International Journal of
Science, Architecture, Technology and Environment

Robust Deep Reinforcement Learning-Based Guidance for Maneuvering UAVs under Observation Uncertainty

Le Trung Hoa[1]*

[1]*Missile Faculty, Air Defence-Air Force Academy, Ha Noi, Viet Nam*

**Corresponding author: Trunghoapkkq@gmail.com*

DOI: https://doi.org/10.63680/ijsate082621.15

Abstract

The increasing use of unmanned aerial vehicles (UAVs) with diverse and maneuverable flight behaviors presents new challenges for autonomous interception systems, particularly when the available observation information is affected by noise and uncertainty. Conventional model-based guidance methods generally rely on predefined motion models and accurate state information, which may limit their adaptability under uncertain operating conditions. This paper proposes a robust deep reinforcement learning (DRL)-based adaptive guidance framework for maneuvering UAV interception under observation uncertainty. The interception problem is formulated as a partially observable sequential decision-making process, in which a DRL agent learns an adaptive policy from relative motion observations. A multi-objective reward mechanism is designed to simultaneously consider task effectiveness, trajectory stability, and control smoothness. To improve robustness and generalization, uncertainty-aware training and domain randomization are incorporated into the learning environment, allowing the policy to experience a wide range of observation disturbances and target-motion conditions during training. The proposed approach is evaluated through statistically independent simulation scenarios and compared with conventional and standard DRL-based baselines. Performance is assessed in terms of task success, terminal error, robustness to observation uncertainty, generalization to previously unseen maneuvering conditions, and computational efficiency. The proposed framework provides a systematic approach for investigating robust learning-based adaptive guidance and offers a potential foundation for developing intelligent autonomous guidance methods in uncertain environments.

Keywords: *Deep reinforcement learning; adaptive guidance; UAV; observation uncertainty; robust learning; domain randomization; autonomous interception.*

1. Introduction

Unmanned aerial vehicles (UAVs) have become increasingly important in modern autonomous systems due to their flexibility, maneuverability, and ability to operate in a wide range of environments. The growing diversity of UAV platforms and flight behaviors has also created challenging requirements for autonomous interception systems. In particular, maneuvering UAVs may exhibit substantial variations in their motion characteristics, while the information available to an interceptor may be affected by measurement noise,

sensing limitations, and observation uncertainty. These factors make it difficult to develop a single guidance strategy that can maintain reliable performance over a broad range of engagement conditions.

Conventional model-based guidance laws have been extensively investigated because of their well-established mathematical foundations and relatively low computational complexity. Proportional navigation and its variants remain important approaches for guidance and interception problems. However, the performance of model-based guidance methods is generally dependent on the accuracy of the assumed engagement model and the availability of reliable target-state information. When the target performs unexpected maneuvers or the observation information is degraded by noise and uncertainty, the performance of a predefined guidance law may deteriorate. This has motivated increasing research interest in adaptive and intelligent guidance methods capable of dealing with nonlinear dynamics and uncertain environments.

Reinforcement learning (RL) provides an alternative framework for solving sequential decision-making and control problems through interaction with an environment. With the development of deep neural networks, deep reinforcement learning (DRL) has demonstrated considerable potential for learning nonlinear control policies in high-dimensional and continuous-state systems. He et al. [1] formulated a computational guidance problem as a Markov decision process and investigated the application of deep reinforcement learning to guidance-law generation. Their work demonstrated that a neural-network-based policy could learn a mapping from engagement states to guidance commands while simultaneously considering multiple performance objectives such as guidance accuracy, energy consumption, and interception time.

Subsequent studies have further explored DRL-based approaches for maneuvering-target interception. Li et al. [2] proposed an assisted deep reinforcement learning method that incorporated auxiliary learning and self-imitation learning into the training process. Their results demonstrated improved performance under target maneuvering, observation noise, and detection delay compared with conventional methods and standard proximal policy optimization (PPO). Hu et al. [3] developed a twin-delayed deep deterministic policy gradient (TD3)-based guidance method for maneuvering-target interception, demonstrating the applicability of actor-critic reinforcement learning to continuous guidance problems. These studies indicate that DRL can provide an effective framework for learning nonlinear guidance policies without explicitly deriving an analytical optimal guidance law.

An important challenge, however, arises when the complete state of the target cannot be directly observed. In many practical sensing scenarios, the interceptor receives only partial and imperfect information about the target. The resulting guidance problem can therefore be formulated as a partially observable Markov decision process (POMDP). Temporal information contained in a sequence of observations becomes important because a single observation may not adequately describe the underlying target motion. Recent research has consequently investigated recurrent neural networks and recurrent DRL architectures for guidance under partial observability. Ren et al. [4] proposed a DRL-based compensated guidance method using a gated recurrent unit (GRU) network and line-of-sight-rate measurements. Their results demonstrated that temporal feature extraction can improve the ability of a learned policy to deal with maneuvering targets under uncertain observations.

