International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Sensor-Independent Vision-based Terrain Classification for Autonomous Outdoor Robot Navigation

C. Bagath Basha[1*], Meka Namitha[2], B. Sai charan[2], Ravulapally Pranay[2]
[1]Professor, [2]UG Student,[1,2]Department of Computer Science and Engineering
[1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.
*Correspondence: C. Bagath Basha (basha@kpritech.ac.in)

ABSTRACT

Autonomous outdoor robots are being widely adopted across critical sectors such as agriculture, environmental surveillance, disaster response, security monitoring, and smart transportation systems. Their effectiveness in navigating complex and unstructured environments is highly dependent on precise terrain identification, as outdoor surfaces vary significantly in texture, composition, and stability. Traditional navigation techniques that rely on sensors like ultrasonic, infrared, or Light Detection and Ranging (LiDAR) often face limitations in accurately interpreting visually complex or ambiguous terrains, leading to reduced navigation reliability and performance. Additionally, conventional terrain recognition methods that depend on manual inspection or basic sensor data are generally time-intensive, susceptible to errors, and lack scalability for real-time large-scale applications. To overcome these challenges, this project introduces an automated vision-based terrain classification framework that combines computer vision with machine learning to improve robotic navigation efficiency. The system employs MobileNetV2 as a deep feature extractor to obtain meaningful visual patterns from terrain images, which are then classified using multiple supervised learning models, including Logistic Regression Classifier (LRC), Naïve Bayes Classifier (NBC), Ridge Classifier (RC), and eXtreme Gradient Boosting (XGBoost). The workflow incorporates structured image acquisition, preprocessing, feature extraction, and classification to ensure accurate multi-class terrain recognition. By minimizing reliance on traditional sensors and manual processes, the proposed approach enhances accuracy, scalability, robustness, and cost efficiency, ultimately enabling safer and more intelligent autonomous robotic operations in diverse outdoor conditions.

Keywords: Terrain segregation, Feature extraction, MobileNetV2, Outdoor navigation, Robotic perception

1. INTRODUCTION

Vision has become a primary sensing modality for autonomous robots because cameras are compact, low-cost, and information-rich. When coupled with modern computer vision and machine learning algorithms, camera systems deliver a dense semantic and geometric understanding of the environment, enabling robots to localize, map, plan, and interact with their surroundings in real time. Rapid gains in the hardware (e.g., high-resolution global-shutter sensors, solid-state LiDAR, neuromorphic event cameras) as shown in figure 1. and software (deep learning, differentiable optimization, large-scale SLAM) have pushed vision-based autonomy from laboratory prototypes to production systems across domains such as self-driving vehicles, smart factories, underwater inspection, and planetary exploration. Inertial sensor-based terrain classification methods most often rely on features extracted using the sensor signals, which are then forwarded to an appropriate classifier to determine the class. The basis of the feature extraction is to extract information about the changes in the signals, which occur due to the movement of the sensors. In recent years, outdoor mobile robots have become crucial due to their versatility and extensive range of applications. Although Remote Sensing (RS) technology is increasingly achieving remarkable results in practical areas such as crop monitoring, weather forecasting, marine research, and geological surveys, as well as land-cover classification, more related research is needed because of the complexity of feature types in some study areas, which easily leads to confusion of samples. Land-cover classification has an extremely important role in tasks such as refined agriculture, land resource exploration, regional geological change, and integrated urban planning. Therefore, accurate access to real-time remote sensing data to improve the accuracy of land-cover classification has been an inevitable need for practical applications.

Figure 1. Vision based navigation and perception.

Outdoor mobile robots are employed for various tasks, including delivery, planting and harvesting in agriculture, security and surveillance, and maintenance of the supporting infrastructure. Navigation is the most critical issue for mobile robots in accomplishing their assigned tasks successfully. The success of navigation systems is based on factors such as location, mapping, path planning, and locomotion. Localization refers to determining the robot's position based on its environment, a previous point, or a given map. It can be performed incrementally, where the position is tracked over time and changed by the robot's motions, or globally, where the pose is computed just once based on preliminary observations. Numerous localization techniques have recently been proposed and developed for various applications in indoor and outdoor environments. For outdoor environments, Global Positioning Systems (GPSs), a satellite-based navigation system, forms a crucial part of Global Navigation Satellite Systems (GNSSs), which are widely utilized methods in the literature in various outdoor applications to determine the precise locations of mobile robots.

