweight and efficient algorithms that can be executed in real-time[14].
Need for Active Perception: Proactive perception-based techniques should be used to identify and evade regions where the autonomous systems are prone to cause navigation errors leading to greater robustness.

Learning-Based Visual Navigation Techniques

Controllers that are based on learning apply data-optimized policies instead of hand-written guidelines. It starts with deep reinforcement learning, which requires extensive simulation but learns navigation from start to finish. Then, to start policy training and lower the hazards of unguided investigation, imitation learning makes use of expert demonstrations. Lastly, hybrid strategies combine learnt residuals or fallback controllers with classical planners. This development demonstrates how subsequent methods maintain generalizability while sacrificing autonomy for effectiveness and safety.

Data Analysis for Mars Visual Navigation

First, the Martian image dataset obtained from HiRISE is used to train and evaluate DB-Net and VIN. Each of the 10,000 high-resolution Martian images in the dataset comprises seven randomly produced ideal trajectory samples. 6/7 datasets make up the training set, whereas 1/7 datasets make up the testing set. Next, the training effectiveness and navigation accuracy of VIN and DB-Net are contrasted. Finally, using model ablation experiments, DB-Net is thoroughly examined. The following queries were particularly looked into:
Problem Formulation: Mars visual navigation is modelled as a Markov Decision Process (MDP) where the rover's state combines the Martian terrain image, target coordinates, and current position. In order to efficiently accomplish the aim, the goal is to develop an optimum policy that assigns each state to the best course of action (out of 8 movement directions).
Input Representation: Each input image has three channels: grayscale terrain, Canny edge-detected image, and target location mask combined into a 128×128×3 tensor. This image shows both visual texture and spatial goal cues, enabling the network to know how hard the terrain is and where to go for navigation.
Network Architecture Dual-Branch Network (DB-Net): DB-Net handles the visual input using specialised branches for global and local feature extraction, which helps it make good navigation judgements even when things are complicated on Mars.
Pre-processing Layers: Two convolutional layers and two max-pooling layers turn the input into a small feature map. This step cuts down on spatial duplication and makes the abstract representation of terrain patterns better..

Branch One Global Feature Extraction: This branch is all about figuring out the target's relative position and the context of the landscape on a broad scale. It makes a global deep feature by using convolutional and residual layers, then pooling and dense layers. Vector $f_1(I_t)$.

Branch Two Local Feature Extraction: This branch focusses on the rover's immediate surroundings and the tiny characteristics of the terrain that are close to It uses convolutional and residual layers to create a local deep feature vector $f_2(I_t, x_{1t}, x_{2t})$.

Fusion and Output: A fully connected layer combines the global and local characteristics to get the action-value function for each of the 8 possible actions, $Q(s_t)$. The rover moves towards the goal by picking the action with the highest Q-value.

Benchmarking   Performance   for   Simulation Environments

To assess the efficacy of learning-based visual navigation techniques under harsh planetary settings, several datasets and simulated scenarios are employed.

Nasa Roams Dataset: The Nasa Roams dataset provides Mars-analogue terrains with realistic lighting, dust, and surface changes. This makes it a great place to test and compare both traditional and learning-based SLAM methods.

Mars-Like Terrain Simulations: Mars-like terrain simulations create controlled settings that mimic the properties of the Martian surface and allow for testing in low-light and dusty circumstances [15], which makes it possible to test the durability of navigation.

Europa Analogue Simulations: Europa analogue simulations provide ice and rough surfaces, which let scientists systematically test navigation, trajectory planning, and pose estimation under very harsh planetary circumstances.

Training Strategy and Evaluation Metrics

The model learns to copy expert trajectories using supervised learning, which is called Imitation Learning. Using L2 regularisation with a cross-entropy loss makes sure that action predictions are correct and that overfitting doesn't happen. The Adam optimiser is used to train on 10,000 Martian terrain photos from the HiRISE dataset, which speeds up convergence. Below are the measures for assessment.

Accuracy tells how often the model gets the right navigation action

The success rate tells how many times the rover achieves its goal without crashing into anything.

Training Efficiency looks at how well a computer works by measuring the time it takes to complete an epoch and the overall pace of convergence.