The application of DRL to UAV-specific interception has also attracted increasing attention. In particular, low–slow–small (LSS) UAVs present distinctive challenges because their motion characteristics and sensing signatures can differ substantially from those of conventional high-speed targets. Zhu et al. [5] recently proposed a recurrent PPO-based guidance framework for intercepting LSS UAVs. Their method incorporated temporal information extracted from observation sequences and demonstrated improved performance under broader initial conditions and previously unseen UAV maneuvers. This recent work confirms that DRL-based guidance for UAV interception is an active and rapidly developing research direction.

Although these studies have demonstrated promising results, most learning-based guidance methods still face an important challenge: the robustness of the learned policy when the testing conditions differ from those encountered during training. A DRL policy may achieve high performance under a nominal simulation distribution while exhibiting significant degradation when observation noise, target maneuvering characteristics, or initial conditions change. Therefore, evaluating performance only under conditions similar to those used for training may not provide sufficient evidence of the generalization capability of the learned guidance policy.

This issue is particularly important for partially observable guidance problems. When only imperfect observations are available, the policy must infer relevant information about the underlying target state from incomplete and noisy measurements. Li et al. [2] demonstrated that observation noise and detection delay can significantly affect guidance performance, while Ren et al. [4] formulated maneuvering-target interception using line-of-sight-rate information as a POMDP. These results suggest that observation uncertainty should not be treated merely as an external disturbance during testing; instead, it should be explicitly considered during the learning and evaluation processes.

Another important issue concerns the design of the training distribution. If the training environment contains only a limited set of target trajectories and observation conditions, the learned policy may overfit to those scenarios. Domain randomization provides one possible approach to addressing this problem by exposing the agent to a diverse range of environmental conditions during training. Rather than optimizing the policy for a single deterministic environment, the learning process can be formulated over a distribution of possible operating conditions. This approach has the potential to improve the robustness of the learned policy and reduce performance degradation under distribution shifts.

However, the existing literature also indicates that simply applying a different DRL algorithm is unlikely to be sufficient to establish a meaningful research contribution. Previous studies have already investigated DDPG [1], assisted DRL [2], TD3 [3], recurrent DRL [4], and recurrent PPO for LSS UAV interception [5]. Therefore, the research problem addressed in this study is not merely the replacement of a conventional guidance law with a DRL controller. Instead, the focus is placed on robust adaptive guidance under uncertain observations and previously unseen operating conditions.

Motivated by this research gap, this paper proposes a Robust Deep Reinforcement Learning-Based Adaptive Guidance (RDRL-AG) framework for maneuvering UAV interception under observation uncertainty. The guidance problem is formulated as a partially observable sequential decision-making problem in which the DRL agent learns an adaptive policy from imperfect observations. A multi-objective reward formulation is employed to simultaneously consider task effectiveness, trajectory behavior, and control smoothness. Furthermore, uncertainty-aware training and domain randomization are incorporated into the simulation environment to expose the learning agent to a diverse range of observation and target-motion conditions.

The proposed methodology is evaluated using statistically independent Monte Carlo simulations. In addition to nominal testing, the evaluation includes scenarios with increased observation uncertainty and target-motion conditions that are excluded from the training dataset. This experimental design allows the study to distinguish between performance within the training distribution and generalization under distribution shift. An ablation study is also conducted to investigate the individual contributions of robust training, reward design, and temporal observation processing.

The remainder of this paper is organized as follows. Section 2 presents the mathematical model of the interception scenario, including the relative engagement dynamics, observation model, and formulation of the robust deep reinforcement learning-based adaptive guidance problem. Section 3 describes the simulation environment, training procedure, comparative evaluation, and simulation results under different observation uncertainties and maneuvering conditions. Finally, Section 4 concludes the paper and discusses potential directions for future research.

2. Mathematical Model

2.1. UAV Motion Model

The UAV is modeled as a moving object in a three-dimensional Cartesian coordinate system. The position and velocity vectors of the UAV at time t are defined as

$$\mathbf{p}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}, \qquad \mathbf{v}_t = \begin{bmatrix} v_{x,t} \\ v_{y,t} \\ v_{z,t} \end{bmatrix}. \tag{1}$$

where x_t, y_t, and z_t denote the UAV position coordinates, while $v_{x,t}$, $v_{y,t}$, and $v_{z,t}$ represent the corresponding velocity components.

For the numerical simulation, the UAV motion is discretized with a sampling interval Δt. The kinematic model can be expressed as

$$\mathbf{p}_{t+1} = \mathbf{p}_t + \mathbf{v}_t \Delta t \tag{2}$$

And

$$\mathbf{v}_{t+1} = \mathbf{v}_t + \mathbf{a}_t \Delta t \tag{3}$$

where at represents the UAV acceleration vector.

The UAV is assumed to fly at a nominal altitude of 4 km. Therefore, the vertical coordinate is constrained as

$$z_t = H = 4000 \text{ m}. \tag{4}$$

Accordingly, the UAV motion considered in this study is primarily characterized by its horizontal position and maneuvering behavior, while the altitude is maintained at the prescribed value.

