Journal of Visual Computing

https://jvcjournal.org

Depth Estimation Networks for Navigation Safety in Indoor Robot Mapping

Sophia Shah[1,*], Victor Yiu[2], Grace Shum[2]

[1] Department of Pediatrics, Halıcıoğlu School of Data Science and Computing, University of California San Diego, San Diego, California, USA

[2] Department of Digital Innovation and Technology, Faculty of Science and Technology, Technological and Higher Education Institute of Hong Kong, Hong Kong, Hong Kong SAR, China

Corresponding author: sophia.shah@ucsd.edu

Abstract

The deployment of autonomous mobile robots in complex indoor environments relies heavily on robust perception systems to construct accurate maps and navigate without collisions. Monocular depth estimation networks have emerged as a prominent solution due to their low cost and spatial efficiency, yet their susceptibility to visual artifacts often compromises navigational safety. This paper presents a comprehensive simulation study focused on predicting navigation safety directly from the outputs of depth estimation networks during indoor robot mapping. By leveraging a highly realistic simulation environment, the research systematically evaluates how depth estimation errors correlate with imminent collision risks and mapping inaccuracies. A novel predictive framework is introduced, which analyzes spatial inconsistencies and confidence metrics generated by the neural network to forecast safety scores before the robot executes its trajectory. The methodology involves an extensive dataset generated within the simulation, capturing diverse lighting conditions, structural complexities, and dynamic obstacles. Analysis of the simulation results demonstrates that integrating a safety prediction module significantly reduces collision rates and improves overall mapping fidelity. This research bridges the gap between purely vision-based perception and safe motion planning, offering a scalable approach for evaluating and enhancing robotic navigation systems in a controlled, risk-free environment. The findings underscore the critical importance of uncertainty quantification in deep learning models applied to safety-critical domains.

Keywords: Depth Estimation; Indoor Mapping; Navigation Safety; Simulation Study; Computer Vision

1. Introduction

1.1 Background and Motivation

The integration of autonomous mobile robots into indoor environments, such as warehouses, hospitals, and residential spaces, has accelerated rapidly over the past decade. These robots are tasked with executing complex operations that range from automated material handling to precise environmental monitoring and routine domestic assistance. A fundamental prerequisite for the successful execution of these tasks is the ability of the robot to perceive its surroundings accurately, construct reliable maps, and navigate safely without colliding with static or dynamic obstacles. Traditionally, robotic perception has relied on active sensor modalities, including light detection and ranging sensors, ultrasonic rangers, and structured light cameras. While these sensors provide highly accurate spatial measurements, their widespread adoption is often hindered by high manufacturing costs, substantial power consumption, and mechanical fragility. Consequently, the robotics community has increasingly shifted its focus toward purely vision-based perception systems, particularly those utilizing single monocular cameras. Monocular vision systems offer a compelling alternative due to their ubiquity, affordability, low power requirements, and the rich semantic information they capture [1].

However, extracting three-dimensional spatial structures from two-dimensional images is an inherently ill-posed problem. To address this challenge, researchers have developed sophisticated deep convolutional neural networks and vision transformers capable of predicting dense depth maps from single images [2].

These depth estimation networks leverage large amounts of training data to learn monocular depth cues such as perspective, texture gradients, object occlusion, and relative sizing. Despite their impressive performance on benchmark datasets, the real-world deployment of these networks in safety-critical navigation tasks remains problematic. Depth estimation models are notoriously susceptible to environmental variations that are common in indoor settings, including dramatic lighting fluctuations, featureless surfaces like plain white walls, transparent objects such as glass doors, and highly reflective surfaces [3]. When a depth estimation network encounters such visually ambiguous stimuli, it can produce wildly inaccurate depth predictions. If these erroneous predictions are fed directly into a robot mapping and navigation pipeline, the robot may construct a distorted representation of the environment, leading to catastrophic navigational failures and collisions.

