Robotica (2025), **43**, pp. 1314–1330
doi:10.1017/S0263574725000219

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Deep learning-based landmark detection and localization for autonomous robots in outdoor settings

M. Anuradha[1], V. Bibin Christopher[2], Francis H. Shajin[3] and S. V. Annlin Jeba[4]

[1]Department of Computer Science and Engineering, S.A. Engineering College, Tamil Nadu, India
[2]Department of Computing Technologies, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nādu, India
[3]Department of Electronics and Communication Engineering, Xpertmindz Innovative Solutions Private Limited, Kuzhithurai, Tamil Nadu, India,
[4]Department of Computer Science and Engineering, Sree Buddha College of Engineering, Pattoor, Kerala, India
Corresponding author: M. Anuradha; Email: anuparini@gmail.com

Received: 12 January 2024; **Revised:** 7 November 2024; **Accepted:** 9 January 2025; **First published online:** 16 April 2025

Keywords: enhanced capsule generation adversarial network; dual interactive Wasserstein generative adversarial network; robot localization; rat swarm optimization; landmark detection

Abstract

Navigation is an important skill required for an autonomous robot, as information about the location of the robot is necessary for making decisions about upcoming events. The objective of the localization technique is "to know about the location of the collected data." In previous works, several deep learning methods were used to detect localization, but none of them gives sufficient accuracy. To address this issue, an Enhanced Capsule Generation Adversarial Network and optimized Dual Interactive Wasserstein Generative Adversarial Network for landmark detection and localization of autonomous robots in outdoor environments (ECGAN-DIWGAN-RSO-LAR) is proposed in this manuscript. Here, the outdoor robot localization dataset is taken from the Virtual KITTI dataset. It contains two phases, which are landmark detection and localization. The landmark detection phase is determined using Enhanced Capsule Generation Adversarial Network for detecting the landmark of the captured image. Then the robot localization phase is determined using Dual Interactive Wasserstein Generative Adversarial Network (DIWGAN) for determining the robot location coordinates as well as compass orientation from identified landmarks. After that, the weight parameters of the DIWGAN are optimized by Rat Swarm Optimization (RSO) algorithm. The proposed ECGAN-DIWGAN-RSO-LAR is implemented in Python. The efficiency of the proposed ECGAN-DIWGAN-RSO-LAR technique shows higher accuracy of 22.67%, 12.45 %, and 8.89% compared to the existing methods.

1. Introduction

Nowadays, the robots' utilization is increasing in several fields due to their applications [1]. The robots are mainly used in the fields of logistics, medical, agriculture, health caring, and housekeeping [2, 3]. To complete the tasks at a particular time, the robots mainly select the navigation process [4, 5]. Due to the improvement of the navigation process, the robots got success in various factors with the localization process [6, 7]. Localization is the process to determine the robot's position in a particular landmark. By knowing their exact location, vehicles can plan their route to the next destination, navigate the environment safely, and ultimately complete their mission successfully [8, 9]. To determine data from the outdoor environment, Global Positioning Systems (GPS) are primarily used with various types of mobile robots. These include robots for higher voltage transmission line inspection, multi-aerial vehicles for autonomous position control, mobile robots for gas level mapping, and vigation systems that utilize GPS along with Inertial Navigation Systems [10, 11]. Additionally, simultaneous localization and mapping techniques are employed for firefighting robots [12, 13]. In the field of large-scale

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

implementation, localization with GPS reduces accuracy during various environmental conditions [14, 15]. Based on government information and satellite signal blockage by huge objects in the environment, like buildings, trees, indoor, or underground use, the signal reflects via buildings or walls[16–18]. So, deep learning-dependent robot localization is used to attain more accuracy [19–21]. In the previous works, several deep learning methods were used to locate and detect the robot localization, but no one method provided sufficient accuracy. These drawbacks have prompted to do this research.

An ECGAN-DIWGAN-RSO-LAR of autonomous robots in outdoor environments is proposed in this manuscript. The outdoor robot localization data is taken through the KITTI database. The landmark detection phase is determined using Enhanced Capsule Generation Adversarial Network (ECGAN) for detecting the landmark in the captured image. Then the robot localization phase is determined using Dual Interactive Wasserstein Generative Adversarial Network (DIWGAN) for determining the coordinates of the robot position with compass orientation from detected landmarks. After that, the weight parameters of the DIWGAN are increased under Rat Swarm Optimization (RSO) [22–23].

The primary contributions of this research are:

- Propose an innovative technique combining ECGAN, DIWGAN, and RSO for outdoor robot localization and landmark recognition.
- The method incorporates ECGAN for accurate landmark detection in captured images and DIWGAN for determining robot location coordinates along compass orientation based on identified landmarks.
- Optimize the weight parameters of DIWGAN using RSO to enhance the accuracy of robot localization.
- Evaluate the proposed approach using performance metrics, and performance is evaluated with other models.

The remaining portion of this manuscript are organized as follows: Section 2 presents the literature survey, Section 3 elucidates the proposed approach, Section 4 exemplifies the outputs, and Section 5 concludes this manuscript.

2. Literature survey

Among various studies associated with robot detection and localization, some recent works are revised here.

Li et al. [24] have suggested a deep sensor fusion for mobile robot localization (RCNN-IMU-LAR) by 2D laser scanner and IMU. The recurrent convolutional neural network was used for fuzing the laser and inertial data to determine the pose in the robot location. RCNN-IMU was presented to localize the robot in an efficient manner in outdoor environments. The weight parameters of the RCNN were optimized by a scan-to-submap optimization approach. It reached less error rate but more computation time.

Chen et al. [25] have presented an enhanced Siamese network to estimate autonomous cars (LIDAR-MCL-LAR) by pairs of LIDAR scans. The deep neural network was used to exploit the dissimilar cues produced from LIDAR data, and it determined the similarities among the cars. The Monte-Carlo localization method was used for global localization to predict the Overlap Net and integrate it for optimal localization of robots. It attained less computation time but more error rates.

Li et al. [26] have presented a multiple task feature extraction network-based deep-learning real-time visual simultaneous localization with mapping (SLAM-CNN-LAR). SLAM methods were used to map and localize the robot in the correct location in unknown environments. CNN was used for predicting the points and descriptors. It provides low cosine proximity (CP) but a high error rate.

Wen et al. [27] have introduced deep reinforcement learning-based path planning for active SLAM. The fully convolutional residual networks were utilized to distinguish the obstruction for getting the depth image. The evasion of obstruction path was intended to duel DQN on the robot's navigation. Concurrently, the two-dimensional map of the environment was built dependent on fast SLAM. It achieved higher accuracy but lower precision.