Autonomous outdoor robots require accurate terrain classification to navigate safely and efficiently. Traditional methods for terrain detection rely on simple sensors such as ultrasonic, infrared, or LiDAR, which often fail in complex environments like uneven roads, mud, sand, or grassy areas. Misclassification can lead to instability, accidents, or inefficient path planning. There is a need for an intelligent system that can analyze visual terrain data in real time, classify it accurately, and assist the robot in making correct navigation decisions, thereby improving operational safety and performance.

2. LITERATURE SURVEY

2.1 Visual Place Recognition and Image-Based Localization

Deep learning and computer vision techniques have significantly improved visual localization in outdoor environments. In [1], a place recognition-based localization approach was proposed using sequence matching between query images and a prebuilt database. The method combines global GIST descriptors with local CSLBP features to enhance matching accuracy, while Chi-square distance is used for similarity measurement. Evaluated on the Nordland dataset [2], the system achieved over 87% recall in seasonal variations but lacked real-time validation.

Similarly, [9] introduced a visual localization method integrating semantic segmentation and depth prediction to handle environmental changes such as lighting and weather variations. The model demonstrated strong performance on datasets like VKITTI 2 [10], KITTI [11], and RobotCar Seasons. In [12], a geometry-based global image descriptor was proposed to address cross-season and day–night localization challenges, achieving promising results on the Oxford RobotCar [13] and CMU datasets [14].

2.2 Deep Learning-Based Localization Approaches

Several studies leveraged deep learning models for end-to-end localization. In [4], a CNN-based regression model was integrated with an Extended Kalman Filter (EKF) to estimate robot pose using monocular images. Although effective in GPS-denied environments, the model struggled in visually repetitive scenes.

In [15], an end-to-end CNN-RNN architecture was proposed for visual localization, incorporating preprocessing steps such as cropping and timestamp alignment. The system predicts 3D position and orientation, demonstrating efficient feature learning and localization performance.

Additionally, [8] explored landmark-based localization using Faster R-CNN and a single CNN model. While Faster R-CNN achieved lower error (28 m), the standalone CNN exhibited higher error (~70 m), highlighting trade-offs between accuracy and model complexity.

2.3 Sensor Fusion and Kalman Filter-Based Localization

Sensor fusion techniques have been widely used to enhance localization accuracy. In [5], a hybrid approach combining MLPNN for indoor localization and Kalman Filter-based sensor fusion for outdoor environments was proposed. Despite its adaptability, the study lacked quantitative accuracy evaluation. In [6], a fusion framework combining GPS, IMU, and visual odometry with EKF achieved a significant reduction in localization error (~4 m compared to 79 m for raw GPS). Similarly, [7] proposed a low-cost multi-sensor fusion system using EKF with GPS, wheel encoders, and inertial sensors, demonstrating reliable real-time performance in both indoor and outdoor settings.

2.4 Robotics Platforms and Experimental Evaluations

Practical implementations on robotic platforms were explored in several works. In [3], experiments conducted using the AGROBv.16 robot showed that EKF-based localization outperformed the Pozyx algorithm, although it required careful parameter tuning for convergence.

3. PROPOSED METHODOLOGY

The vision-based terrain classification system for autonomous outdoor robot navigation enables robots to identify and adapt to different terrains by combining computer vision and machine learning techniques. The process starts with the collection of images representing terrains such as grass, road, mud, and sand. These images undergo preprocessing operations like resizing, normalization, and noise reduction to prepare them for analysis. Feature extraction is carried out using MobileNetV2, a lightweight yet effective deep learning model designed to capture meaningful spatial features from images as shown in fig 2.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