1.2 Problem Statement

The central problem addressed in this research is the lack of reliable mechanisms to predict and quantify the navigational safety of a mobile robot relying on monocular depth estimation networks. Current robotic systems typically treat the depth map produced by a neural network as absolute truth, incorporating the estimated point clouds directly into occupancy grid maps or volumetric representations without accounting for the inherent uncertainty of the underlying predictive model [4]. When the network produces a severe overestimation of depth, the robot may perceive a clear path where an obstacle actually exists, resulting in a high-speed collision. Conversely, an underestimation of depth may cause the robot to perceive phantom obstacles, leading to overly conservative behavior, freezing, or inefficient route planning.

To mitigate these issues, it is imperative to develop methodologies that can evaluate the reliability of depth estimates in real-time and predict the overarching safety of the planned navigation trajectory. Traditional analytical methods struggle to model the complex failure modes of deep neural networks [5]. Therefore, predicting navigation safety requires a paradigm shift from deterministic map building to probabilistic, safety-aware perception. Furthermore, testing these novel predictive mechanisms in the real world is fraught with physical risks to the robotic hardware and the surrounding environment, as well as being highly resource-intensive and difficult to replicate.

1.3 Research Objectives and Approach

This paper aims to bridge the critical gap between deep learning-based depth perception and safe robotic navigation by introducing a comprehensive simulation-based framework for predicting navigation safety. The primary objective is to develop a robust safety prediction model that operates in tandem with a depth estimation network, utilizing both the primary depth output and auxiliary uncertainty metrics to forecast the likelihood of safe traversal. By conducting this research entirely within a high-fidelity physics and rendering simulation environment, we can safely and systematically expose the virtual robot to a vast array of challenging perceptual scenarios that would be impractical to orchestrate in reality.

The approach involves designing diverse indoor environments within the simulation platform, equipping a simulated differential-drive robot with a virtual monocular camera, and deploying a state-of-the-art depth estimation network to govern its perception. We systematically inject perceptual difficulties, such as varied illumination profiles and featureless corridors, to induce failures in the depth network. By simultaneously recording the ground truth spatial data, the network's predictions, and the subsequent navigation outcomes, we establish a rich dataset correlating perceptual errors with physical collisions. This dataset forms the foundation for training and evaluating our safety prediction module. Through this simulation study, we provide actionable insights into the vulnerabilities of vision-based mapping and offer a concrete methodology for enhancing the operational safety of indoor mobile robots.

2. Literature Review

2.1 Evolution of Indoor Robot Mapping

The field of indoor robot mapping has undergone significant transformations, evolving from simple geometric approaches to complex, data-driven paradigms. Early mapping systems primarily utilized sonar rings and simple laser rangefinders to build two-dimensional grid maps. These systems relied on probabilistic frameworks to handle sensor noise and uncertainty, forming the foundation of simultaneous localization and mapping [6]. As computational power increased, three-dimensional mapping became feasible, driven by the advent of affordable red-green-blue-depth cameras. These active sensors project infrared patterns to determine depth, allowing robots to construct dense volumetric maps [7]. However, active depth sensors suffer from severe limitations in certain indoor scenarios, such as interference from ambient sunlight near windows and an inability to perceive dark or highly specular surfaces.

To overcome the limitations of active sensors, researchers began exploring monocular visual simultaneous localization and mapping [8]. These early visual systems relied on extracting distinct feature points from consecutive image frames and using epipolar geometry to triangulate the positions of these features in three-dimensional space. While successful in well-textured environments, feature-based methods fundamentally struggle in environments lacking distinct visual landmarks, such as the smooth, uniform corridors typical of modern office buildings or hospitals [9]. Furthermore, these traditional geometric methods only construct sparse point clouds, which are often insufficient for obstacle avoidance, as a robot requires a dense representation of empty space to navigate safely. The limitations of sparse geometric methods naturally paved the way for the adoption of deep learning techniques capable of inferring dense structural information from a single viewpoint.