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

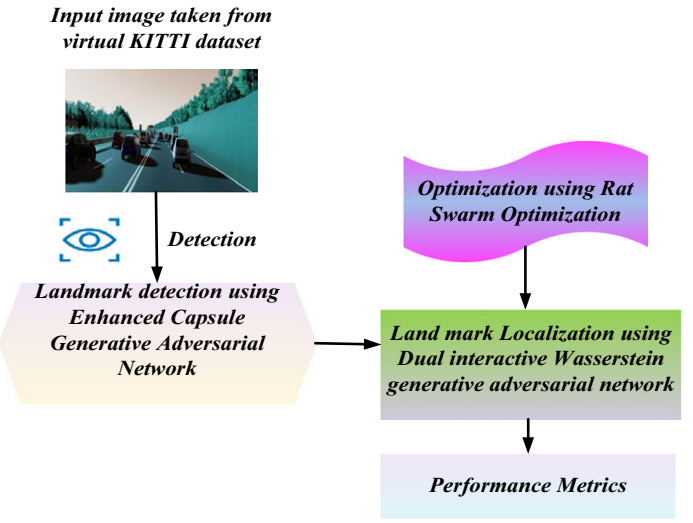


Figure 1. *Proposed ECGAN-DIWGAN-RSO-LAR methodology.*

Chhikara et al. [28] have suggested DCNN optimized with genetic algorithms for autonomous vehicle localization. The main problem of robot localizations was changing the environmental conditions so that the robot could not collect the data in an efficient manner. To localize the robot in correct position was another challenging problem. So DCNN optimized with genetic algorithms was presented to localize the robot in an efficient manner in outdoor environments. Images were used to predict and localize the robot. Then the hyperparameters of the DCNN were optimized using genetic algorithm. It provides a low error rate but high accuracy.

Hu et al. [29] have suggested domain adaptation on the mobile robot localization for semantic with geometric-aware imagery. The aim was to detect the long-term visual localization because it was not captured by the robot due to the changing environment. With the help of the presented method, the robot can capture good quality ground truths without the effort of human using the KITTI dataset. It achieved high accuracy but high mean squared error (MSE).

Tian et al. [30] have suggested 3D quadric landmark construction-based approach for a robust object-oriented SLAM. In this approach, stereo visual SLAM including robust quadric landmark detection method was presented. It contains four parts: deep learning identification, object-oriented data association, dual quadric landmark initialization, and object base pose optimization. To accurately predict the object pose assessment that means local observations, the object data-related approach and object-oriented optimization were used. It provides low mean precision but high absolute error.

3. Proposed ECGAN-DIWGAN-RSO-LAR methodology

In this section, ECGAN-DIWGAN-RSO-LAR is discussed. Here, the robot localization dataset is taken from the Virtual KITTI dataset [31]. Figure 1 portrays the block drawing of the ECGAN-DIWGAN-RSO-LAR technique. Here, the landmark detection phase is determined utilizing ECGAN for detecting the landmark in the captured image. Then the robot localization phase is determined using the DIWGAN for determining the robot position coordinates together with compass orientation from identified landmarks. After that, the weight parameters of the DIWGAN are enhanced using Rat Swarm Optimization.

3.1. Data acquisition

To detect the robot localization, the input dataset is taken from the dataset of Virtual KITTI. This is a photo-realistic synthetic video database to analyze computer vision for object detection, multiple object

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

tracking, scene level, event-level semantic segmentation, optical flow, and depth assessment [32, 33]. This dataset consists of 50 higher-resolution monocular videos created from 5 dissimilar virtual worlds under dissimilar imaging and weather situations in urban settings.

3.2. Enhanced capsule generative adversarial network for landmark detection

In this work, ECGAN is used to detect the landmarks in a particular location. ECGAN offers significant advantages for landmark detection in autonomous robot localization. By combining capsule networks' robustness to spatial transformations and detailed feature representation with the Generative Adversarial Network (GAN) framework's adversarial training benefits, ECGAN enhances the accuracy and resilience of landmark detection. The dynamic routing between capsules allows the network to handle changes in orientation, scale, and perspective effectively, while the GAN component helps the model learn robust features even in noisy or occluded environments. This results in improved localization accuracy, better generalization to diverse outdoor settings, and efficient use of data through possible data augmentation, making ECGAN a powerful and adaptable solution for detecting and localizing landmarks in challenging outdoor scenarios. ECGAN has two networks: (i) discriminator and (ii) generator. It is employed for detecting the landmark in the captured image. Generator is specified as $\hat{H}: M \to D$, where $\hat{H}$ indicates generator, M implies noise space, D refers to the data space, and the discriminator $C: D \to \{0, 1\}$ separates the class of captured images that can be 0 to 1. Therefore, ECGAN method is expressed in Eq. (1):

$$\underset{\hat{H}}{low}\underset{\hat{C}}{high}\grave{w}\left(\hat{C}, \hat{H}\right) = F_{\hat{C}} + F_{\hat{H}} \tag{1}$$

$$\begin{aligned} F_{\hat{C}} &= M_{y \approx z_{data}(y)}\left[\log\left(F_{\hat{C}}\right)\right] \\ F_{\hat{H}} &= M_{x \approx z_x(x)}\left[\log\left(1 - D\left(F_{\hat{H}}\right)\right)\right] \end{aligned} \tag{2}$$

where $z_{data}(y)$ denotes the data distribution and $z_x(x)$ denotes the noise distribution. Each of the several sheets that make up the capsule-net is separated into neurons called capsules. The capsule x and y on layer l and layer above $l + 1$, in which v_x signifies capsule a of vector output and w_y represents capsule y of vector output that has X_y input vector and X_y represented as squash, the w_y can be expressed in Eq. (3):

$$w_y = X_y = \frac{\left\|X_y\right\|^2 X_y}{1 + \left\|X_y\right\|^2 X_y} \tag{3}$$

The ECGAN of discriminator and generator is expressed in Eq. (4):

$$\begin{aligned} \hat{C} &= high\left(\hat{W}j\left(F_{\hat{C}}\right),\ Y_l = 0\right)\} + \max\ \ \hat{W}j\left(F_{\hat{C}}\right),\ Y_{l+1} = 1)\} + \hat{W}_{pv} \\ \hat{H} &= high\ \ \hat{W}j\left(F_{\hat{C}}\right), Y_l = 0\} \end{aligned} \tag{4}$$

A discriminator$(\hat{C}, \hat{H})$ is defined in Eq. (5):

$$\hat{C}, \hat{H} = \sum Y_l\, high\left(0, h^+ - \|n_l\|\right)^2 + \alpha\left(1 - Y_{l+1}\right) high\left(0, \|n_l\| - h^-\right)^2 \tag{5}$$

where n_l represents final output vector, y_l epitomizes input vector created from the generator, T_k epitomizes target value, h^+, h^-represents batch samples, and α denotes batch size.

Then the final landmark detected data is given in Eq. (6):

$$\hat{C}, \hat{H} = \arg\ \ \underset{C}{low}\underset{H}{high}\ \left(\hat{C}, \hat{H}\right) \tag{6}$$

where the discriminator $\hat{C}, \hat{H}$ is determined on training h and $\hat{C}and\ \ \hat{H}$ formulate high-low process where $\hat{H}$ tries to loss $\hat{C}$ when $\hat{H}$ is qualified to discriminate between generated bounding boxes and the labels.

