International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 40-52
https://doi.org/10.70917/ijcisim-2026-4195
Article

Framework for Intelligent navigation in Iraqi urban maps using deep reinforcement learning with LSTM

Ghassan Khazal Ali[1] , Maryam Jassim Mohammed[2], Moayad Awny Sabri[3]

[1,2,] Department of Computer Engineering, College of Engineering, Diyala University, Diyala, Iraq.
[3] Department of Computer Science, Diyala University, Diyala, Iraq
Email: ghassan_khazal@uodiyala.edu.iq, maryamjasim80@gmail.com, moayad1978@gmail.com

Abstract: Autonomous navigation is a complex mission in urban environments and represents an important challenge for intelligent transportation systems and autonomous vehicles, especially in dynamic traffic conditions, complex road topologies, sparse reward problems, and the need for long-term robust decision-making. The recent Deep Reinforcement Learning (DRL) gives promising performance in autonomous navigation in most existing studies that evaluated on a generic simulation environment and showed limited ability on an elastic road network, especially in developing regions. There is a great research gap in developing a navigation framework that is capable of working in complex real-world city environments.

This work aims to develop a navigation framework that improves navigation efficiency , convergence stability , and generation capability in dynamic Iraqi urban environments. The proposed framework integrates Proximal Policy Optimization (PPO) and Long Short-Term Memory (LSTM) with Curriculum learning to address the challenges of sparse rewards, dynamic obstacles, and decision making. To achieve these objectives, a custom dataset named IraqiUnav was constructed using a road network extracted from OpenStreetMap for six major Iraqi cities, such as Baghdad, Mosul, Basra, Najaf, Diyala , and Erbil. The navigation environment was simulated using Carla and the SUMO platform. First, we modeled the traffic conditions , obstacle distribution , and road network . The navigation problem was formulated as a goal-conditioned Markov Decision Process, and the proposed Curriculum LSTM-PPO was trained and evaluated across multiple navigation scenarios.

The experimental results showed that the designed framework achieved 91.4% success rate and reduced the collision rate to 3.5%. our work performs better than PPO, PPO-LSTM, SAC and classical A*. the integration of curriculum and LSTM improved temporal reasoning and navigation consistency in dynamic environments and cross city evaluation proven strong generalize ability across invisible cites maps.

The results indicate that combining oof deep learning and temporal mechanisms and curriculum training can significantly improve autonomous navigation performance in complex environments. The proposed work offers a practical and scalable solution for future intelligent transport systems and autonomous vehicles, especially in irregular road areas and dynamic conditions.

Keywords: NA

1. Introduction

One of the core abilities for intelligent transport system is autonomous navigation and mobile robot applications. The structured environment can compute feasible routes by searching an known map. In Iraq cities the environment introduce complex setting like traffic flow is often heterogenous , road can be narrow or irregular, instruction may be dense and dynamic obstacles like vehicle , motorcycle, pedestrian , and road work can change the usable path. These factors make navigation a sequential decision making problem rather than pure path problem. The agent must decide not only where to go but also compute path and how decide his move over time by minimizing collision and unnecessary convolution. Classical method like A* , Dijkstra and RRT remain useful baseline but they depend on static maps and not respond effectively to the changing conditions. DRL provide a data-driven by allowing the agent to learn navigation behavior through interaction with the simulated city map. DRL require many samples

and can fail when reword sparse or when the agent lacks memory of previous observation this work address this issue by integrating LSTM memory and curriculum target generation and curriculum controls the order in which destination are sampled by making LSTM enable temporal reasoning across routs.

Several studies proposed hybrid approaches that combine reinforcement learning with classical techniques to improve navigation safety and stability in dynamic urbans. These architectures have shown superior performance compared to learning based or rule-based methods.

Therefore , This work aim how can performance robustness of DRL for autonomous navigation be improved in complex and uncertain urban using curriculum learning strategies and how design effective reward function?

To address this curriculum learning has been introduced and the agents trained form simple tasks to complex. This paper demonstrates that spare to dense rewards transition can improve learning efficiency and enhance generalization and stability policy in complex environments.

2. Related work

Autonomous navigation has traditional relies on classical mapping and planning control. Thes methos usually construct geometric map the compute collision path using planner like PRM, RTT A* and Dijkstra. These approach are effective in structured and static environments the often depend on accurate maps and stable reading sensors. As result the performance can downgrading in dynamic environments and the sensors riding can be nosy.

DRL has emerged as alternative approach for navigation because it enables agent to learn navigation policy directly by interacting with the dynamic environment by mapping sensory observation to action without requiring large amount of data and fully predefined models.

Several recent studies have investigated DRL for navigation under realistic conditions. benchmarked DRL algorithm such as Twin Delayed DDPG (TD3), Proximal Policy Optimization (PPO-LSTM) and dreamrV3 in sensor denied environment and show that sensor noise, sensor failure and incomplete observation in navigation performance.