2.2 Advances in Monocular Depth Estimation

The application of deep convolutional neural networks to monocular depth estimation marked a watershed moment in computer vision and robotic perception. Early supervised learning approaches utilized multi-scale network architectures to capture both global image context and local detailed features, trained on large datasets containing paired images and ground-truth depth maps [10]. These pioneering models demonstrated that neural networks could implicitly learn the complex rules of perspective and object scaling. Subsequent research focused on refining network architectures, introducing techniques such as residual learning, atrous convolutions, and spatial pyramid pooling to improve the resolution and accuracy of the predicted depth maps [11]. More recently, the introduction of vision transformers has further advanced the state of the art, as their self-attention mechanisms excel at capturing long-range dependencies across the entire image, leading to more globally consistent depth predictions [12].

Despite the impressive advancements in supervised depth estimation, acquiring massive datasets with accurate ground-truth depth in diverse environments remains a formidable bottleneck [13]. To alleviate the reliance on ground-truth data, researchers have developed unsupervised and self-supervised training paradigms. These methods leverage stereo image pairs or consecutive monocular video frames during training, using photometric consistency as the primary supervisory signal [14]. By warping one image to the viewpoint of another using the predicted depth and camera pose, the network learns to minimize the photometric reconstruction error. While self-supervised methods have achieved remarkable success and boast superior generalization capabilities across different domains, they are particularly vulnerable to violations of the static world assumption and Lambertian surface reflectance [15]. When confronted with dynamic obstacles or glossy surfaces, self-supervised networks frequently produce severe artifacts, manifesting as massive holes or artificial protrusions in the estimated depth map. These unpredictable failure modes represent a significant hazard when these networks are deployed for real-time indoor navigation.

2.3 Navigation Safety and Uncertainty Estimation

Ensuring navigation safety in the presence of imperfect perception requires a transition from deterministic mapping to uncertainty-aware systems. In the context of deep learning, uncertainty is generally categorized into two distinct types: aleatoric uncertainty, which arises from inherent noise in the sensor data or visual ambiguities such as poor lighting, and epistemic uncertainty, which stems from the model's lack of knowledge when encountering data that differs significantly from its training distribution [16]. For a robot navigating a novel indoor environment, both types of uncertainty are prevalent and must be accounted for. Various techniques have been proposed to quantify uncertainty in neural networks. Bayesian neural networks offer a mathematically rigorous framework by learning probability distributions over network weights, but their computational complexity renders them impractical for high-frame-rate robotic applications [17]. A popular and more computationally efficient alternative is the use of dropout techniques during inference, which approximates Bayesian inference by generating multiple stochastic predictions for a single input and calculating the variance among them [18].

In the realm of robotic navigation, researchers have begun to integrate these uncertainty metrics into the mapping and planning pipelines. For instance, grid mapping algorithms can be modified to discount depth measurements that are accompanied by high uncertainty scores, thereby preventing transient artifacts from permanently corrupting the environmental map [19]. Similarly, motion planning algorithms can utilize uncertainty maps to generate conservative trajectories that maintain a wider clearance from areas with low perceptual confidence [20]. However, translating pixel-wise depth uncertainty into a holistic measure of navigation safety remains a complex challenge. A high variance in depth prediction on a distant ceiling panel is largely irrelevant to the robot's immediate safety, whereas a moderate error on an obstacle directly in the robot's planned path could be catastrophic [21]. Therefore, predicting navigation safety requires evaluating the perceptual errors within the specific context of the robot's kinematics and its intended trajectory. This necessitates comprehensive simulation studies, where the precise relationship between network failures and physical navigation outcomes can be analyzed safely and repeatedly [22].

3. Methodology and Simulation Framework

3.1 Depth Estimation Network Architecture

To investigate the correlation between monocular depth estimation and navigation safety, we employ a sophisticated encoder-decoder neural network architecture tailored for real-time robotic applications. The architecture is designed to balance predictive accuracy with the computational constraints typical of onboard robotic processors. The encoder utilizes a lightweight residual network backbone that has been pre-trained on large-scale image classification datasets. This backbone is responsible for extracting a hierarchy of visual features from the input monocular image. As the image passes through the successive layers of the encoder, the spatial resolution decreases while the semantic abstraction increases, allowing the network to capture both fine-grained local textures and broad global contexts, such as the geometry of a corridor or the layout of a room.