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

3.3. Dual interactive Wasserstein generative adversarial network for landmark localization

DIWGAN is employed to localize the labels of robot location to produce the data location. The data location contains location coordinates and compass direction for locating the coordinates, and the direction is in a magnetic-referenced compass orientation manner. DIWGAN offers distinct advantages for landmark localization in autonomous robot navigation. Its robustness to input variability ensures reliable detection across diverse environmental conditions, while its generative modeling capabilities enable precise representation of landmarks, crucial for accurate localization. The dual discriminator architecture enhances training stability and convergence, refining localization predictions effectively. When optimized with techniques like RSO, DIWGANs further improve performance by fine-tuning parameters for optimal results. They inherently capture spatial relationships and orientation information, essential for precise navigation planning. Compatible with deep learning frameworks, DIWGANs facilitate seamless integration and scalability, supporting advanced applications in real-world autonomous systems. DIWGAN classifier transmits the input images toward the spike train under the time-to-first-spike code procedure. After that, the network receives these spikes. In this case, the output layer's binary neurons in the invisible layer won't produce more than one image, causing a system error. To prevent this, the temporal backpropagation algorithm is used in conjunction with DIWGAN to update the synaptic weights, which localizes the labeled bounding box and the location landmark data. Data location contains latitude and longitude angles with magnetic compass orientation for determining the robot location coordinates as well as compass orientation from identified landmarks. For intensity of location $[0;L_{high}]$, input neuron time j_k, related to pixel intensity jth, L_j, is exhibited in Eq. (7):

$$j_k = \left[\frac{L_{high} - L_j}{L_{high}} - h_{high} \right] \tag{7}$$

where the term h_{high} is the high striking time. The input neurons with high pixel intensities have minimal spike delay. Hence the spike pattern of the input neuron jth is derived in Eq. (8):

$$Q_j^0 (j) = \begin{cases} 1 \ if \ j = j_k \\ 0 \ otherwise \end{cases} \tag{8}$$

Let ith neuron of kth layer gets arriving spikes by -1 or + 1 weights:

$$W_i^k (j) = W_i^k (j-1) + \beta^k \sum_j C_{ij}^k Q_j^{k-1} \tag{9}$$

where Q_j^{k-1} symbolizes input spike pattern,W_i^k represents vas potential membrane,C_{ij}^k represents binary synaptic weight associated with neuron j, and β^k is a scaling parameter shared between all kth layer neurons. The neuron strikes once the threshold ϑ crossed through possible membrane $Q_i^k(h)$:

$$Q_i^k (h) = \begin{cases} 0 & if \ W_i^k (h) \geq \vartheta \& Q_i^k (< h) \neq 1 \\ 1 & otherwise \end{cases} \tag{10}$$

where $Q_i^k(< h) \neq 1$ denotes neuron non-strikes at any previous equations.

$$f_i = \left(H_i^0 - j_k \right) / h_{high} \tag{11}$$

where h_i^0 specifies real striking time, H_i^0 specifies target striking time for the output neuron ith, reset voltages as 0 for input image, and perform the simulation in h_{high} is expressed in Eq. (12):

$$H_i^0 = \begin{cases} \varepsilon - \eta & if \ i = j, \\ \varepsilon + \eta & if \ i \neq j \& j_k < \varepsilon + \eta \\ h_i^0 & if \ i \neq j \& j_k \geq \varepsilon + \eta \end{cases} \tag{12}$$

where η is an optimistic constant, $\varepsilon + \eta$, and $\varepsilon = low \left\{ h_i^0 \ 1 \leq i \leq C \ \right\}|$ denotes a minimal striking period on the output layer belonging to the $i$thcategory. In Eq. (13), the scaling factor φ^k is labeled:

$$\phi^k = \phi^k - \delta \frac{\psi G}{\psi \beta^k} \tag{13}$$

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

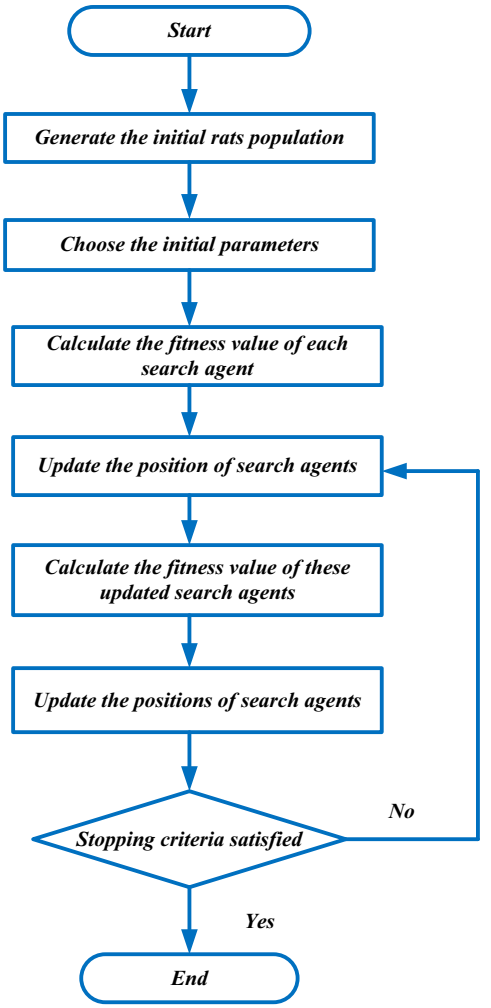


Figure 2. *Flowchart of Rat Swarm Optimization algorithm.*

where δ implicates the learning ratio parameter and $\frac{\psi G}{\psi \beta^k}$ denotes gradient value loss functioning of weights. The weight parameters are used to increase the accuracy by reducing the error rate based on the RSO algorithm.

3.4 Optimize parameters of DIWGAN using Rat Swarm Optimization

RSO is a new nature-inspired meta-heuristic approach that is used for dealing with constrained and global optimization issues. The RSO is active to tune fighting and chasing behaviors based on the DIWGAN classifier. Hence, RSO is proposed to identify fault categorization with maximal accuracy of the turning process. Figure 2 shows the flowchart of the RSO algorithm. The stepwise process of DIWGAN with the help of the RSO approach is delineated in the following:

***Step 1*:** Initialization

The rat population $\hat{P}_j$ is initialized and characterized the solution of DIWGAN, j ranging from $(1to\ n)n$ denotes the population size and $\hat{P}_j$ represents rat position.

***Step 2*:** Random location

The input parameters are formed randomwise after that initialization. Then the greater fitness values depends on specific behavior setting $\hat{P}_s$, the best optimal solution is represented as $\hat{P}_s$.

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

***Step 3*:** Fitness function for optimizing the parameters of $\frac{\psi G}{\psi \beta^k}$

The random solution count is created from initialization. It is used to decrease the objective function, optimizing the weight $\frac{\psi G}{\psi \beta^k}$ of DIWGAN, as given in Eq. (14):

$$Fitnessfunction = optimize\left(\frac{\psi G}{\psi \beta^k}\right) \tag{14}$$

From Eq. (14), new behavior is enhanced over the current best behavior.