Simultaneous Localization and Mapping (Slam) Integration

Simultaneous Localisation and Mapping (SLAM) let robots explore new places while figuring out where they are and making maps at the same time. Using landmarks, data association, and loop closure[16]. It makes things less ambiguous. Kalman Filters (KF) and Extended Kalman Filters (EKF) perform well when there is Gaussian noise, but not so well when there is non-Gaussian or very nonlinear noise. Particle Filters, on the other hand, are more flexible but cost more to compute.

Visual SLAM (V-SLAM) tracks camera movement using sequences of images. This gives robotics, AR, and VR apps a lot of spatial information. Recent advancements amalgamate semantic comprehension and learning-based methodologies, merging deep learning with geometry-centric strategies to augment robustness and accuracy in challenging conditions, such as low illumination, dust, or irregular terrain, realizing enhancements of up to 20% in robustness relative to conventional SLAM techniques, including:

Feature-Based SLAM

There are two types of feature-based SLAM: filter-based and bundle adjustment-based. Early systems employed Extended Kalman Filters (EKFs) to figure out where things were and where they were going, but they had problems with scalability and non-linearity. FastSLAM made things more efficient by integrating particle filters with EKFs, although sample degeneracy still happened. Parallel Tracking and Mapping (PTAM) improved real-time performance by splitting posture estimation and mapping into several threads that run at the same time [17]. Later, ORB-based systems used characteristics that were rotationally invariant and scale-aware in a Bag-of-Words framework. This made relegalization and loop closure better for monocular setups. RGB-D SLAM made it possible to make rich 3D maps and track motion more accurately. Object- and planar-based SLAM added semantic understanding to make localisation better, although it usually needed supervised learning to recognise objects first.

Direct SLAM

Direct SLAM approaches work directly on the raw image data, unlike feature-based systems, which rely on feature detectors or descriptors. These methods use photometric constancy to line up images that come one after the other. They work especially well when feature extraction is not dependable or not enough[18]. With the emergence of deep learning (DL), direct SLAM has changed to include learnt representations of the environment. Models like DTAM, LSD-SLAM, and SVO did very well in real-time dense mapping and tracking applications. After that, DSO made accuracy even better by adding a bundle adjustment process based on

temporal multi-view stereo. More subsequent methods, such as Code SLAM and CNN-SLAM, used DNN to get dense representations of the environment, which made direct SLAM more reliable in difficult visual situations.

Graph-Based SLAM

Based on graphs in SLAM's graph representation of the environment, nodes represent robot postures or landmarks, and edges reflect spatial limits that come from sensor data. The goal is to improve this graph such that the locations and landmarks are as reliable as possible. This method treats SLAM as a nonlinear least-squares problem and uses optimization techniques like the Levenberg-Marquardt or Gauss-Newton algorithms to solve it. Graph-based SLAM is more accurate and scalable than filter-based approaches because it re-optimizes the whole trajectory when new restrictions, such loop closures, are found. Graph-based SLAM is now useful for real-time applications thanks to popular frameworks like Ceres Solver, GTSAM, and g2o.

Integration with Learning-based and Sensor-Fusion Models

Recent developments in DL have brought Vision Transformers (Vitis) to the fore as effective tools for extracting visual features for autonomous navigation tasks. The Vita takes in raw images and turns them into context-aware, strong feature representations. These are then provided to the SLAM module to improve posture estimation and make sure the map is always accurate[19]. ViTs are different from standard convolution-based networks because they capture global visual cues and long-range dependencies. This is especially helpful in planetary conditions with little light, dust, or roughness. This hybrid Vision Transformer–SLAM pipeline let's find way around more reliably and make accurate 3D maps even when the circumstances are bad for feature-based or direct SLAM approaches.

Sensor Fusion for Dense 3D Perception

There is no one sensor type that works best in all planetary settings. Because of this, new robotic navigation systems are using sensor fusion more and more to improve depth perception and environmental awareness[20]. Visual-inertial odometry combines data from a camera and an IMU to give high-frequency motion estimates and a metric scale. LiDAR, on the other hand, gives long-range geometric information that doesn't vary with light levels[21]. Camera-LiDAR depth completion networks efficiently match dense image gradients with sparse point clouds, achieving centimetre-level mapping accuracy even in difficult conditions like fog or dust[22]. Depth data from stereo vision, RGB-D sensors, and LiDAR are commonly integrated through probabilistic occupancy fusion methods for creating reliable volumetric maps for route planning and obstacle avoidance, such as Euclidean Signed Distance Field (ESDF) or Truncated Signed Distance Field (TSDF).