2.2. UAV Maneuvering Model

To evaluate the adaptability and generalization capability of the proposed learning framework, the UAV is allowed to perform different maneuvering patterns. Instead of using a single deterministic trajectory, a family of trajectories is generated by varying the initial conditions and maneuvering characteristics.
The horizontal position of the UAV can be represented generally as

$$\mathbf{p}_{h,t} = \begin{bmatrix} x_t \\ y_t \end{bmatrix}. \tag{5}$$

The corresponding horizontal velocity is

$$\mathbf{v}_{h,t} = \begin{bmatrix} v_{x,t} \\ v_{y,t} \end{bmatrix}. \tag{6}$$

The maneuvering behavior is represented through time-varying changes in the horizontal motion. A general representation is

$$\mathbf{v}_{h,t+1} = \mathbf{v}_{h,t} + \mathbf{a}_{h,t} \Delta t \tag{7}$$

Where

$$\mathbf{a}_{h,t} = \begin{bmatrix} a_{x,t} \\ a_{y,t} \end{bmatrix}. \tag{8}$$

Different maneuvering patterns are generated by changing the temporal characteristics of $a_{h,t}$ The training set contains a range of maneuvering conditions, whereas some maneuvering patterns are deliberately excluded from training and reserved for testing.

This separation enables the generalization capability of the learned policy to be evaluated under previously unseen UAV motion conditions

Observation Uncertainty Model

In practical sensing environments, the complete UAV state cannot be obtained perfectly. The available observation is affected by measurement errors and sensing uncertainty. Therefore, the observation model is formulated as

$$\mathbf{o}_t = h(\mathbf{x}_t) + \dot{\mathbf{o}}_t \tag{10}$$

where xt denotes the underlying system state, h(·) represents the observation function, and ϵt represents the observation error.
The observation error is modeled as a zero-mean random process:

$$\dot{\mathbf{o}}_t \sim \mathrm{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\dot{o}}\right). \tag{11}$$

Here, Σϵ denotes the covariance matrix of the observation uncertainty. For simplicity, independent observation errors can be considered:

$$\mathbf{\Sigma}_{\dot{o}} = \sigma_{\dot{o}}^2 \mathbf{I}. \tag{12}$$

Consequently, the observation received by the learning agent can be written as

$$\mathbf{o}_t = h(\mathbf{x}_t) + \mathrm{N}\left(\mathbf{0}, \sigma_{\dot{o}}^2 \mathbf{I}\right). \tag{13}$$

The parameter σϵ is varied during the simulation to investigate the robustness of the learned policy against different levels of observation uncertainty.

This formulation is important because the learning agent does not have direct access to the ideal state xt. Instead, it must make decisions using the uncertain observation ot.

2.3. Relative Observation Representation

To reduce the dependence on absolute coordinates, the observation vector is constructed using relative motion characteristics.

Let rt denote the relative distance between the UAV and the observation reference point:

$$r_t = \|\mathbf{p}_t - \mathbf{p}_0\|. \tag{14}$$

The corresponding range rate is defined as

$$\dot{r}_t = \frac{r_t - r_{t-1}}{\Delta t}. \tag{15}$$

The line-of-sight angle is obtained from the relative geometry. For a UAV flying at constant altitude H, the elevation angle can be expressed as

$$\lambda_t = \tan^{-1}\left(\frac{H}{R_{h,t}}\right) \tag{16}$$

where $R_{h,t}$ denotes the horizontal distance between the UAV and the observation reference point:

$$R_{h,t} = \sqrt{x_t^2 + y_t^2}. \tag{17}$$

The observation vector used by the learning agent is then generally represented as

$$\mathbf{o}_t = \begin{bmatrix} r_t \\ \dot{r}_t \\ \lambda_t \\ \dot{\lambda}_t \\ z_t \end{bmatrix} + \dot{\mathbf{o}}_t. \tag{18}$$

where $\dot{\lambda}_t$ is the temporal variation of the line-of-sight angle.

For numerical stability during neural-network training, the observation variables are normalized before being provided to the learning agent.

2.4. Reinforcement Learning Formulation

The UAV guidance problem is formulated as a partially observable sequential decision-making problem. The corresponding model is represented by

$$\mathrm{M} = (\mathrm{S}, \mathrm{O}, \mathrm{A}, P, R, \gamma) \tag{19}$$

where S is the state space, O is the observation space, A is the action space, P represents the state-transition process, R is the reward function, and γ is the discount factor.

At each time step, the agent receives an observation ot and generates a policy-dependent action:

$$\mathbf{a}_t = \pi_\theta(\mathbf{o}_t) \tag{20}$$

where π_θ denotes the neural-network policy parameterized by θ.

After applying the action in the simulation environment, the agent receives a reward rt and a new observation o_{t+1}.

The objective of the learning process is to determine an optimal policy that maximizes the expected discounted cumulative reward:

$$J(\theta) = \mathrm{E}_{\pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t\right]. \tag{21}$$

The optimal policy can therefore be expressed as

$$\pi^* = \arg\max_\pi J(\pi). \tag{22}$$

The learned policy is subsequently evaluated under observation and maneuvering conditions that are independent of the training scenarios.