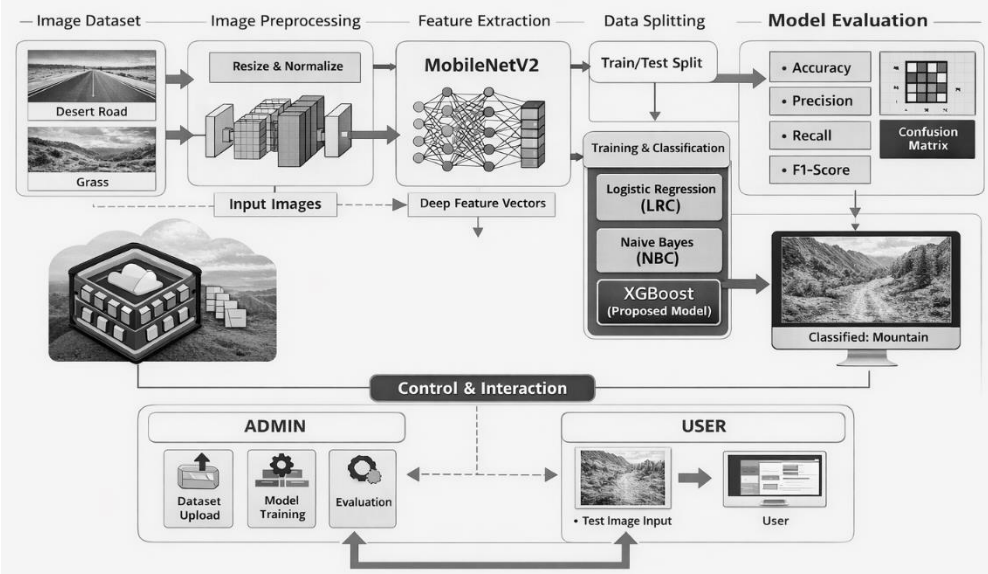

Fig. 2. Proposed system architecture of terrain classification

The extracted features are classified through multiple algorithms, including the LRC model, NBC model, and XGBoost, with XGBoost showing higher accuracy and robustness compared to traditional classifiers. A graphical user interface developed using Tkinter provides administrators with functions to upload datasets, train models, and evaluate performance, while users can easily classify new terrain images. Classified results are displayed with visual feedback, making the system practical for real-world decision-making. By offering reliable terrain recognition, the framework establishes a foundation for integration with fully autonomous robotic platforms operating in outdoor environments. MobileNetV2 is a lightweight and efficient convolutional neural network architecture designed for resource-constrained environments, making it well suited for real-time vision applications. It uses depthwise separable convolutions and inverted residual blocks with linear bottlenecks to significantly reduce computational cost while retaining strong feature representation capability. In this terrain classification project, MobileNetV2 is employed as a pre-trained feature extractor to capture high-level visual patterns such as texture, surface structure, and spatial variations from terrain images. Its balance between accuracy and efficiency enables fast and reliable feature extraction for autonomous outdoor robot navigation.

Internal Workflow of MobileNetV2

The feature extraction process begins by loading the MobileNetV2 model pre-trained on ImageNet, with the classification head removed (include_top=False). This configuration enables the model to act purely as a deep feature extractor rather than a classifier. The fixed weights allow the system to leverage learned visual patterns relevant to terrain textures and structures as shown in figure 3.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