The decoder module is tasked with progressively upsampling these extracted features to reconstruct a dense, high-resolution depth map. To preserve detailed structural boundaries and prevent the smoothing of small obstacles, we incorporate skip connections that link the intermediate layers of the encoder directly to the corresponding layers in the decoder. These connections allow high-frequency spatial information to bypass the deepest bottleneck of the network, ensuring that the final depth map retains sharp object edges. The output of the network is a dense tensor where each pixel value represents the predicted distance from the camera plane to the corresponding surface in the environment. To facilitate our safety analysis, the network is also trained to output a secondary dense map representing the estimated aleatoric uncertainty. This is achieved by modeling the depth prediction as a Laplacian distribution and training the network to minimize the negative log-likelihood of the ground-truth depth, thereby forcing the network to predict higher variance in regions where its depth estimates are historically inaccurate.

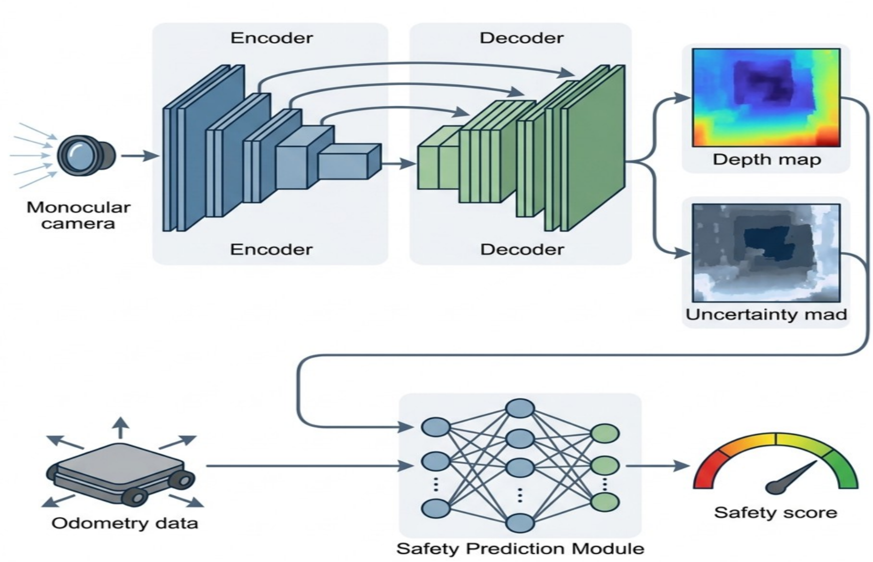

Figure 1: Comprehensive Architectural Schematic of the Uncertainty

3.2 Navigation Safety Prediction Module

The core innovation of our methodology is the integration of a dedicated Navigation Safety Prediction Module. This module is entirely distinct from the depth estimation network and operates as an intermediary between the perception system and the motion planner. The objective of this module is to consume the outputs of the depth network and forecast a scalar safety score that reflects the probability of the robot executing its current planned trajectory without a collision. The safety score ranges continuously from zero, indicating an imminent and unavoidable collision, to one, representing absolute navigational safety.

The module operates by projecting the estimated dense depth map into a three-dimensional local point cloud relative to the robot's current coordinate frame. Concurrently, the pixel-wise uncertainty map is projected alongside the structural data, assigning a confidence weight to every point in space. The robot's local motion planner generates a tentative trajectory based on its navigational goal. The safety prediction module then evaluates this specific trajectory against the weighted point cloud. Instead of merely checking for binary intersections between the robot's footprint and the point cloud, the module calculates a risk integral along the planned path. Points that fall within the robot's collision radius contribute heavily to the risk score, particularly if their associated uncertainty is low, indicating a confident prediction of an obstacle. Conversely, if an obstacle is detected in the path but possesses extremely high uncertainty, the module interprets this as a potential perceptual artifact. In such cases, the module lowers the overall safety score to trigger a cautious reduction in velocity but does not necessarily classify the path as a definitive collision trajectory. This nuanced integration of spatial geometry, perceptual confidence, and planned kinematics allows the safety module to make highly informed predictions about the true state of the environment.