Quantitative Impact of Sensor Fusion on 3D Perception

Recent benchmarks highlight the substantial advantages of combining complementary sensors for robust 3D perception. RGB–LiDAR fusion techniques, such as PENet, preserved real-time inference (20–25 FPS) while reducing root mean square depth error by around 30% on the KITTI depth completion benchmark when compared to LiDAR-only interpolation. In comparison to vision-only systems that drift 1-2% under fast motion, combining Visual-inertial odometry evaluations utilizing A monocular camera with an IMU reduced translational drift to less than 0.5% of trajectory length, according to the TUM VIE and EuRoC MAV datasets. For high-speed outdoor navigation, stereo–event camera fusion (as validated on DSEC) enabled dense, blur-resistant depth reconstruction at high temporal resolutions. Reconstruction completeness was increased by about 10 IoU points using indoor RGB-D fusion on NYUv2, particularly on low-texture surfaces.

Learning-Based Visual Navigation Techniques Analysis and Discussion

This section analyses and discusses the performance, efficiency, and adaptability of learning-based visual navigation models (DB-Net, VIN, and others) for planetary exploration under diverse Martian terrains.

Algorithm Selection for Planetary Visual Navigation

Both DB-Net and VIN are used in Mars-like terrain simulations for planetary visual navigation. DB-Net enhances perception through global–local feature fusion, while VIN enables internal planning for efficient path navigation in unstructured environments, as follows:

DB-Net (Double-Branch Network) Algorithm

The Double-Branch Network (DB-Net) serves to improve decision-making in visual navigation as it consists of global and local feature extraction. A parallel convolutional processing of an input image is done in two branches. The global branch takes big picture information such as terrain plan whereas the local branch takes information regarding local barriers. The fused feature maps are processed by fully connected layers to create an optimal policy of navigation with supervised learning on expert trajectories. This bifurcated architecture strikes a balance between awareness of the context and the awareness of the obstacles, resulting in the increase of

robustness and flexibility in the presence of different conditions on the planet.

Value Iteration Network (VIN) Algorithm

The value iteration planning process serves as the foundation for the VIN, a deep learning model for end-to-end navigation policy learning. It represents the environment as a grid where there is a reward and a transition probability in each cell. VIN trains the value function using convolutional layers to update the value function over and over again using differentiable operations. This helps it get closer to the best state-action values. The network learns a policy that gets the most total rewards, which lets it plan ahead while it's making decisions. VIN is good for autonomous planetary navigation in map-sparse and dynamic settings because it combines perception and control.

Analysis of both Techniques

This section tested DB-Net and VIN using a collection of Martian photos. Success was determined as reaching the target without wandering into unsafe areas. Table I shows that DB-Net did better than VIN. It had higher training/testing accuracy (96.4%/95.6% vs. 90.0%/89.8%) and success rates (96.0%/93.3% vs. 81.1%/79.4%), and it took less time per epoch (52.8 s vs. 97.5 s). This shows that it is more efficient and reliable for Mars visual navigation.

Table 1 Comparative analysis on Mars Visual Navigation

Performance	Model	
28x28 grid map	DB-Net	VIN
Training accuracy	96.4%	90.0%
Testing accuracy	95.6%	89.8%
Training success rate	96.0%	81.1%
Testing success rate	93.3%	79.4%
Average time cost (each epoch)	52.8s	97.5s

(a) Training loss   (b) Training error
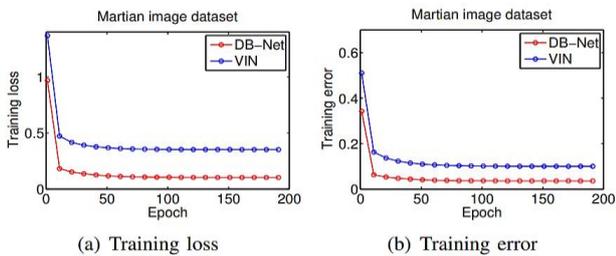
Figure 2 Performance on DB-Net and VIN