***Step 4*:** Balanced exploration and exploitation

This step aims to balance the exploration of new solutions and the exploitation of known good solutions. While both exploration and exploitation are crucial, their effectiveness may vary depending on the iteration. The balanced exploration and exploitation is expressed in Eq. (15):

$$Z' = P - x \times \left(\frac{z}{Max_{Iteration}}\right) \tag{15}$$

where z indicates the location of prey, Z' represents a parameter for balanced exploration with exploitation over the sequence of iterations, and P is a random number between $(1, 5)$, $x = 0, 1, 2, \ldots, Max_{Iteration}$.

***Step 5*:** Update the positioning of rat to optimize the weight and bias parameter $\frac{\psi G}{\psi \beta^k}$

The updated position is depending on rat positions articulated in Eq. (16):

$$Q_s\,(x+1) = \left|\hat{Q}_s\,(x) - q\right| \tag{16}$$

The updated rat's next position denotes $Q_s(x+1)$, which protects the best solution and informs the position of extra search agents.

Step 6: Termination

RSO enhances the value of the DIWGAN parameter for enhancing the accuracy without error. The process iteratively repeats step 3 until it satisfies the halting criteria time iteration $T = T + 1$, which includes either reaching the maximum number of iterations or meeting the convergence threshold. The process stops after 100 iterations or when the halting criteria are fulfilled.

4. Result and discussion

In this segment, ECGAN-DIWGAN-RSO-LAR of autonomous robots in outdoor environments is described. The aim is to emulate a landmark detection and localization, typically with lower heights. The neural network is trained with 0.00025 learning rate for 4000 epochs, using a batch size of two images, on a PC with a 2.50 GHz CPU, Intel Core i5, 8 GB RAM, and Windows 7. Epochs generally refer to training/fine-tuning performance over successive complete passes through the training dataset, where the model updates its parameters to improve its performance on the entire dataset collectively. The proposed ECGAN-DIWGAN-RSO-LAR is implemented in Python utilizing the Virtual KITTI benchmark dataset. The effectiveness of the ECGAN-DIWGAN-RSO-LAR is analyzed under the performance metrics. The obtained outcomes of the ECGAN-DIWGAN-RSO-LAR method are compared with three existing approaches, such as deep sensor fusion among two-dimensional laser scanners with IMU for mobile robot localization (RCNN-IMU-LAR), Siamese network for calculating the similarity of LiDAR scan including applications to loop closing along localization (LIDAR-MCL-LAR), and deep-learning real-time visual SLAM depending on multiple task feature extraction network with self-supervised feature points (SLAM-CNN-LAR), respectively. Figure 3 shows the output of robot detection and localization using the Virtual KITTI dataset.

Figure 3 illustrates the output results of robot detection and localization using the Virtual KITTI dataset, comparing the proposed ECGAN-DIWGAN-RSO-LAR method with three existing approaches: RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR. It visually demonstrates that the ECGAN-DIWGAN-RSO-LAR method achieves superior performance compared to RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR in landmark detection and localization tasks. The results

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

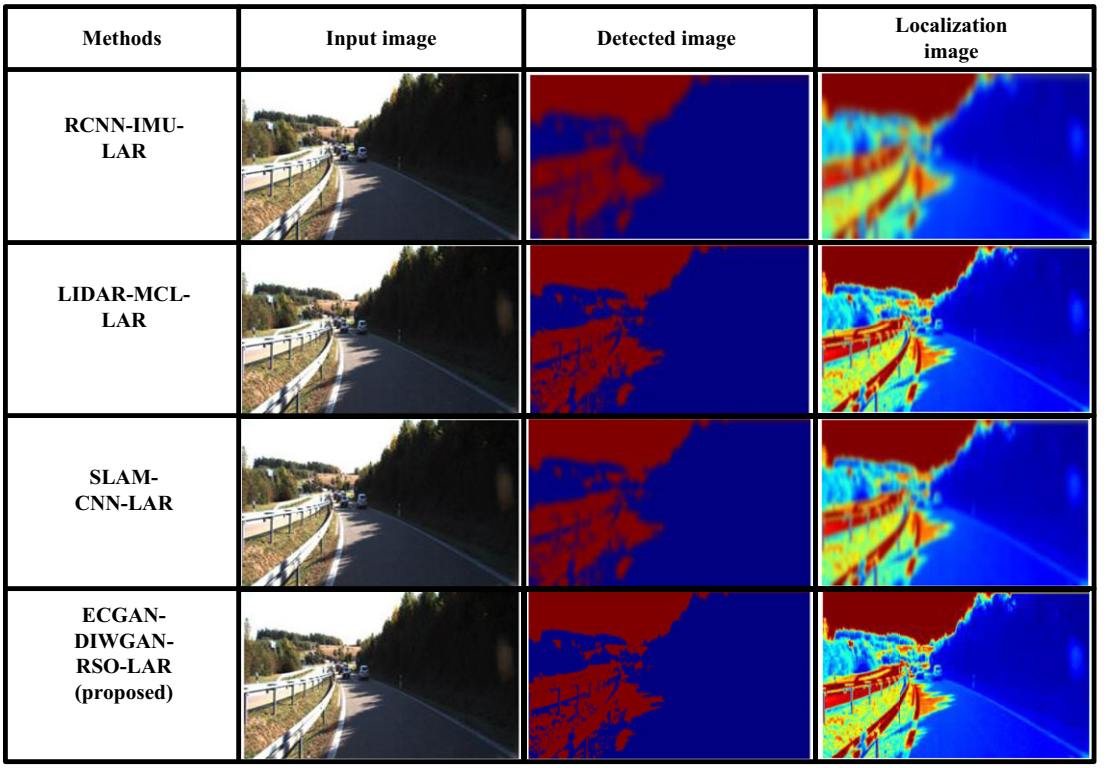


Figure 3. *Output results of robot detection and localization using Virtual KITTI dataset.*

highlight the proficiency of the ECGAN-DIWGAN-RSO-LAR approach in enhancing the robustness and accuracy of autonomous robot navigation systems in complex outdoor environments, establishing it as a promising advancement over existing state-of-the-art methods.

4.1. Performance metrics

The performance metrics, like MSE, CP, accuracy, precision, recall, Intersection over Union (IoU), computational efficiency, error rate, computational time, and F1-score are examined to analyze the efficiency of the ECGAN-DIWGAN-RSO-LAR method.

4.1.1. Mean squared error (MSE)

This is often employed in regression analysis to assess the quality of a predictive mode. The average squared variation amid the estimated and real values. This is calculated by Eq. (17):

$$MSE = \frac{1}{n} \times \sum_{i=1}^{n} \left| (k_i - l_i)^2 \right| \tag{17}$$

where k is represented as the outcome from localization, l denotes the reference value at the test database, n denotes the number of data at the testset, and i implies the number of images.

4.1.2. Cosine proximity (CP)

CP is the measure of similarity among two vectors that is reference value and the count of data in the testset. It is computed by Eq. (18):

$$CP = \frac{\sum_{i=1}^{n} k_i * l_i}{\sqrt{\sum_{i=1}^{n} k_i^2} \sqrt{\sum_{i=1}^{n} l_i^2}} \tag{18}$$

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

where k is represented as outcome from localization, l denotes reference value at the test database, n denotes number of data at the testset, and i implies number of images.

