Disturbance-Aware Hybrid Learning for Robust and Adaptive UAV Flight in Extreme Winds

Huidong Liu[1], Jiarui Dou[1], Jiangshan Ai[1], Enwen Hu[2*], Xianlei Long[1], Mingyan Li[1], Chao Chen[1], Fuqiang Gu[1*]
[1]College of Computer Science, Chongqing University
[2]College of Electrical Engineering, Beijing University of Posts and Telecommunications
{liuhuidong,ajs}@stu.cqu.edu.cn, { doujr, xianlei.long, limy2021, cschaochen, gufq}@cqu.edu.cn, owen.hu@bupt.edu.cn

Abstract

Safe and precise maneuvering of quadrotor unmanned aerial vehicles (UAVs) in high-speed wind environments remains a critical challenge. Wind disturbances are nonlinear, time-varying, and difficult to model, causing traditional controllers to struggle with perception and compensation, especially under unseen wind distributions. To address these limitations, we introduce WA-TD3, a data-driven control framework that enables real-time wind disturbance perception and adaptive compensation without dedicated wind sensors. WA-TD3 employs a deep residual network to extract wind characteristics from temporal patterns in state deviations, forming a dynamics residual-driven perception mechanism that implicitly models and compensates for unknown winds. This residual is integrated into a perception-augmented reinforcement learning architecture, providing the policy with enhanced state information for proactive disturbance-aware control. Extensive experiments on complex trajectories under varying wind intensities demonstrate that WA-TD3 consistently outperforms state-of-the-art methods, achieving over 62% improvement in tracking accuracy under strong winds.

1 Introduction

Quadrotor unmanned aerial vehicles (UAVs) have been increasingly deployed in a wide range of applications, including logistics delivery [Shao *et al.*, 2025], surveillance [Gu *et al.*, 2018], and search-and-rescue operations [Abdellatif *et al.*, 2025]. These applications place stringent requirements on flight stability and control robustness, particularly under challenging environmental conditions [Gu *et al.*, 2025]. Among various external disturbances, wind remains one of the most critical factors degrading autonomous flight performance in outdoor environments, where wind fields are inherently stochastic, time-varying, and difficult to model [Wang *et al.*, 2025]. Due to their lightweight structure and underactuated dynamics, quadrotors are especially vulnerable to aerodynamic disturbances, where even moderate wind gusts can lead to significant tracking errors or loss of stability [O'Connell *et al.*, 2022; Sun *et al.*, 2025; Cioffi *et al.*, 2025].

Conventional approaches to wind disturbance rejection primarily rely on classical feedback control techniques, such as PID controllers with feedforward compensation [Liu *et al.*, 2025b] or model predictive control (MPC) [Hanover *et al.*, 2021; Pravitra *et al.*, 2020]. While effective under nominal conditions, these methods exhibit inherent limitations when confronted with the nonlinear, stochastic, and spatially distributed nature of real-world wind fields. Accurate modeling of wind dynamics remains challenging, as wind speed, direction, and turbulence exhibit significant uncertainty and variability. Consequently, controllers based on simplified disturbance models often suffer from degraded performance in out-of-distribution scenarios [Huang *et al.*, 2023; Kong *et al.*, 2025]. Furthermore, uncertainty in aerodynamic parameters can necessitate frequent manual retuning, increasing computational overhead and limiting deployment on resource-constrained UAV platforms [Lee *et al.*, 2025].

Recent advances in machine learning, particularly reinforcement learning (RL), have provided promising alternatives for UAV control under uncertainty [Liu *et al.*, 2025a; Vaidya and Keshavan, 2025]. Deep RL methods can learn control policies directly from interaction data, enabling them to capture complex nonlinear dynamics without relying on explicit system models [Zhang *et al.*, 2025a]. However, most existing RL-based UAV controllers implicitly treat wind disturbances as unstructured noise [Yang *et al.*, 2025; Huang *et al.*, 2023], without explicitly reasoning about their physical effects on flight dynamics. As a result, learned policies often exhibit limited generalization capability and may fail under wind conditions that differ from those encountered during training.

To address these limitations, we propose WA-TD3 (Wind-Aware Twin Delayed Deep Deterministic Policy Gradient), a novel control method that explicitly accounts for wind disturbances by inferring wind field characteristics from dynamics residuals and integrating this information into RL-based control, as illustrated in Fig. 1. The key idea is to leverage temporal deviations between predicted and observed UAV dynamics to construct a residual-driven wind perception model. By exploiting both current and historical state information, the proposed model captures latent wind-induced forces with-

*Corresponding author

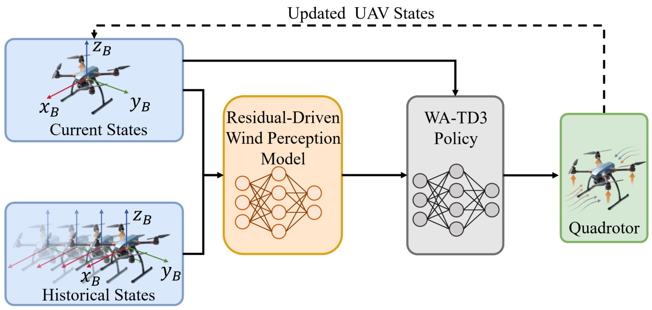

Figure 1: Conceptual overview of WA-TD3, where wind disturbances are implicitly inferred from dynamics residuals and incorporated into a wind-aware reinforcement learning policy for robust quadrotor control.