Figure 2 shows that DB-Net's training loss and error converge quicker than VIN's. DB-Net is more accurate than VIN; after 200 training epochs, it had a training accuracy of 96.4% and a testing accuracy of 95.4%. DB-Net further shows that it is more efficient by cutting the average time per training period by 45.8% compared to VIN. DB-N*et al*so does quite well on the data used for testing and training. It's important to note that the testing data includes photos of the Martian environment that were never viewed during

training. Still, DB-Net keeps up its high-precision visual navigation. VIN, on the other hand, has a testing success rate of less than 80%, which shows that it doesn't work as well as it should.

Table 2 Learning-Based Visual Navigation Results Summary

| Model	Accuracy (Testing)	Success Rate (Testing)
DB-Net	95.6%	93.3%
VIN (Value Iteration Network) [23]	89.8%	79.4%
RF Classifier (Terrain Features)[24]	94.66%	>92% per terrain type

(a)   (b)   (c)
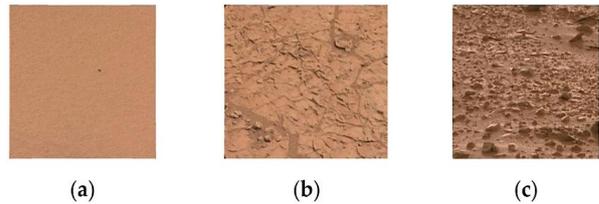
Figure 3 Mars Terrain (a) Sandy Terrain, (b) Hard Terrain, and (c) Gravel Terrain.

Table II summarizes the performance of learning-based visual navigation models. DB-Net achieves 95.6% accuracy and 93.3% success rate, VIN achieves 89.8% accuracy and 79.4% success rate, and the RF classifier achieves 94.66% accuracy with over 92% success per terrain type. Figure 3 highlights three key terrain types for safe rover operation: sandy terrain (ST), hard terrain (HT), and gravel terrain (GT), which together represent the majority of surfaces encountered. ST is soft sand, HT is rigid bedrock or slate, and GT is hard gravel.

(a)   (b)   (c)
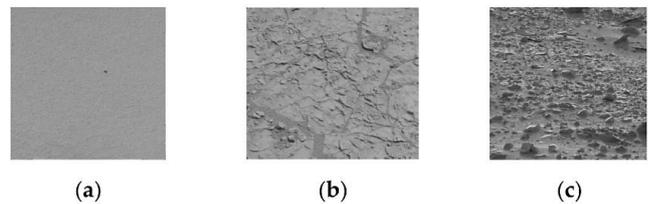
Figure 4 Multiscale Grey Gradient-Grade Features

As shown in Figure 4, GT has the largest grayscale shifts, followed by HT and ST. The gray gradient increases with the strength of the gray value variations. The grayscale gradient levels in this investigation were indicated by a number of thresholds. A gray gradient-based feature is extracted from a picture by figuring out each gradient level's pixel ratio.


(a)   (b)   (c)
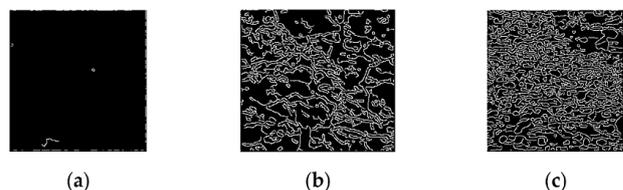
Figure 4 Extraction Results pf Strong Edges: (a) ST, (b) HT, and (c) GT

Figure 5 illustrates the extraction of strong edges for (a) ST, (b) HT, and (c) GT. The Canny technique, which is gradient-based, is used to detect edges in the study's sample image. The image with the greatest number of strong edges is GT, followed by HT, and the image with the fewest strong edges is ST.
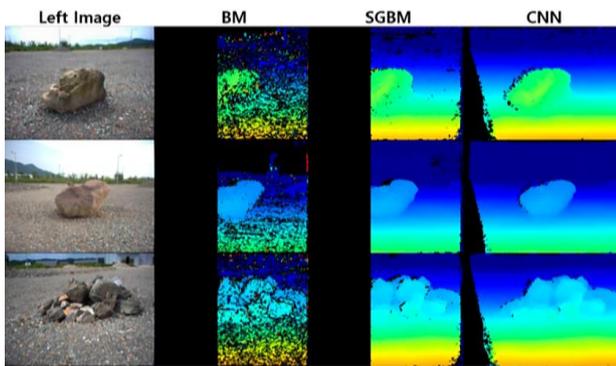
Figure 5 Comparison of Dense Disparity Map.