RUMOR introduced a DRL-based planner for navigation in dynamic environments, emphasized that real-world navigation is difficult because agents cannot be trained on every possible scenario , also showed limitations in many existing DRL systems including partial observability and insufficient consideration of robot dynamic constraints. RUMOR combines a descriptive model of the dynamic environment with DRL to improve robustness and interpretability.

Morad et al. proposed Curriculum learning has also become an important strategy for improving DRL training . an automatic curriculum learning method for visual navigation, showing that task selection can improve generalization and sample efficiency in real-world transfer. Their work supports the idea that navigation agents should not be trained using random or uniform goal sampling only and should instead receive tasks in organized progression difficulty.

Other studies focused on reward design and sparse reward problems. Freitag et al. proposed a two-stage reward curriculum to improve learning in environments with complex reward functions, shows that gradually moving from simpler reward structures to more complete reward objectives can improve stability and reduce undesirable local optimum. the reward formulation used in the proposed IraqiUrbanNav framework, where goal-reaching, collision avoidance, path efficiency, and smooth movement must be balanced.

park et al. studied sparse-to-dense reward transition in goal-oriented reinforcement learning. the results indicate that gradually shifting from sparse exploration-based rewards to denser goal-directed rewards improves success rate, sample efficiency, and generalization . This is directly relevant to large-scale Iraqi city maps, where sparse rewards may slow learning because the agent receives positive feedback only after reaching distant goals.

Hybrid learning-control approaches have also been proposed to improve safety and practical deployment. Gutiérrez-Moreno et al. developed a hybrid autonomous driving architecture that combines DRL-based decision-making with classical control methods. the results in CARLA show that combining learning-based decisions with reliable control modules can improve safety, comfort, and completion time in urban scenarios. This supports the use of hybrid simulation and control tools such as CARLA in the proposed methodology.

Large-scale city navigation has been studied in DeepNav Brahmbhatt and Hays developed a CNN-based navigation method using street-view images organized as city graphs. the dataset included ten city graphs and more than one million street-view images, and A* search was used to generate supervision labels. This work is important

for the present study because it demonstrates the value of graph-based city representations for training navigation agents in large urban environments.

we the proposed framework Compared with previous studies, our work focuses on Iraqi urban maps and introduces a curriculum-based LSTM-PPO framework for navigation in cities such as Baghdad, Basra, Diyala, Mosul, Erbil, and Najaf. Unlike studies that focus only on generic simulated maps and indoor environments. this research emphasizes Iraqi road topology, irregular street layouts, congestion, dynamic obstacles, and map uncertainty. The integration of LSTM memory, curriculum-based goal generation, and PPO optimization aims to improve long-term decision-making, reduce sparse reward problems, and enhance generalization across Iraqi city scenarios.

Table 1 summarizes the most relevant studies and explains how each one informs the proposed Iraqi-map navigation framework.

Ref.	**Study / Year**	**Core method**	**Environment / dataset**	**Main contribution**	**Limitation**	**Relevance to this work**
[1]	2017	CNN + A* supervision	10 city graphs, >1M street-view images	City-scale navigation using graph-structured visual data	Supervised setting, not Iraqi maps	Supports city-scale graph navigation motivation
[7]	2024	Confidence-based curriculum for MAPF	Multi-agent maps	Expands goal radius as agents improve	Multi-agent focus; no Iraqi case study	Supports adaptive target difficulty progression
[8]	2023	Cumulative curriculum DRL	Grid-based exploration maps	Improves sample efficiency and generalization	Exploration focus, not destination navigation	Supports cumulative curriculum design
[6]	2025	Reward curriculum	Goal-oriented RL tasks	Improves sample efficiency and generalization	Not urban-map specific	Supports reward shaping and sparse-to-dense training
[9]	2024	PPO with curriculum learning	Webots E-puck robot	Improves path planning and adaptability	Small robotic simulator	Supports PPO curriculum baseline
[10]	2023	BO-based curriculum	Autonomous racing	Improves robustness to environment variation	Racing domain only	Supports automated curriculum search
[11]	2021	RL + temporal occupancy grids	Narrow-lane autonomous driving	Models interaction with traffic participants	Focused on negotiation, not city maps	Supports temporal sequences and curriculum
[12]	2024	Off-policy RL + curriculum prioritization	Unknown robot environments	Improves sample efficiency and safety	Robot navigation rather than urban graphs	Supports curriculum prioritization
[13]	2021	Automatic curriculum navigation	Real/sim embodied navigation	Selects tasks using geometric features	Indoor/semantic targets	Supports geometric target difficulty

[14]	2025	Two-stage reward curriculum	Control and mobile robot tasks	Balances complex reward terms	Reward focus rather than map curriculum	Supports multi-term reward design
[15]	2024	DRL simulation analysis	Garage and obstacle-dense Unity tasks	Shows CL helps difficult DRL training	Low-speed AD only	Supports simulation/training design
[16]	2025	DRL with dynamic environment model	Dynamic robot navigation	Robustness in unseen crowded scenarios	Not curriculum-based	Supports DRL for dynamic environments
[17]	2022	Generative imitation + collision prediction	Indoor scenes and TurtleBot	Improves generalization and safety	Imitation learning, indoor focus	Supports target-driven navigation and safety modules

3. Methodology

Our methodology consists of five interconnected modules:

First: a map of Iraqi cities is converted into a grid or graphical representation, where nodes represent correct locations or intersections, and edges represent sections of roads.