4.1.3 Accuracy

The accurate prediction rate to the total proceedings in the dataset using Eq. (19):

$$Accuracy = \frac{D_P + D_N}{D_P + D_N + R_P + R_N} \tag{19}$$

where (D_P)symbolizes true positive, (D_N) symbolizes true negative, (R_P) epitomizes the false positive, and (R_N) epitomizes false negative.

4.1.4 Precision

This is scaled as the classifier capacity to calculate normal data without conditions using Eq. (20):

$$\Pr ecision = \frac{D_P}{D_P + R_P} \tag{20}$$

where (D_P)denotes true positive and (R_P) represents false positive.

4.1.5. Recall

Recall is otherwise called sensitivity; it evaluates the efficacy of categorization or detection mode. It calculates the percentage of real positive examples among all positive instances in the dataset that the model correctly detected. It is calculated using Eq. (21):

$$\text{Re } call = \frac{D_P}{D_P + R_N} \tag{21}$$

where (D_P)denotes true positive and (R_N) describes the false negative.

4.1.6. F1-score

It assesses the accuracy of the method especially in classification tasks. This is the harmonic mean of precision and recall and balance the tradeoff among these two measurements and is determined by Eq. (22):

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{22}$$

4.1.7. Intersection over Union (IoU)

IoU evaluates the accuracy of object detection by scaling the overlap among the estimated bounding box and ground truth bounding box. The rate at which the expected and ground truth boxes' areas of intersection and union are equal is known as the IoU. It is expressed in Eq. (23):

$$IoU = \frac{Area\ of\ \text{intersection}}{Area\ of\ Union} \tag{23}$$

4.1.8. Computational efficiency

It is used to evaluate the effectiveness of a proposed model in terms of its accuracy relative to the computational resources it consumes. It is calculated by Eq. (24):

$$Computational\ \ efficiency = \frac{Accuracy}{Computation\ time} \tag{24}$$

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

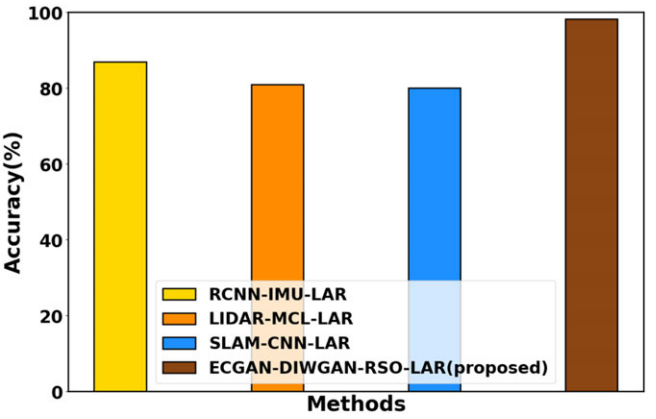


Figure 4. *Accuracy analysis.*

4.1.9. Mean absolute percentage error (MAPE)

MAPE computes the accuracy of a predictive model. This is the mean of the absolute percentage errors among the predicted and real values. It is computed by Eq. (25):

$$MAPE = \frac{1}{n} \times \sum_{i=1}^{n} \left| \frac{k_i - l_i}{k_i} \right| \tag{25}$$

where n denotes the number of observations, k_i specifies the actual value, and l_i specifies the forecasted value.

4.2. Experimental results

Figures 4–12 depict the simulation results of the proposed ECGAN-DIWGAN-RSO-LAR approach. Here, the proposed approach is analyzed with the existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR models.

Figure 4 presents the accuracy analysis of the ECGAN-DIWGAN-RSO-LAR framework, highlighting its robust performance in autonomous robot localization. By employing ECGAN for landmark detection and optimizing DIWGAN with RSO for localization, the framework achieves high accuracy in identifying landmarks and determining robot coordinates and orientation relative to these landmarks. The proposed method demonstrates substantial improvements, achieving 21.78%, 24.13%, and 31% better accuracy compared with existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR models, respectively. While the accuracy results indicate the system's strong performance in detecting landmarks and improving localization, further evaluation using additional metrics like localization error and orientation accuracy is necessary to fully justify the precise determination of robot coordinates and orientation relative to the detected landmarks.

Figure 5 displays precision evaluation. The precision of the ECGAN-DIWGAN-RSO-LAR method is systematically evaluated to gauge its capability in accurately identifying landmarks and localizing robots while minimizing false positives. Here, the proposed ECGAN-DIWGAN-RSO-LAR method attains 24.86%, 22.75%, and 26.98% better precision compared with the existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR, respectively.

Figure 6 depicts recall analysis, which assesses the method's ability to correctly identify landmarks and localize robots relative to the total number of actual positive instances. By calculating the ratio of true positive detections to the total of both false negatives and true positives, recall offers a measure of the method's capability to capture all relevant landmarks and determine robot locations. This evaluation is crucial for ensuring that the method effectively identifies landmarks critical for navigation in outdoor environments. The proposed ECGAN-DIWGAN-RSO-LAR method attains 24.01%,

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

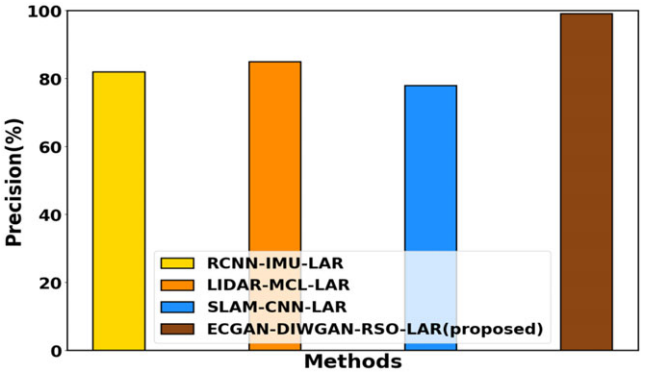


Figure 5. *Precision evaluation.*

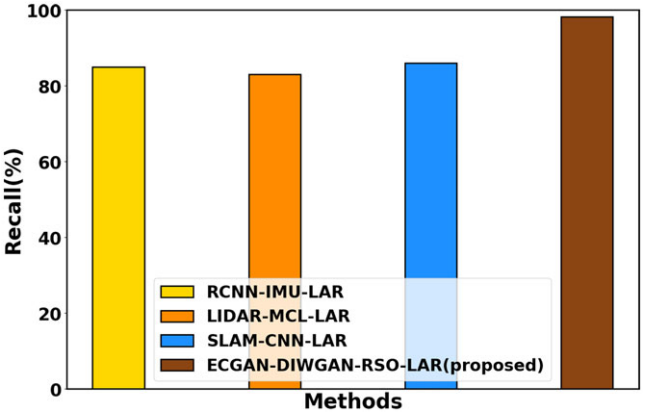


Figure 6. *Recall analysis.*

19.03%, and 24.04% greater recall compared to the existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR methods, respectively.