Dense disparity maps produced by the stereo CNN model and conventional techniques, Figure 6 presents a qualitative comparison of techniques like block matching (BM) and semi-global block matching (SGBM). The BM-based disparity map shows a lot more noise in comparison to the maps generated by CNN and SGBM, even with careful parameter tweaking. Only differences inside the overlapping area of the stereo image are described by BM and SGBM. By predicting differences outside of the overlapped region, the stereo CNN, on the other hand, minimizes the quantity of black pixels. This indicates that the rover's mobility needs can be decreased by using denser 3D point clouds using the stereo CNN model.
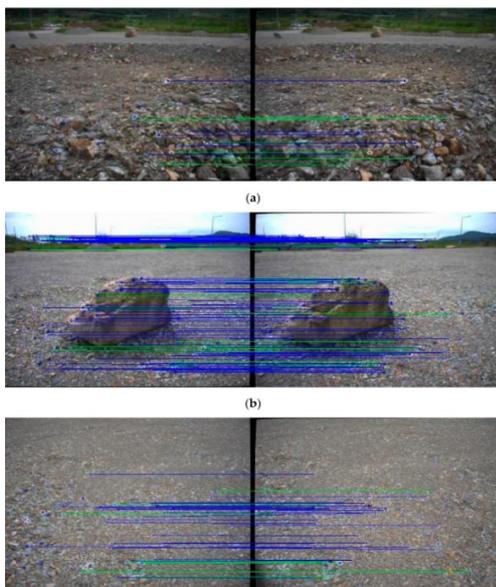
Figure 6 Feature-matching results from terrain features with AKAZE (threshold). (a) Crater with AKAZE (0.005); (b) rock with AKAZE (0.002); (c) ground with AKAZE (0.002)

The arbitrary use of AKAZE with wide thresholds for visual analysis is depicted in Figure 7. Green lines indicate feature matches that are approved by the disparity map but not by the NNDR constraint, whereas blue lines indicate feature matches that meet both requirements. The disparity map constraint in each keyframe increases the amount of extracted map points and may strengthen the robustness of the localization module in the suggested method, as seen by feature-matching results that show all of the green and blue lines are exactly matched.

Literature Review

This literature review on Visual Navigation in Extreme Conditions for Planetary Exploration emphasises significant trends, empirical results, and technology advancements, providing insights to inform future research and practical applications.

Savinykh *et al*. (2022) provide a novel SLAM technique that can operate in low light by enhancing the light conditions on input image with a Generative Adversarial Network (GAN) pre-treatment module, which increases the localization resilience. The suggested approach was tested using a motion capture device to get ground truth data and a specially created indoor dataset consisting of 14 sequences with different levels of light intensity. Unlike the current techniques, which only trace 0.6% of the sequence time, the experimental results indicate that the suggested approach maintains exceptional dependability even at extremely low light levels, with the darkest sequences tracking at 25.1%[25].

Quan *et al*. (2022) Provide a technique that improves the low-light image by utilizing the temporal data of a succession of input images. Then, use the improved outcome to boost the visual SLAM feature extraction and matching quality. Their method trains a three-dimensional CNN to anticipate pixel-wise grayscale transformation curves, which results in a low-light improved image. Additionally, they included their approach into VINS-Mono and evaluated it against comparable techniques for enhancing low-light images using the public TUM-VI dataset. The suggested approach offers a reduced positioning inaccuracy. In low light, Zero-DCE++ has a 19.83% higher positioning root-mean-squared error than the method[26].

Hong *et al*. (2021) present The visual simultaneous localisation and mapping (SLAM)-based robotic mapping approach, which uses a rover's stereo camera system. The technique uses S-PTAM as a foundational framework to improve mapping abilities in the homogenous and unstructured settings of planetary terrains by combining the disparity map from the self-supervised DL. Recent planetary exploration has concentrated on rover (or lander)-based surface missions towards the base construction for long-term human exploration and habitation because of the discovery of water-ice and lava tubes on the Moon and

Mars, as well as the advancement of in-situ resource utilisation (ISRU) technology[27].