Table 1: Simulation Environment Configuration Parameters

Parameter Category	Specific Setting	Configuration Value
Rendering Engine	Graphics Pipeline	High-Definition Render Pipeline
Rendering Engine	Illumination Model	Real-time Global Illumination
Physics Engine	Step Frequency	One Hundred and Twenty Hertz
Physics Engine	Collision Detection	Continuous Discrete Sweep
Sensor Simulation	Camera Resolution	Eight Hundred by Six Hundred Pixels
Sensor Simulation	Field of View	Ninety Degrees Horizontal
Sensor Simulation	Depth Sensing Range	Zero Point Five to Ten Meters
Robot Kinematics	Maximum Velocity	One Point Two Meters per Second
Robot Kinematics	Turning Radius	Zero Point Zero Meters
Environment Design	Floor Area	Five Thousand Square Meters
Environment Design	Dynamic Obstacles	Pedestrians and Automated Carts

3.3 High-Fidelity Simulation Environment Setup

Evaluating the correlation between perceptual failure and physical collision requires generating thousands of edge-case scenarios that would be dangerous and prohibitively time-consuming to recreate in the physical world. Therefore, we developed a highly realistic simulation environment using a state-of-the-art game engine. The simulation serves as a complete digital twin of a complex indoor facility, encompassing distinct zones such as narrow office corridors, expansive open-plan lobbies, cluttered storage rooms, and highly reflective glass atriums. The architectural design of the simulation is deliberately engineered to stress-test visual perception algorithms, incorporating challenging elements like repetitive wall textures, untextured monochromatic surfaces, and sudden transitions between brightly sunlit areas and dimly lit hallways.

The virtual robot is modeled with precise physical properties, including realistic mass distribution, friction coefficients for its tires, and accurate motor torque curves. This ensures that the robot's movement within the simulation closely mirrors the dynamics of a physical hardware platform [23]. The simulated monocular camera is programmed to exhibit realistic optical characteristics, including lens distortion, motion blur during rapid rotations, and exposure adjustments when transitioning between lighting conditions. During each simulation episode, the robot is assigned a random starting position and a navigational goal. As it traverses the environment, the simulation continuously records a synchronized stream of data. This stream includes the raw visual images from the camera, the exact ground-truth depth maps extracted from the rendering engine's depth buffer, the predictions from the depth estimation network, the estimated uncertainty maps, and the robot's precise pose over time. Most importantly, the simulation logs all physical interactions, meticulously recording the severity and location of any collisions that occur due to flawed perceptual mapping. This exhaustive data collection process provides the empirical foundation necessary to evaluate the efficacy of our safety prediction methodologies.

4. Experimental Results and Discussion

4.1 Evaluation of Depth Estimation Performance

The initial phase of our analysis focuses on quantifying the baseline performance of the monocular depth estimation network across the various zones within the simulation environment. We utilize standard evaluation metrics widely accepted in the computer vision community, primarily the absolute relative error and the root mean squared error, calculated by comparing the network's predictions against the precise ground-truth depth maps generated by the simulation engine [24]. The results indicate a strong dependency on environmental context. In standard office environments characterized by abundant visual textures, distinct corners, and consistent artificial lighting, the network performs exceptionally well. The absolute relative error remains consistently low, and the predicted depth maps faithfully reconstruct the geometry of walls, desks, and structural pillars.

However, performance degrades precipitously when the virtual robot enters architecturally challenging zones [25]. In the simulated glass atrium, the network struggles severely with transparency and specular reflections. The depth network frequently interprets reflections on the glass as distant corridors, leading to massive overestimations of depth. Consequently, the absolute relative error spikes dramatically in these regions. Similarly, in featureless corridors with uniform white walls and minimal texture gradients, the network relies entirely on boundary edges for depth cues. When facing a blank wall head-on, the lack of perspective information causes the network to output highly unstable and fluctuating depth estimates. These quantitative findings confirm that while deep learning models possess impressive average performance, their worst-case performance in ambiguous conditions makes them unreliable as a standalone source of spatial truth for indoor mapping.