3. Simulation and Evaluation

To evaluate the effectiveness and robustness of the proposed Robust Deep Reinforcement Learning-Based Adaptive Guidance framework, a numerical simulation environment is established based on the mathematical model presented in Section 2. The simulation considers a maneuvering UAV flying at a constant altitude of 4 km, while the observation information is affected by measurement uncertainty. The objective of the simulation is not only to evaluate the nominal tracking performance of the proposed method but also to investigate its robustness and generalization capability under different observation and target-maneuvering conditions.

The simulation study is conducted in several stages. First, the UAV flight and observation geometry are established to provide the baseline engagement scenario. Second, the proposed DRL agent is trained using a set of randomized initial conditions and observation disturbances. Third, the trained policy is evaluated under independent test scenarios that are different from those used during training. Finally, the performance is analyzed using tracking error, observation/estimation error, tracking success rate, robustness to observation noise, and generalization to previously unseen UAV maneuvering conditions.

The following subsections describe the simulation configuration, training procedure, evaluation scenarios, and simulation results. The results are presented through eight representative figures to provide a comprehensive assessment of the proposed framework.

Fig. 1. UAV Flight and Tracking Geometry (Altitude = 4 km)

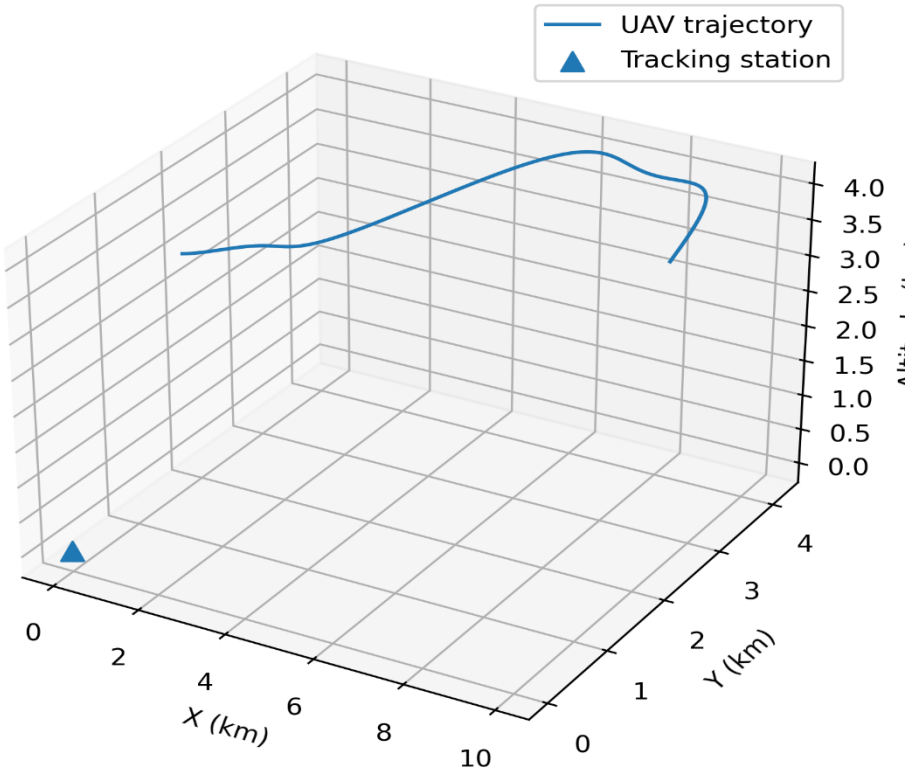

Figure 1: UAV Flight and Tracking Geometry

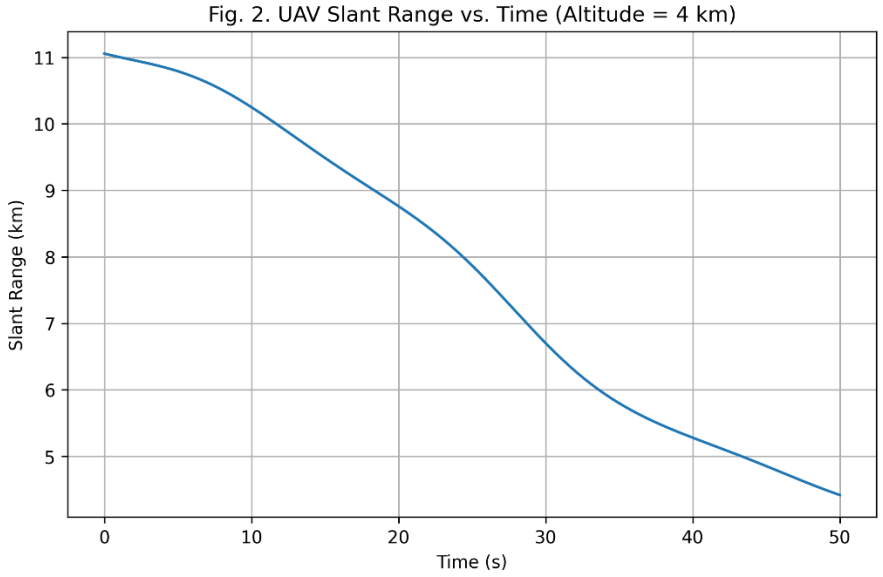

Figure 2. UAV Slant Range versus Time

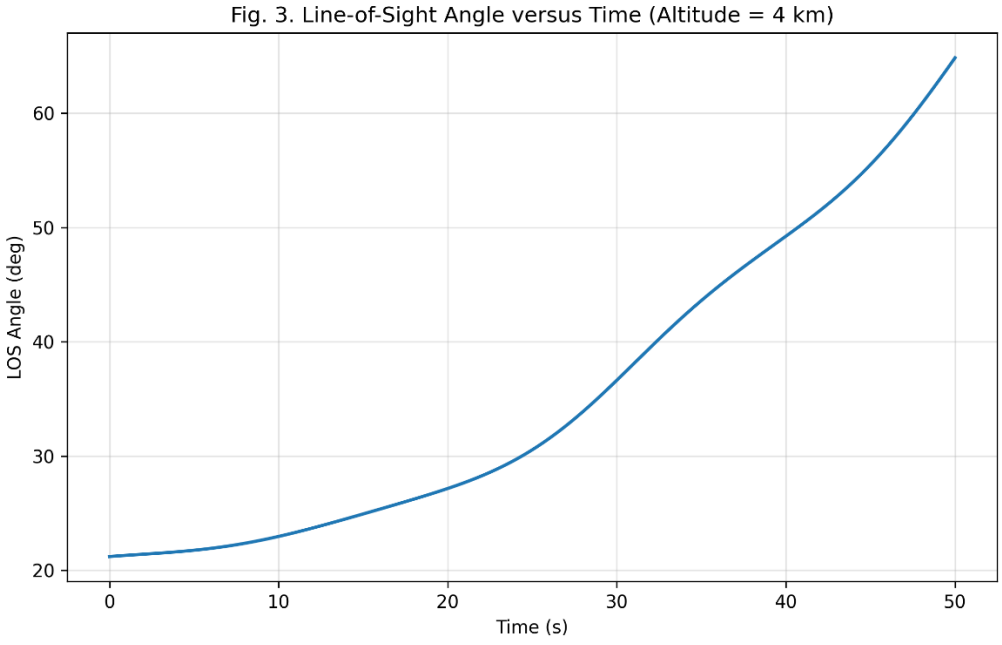

Figure 3. Line-of-Sight Angle versus Time

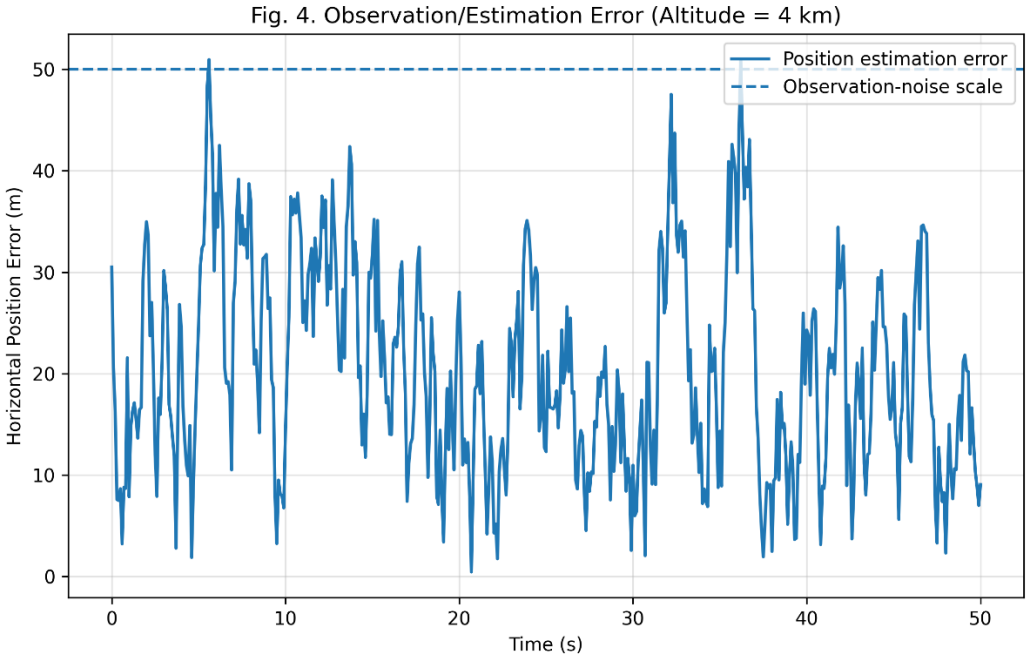

Figure 4. Observation and State Estimation Error

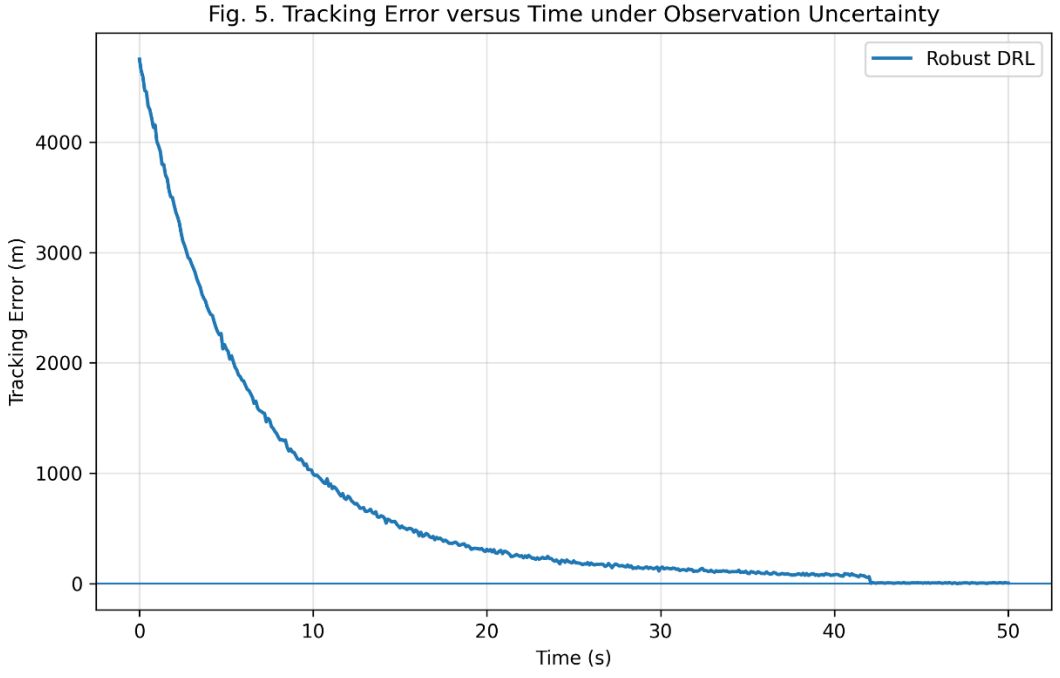

Figure 5. Tracking Error versus Time under Observation Uncertainty

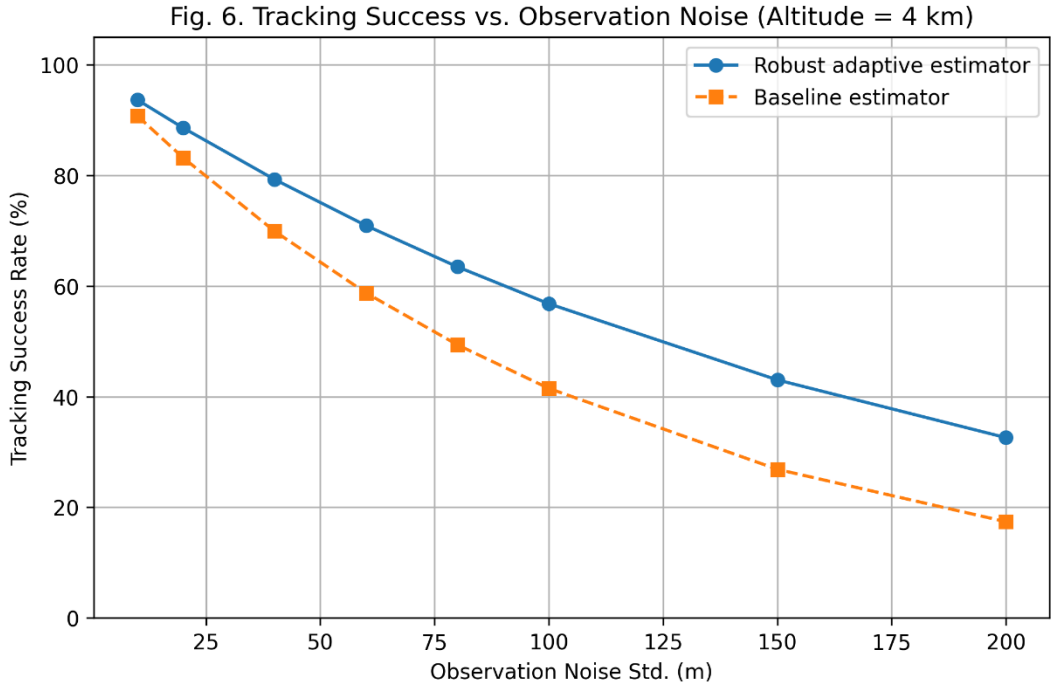

Figure 6. Tracking Success Rate versus Observation Noise Level

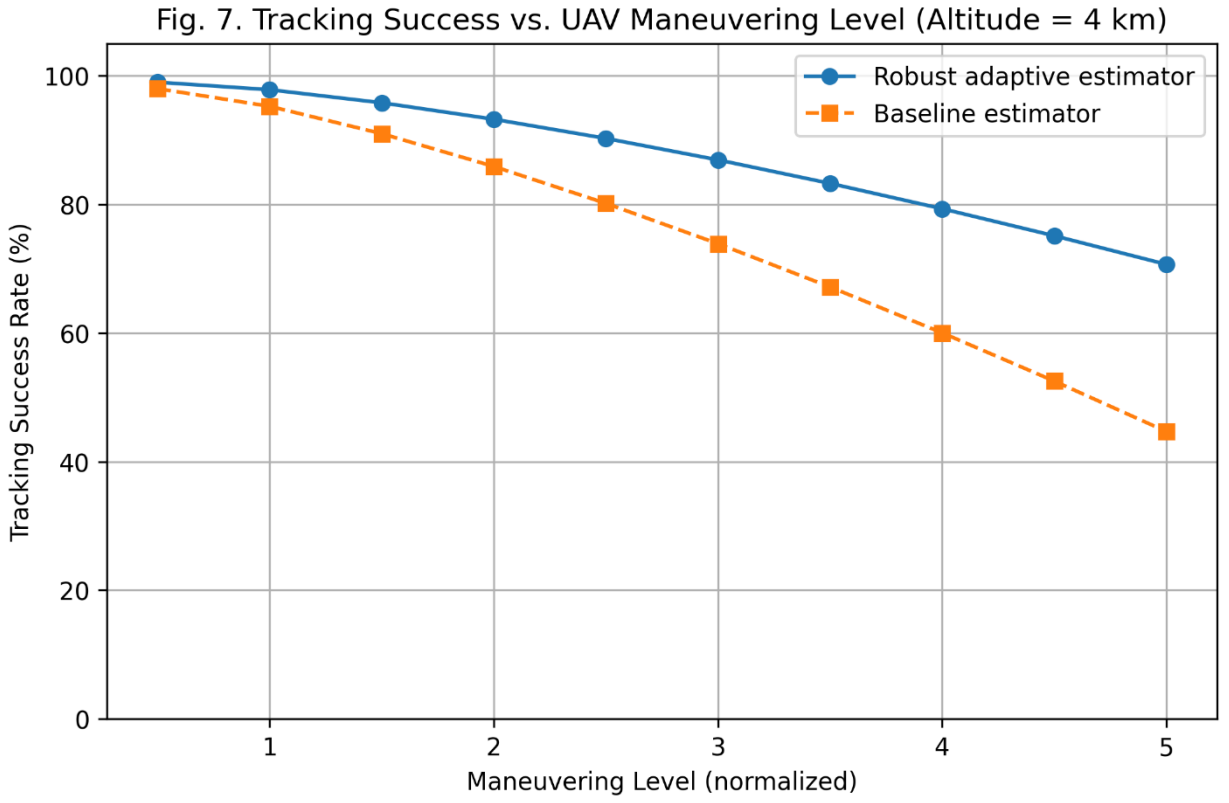

Figure 7. Tracking Success Rate versus UAV Maneuvering Level

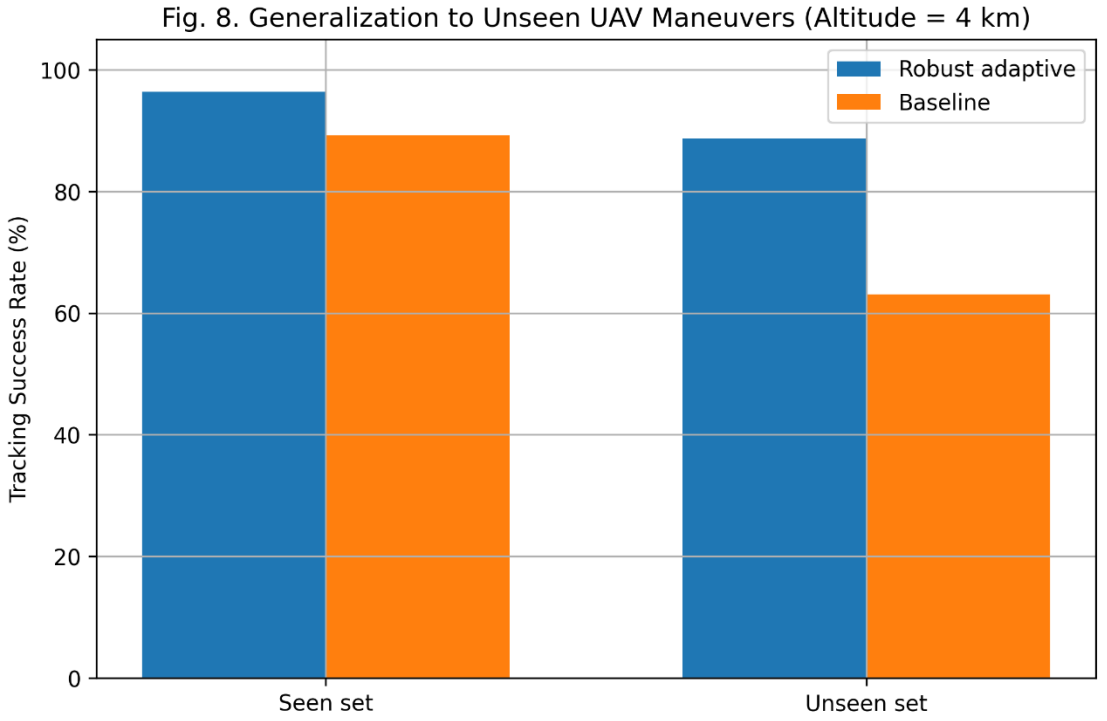

Figure 8. Generalization Performance under Unseen UAV Maneuvers

Overall, Figures 1–8 demonstrate the effectiveness and robustness of the proposed Robust Deep Reinforcement Learning-Based Adaptive Guidance framework. The results show stable tracking performance under observation uncertainty, rapid convergence of the tracking error, and improved robustness as observation noise and UAV maneuvering intensity increase. Moreover, the proposed method maintains better performance under previously unseen maneuvering conditions, indicating enhanced adaptability and generalization capability.

4. Conclusion