Karkus, Cai and Hsu (2021) present a navigation architecture and the Differentiable SLAM Network (SLAM-net), which allow planar robot navigation in interior situations that have never been seen before. SLAM-net backpropagates through to train task-oriented neural network components, a particle filter-based SLAM method is encoded in a differentiable computation graph. The capacity to collaboratively tune every model component for the ultimate objective makes SLAM-net robust in challenging situations. They use a variety of real-world RGB and RGB-D datasets to perform experiments on the Habitat platform. The popular ORB-SLAM is significantly outperformed by SLAM-net in noisy environments. With a wide margin of success (37% to 64%), their navigation architecture with SLAMnet greatly surpasses the most advanced solution for the Habitat Challenge 2020 PointNav problem[28].

Chiodini *et al.* (2020) provide a technique for creating accurate semantic maps in three dimensions for the environment of Mars. The method's input is a stereo image captured by a rover-mounted camera. First, DeepLabv3+, an encoder-decoder CNN, is used to identify the images. Following semantic segmentation, the labels are concatenated to create stereo depth maps in a Voxel form. Utilize the ESA Katwijk Beach Planetary Rover Dataset to assess the methodology. For autonomous exploration rovers, terrain evaluation is essential. Recognizing the surrounding environment is necessary for a number of reasons, including autonomous target identification and the best possible trajectory planning[29].

Zhou *et al.* (2019) displays a multi-layered grid map with basic geometric and mechanical components after extracting the significant parameters from terrain bearing and shearing models. Using dense visual input, a matching mapping approach is developed to predict the terrain mechanical features of the whole visual field and reconstruct elevation on the map. A rover prototype is used in experiments to confirm the feasibility of the suggested method in a Mars simulation yard. The terrain map is the result of an autonomous scientific exploration system's mapping procedure, which is made of sensory information of numerous origins and is the basis of decision-making processes[30].

The latest developments on learning-based visual navigation and SLAM have significantly enhanced autonomous rovers in adverse and low-light conditions. DL algorithms based on GAN-based pre-processing, time optimization with 3D CNNs, and the differentiable SLAM networks improve the accuracy of localization, fidelity of mapping, and feature encoding. The SLAM-based frameworks are enhanced by stereo vision and improved terrain perception and 3D mapping, and the double-branch neural networks can facilitate effective global local feature fusion as part of the profile-based robust navigation. All in all, these approaches are vital for making sensing, mapping, and path planning better. They also let the rovers work in the complicated environment of the planets.

Conclusion and Future Work

This work has shown the significance of visual navigation through learning for planetary rovers that have to move across rough and unstructured terrain. Combining Dual-Branch Networks (DB-Net) and Value Iteration Networks (VIN) with SLAM and sensor-fusion systems makes sure that rovers can accurately estimate their position, create accurate maps, and plan efficient paths in terrain that is similar to Mars. The experimental results using HiRISE data and the Athena rover platform show that DB-Net is better than VIN when it comes to navigation accuracy, success rates, and the speed at which training converges. The rover can generalize to new terrains without tripping over boundaries and counterbalancing low-texture or dusty conditions since it can use both global and local visual features and learn by imitating others. Adding visual-inertial and LiDAR-based sensor fusion also makes 3D perception and navigation more accurate. In general, our research shows that hybrid DL and SLAM systems can make it easier for robots to explore planets on their own and reduce the need for people to manage them. These findings establish a foundation for the integration of intelligent navigation systems in forthcoming Mars missions and demonstrate the capacity of learning-based techniques to adjust to diverse environmental circumstances, therefore enhancing rover safety and operational efficiency. In terms of ethics, AI-powered autonomous systems for space exploration should carefully follow planetary protection rules, stop biological contamination, and encourage long-term exploration methods. To make sure that technology progress is in line with scientific integrity and environmental protection, need to be open about how algorithms make judgements, hold people accountable for how well they do their jobs, and utilize data responsibly.