Second: potential target destinations are extracted from the navigable nodes and filtered to remove inaccessible or simple targets.

Third: each target is assigned a difficulty level based on distance, route complexity, obstacle density, and topological bottlenecks.

Fourth: the targets are categorized into training levels from easiest to hardest.

Fifth: the system is trained using a convolutional neural network/graph encoder, followed by an LSTM layer and PPO vertices for the representative-critic.

During training, targets are selected from the current training level, and when the system achieves a sufficient success rate, it advances to the next level. This design allows the training distribution to adapt to the system's capabilities and minimizes early exposure to extremely difficult tasks.

Dataset construction

The road maps were transformed into graph based representations, the nodes represent intersections and edges represent navigable roads. The dataset includes GPS coordinates, traffic density, obstacle locations, RGB observations, and target destinations. More than 60,000 navigation states were generated for training and evaluation.

Table 2 Dataset statistics:

city	node	Edges
Baghdad	15,200	21,400
Basra	10,300	19,000
Mosul	12,700	16,600
Diyala	9,500	12,900
Najaf	9,400	14,600
Erbil	13,900	20,500

Simulation environment

The CARLA simulator was used to generate realistic autonomous driving and urban navigation scenarios. The environment consist of dynamic vehicles, pedestrians, traffic lights, RGB cameras, LiDAR sensors, and GPS observations. also the traffic simulator was integrated to simulate traffic congestion and dynamic urban traffic flow.

The environment is represented mathematically as:

$G = (V, E)$..... 1

Where:

V: intersections and navigation nodes.

E: roads and graph connections.

This part shows the methodical formulation used by the propsed framework. The proposed framework employs the (PPO) algorithm because of its stability and effectiveness in autonomous navigation tasks. The navigation problem is formulated as a Goal-Conditioned Markov Decision Process:

$S_t = [X_t, Y_t, V_t, O_t, G_t]$..... 2

Where :

Xt, Yt: current coordinates

Vt: vehicle velocity

Ot: obstacle information

Gt: target destination

the navigation task formulated as goal conditioned Markov decision process where S is the state space, P is the transition , R is the reward function , gamma is the discount factor and G is set of destination target.

The optimization objective is at each timestep t, the agent observation state St receive goal g ∈ G and select the action :

$a_t \sim \pi_\theta(a_t|S_t, g)$..... 3

The policy select an action on the current state and destination goal at time t.

$J(\theta) = E_g \sim p(g\}k), \pi_\theta[\sum_{t=0}^{T} \gamma^t(s_t, a_t, g)]$..... 4

The optimization objective is to maximize discounted return under the current curriculum goal distribution.

$$R(s_t, a_t, g) = +R_{goal} \text{ if } (d(s_t, g) \leq \varepsilon \ ; \ r_{step} \text{ pre step; } -r_{collision} \text{ if collision}$$

Then reward function combines successful goal arrival, step cost, and collision penalty.

$R_{shaping}(s_t, g) = -\tau d_{geo}(s_t, g)$..... 5

the distance shaping term provides denser learning feedback while preserving the goal *nature of the task.*

$R_{total} = R(s_t, a_t, g) + R_{shaping}(s_t, g) + R_{smooth}$..... *6*

The total reward includes task reward, distance shaping, and a smoothness term for reducing unstable movement.

$D(g) = \propto d(g) + \beta c(g) + \eta o(g) + \delta m(g)$..... *7*

difficulty of the target destination is computed using geodesic distance, path complexity, obstacle density, and topological score.

$G = G_1 \cup G_2 \cup \ldots \cup G_k$ *8*

Then selected targets are partitioned into curriculum levels from easy to difficult.

$g \in G_k \ , \quad T_k \leq D(g) < T_{k+1}$ *9*

A goal belongs to curriculum level k if its difficulty lies within the corresponding interval.

$P(g|K) = \frac{exp(-\tau|D(g)-\mu_k|)}{\sum_{g_j} exp(-\tau|D(g_j)-\mu_k|)}$..... *10*

The process of sampling objectives within the curriculum level gives priority to destinations close to the average difficulty of that level.

$X_t = [\emptyset(S_t), \omega(g), a_{t-1}, r_{t-1}]$..... *11*

The LSTM entry combines features of encrypted state, target inclusion, previous action, and previous reward.

$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$..... 12

Input gate of the LSTM.

$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$..... 13

Forget gate of the LSTM.

$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$..... *14*

Output gate of the LSTM.