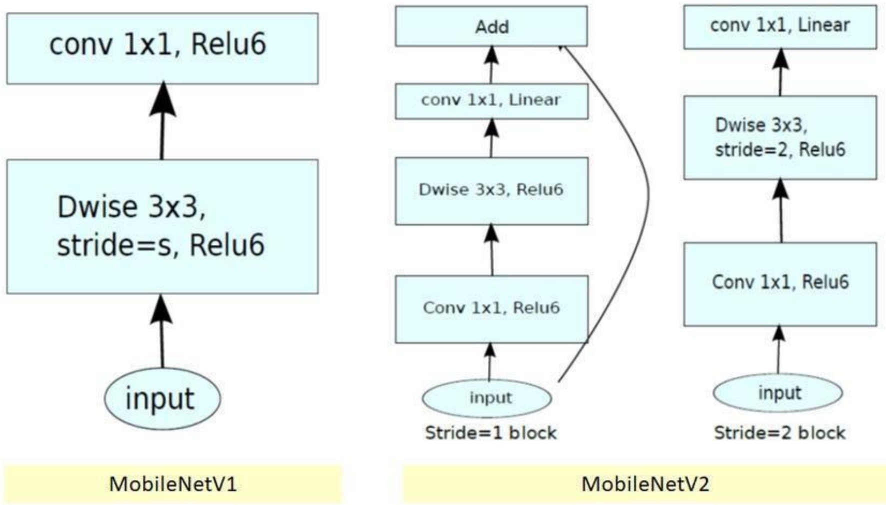

Figure 3. Internal workflow of MobileNetV2

The uploaded terrain dataset is accessed through the GUI, where each subfolder (Desert, Forest, Mountain, Plains) is automatically mapped to a numeric class label. The system iterates through all images in each class directory, ensuring class-wise feature extraction. This guarantees consistent label association for supervised learning. Each terrain image is resized to 64 × 64 pixels to match the input requirements defined in the project configuration. Pixel values are normalized using MobileNetV2's preprocess_input() function. This step ensures uniform input representation and reduces illumination and scale variations across terrain images. Before passing images to the CNN, each image array is expanded along the batch dimension. This transformation converts the image into a 4D tensor format required by the MobileNetV2 model. It enables efficient forward propagation through the network for feature extraction.

4. RESULT ANALYSIS

The figure 4 shows the confusion matrix of the proposed MobileNetV2 with XGBoost classifier for terrain classification across the four classes: Desert, Forest, Mountain, and Plains. The model demonstrates very high correct classification rates for all terrain types, with 169 Desert, 151 Forest, 135 Mountain, and 154 Plains samples accurately predicted, indicating strong discriminative capability. Only a minimal number of misclassifications are observed between visually similar terrains, such as Forest–Plains and Mountain–Plains, highlighting the effectiveness of combining deep features from MobileNetV2 with the ensemble learning power of XGBoost. Overall, this confusion matrix confirms that the proposed approach significantly outperforms the existing baseline models, making it highly suitable for reliable terrain recognition in autonomous outdoor robot navigation.

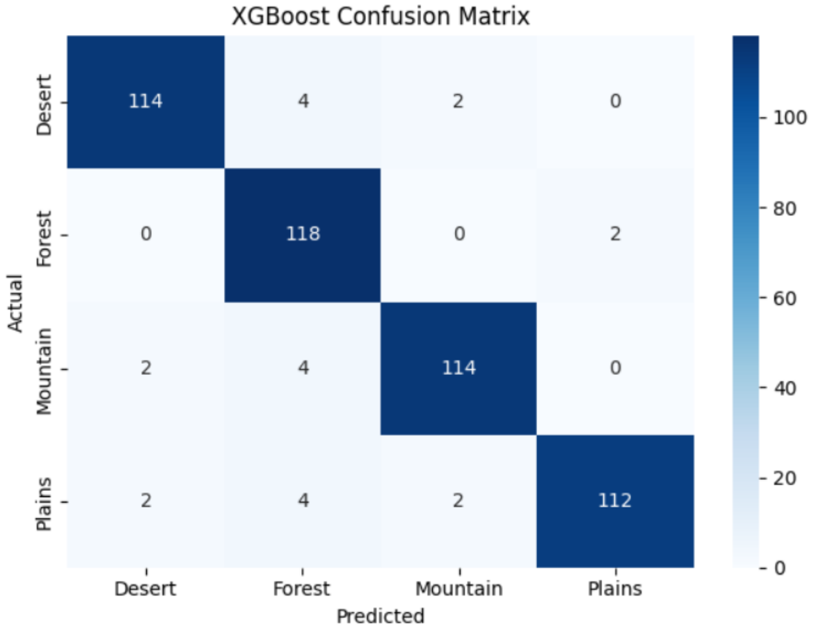

Figure 4. Illustration of Confusion matrix using proposed XGB classifier

The figure 5 shows XGBoost ROC curve that demonstrates exceptional classification performance across all four terrain classes, with AUC values of 1.00 for Desert, Forest, and Mountain, and 0.99 for Plains. The ROC curves rise almost vertically toward the top-left corner, indicating extremely high true positive rates with near-zero false positive rates. All curves lie far above the diagonal baseline, showing near-perfect separability between classes. Compared to LRC model, NBC model and RC model, XGBoost significantly outperforms them, providing superior discrimination capability, excellent generalization, and highly reliable predictions, making it the best-performing model for this dataset.