Future efforts will concentrate on enhancing the navigation framework to synchronize multi-rover systems within diverse planetary conditions. Adding reinforcement learning to make things more stable and generalizable, boosting real-time performance on hardware with limited resources, and adding multimodal sensor fusion with event cameras will all help longer, more flexible autonomous missions.

References

[1] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Apr. 2021, doi: 10.1109/TRO.2021.3075644.

[2] T. S. Balint, E. A. Kolawa, J. A. Cutts, and C. E. Peterson, "Extreme environment technologies for NASA's robotic planetary exploration," *Acta Astronaut.*, vol. 63, no. 1–4, pp. 285–298, Jul. 2008, doi: 10.1016/j.actaastro.2007.12.009.

[3] J. Galvez-Serna, F. Vanegas, S. Brar, J. Sandino, D. Flannery, and F. Gonzalez, "UAV4PE: An Open-Source Framework to Plan UAV Autonomous Missions for Planetary Exploration," *Drones*, vol. 6, no. 12, p. 391, Dec. 2022, doi: 10.3390/drones6120391.

[4] T. S. Balint, J. A. Cutts, E. A. Kolawa, and C. E. Peterson, "Extreme environment technologies for space and terrestrial applications," in *Space Exploration Technologies*, W. Fink, Ed., Apr. 2008, p. 696006. doi: 10.1117/12.780389.

[5] R. Del Prete and A. Renga, "A Novel Visual-Based Terrain Relative Navigation System for Planetary Applications Based on Mask R-CNN and Projective Invariants," *Aerotec. Missili Spaz.*, vol. 101, no. 4, pp. 335–349, Dec. 2022, doi: 10.1007/s42496-022-00139-0.

[6] F. Mahlknecht *et al.*, "Exploring Event Camera-Based Odometry for Planetary Robots," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 8651–8658, Oct. 2022, doi: 10.1109/LRA.2022.3187826.

[7] B. Kuang, C. Gu, Z. A. Rana, Y. Zhao, S. Sun, and S. G. Nnabuife, "Semantic Terrain Segmentation in the Navigation Vision of Planetary Rovers—A Systematic Literature Review," *Sensors*, vol. 22, no. 21, p. 8393, Nov. 2022, doi: 10.3390/s22218393.

[8] D. S. Chaplot, D. Gandhi, S. Gupta, A. Gupta, and R. Salakhutdinov, "Learning to Explore using Active Neural SLAM," *8th Int. Conf. Learn. Represent. ICLR 2020*, Apr. 2020.

[9] R. Mu, P. Wu, Y. Deng, and H. Song, "Optical Navigation Method and Error Analysis for the Descending Landing Phase in Planetary Exploration," *Aerospace*, vol. 9, no. 9, p. 496, Sep. 2022, doi: 10.3390/aerospace9090496.

[10] M. Maimone, Y. Cheng, and L. Matthies, "Two years of Visual Odometry on the Mars Exploration Rovers," *J. F. Robot.*, vol. 24, no. 3, pp. 169–186, Mar. 2007, doi: 10.1002/rob.20184.

[11] U. A. Korat and A. Alimohammad, "A Reconfigurable Hardware Architecture for Principal Component Analysis," *Circuits, Syst. Signal Process.*, vol. 38, no. 5, pp. 2097–2113, 2019, doi: 10.1007/s00034-018-0953-y.

[12] J. Strader, K. Otsu, and A. Agha-mohammadi, "Perception-aware autonomous mast motion planning for planetary exploration rovers," *J. F. Robot.*, vol. 37, no. 5, pp. 812–829, Aug. 2020, doi: 10.1002/rob.21925.

[13] C. Papachristos, S. Khattak, F. Mascarich, and K. Alexis, "Autonomous Navigation and Mapping in Underground Mines Using Aerial Robots," in *2019 IEEE Aerospace Conference*, IEEE, Mar. 2019, pp. 1–8. doi: 10.1109/AERO.2019.8741532.

[14] S. Gupta, J. Davidson, S. Levine, R. Sukthankar, and J. Malik, "Cognitive Mapping and Planning for Visual Navigation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 7272–7281. doi: 10.1109/CVPR.2017.769.