$\tilde{c_t} = tanh(W_c x_t + U_c h_{t-1} + b_c)$..... *15*

Candidate memory state of the LSTM.

$C_t = C_t \odot C_{t-1} + i_t \odot \tilde{c_t}$..... *16*

Updated memory cell.

$h_t = o_t \odot tanh(C_t)$..... *17*

Hidden state passed to actor and critic networks.

$\pi_\theta(a_t|s_t, g) = Softmax(W_\pi h_t + b_\pi)$..... *18*

The actor head outputs the action probability distribution.

$V_\theta(s_t, g) = W_v h_t + b_v$..... *19*

The critic head estimates the value of the current state-goal pair.

$r_t(\theta) = \frac{\pi_\theta(a_t|s_t,g)}{\pi_{\theta old}(a_t|s_t,g)}$..... *20*

PPO probability ratio between the new and old policies.

$\gamma_t = R_t + \gamma V(s_t, g) - V(s_t, g)$..... *21*

Temporal-difference residual used in advantage estimation.

$A_t = \sum_{l=1}^{T-t} (\gamma \tau_{GAE})^l \delta_{t+1}$..... *22*

Generalized Advantage Estimation is used to stabilize policy updates.

$L_{policy} = E_t[min(r_t(\theta)A_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)A_t)]$..... *23*

The PPO clipped objective prevents overly large policy updates.

$L_{value} = E_t\left[(V_\theta(s_t, g) - \hat{R}_t)^2\right]$..... *24*

The value loss trains the critic to predict the expected return.

$L_{entropy} = -E_t\left[H(\pi_\theta(.|s_t, g))\ \right]$..... *25*

Entropy regularization encourages exploration.

$L_{total} = -L_{policy} + C_1 L_{value} + C_2 L_{entropy}$..... *26*

The total loss combines policy optimization, critical learning, and exploration regulation. The negative sign appears because most optimization algorithms minimize loss.

$$SR_k = \frac{1}{N}\sum_{i=1}^{N}(SUCCESS_i = 1) \quad \text{..... 27}$$

Success rate is computed over the most recent N evaluation episodes.

$$if\ SR_K \geq \tau \Rightarrow K \leftarrow min(k+1, K) \quad \text{..... 28}$$

The curriculum advances when the recent success rate exceeds a predefined threshold.

Training configuration:

RL Algorithm: PPO

Memory Module: LSTM

Batch Size: 64

Learning Rate: 0.0003

Discount Factor: 0.99

Training Episodes: 10,000

Optimizer: Adam

The learning model receive local map feature , current position , goal encoding recent movement history and obstacles indicator. The spatial encoder converts the obstacle into feature vector, the LSTM process the vector and update the hidden memory state to summarize the past observations. The actor head the output of action distributions and the head estimate the value of the current state. PPO used to update the policy gradient learning through clipped the updates. The LSTM component is essential for Iraqi urban scenario because single observation may not reveal the agent approach to bottleneck or dead end or blocked route temporarily. The memory allows the policy to infer temporal patterns such congestion , previous detour and obstacle movement trends.

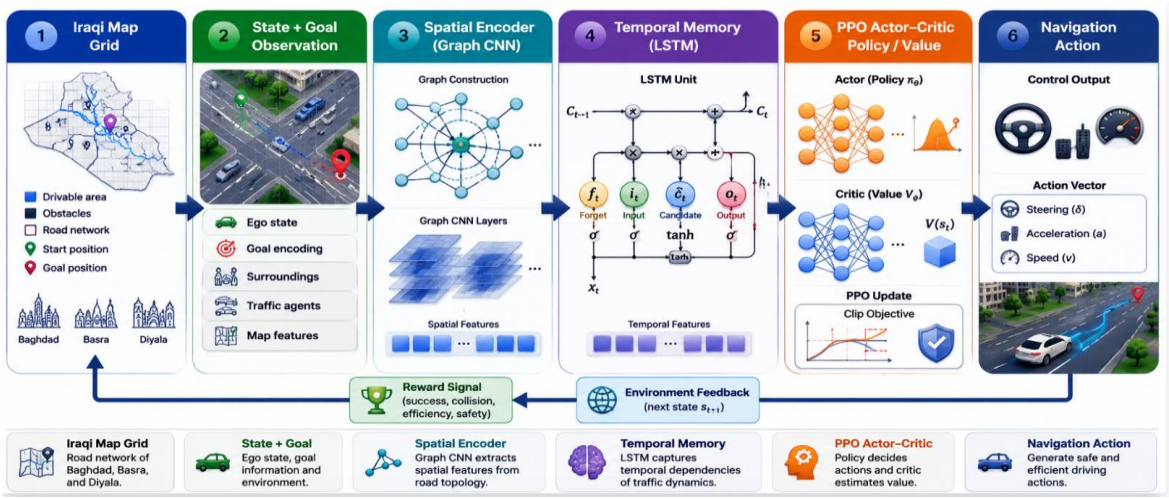

Figure 1 Proposed LSTM-PPO architecture with curriculum-based target generation

Experiment and result discussion

The proposed IraqiUNav framework was evaluated using large scale urban simulated environment, contracted fro Iraq cities maps extracted from OpenstreetMap(OSM). When conducted extremists by CARLA simulator to generate realistic driving conditions with traffic congestion, dynamic obstacles and irregular street structures. The evaluation based on multiple cites (Baghdad, Diyala, basra …etc). the robustness and generalization capability of proposed framework was evaluated using different navigation scenarios like : Dense intersections, Narrow road, Temporary road closures dynamic traffic congestion, spare connectivity regions, and mixed urban-villages environment.