Figure 7 delineates the CP analysis. The performance analysis of CP in the context of landmark identification with localization of autonomous robots in outdoor environments serves to evaluate the method's ability to accurately determine the similarity or proximity between detected landmarks and reference landmarks in the environment. By quantifying the cosine of the angle between the vectors representing the features of detected landmarks and reference landmarks, CP provides a measure of how closely the detected landmarks match the expected landmarks. Here, the performance of the ECGAN-DIWGAN-RSO-LAR attains 27.12%, 22.65%, and 13.98% higher CP compared with existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR, respectively.

Figure 8 implicates the MSE analysis. MSE evaluates the accuracy of the proposed method. MSE quantifies the average squared difference among the predicted and ground truth landmark locations or robot coordinates, providing insights into the precision of the localization process. By computing MSE on a separate test dataset, the method's ability to accurately determine landmark positions and robot locations in outdoor environments can be thoroughly assessed. Here, the proposed ECGAN-DIWGAN-RSO-LAR method attains 23.78%, 28.72%, and 19.21% lower MSE compared with the existing methods like RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR, respectively.

Figure 9 displays the computation time analysis. Computation time analysis of the proposed ECGAN-DIWGAN-RSO-LAR method focuses on evaluating the efficiency of both landmark detection and robot localization tasks. In the landmark detection phase utilizing ECGAN and robot localization

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

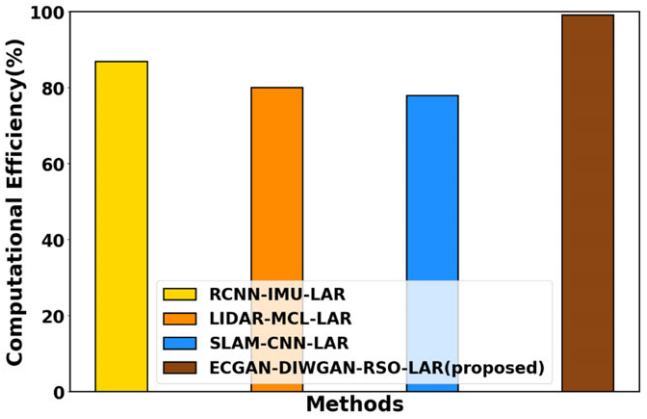


Figure 7. *Cosine proximity analysis.*

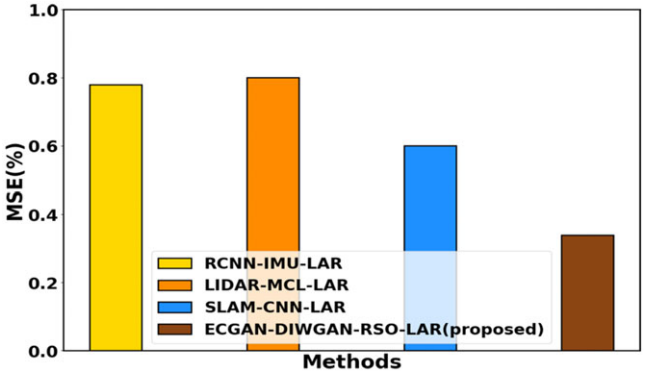


Figure 8. *Performance of MSE.*

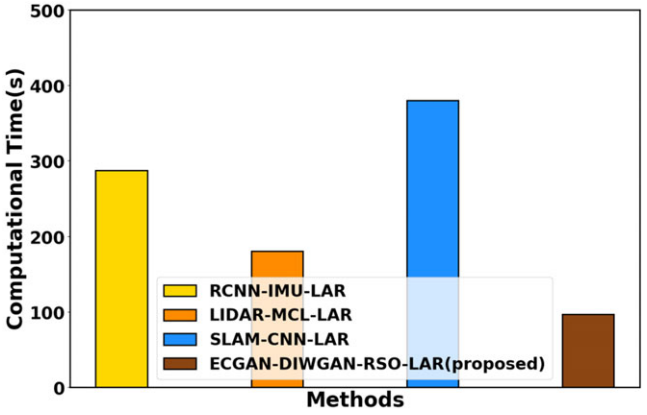


Figure 9. *Computational time assessment.*

using DIWGAN optimized with RSO, the computation time includes estimating the robot's position and orientation based on detected landmarks. Here, the proposed ECGAN-DIWGAN-RSO-LAR method attains 11.43%, 27.74%, and 30.67% less computation time than the existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR, respectively.

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

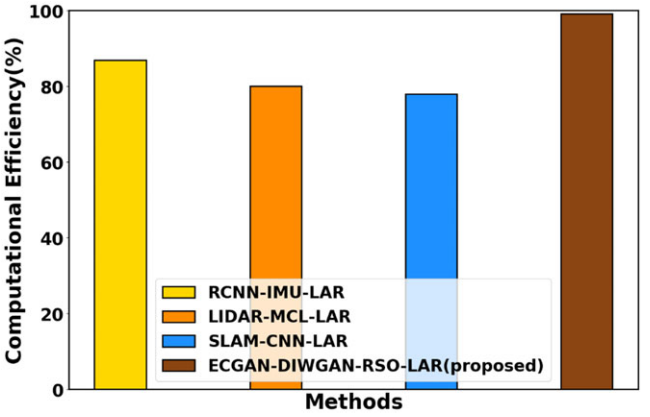


Figure 10. *Computational efficiency analysis.*

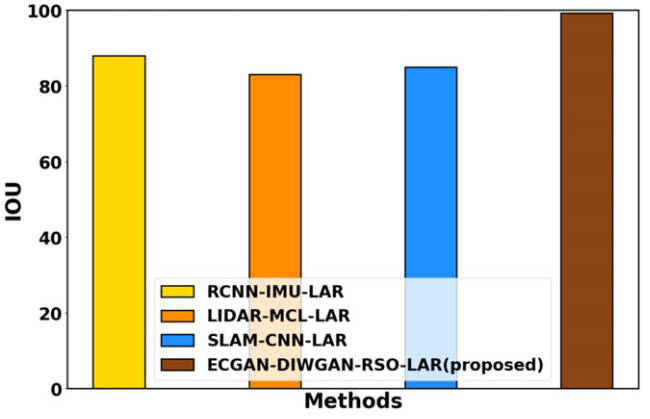


Figure 11. *Performance analysis of IoU.*

Figure 10 displays the computational efficiency analysis. Computational efficiency in the proposed ECGAN-DIWGAN-RSO-LAR framework refers to its ability to achieve accurate and reliable landmark detection and robot localization using optimized computational resources. This includes minimizing processing time, memory usage, and energy consumption while maintaining high performance levels. Techniques such as RSO for parameter tuning in DIWGAN and the efficient architecture of ECGAN contribute to streamlined operations and faster convergence, making the method suitable for real-time applications on autonomous robot platforms operating in dynamic outdoor environments. The proposed ECGAN-DIWGAN-RSO-LAR method achieves 12.98%, 28.04%, and 30% higher computational efficiency compared with the existing methods, like RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR, respectively.

Figure 11 shows IoU. In the context of the proposed approach, IoU quantifies how well the detected landmarks align with their actual positions. A higher IoU score indicates a greater overlap between the predicted and true landmark locations, reflecting the model's precision in localization. For instance, if a detected landmark's bounding box closely matches the ground truth location, IoU is high, signifying accurate detection. The proposed ECGAN-DIWGAN-RSO-LAR method achieves 23.15%, 21.04%, and 15.32% higher IoU compared with existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR models, respectively.