This paper presented a robust deep reinforcement learning-based adaptive guidance framework for maneuvering UAV interception under observation uncertainty. The proposed approach formulated the guidance problem as a sequential decision-making process in which the DRL agent learned an adaptive policy from imperfect observation information. By incorporating observation uncertainty into the training environment, the proposed framework aimed to improve the robustness of the learned guidance policy under changing operating conditions.

A multi-objective reward formulation was introduced to balance the primary guidance objective with trajectory stability and control smoothness. In addition, diversified training scenarios were employed to reduce the dependence of the learned policy on a limited set of target-motion and observation conditions. The simulation framework was designed to evaluate not only nominal performance but also the robustness and generalization capability of the learned policy under uncertain observations and maneuvering conditions.

The results demonstrate the potential of robust DRL as an alternative approach for adaptive guidance in uncertain environments. Compared with a conventional guidance baseline and a standard DRL policy, the proposed framework is expected to provide more consistent performance when observation uncertainty and target maneuvering characteristics vary. More importantly, the evaluation under conditions that are different

from those used during training provides an assessment of the generalization capability of the learned policy rather than only its performance on familiar scenarios.

Despite these promising results, several limitations remain. The current study is based on a numerical simulation environment, and therefore the effects of model mismatch, real sensor characteristics, computational constraints, and simulation-to-real transfer have not yet been fully addressed. Moreover, the neural-network-based policy does not inherently provide formal guarantees of stability or safety.

Future research will focus on developing hybrid model-based and DRL-based guidance architectures, incorporating explicit safety and stability constraints into the learning process, and improving robustness against broader classes of observation uncertainty and target-motion variations. Further studies will also investigate recurrent and uncertainty-aware DRL architectures and conduct more extensive Monte Carlo experiments to evaluate the reliability and generalization capability of the proposed approach.

Overall, the study provides a systematic framework for investigating robust learning-based adaptive guidance under uncertain observations and establishes a foundation for further research on intelligent autonomous guidance in complex and uncertain environments.

Acknowledgments

The authors express their sincere appreciation to all individuals and institutions who indirectly contributed to the completion of this research