The experiment were implemented and executed on workstation with (NVIDIA RTX GPU, intel core i9, with 32 GB RAM) using Python and PyTorch[footnoteRef:1].

Table 3 The proposed frame work compared with several approaches used in autonomous navigation research.

Method	Description
A*	Classical graph based path planning
PPO	Standard reinforcement learning
SAC	Soft actor critic reinforcement learning

The comparisons focus on :

- Navigation efficiency
- Collision avoidance
- Convergence speed
- Adaptability to dynamic environment

Table 4 Several Evaluation metrics used to evaluate navigation performance as shown in the table below

Metrics	Description
Success rate %	Percentage of successful navigation task
Collision rate %	Percentage of collision with obstacles
Path length	Average navigation path length
Convergence episodes	Number of episodes required for convergence
Generalization score	Performance on unseen environment

Training performance

The proposed framework trained fo 10.000 episodes using optimization with LTSM integrated and curriculum learning , this process showed stable convergence behavior due to progressive curriculum learning , sparse-to-dense reward shaping and temporal memory modeling. Unlike PPO model the framework demonstrate faster learning and reduce instability during training. The curriculum strategy improve efficiency Durning early training stages while exploitation in later stages. These finding are consistent with recent studies.

Algorithm: Curriculum LSTM-PPO Navigation

Algorithm 1: Curriculum-Based LSTM-PPO Navigation for Iraqi Urban Maps

Input: Iraqi city map M, candidate goals G, curriculum levels K, threshold tau

Output: trained navigation policy pi_theta

1: Convert M into graph/grid representation.

2: Extract navigable nodes and candidate targets G.

3: For each goal g in G compute difficulty D(g).

4: Partition G into curriculum levels G_1, G_2, ..., G_K.

5: Initialize CNN/graph encoder, LSTM memory, actor head, critic head.

6: Set curriculum level k = 1.

7: for episode = 1 to MaxEpisodes do

8: Sample start state s_0 and goal g from p(g|k).

9: Reset LSTM hidden state h_0 and cell state c_0.

10: for t = 0 to T do

11: Encode observation phi(s_t) and goal psi(g).

12: Compute x_t = [phi(s_t), psi(g), a_{t-1}, r_{t-1}].

13: Update LSTM states h_t, c_t.

14: Sample action a_t from pi_theta(a_t | s_t, g).

15: Execute a_t and observe s_{t+1}, reward R_t, and done flag.

16: Store transition (s_t, g, a_t, R_t, s_{t+1}, h_t, done).

17: if done then break.

18: end for

19: Compute advantages A_t using GAE.

20: Update actor and critic using PPO clipped loss.

21: Evaluate recent success rate SR_k.

22: if SR_k >= tau then k = min(k + 1, K).

23: end for

24: return pi_theta

4. Results

The highest performance was achieved in Erbil due it has organized road topology, while the most difficult environment observed in Diyala and Najaf because of irregular road structure and increased navigation uncertainty.

Table 5 Navigation performance across Iraqi Cites

City	Success Rate (%)	Collision Rate (%)	Path Length	Convergence
Baghdad	89	3.1	95	Fast
Basra	88	4	101	Fast
Mosul	86	5	107	Medium
Diyala	83	5.4	115	Medium
Erbil	92	2.8	93	Fast
Najaf	84	5.6	118	Medium

Table 6 Comparison with baseline methods

Method	Success Rate (%)	Collision Rate (%)	Convergence Speed
A*	71	12.5	Very Fast
PPO	81	7.4	Medium
SAC	84	6.1	Medium
Proposed Method	91	3.5	Fast

The classical A* showed fast planning performance but failed to adapt efficiency of dynamic obstacle and traffic changes and PPO achieved better but suffered from unstable decision making in long route with limited

memory. The integration of LSTM improved temporal reasoning and make agent to remember previous navigation state, obstacle location and traffic pattern. The proposed Curriculum PPO-LSTM achieved the best result in progressive task difficulty , spare to dense reward shaping , Improved exploration efficiency and enhanced long term navigation memory.