Figure 12 displays F1-score analysis. In the context of the proposed ECGAN-DIWGAN-RSO-LAR approach for autonomous robot localization, the F1-score is particularly relevant for evaluating the

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

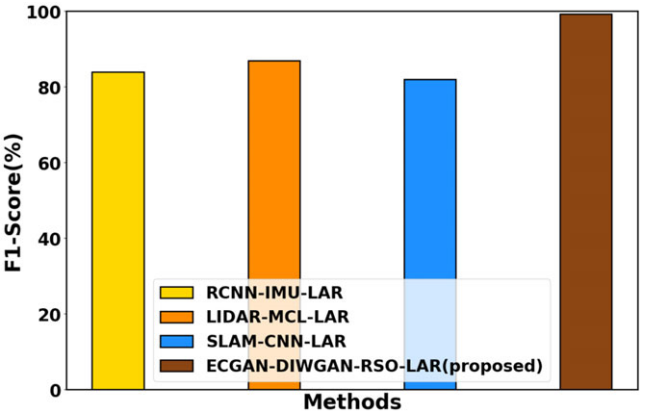


Figure 12. *F1-score analysis.*

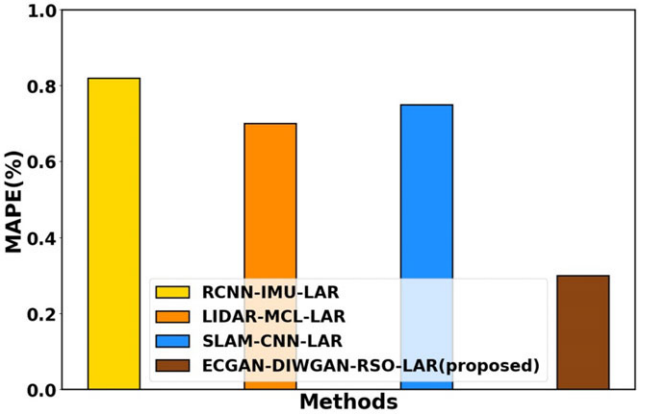


Figure 13. *MAPE analysis.*

model's effectiveness in detecting landmarks and accurately localizing robots in outdoor environments. Achieving a maximal F1-score in this context specifies that the proposed approach efficiently balances precision and recall, leading to robust and accurate autonomous navigation capabilities in complex outdoor environments. Here, the ECGAN-DIWGAN-RSO-LAR method achieves 20.15%, 19.04%, and 25.32% better F1-score analyzed to the existing RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR.

Figure 13 displays MAPE analysis. The lower MAPE achieved by the ECGAN-DIWGAN-RSO-LAR framework underscores its effectiveness in minimizing prediction errors and improving the overall reliability of autonomous robot localization in diverse outdoor settings. This comparative analysis emphasizes the advancements brought by the proposed method over existing techniques. Here the proposed ECGAN-DIWGAN-RSO-LAR method attains 23%, 29.07%, and 30% lower MAPE compared with existing methods, like RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR, respectively.

4.3 Discussion

The proposed ECGAN-DIWGAN-RSO-LAR framework represents a substantial advancement in autonomous robot localization by combining ECGAN for accurate landmark detection and DIWGAN for

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

precise robot localization. The proposed methodology is designed to handle various environmental conditions, as indicated by the use of diverse virtual worlds for training. The proposed ECGAN-DIWGAN-RSO-LAR method is compared with existing methods such as RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR. The existing models tend to fail in scenarios where environmental complexity and variability are high. One significant reason for their shortcomings is their reliance on traditional methods that do not effectively handle occlusions or indistinct landmarks. For instance, these models often misclassify landmarks that are partially obscured by obstacles or those that lack distinct features, leading to a higher rate of false positives and negatives. In contrast, the proposed ECGAN-DIWGAN-RSO-LAR method excels in these situations due to its innovative architecture that combines ECGAN for robust landmark detection and DIWGAN for precise localization. The ECGAN's ability to generate high-quality landmark representations allows it to better identify and differentiate between landmarks, even in cluttered environments. Moreover, RSO optimizes DIWGAN parameters, accelerating convergence and further boosting performance metrics. This is reflected in the precision metrics, where the proposed model achieves 24.86%, 22.75%, and 26.98% greater precision compared to the existing methods. The recall assessment further supports the proposed method's effectiveness in correctly identifying landmarks. Thus the ECGAN-DIWGAN-RSO-LAR method demonstrates significant improvements in localization accuracy compared to existing methods such as RCNN-IMU-LAR, LIDAR-MCL-LAR, and SLAM-CNN-LAR.

5. Conclusion

The proposed ECGAN-DIWGAN-RSO-LAR technique was successfully implemented in this manuscript. The proposed ECGAN-DIWGAN-RSO-LAR was implemented in Python. The ECGAN-DIWGAN-RSO-LAR technique attains 30.92%, 21.22%, and 12.35% less MSE and 25.96%, 26.17%, and 2.93% less MAPE, compared to the existing RCNN-IMU-LAR, LIDAR-MCL-LAR and SLAM-CNN-LAR models, respectively. Despite the advancements of the ECGAN-DIWGAN-RSO-LAR framework in autonomous robot localization, several limitations remain. Enhancing the proposed model's adaptability to real-time conditions and varying scenarios remains a challenge that needs to be addressed for improved practical applicability. Future work will focus on improving the model's adaptability to dynamic environments by incorporating online learning techniques and investigating hybrid optimization techniques to enhance the performance and efficiency of the proposed system. Also, future studies will explore the integration of edge computing solutions to further improve the real-time processing capabilities of the system.

Ethical approval and consent to participate. This article does not contain any studies with human participants performed by any of the authors.

Human and animal ethics. Not applicable.

Consent for publication. Not applicable.

Availability of supporting data. Data sharing does not apply to this article as no new data has been created or analyzed in this study.

Authors' contributions. Dr. M. Anuradha (corresponding author): Conceptualization, methodology, writing – original draft preparation.
Dr. V. Bibin Christopher: Supervision
Mr. Francis H Shajin: Supervision
Dr. S. V. Annlin Jeba: Supervision

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

Competing interests. Not applicable.

References

[1] P. K. Panigrahi and S. K. Bisoy, "Localization strategies for autonomous mobile robots: A review," *J. King Saud Univ.-Comput. Inf. Sci.* (2021).

[2] L. Wan, Y. Sun, L. Sun, Z. Ning and J. J. Rodrigues, "Deep learning based autonomous vehicle super resolution DOA estimation for safety driving," *IEEE Trans. Intell. Transp.* **22**(7), 4301–4315 (2020).

[3] Y. Ma, Z. Wang, H. Yang and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: A survey," *IEEE/CAA J. Automat. Sinica* **7**(2), 315–329 (2020).

[4] S. Chen, J. Dong, P. Ha, Y. Li and S. Labi, "Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles," *Comput.-AIDED Civil Infrast Inf.* **36**(7), 838–857 (2021).