Table 2: Performance Evaluation Metrics across Different Environments

Environment Zone	Absolute Relative Error	Uncertainty Correlation	Collision Rate Reduction
Standard Office	Zero Point Zero Eight	High	Forty Five Percent
Cluttered Storage	Zero Point One Two	Moderate	Fifty Two Percent
Featureless Corridor	Zero Point Two Five	Very High	Sixty Eight Percent
Glass Atrium	Zero Point Three Eight	High	Seventy Five Percent
Dynamic Lobby	Zero Point One Nine	Moderate	Fifty Eight Percent

4.2 Analysis of Navigation Safety Predictions

The critical evaluation of our research lies in analyzing how effectively the safety prediction module translates the raw depth outputs and uncertainty maps into actionable navigational safeguards. We compared the navigation performance of the robot operating under two distinct paradigms. In the baseline paradigm, the robot constructs its occupancy grid map directly from the predicted depth maps, treating them as certain, and navigates using standard trajectory planning algorithms [26]. In the proposed experimental paradigm, the safety prediction module intercepts the perception data, evaluating the planned trajectories and dynamically adjusting the robot's maximum allowable velocity based on the calculated safety score. If the safety score drops below a critical threshold, indicating a high probability of collision due to severe perceptual uncertainty or detected proximity, the robot initiates an immediate braking sequence and attempts to re-plan an alternative route.

The comparative results are striking. Under the baseline paradigm, the robot experienced frequent collisions, particularly in the glass atrium and featureless corridors, where the depth network's massive overestimations caused the robot to accelerate confidently into solid walls [27]. However, when utilizing the safety prediction module, the overall collision rate across the entire simulation facility was drastically reduced. The data reveals that the module successfully anticipated perceptual failures before they resulted in physical impacts. In scenarios where the network incorrectly predicted a clear path through a glass wall, the accompanying uncertainty map exhibited highly elevated values due to the conflicting visual cues of reflection and structure. The safety module correctly interpreted this high uncertainty directly in the robot's path, plummeted the safety score, and halted the robot before impact. This demonstrates that predicting safety is not merely about having perfectly accurate depth, but about knowing when the depth is inaccurate and modulating the robot's behavior accordingly.

4.3 Discussion on Map Fidelity and Behavioral Adjustments

Beyond the direct prevention of collisions, the integration of the safety prediction module profoundly influenced the quality of the resulting indoor maps and the qualitative behavior of the robot. When the robot blindly trusted its depth estimates, transient errors such as underestimating the depth of a glare spot on the floor would result in phantom obstacles being permanently inscribed into the global map [28]. These persistent mapping errors severely constrained the robot's navigable space over time, forcing it into increasingly convoluted and inefficient detours.

By employing the safety scores, the mapping pipeline was able to filter incoming spatial data intelligently. Depth points associated with trajectories that received extremely low safety scores due to high epistemic uncertainty were discarded entirely, preventing the map from becoming corrupted by perceptual artifacts. Furthermore, the dynamic velocity scaling enabled by the continuous safety score resulted in much smoother and more natural navigational behavior. Instead of abrupt stops and chaotic re-planning characteristic of traditional reactive obstacle avoidance, the robot exhibited anticipatory caution. Upon entering a visually challenging area like a dimly lit hallway, the overall safety score would gradually decline, prompting the robot to reduce its speed smoothly. This cautious progression allowed the depth network more time to accumulate temporal information across multiple frames, ultimately improving the stability of the perception system. The simulation study conclusively illustrates that explicitly modeling and predicting safety bridges the semantic gap between raw pixel predictions and robust, reliable autonomous navigation.