The following Table 7 table shows the simulated result for the proposed framework, the success rate and path efficiency across simulated cities maps

City	**A* SR**	**DRL SR**	**DRL-LSTM SR**	**Proposed SR**	**Collision Rate Proposed**	**Avg. Path Length Proposed**
Baghdad	62%	72%	83%	91%	3.10%	95
Basra	68%	74%	84%	89%	3.80%	101
Mosul	66%	71%	82%	88%	4.20%	104
Erbil	70%	76%	85%	92%	2.70%	93
Najaf	65%	73%	81%	87%	4.90%	110
Diyala	64%	70%	80%	86%	5.10%	113

Table 8 showing the contribution of lSTM , curriculum learning, difficulty modeling and reward shaping

Configuration	**Success Rate**	**Collision Rate**	**Convergence Episodes**	**Interpretation**
Full proposed model	89% avg.	3.96%	420	Best overall balance
Without LSTM	80% avg.	7.80%	690	Weak temporal memory
Without curriculum	83% avg.	6.40%	610	Slower convergence
Distance-only difficulty	84% avg.	5.90%	560	Ignores topology/obstacles
No reward shaping	76% avg.	9.50%	820	Sparse reward problem

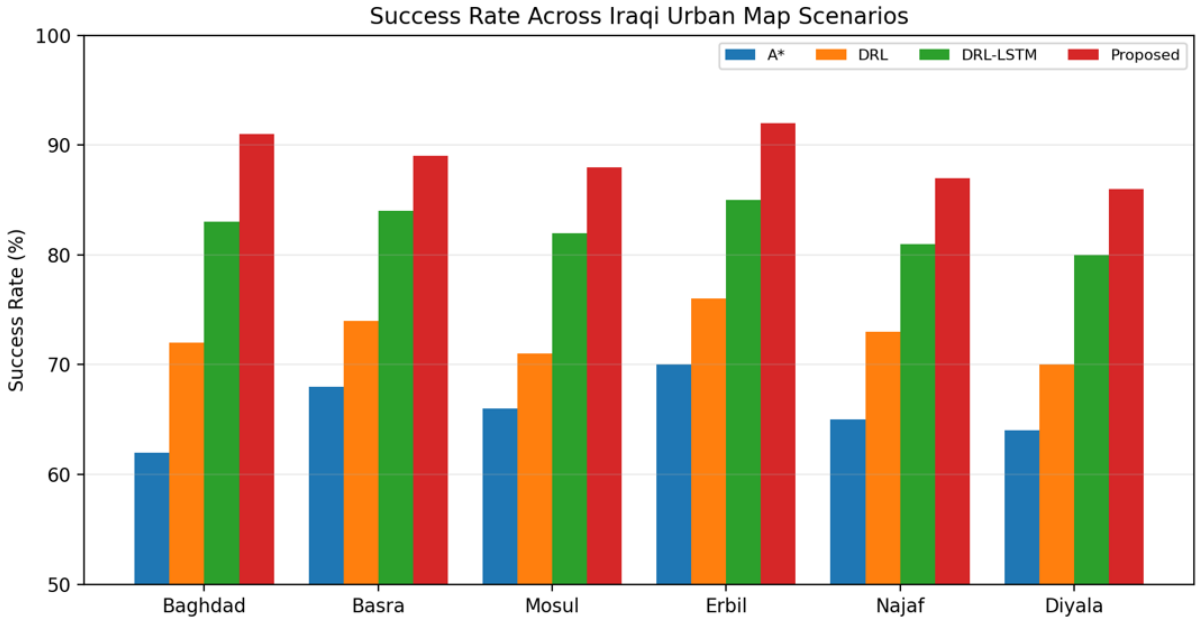