[5] Z. Xiong, Z. Cai, Q. Han, A. Alrawais and W. Li, "ADGAN: Protect your location privacy in camera data of auto-driving vehicles," *IEEE Trans. Ind. Inf.* **17**(9), 6200–6210 (2020).

[6] Q. Luo, Y. Cao, J. Liu and A. Benslimane, "Localization and navigation in autonomous driving: Threats and countermeasures," *IEEE Wirel Commun.* **26**(4), 38–45 (2020).

[7] C. Li, Y. Fu, F. R. Yu, T. H. Luan and Y. Zhang, "Vehicle position correction: A vehicular blockchain networks-based GPS error sharing framework," *IEEE Trans. Intell. Transp.* **22**(2), 898–912 (2020).

[8] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez and C. Wellington, "3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Proc. Mag.* **38**(1), 68–86 (2020).

[9] M. A. Esfahani, H. Wang, K. Wu and S. Yuan, "AbolDeepIO: A novel deep inertial odometry network for autonomous vehicles," *IEEE Trans. Intell. Transp.* **21**(5), 1941–1950 (2019).

[10] H. Taghavifar, C. Hu, Y. Qin and C. Wei, "EKF-neural network observer based type-2 fuzzy control of autonomous vehicles," *IEEE Trans. Intell. Transp.* **22**(8), 4788–4800 (2020).

[11] C. Hu, Y. Chen and J. Wang, "Fuzzy observer-based transitional path-tracking control for autonomous vehicles," *IEEE Trans. Intell. Transp.* **22**(5), 3078–3088 (2020).

[12] I. Yaqoob, L. U. Khan, S. A. Kazmi, M. Imran, N. Guizani and C. S. Hong, "Autonomous driving cars in smart cities: Recent advances, requirements, and challenges," *IEEE Network* **34**(1), 174–181 (2019).

[13] L. Wang, X. Fan, J. Chen, J. Cheng, J. Tan and X. Ma, "3D object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities," *Sustain. Cities Soc.* **54**, 102002 (2020).

[14] C. Toft, W. Maddern, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, T. Pajdla and F. Kahl, "Long-term visual localization revisited," *IEEE Trans. Pattern Anal.* **44**(4), 2074–2088 (2020).

[15] Y. Guo, Y. Xu and S. Li, "Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network," *Automat. Constr.* **112**, 103124 (2020).

[16] B.R. Annapoorna, and D.R. Babu. "Detection and localization of cotton based on deep neural networks." Materials Today: Proceedings, **80** (2023) pp.3328-3332.

[17] B.R, Annapoorna; M. B, Mrs. Shanthi; and Mungara, Dr. Jitendranath "A Secure Packet Hiding Technique For Preventing Jamming Attacks," International Journal Of Smart Sensor And Adhoc Network: 3(1), (2012) Article 7.

[18] G.N. Nimeshika, and D., Subitha. "Enhancing Alzheimer's disease classification through split federated learning and GANs for imbalanced datasets." PeerJ Computer Science, 10 (2024) p.e2459.

[19] S. J. Fusic, G. Kanagaraj, K. Hariharan and S. Karthikeyan, "Optimal path planning of autonomous navigation in outdoor environment via heuristic technique," *Transp. Res. Interdiscip. perspect.* **12**, 100473 (2021).

[20] dataset-https://europe.naverlabs.com/research/computer-vision/proxy-virtual-worlds-vkitti-2/.

[21] D. P. Grabowsky, J. M. Conrad and A. F. Browne, "A Breadcrumb System for Assisting Outdoor Autonomous Robots with Path Identification and Localization"," In: *SoutheastCon* (IEEE, 2021) pp. 1–6.

[22] P. Shamsolmoali, M. Zareapoor, L. Shen, A. H. Sadka and J. Yang, "Imbalanced data learning by minority class augmentation using capsule adversarial networks," *Neurocomputing* **459**, 481–493 (2021).

[23] Z. Hu, H. Xue, Q. Zhang, J. Gao, N. Zhang, S. Zou, Y. Teng, X. Liu, Y. Yang, D. Liang and X. Zhu, "DPIR-Net: Direct PET image reconstruction based on the Wasserstein generative adversarial network," *IEEE Trans. Radiat. Plasma Med. Sci.* **5**(1), 35–43 (2020).

[24] C. Li, S. Wang, Y. Zhuang and F. Yan, "Deep sensor fusion between 2D laser scanner and IMU for mobile robot localization," *IEEE Sens. J.* **21**(6), 8501–8509 (2019).

[25] X. Chen, T. Läbe, A. Milioto, T. Röhling, J. Behley and C. Stachniss, "OverlapNet: A Siamese network for computing LiDAR scan similarity with applications to loop closing and localization," *Auton. Robot.* **46**(1), 61–81 (2022).

[26] G. Li, L. Yu and S. Fei, "A deep-learning real-time visual SLAM system based on multi-task feature extraction network and self-supervised feature points," *Measurement* **168**, 108403 (2021).

[27] S. Wen, Y. Zhao, X. Yuan, Z. Wang, D. Zhang and L. Manfredi, "Path planning for active SLAM based on deep reinforcement learning under unknown environments," *Intell. Serv. Rob.* **13**(2), 263–272 (2020).

[28] P. Chhikara, R. Tekchandani, N. Kumar, V. Chamola and M. Guizani, "DCNN-GA: A deep neural net architecture for navigation of UAV in indoor environment," *IEEE Internet Things J.* **8**(6), 4448–4460 (2020).

[29] H. Hu, Z. Qiao, M. Cheng, Z. Liu and H. Wang, "Dasgil: Domain adaptation for semantic and geometric-aware image-based localization," *IEEE Trans. Image Process.* **30**, 1342–1353 (2020).

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press

[30] R. Tian, Y. Zhang, Y. Feng, L. Yang, Z. Cao, S. Coleman and D. Kerr, "Accurate and robust object-oriented SLAM with 3D quadric landmark construction in outdoor environment. arXiv preprint 2110.08977 (2021).
[31] G. Dhiman, M. Garg, A. Nagar, V. Kumar and M. Dehghani, "A novel algorithm for global optimization: Rat swarm optimizer," *J. Amb. Intell. Hum. Comput.* **12**(8), 8457–8482 (2021).
[32] A. J. Wilson, W. S. Kiran, A. S. Radhamani and M. P. Bharathi, "Optimizing energy-efficient cluster head selection in wireless sensor networks using a binarized spiking neural network and honey badger algorithm," *Knowl-BASED Syst.* 112039 (2024).
[33] R. R. Kanth and T. P. Jacob, "Enhanced Capsule Generative Adversarial Network with Blockchain Fostered Intrusion Detection System for Enhancing Cyber security in Cloud," In: *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)* (IEEE, 2023) pp. 1–6.

Cite this article: M. Anuradha, V. Bibin Christopher, F. H. Shajin and S. V. Annlin Jeba (2025). "Deep learning-based landmark detection and localization for autonomous robots in outdoor settings", Robotica **43**, 1314–1330. https://doi.org/10.1017/S0263574725000219

https://doi.org/10.1017/S0263574725000219 Published online by Cambridge University Press