[15] W. Shao, L. Cao, W. Guo, J. Xie, and T. Gu, "Visual navigation algorithm based on line geomorphic feature matching for Mars landing," *Acta Astronaut.*, vol. 173, pp. 383–391, Aug. 2020, doi: 10.1016/j.actaastro.2020.04.055.

[16] S. Hong, P. Shyam, A. Bangunharcana, and H. Shin, "Robotic Mapping Approach under Illumination-Variant Environments at Planetary Construction Sites," *Remote Sens.*, vol. 14, no. 4, p. 1027, Feb. 2022, doi: 10.3390/rs14041027.

[17] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A Comprehensive Survey of Visual SLAM Algorithms," *Robotics*, vol. 11, no. 1, p. 24, Feb. 2022, doi: 10.3390/robotics11010024.

[18] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, and H. Voos, "Visual SLAM: What Are the Current Trends and What to Expect?," *Sensors*, vol. 22, no. 23, p. 9297, Nov. 2022, doi: 10.3390/s22239297.

[19] Y. Dai, T. Zheng, C. Xue, and L. Zhou, "SegMarsViT: Lightweight Mars Terrain Segmentation Network for Autonomous Driving in Planetary Exploration," *Remote Sens.*, vol. 14, no. 24, p. 6297, Dec. 2022, doi: 10.3390/rs14246297.

[20] T. Lai, "A Review on Visual-SLAM: Advancements from Geometric Modelling to Learning-Based Semantic Scene Understanding Using Multi-Modal Sensor Fusion," *Sensors*, vol. 22, no. 19, p. 7265, Sep. 2022, doi: 10.3390/s22197265.

[21] C. F. Olson, L. H. Matthies, J. R. Wright, R. Li, and K. Di, "Visual terrain mapping for Mars exploration," *Comput. Vis. Image Underst.*, vol. 105, no. 1, pp. 73–85, Jan. 2007, doi: 10.1016/j.cviu.2006.08.005.

[22] V. M. L. G. Nerella, "Architecting Secure, Automated Multi-Cloud Database Platforms Strategies for Scalable Compliance.," *Int. J. Intell. Syst. Appl. Eng.*, vol. 9, pp. 128–138, 2021.

[23] J. Zhang, Y. Xia, and G. Shen, "A Novel Deep Neural Network Architecture for Mars Visual Navigation," Aug. 2018.

[24] F. Lv *et al.*, "Highly Accurate Visual Method of Mars Terrain Classification for Rovers Based on Novel Image Features," *Entropy*, vol. 24, no. 9, p. 1304, Sep. 2022, doi: 10.3390/e24091304.

[25] A. Savinykh *et al.*, "DarkSLAM: GAN-assisted Visual SLAM for Reliable Operation in Low-light Conditions," *2022 IEEE 95th Veh. Technol. Conf.*, vol. 2022-June, pp. 1–6, Jun. 2022, doi: 10.1109/VTC2022-Spring54318.2022.9860754.

[26] Y. Quan, D. Fu, Y. Chang, and C. Wang, "3D Convolutional Neural Network for Low-Light Image Sequence Enhancement in SLAM," *Remote Sens.*, vol. 14, no. 16, p. 3985, Aug. 2022, doi: 10.3390/rs14163985.

[27] S. Hong, A. Bangunharcana, J.-M. Park, M. Choi, and H.-S. Shin, "Visual SLAM-Based Robotic Mapping Method for Planetary Construction," *Sensors*, vol. 21, no. 22, p. 7715, Nov. 2021, doi: 10.3390/s21227715.

[28] P. Karkus, S. Cai, and D. Hsu, "Differentiable SLAM-net: Learning Particle SLAM for Visual Navigation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2021, pp. 2814–2824. doi: 10.1109/CVPR46437.2021.00284.

[29] S. Chiodini, L. Torresin, M. Pertile, and S. Debei, "Evaluation of 3D CNN Semantic Mapping for Rover Navigation," in *2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, IEEE, Jun. 2020, pp. 32–36. doi: 10.1109/ MetroAero Space 48742.2020.9160157.

[30] R. Zhou, L. Ding, H. Gao, W. Feng, Z. Deng, and N. Li, "Mapping for Planetary Rovers from Terramechanics Perspective," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Nov. 2019, pp. 1869–1874. doi: 10.1109/IROS40897.2019.8967984.