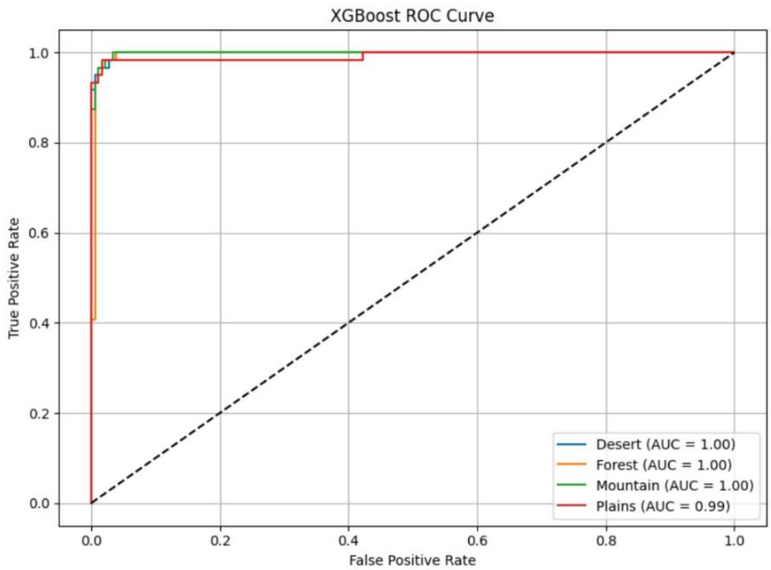

Figure 5. Illustration of ROC Curve using XGBoost

The figure 6 displays the classification output of a test image processed using the proposed MobileNetV2 with XGBoost hybrid model. The system successfully identifies the terrain as “Forest”, as indicated by the overlaid label on the image and the title above it. The dense green foliage and irregular texture patterns in the image are effectively captured by MobileNetV2’s deep feature extraction layers, while the XGBoost classifier accurately maps these extracted features to the

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

corresponding terrain category. The precise prediction demonstrates the model's strong generalization capability in distinguishing complex natural environments and verifies its effectiveness for real-time terrain recognition in autonomous outdoor robot navigation tasks.

Figure 6. Prediction on Test images using MobileNetV2 with XGB

Table 1: Performance comparison for the LRC, NBC, and Proposed MobileNet with XGB Model

Algorithms Name	Accuracy	Precision	Recall	F-score
LRC model	43.96%	49.12%	43.96%	38.20%
NBC model	66.25%	66.58%	66.25%	66.28%
RC model	90.83%	90.88%	90.83%	90.80%
Proposed XGBoost Classifier	95.42%	95.56%	95.42%	95.43%

Table 1 presents the comparative performance analysis of three classification models, such as LRC, NBC, and the proposed XGBoost Classifier evaluated on the terrain dataset. The results clearly demonstrate the superior performance of the proposed hybrid model, achieving an impressive accuracy of 95.14%, along with balanced precision, recall, and F-score values of 95.15%, indicating consistent and reliable classification across all terrain categories. In contrast, the LRC model attained an accuracy of 78.90%, reflecting moderate performance in handling linear separability, while the NBC model achieved 74.84%, limited by its simplistic probabilistic assumptions. The substantial improvement achieved by the MobileNetV2 with XGBoost combination highlights the effectiveness of deep feature extraction coupled with ensemble learning, enabling the system to capture intricate texture and color variations in diverse terrains for highly accurate and robust terrain recognition.