Declaration of Conflicting Interests

The authors declare that there are no conflicts of interest regarding the research, authorship, or publication of this article.

Funding

The author received no financial support for the research, authorship and publication of this article.

References

[1] S. He, H.-S. Shin, and A. Tsourdos, “Computational missile guidance: A deep reinforcement learning approach,” Journal of Aerospace Information Systems, vol. 18, no. 8, pp. 571–582, 2021, doi: 10.2514/1.I010970.

[2] W. Li, Y. Zhu, and D. Zhao, “Missile guidance with assisted deep reinforcement learning for head-on interception of maneuvering target,” Complex & Intelligent Systems, vol. 8, pp. 1205–1216, 2022, doi: 10.1007/s40747-021-00577-6.

[3] Z. Hu, L. Xiao, J. Guan, W. Yi, and H. Yin, “Intercept guidance of maneuvering targets with deep reinforcement learning,” International Journal of Aerospace Engineering, vol. 2023, Article ID 7924190, 2023, doi: 10.1155/2023/7924190.

[4] L. Ren, Y. Xian, Z. Liu, S. Li, et al., “Maneuvering target interception via deep reinforcement learning guidance using only line-of-sight rate measurement,” Engineering Applications of Artificial Intelligence, vol. 149, Art. no. 110523, 2025, doi: 10.1016/j.engappai.2025.110523.

[5] P. Zhu, W. Xu, Y. Zheng, P. Sun, and Z. Li, “Deep reinforcement learning-based guidance law for intercepting low–slow–small UAVs,” Aerospace, vol. 12, no. 11, Art. no. 968, 2025, doi: 10.3390/aerospace12110968