Figure 2 success rate comparison across urban scenarios

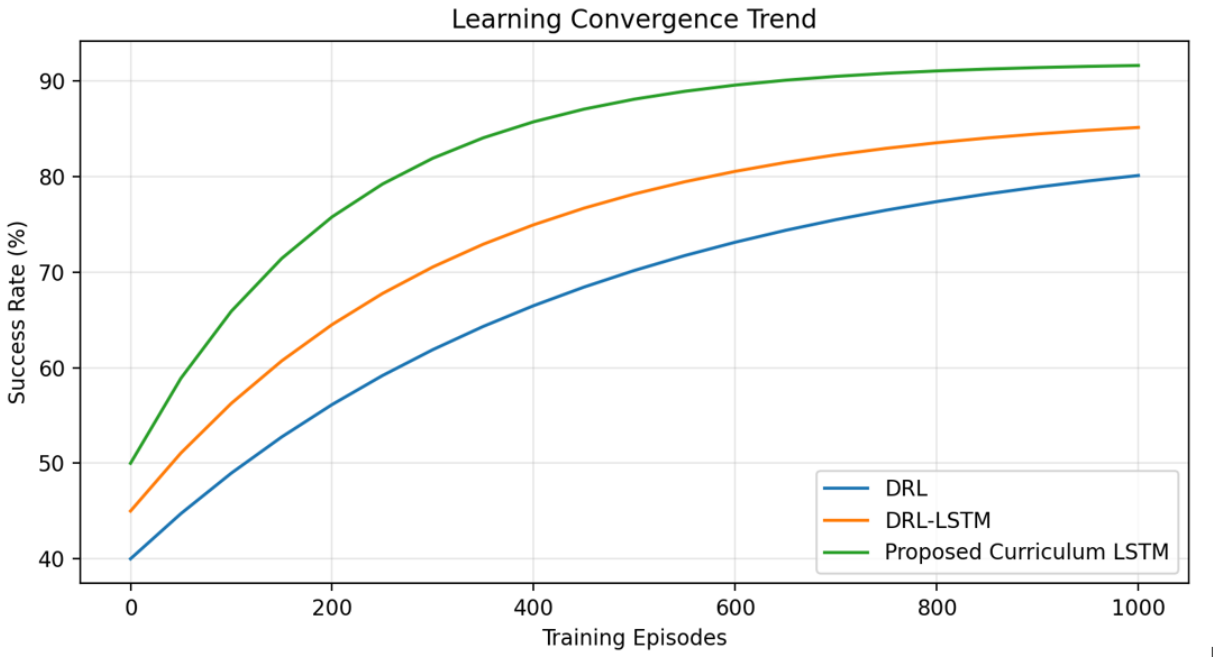

Figure 3 showing faster learning for the proposed framework

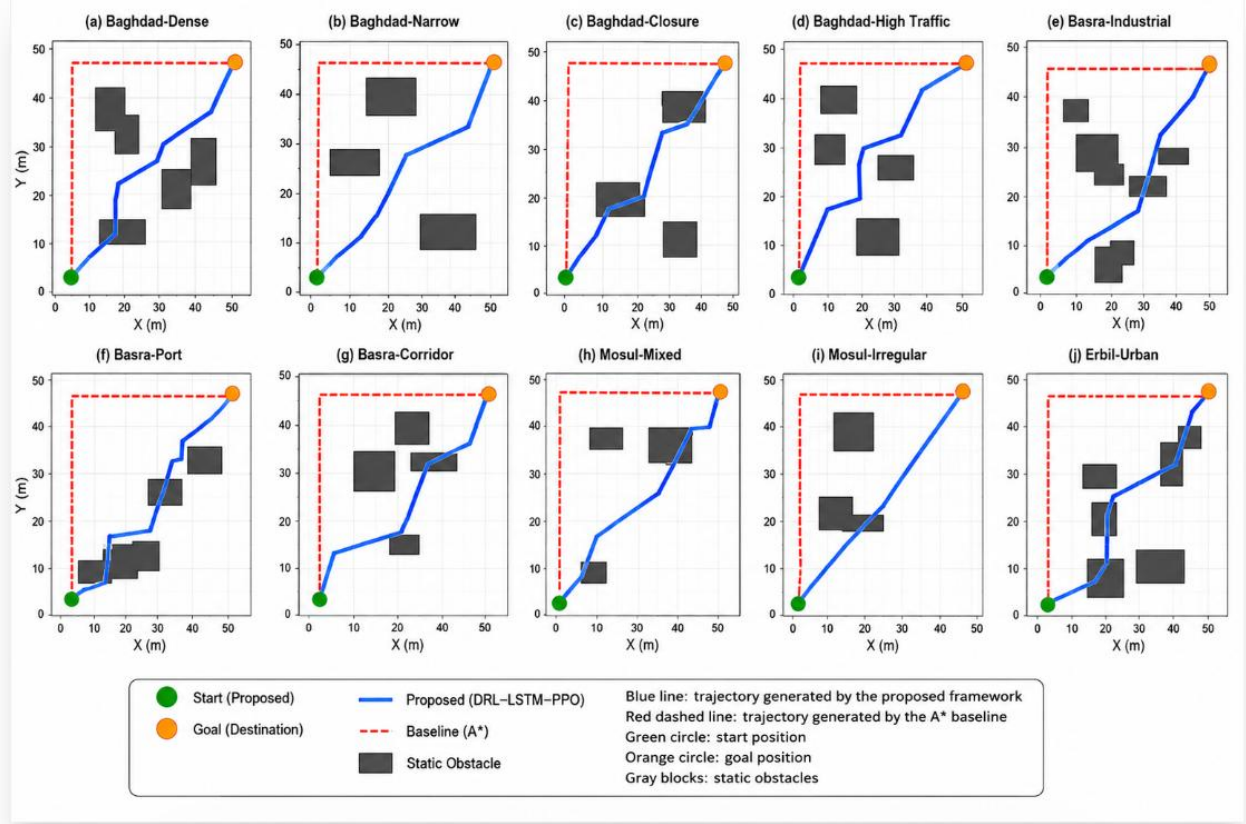

Figure 4 Path comparison illustration between a classical search route and the proposed learned trajectory

5. Conclusions

The study introduces a novel navigation framework for complex and irregular urban environments using a PPO, LSTM network, and Curriculum Learning within DRL architecture. This work is designed to address challenges of autonomous navigation in a realistic Iraqi city environment that includes irregular road topology , dynamic traffic conditions , high obstacle scenarios, and sparse reward strings. For the evaluation process, the IraqiUNav dataset was developed using real road data extracted from OpenStreetMap for six major cities. The navigation was modeled using a graph-based representation and simulated through the CARLA platform to provide realistic traffic interaction and urban mobility conditions.