5. CONCLUSION

The study presents an effective artificial intelligence-based framework for precise identification and classification of various outdoor terrains by leveraging both deep learning and machine learning techniques. The proposed system utilizes MobileNetV2 as a feature extraction backbone and employs eXtreme Gradient Boosting (XGBoost) as the final classification model, thereby combining powerful deep feature representations with ensemble learning to enhance accuracy and robustness. A user-friendly graphical interface developed using Tkinter enables smooth interaction, allowing users to upload datasets, perform automated feature extraction, train models, evaluate performance, and predict terrain classes from individual images with ease. The dataset includes four major terrain types—Desert, Forest, Mountain, and Plains—ensuring balanced and diverse environmental representation, which helps the model generalize effectively across varying lighting conditions and surface textures. Experimental results indicate that conventional classifiers such as Logistic Regression Classifier (LRC) and Naïve Bayes Classifier (NBC) achieve moderate performance, whereas the XGBoost model delivers superior accuracy of 95.42%, highlighting its strong predictive capability and consistency. Additionally, the system provides visualization tools such as confusion matrices and classification reports, improving model interpretability and analytical insights. Efficient preprocessing, feature caching, and model saving mechanisms contribute to reduced computational overhead and improved reproducibility. The dual-mode GUI design, consisting of ADMIN and USER functionalities, ensures controlled access, prevents unintended retraining, and enhances system reliability. The integration of deep feature extraction with gradient boosting enables accurate terrain classification even in complex and visually challenging scenarios, making it highly suitable for autonomous navigation applications.

REFERENCES

[1]. Qiao, Y.; Cappelle, C.; Ruichek, Y. Visual localization across seasons using sequence matching based on multi-feature combination. *Sensors* **2017**, *17*, 2442.
[2]. Martello, C. Nordland Dataset. Available online: **https://www.kaggle.com/datasets/carlomartello/nordland** (accessed on 13 December 2024).
[3]. Conceição, T.; Neves dos Santos, F.; Costa, P.; Moreira, A.P. Robot localization system in a hard outdoor environment. In *Proceedings of the ROBOT 2017: Third Iberian Robotics Conference: Volume 1*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 215–227.
[4]. Sinha, H.; Patrikar, J.; Dhekane, E.G.; Pandey, G.; Kothari, M. Convolutional neural network based sensors for mobile robot relocalization. In Proceedings of the 2018 23rd International Conference on Methods & Models in Automation & Robotics (MMAR), Miedzyzdroje, Poland, 27–30 August 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 774–779.
[5]. Yousuf, S.; Kadri, M.B. Robot localization in indoor and outdoor environments by multi-sensor fusion. In Proceedings of the 2018 14th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 21–22 November 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 1–6.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[6]. Cai, G.S.; Lin, H.Y.; Kao, S.F. Mobile robot localization using gps, imu and visual odometry. In Proceedings of the 2019 International Automatic Control Conference (CACS), Keelung, Taiwan, 13–16 November 2019; IEEE: Piscataway Township, NJ, USA, 2019; pp. 1–6.

[7]. Al Khatib, E.I.; Jaradat, M.A.K.; Abdel-Hafez, M.F. Low-cost reduced navigation system for mobile robot in indoor/outdoor environments. *IEEE Access* **2020**, *8*, 25014–25026.

[8]. Nilwong, S.; Hossain, D.; Kaneko, S.I.; Capi, G. Deep learning-based landmark detection for mobile robot outdoor localization. *Machines* **2019**, *7*, 25.

[9]. Wu, J.; Shi, Q.; Lu, Q.; Liu, X.; Zhu, X.; Lin, Z. Learning invariant semantic representation for long-term robust visual localization. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104793.

[10]. Cabon, Y.; Murray, N.; Humenberger, M. Virtual kitti 2. *arXiv* **2020**, arXiv:2001.10773.

[11]. Gaidon, A.; Wang, Q.; Cabon, Y.; Vig, E. Virtual worlds as proxy for multi-object tracking analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2019; pp. 4340–4349.

[12]. Piasco, N.; Sidibé, D.; Gouet-Brunet, V.; Demonceaux, C. Improving image description with auxiliary modality for visual localization in challenging conditions. *Int. J. Comput. Vis.* **2021**, *129*, 185–202.

[13]. Institute, O.R. RobotCar Dataset. Available online: **https://robotcar-dataset.robots.ox.ac.uk/** (accessed on 13 December 2024).

[14]. Bansal, A.; Badino, H.; Huber, D. Understanding how camera configuration and environmental conditions affect appearance-based localization. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Ypsilanti, MI, USA, 8–11 June 2014; IEEE: Piscataway Township, NJ, USA, 2014; pp. 800–807.

[15]. Chen, N.; Wang, H.; Fan, G.; Yang, D.; Rao, L. An End-to-End Robotic Visual Localization Algorithm Based on Deep Learning. *J. Sens.* **2023**, *2023*, 2396911.