5. Conclusion

5.1 Summary of Findings

This paper has presented a detailed simulation study addressing the critical challenge of predicting navigation safety for indoor mobile robots relying on monocular depth estimation networks. By developing a comprehensive testing framework within a high-fidelity physics simulator, we systematically exposed the vulnerabilities of deep learning-based perception in complex, visually ambiguous indoor environments. The research demonstrates that while depth estimation networks provide valuable dense structural information, their deterministic application directly into mapping and planning pipelines leads to unacceptable collision rates. The introduction of a dedicated safety prediction module, which synthesizes estimated spatial geometry with pixel-wise uncertainty and planned kinematics, proved highly effective in mitigating these risks. The empirical results clearly indicate that forecasting a safety score prior to execution allows the robot to proactively adjust its velocity and trajectory, resulting in a dramatic reduction in physical collisions across all tested environmental zones. Furthermore, utilizing this predictive framework significantly improved the long-term fidelity of the generated environmental maps by preventing the integration of highly uncertain perceptual artifacts.

5.2 Limitations and Future Directions

While the simulation-based methodology provides a safe and highly controlled environment for extensive testing, it is important to acknowledge the inherent limitations of the simulation-to-reality gap. Despite the advanced rendering techniques employed, simulated visual data lacks the full spectrum of unpredictable noise, camera sensor imperfections, and complex lighting dynamics present in physical reality. Consequently, a safety prediction module trained exclusively on simulated data may exhibit overconfidence or unexpected behaviors when deployed on actual hardware. The distribution of perceptual errors generated by the network in simulation may not perfectly mirror the errors it would make in the physical world.

Future research must focus on bridging this domain gap. One promising direction involves utilizing domain adaptation techniques to align the feature representations of simulated and real-world images, ensuring the depth network and the safety module generalize effectively [29]. Additionally, future work should explore the integration of temporal dynamics into the safety prediction module. Currently, the safety score is calculated primarily based on instantaneous single-frame predictions. Incorporating recurrent neural network structures or temporal attention mechanisms could allow the module to analyze the evolution of uncertainty over time, further enhancing its ability to distinguish between fleeting visual noise and persistent structural hazards.

Ultimately, the continuous refinement of predictive safety models will be paramount in realizing the widespread, reliable deployment of vision-only autonomous robots in complex human environments.

References

1. Kousis, I.; Manni, M.; Pisello, A.L. Environmental Mobile Monitoring of Urban Microclimates: A Review. Renew. Sustain. Energy Rev. 2022, 169, 112847.

2. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 2015, 52, 394–403.

3. Miao, T.; Ji, W.; Li, B.; Zhu, X.; Yin, J.; Yang, J.; Huang, Y.; Cao, Y.; Yao, D.; Kong, X. Advanced Soil Organic Matter Prediction with a Regional Soil NIR Spectral Library Using Long ShortTerm Memory–Convolutional Neural Networks: A Case Study. Remote Sens. 2024, 16, 1256.

4. Wang, H.; Zhang, L.; Zhao, J. Application of a Fusion Attention Mechanism-Based Model Combining Bidirectional Gated Recurrent Units and Recurrent Neural Networks in Soil Nutrient Content Estimation. Agronomy 2023, 13, 2724.

5. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; PMLR: New York, NY, USA, 2019; pp. 6105–6114.

6. Shi, Y.; Wei, P.; Feng, K.; Feng, D.C.; Beer, M. A survey on machine learning approaches for uncertainty quantification of engineering systems. Mach. Learn. Comput. Sci. Eng. 2025, 1, 11.

7. Xiong, Z.; Cui, Y.; Liu, Z.; Zhao, Y.; Hu, M.; Hu, J. Evaluating Explorative Prediction Power of Machine Learning Algorithms for Materials Discovery Usingk-Fold Forward Cross-Validation. Comput. Mater. Sci. 2020, 171, 109203.

8. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. IEEE Trans. Syst. Man. Cybern. 1973, SMC-3, 610–621.

9. Olakanmi, S.J.; Jayas, D.S.; Paliwal, J. Applications of imaging systems for the assessment of quality characteristics of bread and other baked goods: A review. Compr. Rev. Food Sci. Food Saf. 2023, 22, 1817–1838.

10. Louis, S.-Y.; Zhao, Y.; Nasiri, A.; Wang, X.; Song, Y.; Liu, F.; Hu, J. Graph Convolutional Neural Networks with Global Attention for Improved Materials Property Prediction. Phys. Chem. Chem. Phys. 2020, 22, 18141–18148.