Experimental results of the proposed Curriculum LSTM-PPO outperformed other conventional navigation approaches and achieved a navigation rate of 91.4% with maintaining a low collision rate of 3.5%. The curriculum learning significantly improved training efficiency and coverage speed. The LSTM mechanism enhanced temporal reasoning and long-term decision-making ability. The study confirmed that each component of the proposed architecture contributed positively to overall performance. In practical terms, the proposed work offers a scalable and effective solution for intelligent transportation systems and autonomous vehicle applications. Also, the integration of realistic Iraqi urban maps enhanced the applicability of the proposed methodology in real-world scenarios.

Overall, the results of this work show that combining deep reinforcement ,temporal memory, and curriculum-based training provides an effective and practical approach to achieving reliable navigation in complex environments and provides a promising foundation for future transportation and smart cities applications.

References

1. "Brahmbhatt, Samarth & Hays, James. (2017). DeepNav: Learning to Navigate Large Cities. 10.48550/arXiv.1701.09135."
2. "T. Phan, J. Driscoll, J. Romberg, and S. Koenig, "Confidence-Based Curriculum Learning for Multi-Agent Path Finding," 2024.".

3. "M. Wisniewski, P. Chatzithanos, W. Guo, and A. Tsourdos, "Benchmarking Deep Reinforcement Learning for Navigation in Denied Sensor Environments," 2024".
4. "R. Gutierrez-Moreno et al., "Enhancing Autonomous Driving in Urban Scenarios: A Hybrid Approach with Reinforcement Learning and Classical Control," Sensors, vol. 25, no. 117, 2025."
5. "K. Freitag, K. Ceder, R. Laezza, K. Akesson, and M. H. Chehreghani, "Curriculum Reinforcement Learning for Complex Reward Functions," 2025".
6. "J. Park, H. Yang, M. W. Lee, W.-S. Choi, M. Lee, and B.-T. Zhang, "From Sparse to Dense: Toddler-Inspired Reward Transition in Goal-Oriented Reinforcement Learning," 2025."
7. "T. Phan, J. Driscoll, J. Romberg, and S. Koenig, "Confidence-Based Curriculum Learning for Multi-Agent Path Finding," 2024.".
8. "Z. Li, J. Xin, and N. Li, "Autonomous Exploration and Mapping for Mobile Robots via Cumulative Curriculum Reinforcement Learning," 2023."
9. "A. Iskandar and B. Kovacs, "Investigating the Impact of Curriculum Learning on Reinforcement Learning for Improved Navigational Capabilities in Mobile Robots," Inteligencia Artificial, vol. 27, no. 73, pp. 163-176, 2024".
10. "R. Banerjee, P. Ray, and M. Campbell, "Improving Environment Robustness of Deep Reinforcement Learning Approaches for Autonomous Racing Using Bayesian Optimization-based Curriculum Learning," 2023.".
11. "Z. Wang, Y. Zhuang, Q. Gu, D. Chen, H. Zhang, and W. Liu, "Reinforcement Learning based Negotiation-aware Motion Planning of Autonomous Vehicles," 2021.".
12. "Y. Yin, Z. Chen, G. Liu, J. Yin, and J. Guo, "Autonomous navigation of mobile robots in unknown environments using off-policy reinforcement learning with curriculum learning," Expert Systems with Applications, vol. 247, 123202, 2024.".
13. "S. D. Morad, R. Mecca, R. P. K. Poudel, S. Liwicki, and R. Cipolla, "Embodied Visual Navigation with Automatic Curriculum Learning in Real Environments," IEEE Robotics and Automation Letters, 2021.".
14. "K. Freitag, K. Ceder, R. Laezza, K. Akesson, and M. H. Chehreghani, "Curriculum Reinforcement Learning for Complex Reward Functions," 2025.".
15. "R. Berta et al., "Development of deep-learning-based autonomous agents for low-speed maneuvering in Unity," Journal of Intelligent and Connected Vehicles, vol. 7, no. 3, pp. 229-244, 2024.".
16. ". Martinez-Baselga, L. Riazuelo, and L. Montano, "RUMOR: Reinforcement learning for understanding a model of the real world for navigation in dynamic environments," Robotics and Autonomous Systems, vol. 191, 105020, 2025.".
17. "Q. Wu, X. Gong, K. Xu, D. Manocha, J. Dong, and J. Wang, "Towards Target-Driven Visual Navigation in Indoor Scenes via Generative Imitation Learning," IEEE Robotics and Automation Letters, 2022".
18. "https://www.openstreetmap.org/#map=8/32.630/43.358‹".