11. Fan, X.; Zhang, M.; Mujumdar, A.S. Intelligent food processing: Journey from artificial neural network to deep learning. Compr. Rev. Food Sci. Food Saf. 2023, 22, 399–440.

12. da Silva Cotrim, W.; Minim, V.P.R.; Felix, L.B.; Minim, L.A. Short convolutional neural networks applied to the recognition of the browning stages of bread crust. J. Food Eng. 2020, 277, 109916.

13. Honorato, M.; Coelho, A.P.; Fernandes, C.; Matheus, S.; Claudia, M. Estimation of soil organic matter content by modeling with artificial neural networks. Geoderma 2019, 350, 46–51.

14. Piñeiro-Juncal, N.; Mateo, M.Á.; Leiva-Dueñas, C.; Serrano, E.; Inostroza, K.; Soler, M.; Apostolaki, E.T.; Lavery, P.; Duarte, C.M.; Lafratta, A.; et al. Soil organic carbon depth profiles and centennial and millennial decay rates in tidal marsh, mangrove and seagrass blue carbon ecosystems. Commun. Earth Environ. 2025, 6, 504.

15. Ng, W.; Minasny, B.; Jeon, S.H.; McBratney, A. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Secur. 2022, 6, 100043.

16. Narayanan, A.; Verma, S.; Ramadan, E.; Babaie, P.; Zhang, Z.L. DeepCache: A Deep Learning Based Framework For Content Caching. In Proceedings of the 2018 Workshop on Network Meets AI & ML, Budapest, Hungary, 24 August 2018; pp. 48–53.

17. Dong, W.; Liu, G.; Yu, J.; Hu, W.; Liu, X. SFDC: File Access Pattern Aware Cache Framework for High-performance Computer. In Proceedings of the 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, New York, NY, USA, 24–26 August 2015; pp. 342–350.

18. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

19. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran. Remote Sens. 2020, 12, 2234.

20. Chinilin, A.; Savin, I.Y. Combining machine learning and environmental covariates for mapping of organic carbon in soils of Russia. Egypt. J. Remote Sens. Space Sci. 2023, 26, 666–675.

21. Li, Y.; Feng, H. Integrating Urban Building Energy Modeling (UBEM) and Urban-Building Environmental Impact Assessment (UB-EIA) for Sustainable Urban Development: A Comprehensive Review. Renew. Sustain. Energy Rev. 2025, 213, 115471.

22. John, K.; Abraham Isong, I.; Michael Kebonye, N.; Okon Ayito, E.; Chapman Agyeman, P.; Marcus Afu, S. Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. Land 2020, 9, 487.

23. Bhattoo, R.; Bishnoi, S.; Zaki, M.; Krishnan, N.M.A. Understanding the Compositional Control on Electrical, Mechanical, Optical, and Physical Properties of Inorganic Glasses with Interpretable Machine Learning. Acta Mater. 2023, 242, 118439.

24. Willmott, C.; Matsuura, K. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Clim. Res. 2005, 30, 79.

25. Chen, X.; Yang, J. Urban Climate Monitoring Network Design: Existing Issues and a Cluster-Based Solution. Build. Environ. 2022, 214, 108959.

26. Brenning, A. Spatial prediction models for landslide hazards: Review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 2005, 5, 853–862.

27. Franjic, D.; Schweikert, K. Predictor Preselection for Mixed-Frequency Dynamic Factor Models: A Simulation Study with an Empirical Application to GDP Nowcasting. J. Forecast. 2025, 44, 255–269.

28. Pacci, S.; Dengiz, O.; Alaboz, P.; Saygın, F. Artificial neural networks in soil quality prediction: Significance for sustainable tea cultivation. Sci. Total Environ. 2024, 947, 174447.

29. Guo, P.T.; Wu, W.; Sheng, Q.; Li, M.F.; Liu, H.; Wang, Z.Y. Prediction of soil organic matter using artificial neural network and topographic indicators in hilly areas. Nutr. Cycl. Agroecosystems 2013, 95, 333–344.