out requiring dedicated wind sensors. The inferred wind disturbance is then incorporated into a perception-augmented actor–critic architecture, where wind-aware state representations enable the policy to proactively compensate for wind effects rather than merely reacting to disturbances. Furthermore, we adopt a joint training paradigm that simultaneously optimizes wind perception and control performance within a closed-loop system. This end-to-end learning strategy improves robustness and generalization to unseen and complex wind conditions. Unlike prior approaches that treat wind as an unstructured disturbance, WA-TD3 implicitly learns wind dynamics and integrates them into decision-making, resulting in improved adaptability and control stability under severe wind disturbances.

Our main contributions are as follows:

- We propose a dynamics residual–driven wind perception mechanism that extracts wind field characteristics from temporal UAV state deviations, enabling real-time wind estimation without external sensors.
- We design a perception-augmented reinforcement learning framework that integrates estimated wind disturbances into the control policy, allowing proactive and wind-aware decision making.
- We demonstrate through extensive experiments that WA-TD3 consistently outperforms state-of-the-art methods across diverse wind conditions and trajectory tracking tasks, achieving over 62% improvement in tracking accuracy under strong wind disturbances.

2 Related Work

2.1 Traditional Wind Disturbance Control Approaches

Wind disturbance rejection for UAVs has traditionally relied on model-based and classical control. Early methods used PID controllers with feedforward compensation, modeling wind as external disturbances mitigated via observer-based estimation [Rosales *et al.*, 2018]. These approaches work under mild conditions but assume simplified disturbances and require accurate system models and extensive manual tuning.

Adaptive control methods, such as $\mathcal{L}_1$ adaptive control combined with MPC [Pereida and Schoellig, 2018] or MPPI [Pravitra *et al.*, 2020], improve robustness under model uncertainty and enable fast online adaptation. However, they often rely on restrictive assumptions, such as constant or slowly varying wind disturbances, and require continuous parameter adaptation, resulting in increased computational complexity and limited scalability to resource-constrained platforms.

Wind estimation techniques provide explicit disturbance awareness [Hollenbeck *et al.*, 2018]. Direct sensors, such as pitot tubes or ultrasonic anemometers, measure wind but increase payload and system complexity. Sensor-free methods infer wind from discrepancies between commanded and observed motion using EKF/UKF [Ahmed and Woolsey, 2025]. More recently, learning-based estimators have been proposed to infer wind characteristics from flight data. While effective in structured settings, these approaches remain sensitive to modeling assumptions and often degrade under highly nonlinear or turbulent wind conditions [Sikkel *et al.*, 2016; Ilyas *et al.*, 2025].

Overall, traditional control and estimation methods exhibit limited robustness in complex, unstructured wind environments. Their reliance on explicit disturbance models, restrictive assumptions, and manual tuning motivates the exploration of more flexible learning-based approaches.

2.2 Learning-Based Control and Perception Integration

Learning-based control, particularly reinforcement learning (RL), has emerged as a promising approach for quadrotor flight under uncertainty. Early model-free RL methods demonstrate agility but treat wind as unstructured noise, relying on domain randomization and showing limited out-of-distribution (OoD) generalization [Ma *et al.*, 2023].

Hybrid learning–control frameworks aim to improve robustness. DATT [Huang *et al.*, 2023] integrates deep RL with feedforward, feedback, and $\mathcal{L}_1$ adaptive control [Hovakimyan and Cao, 2010] for fast trajectory tracking, but is sensitive to hyperparameters and computationally heavy. Neural-IMC [Gao *et al.*, 2025] combines a simplified dynamics predictor with RL using prediction error as an auxiliary signal, but this is only available during training. ProxFly [Zhang *et al.*, 2025b] uses residual RL for aerodynamic compensation in proximity flight, yet residual corrections may induce high-frequency oscillations that reduce stability.

Recent methods focus on OoD generalization. OoD-Control [Ye *et al.*, 2024] uses noise-induced strategies to bound control errors but does not capture temporal wind structure. Domain randomization [Andres *et al.*, 2025; Xue *et al.*, 2023] improves robustness across conditions but incurs high computational cost and neglects structured, time-varying aerodynamic disturbances.

In contrast, WA-TD3 explicitly incorporates physically grounded acceleration residuals as real-time policy inputs, enabling sensor-free perception of wind-induced dynamics mismatches. By exploiting the temporal structure of these residuals within a perception-augmented RL framework, WA-TD3 allows policies to leverage structured disturbance information for robust performance and rapid adaptation under extreme, previously unseen wind conditions.

3 Problem Formulation

We consider trajectory tracking for a quadrotor UAV under stochastic wind disturbances. The objective is to follow a reference trajectory while satisfying system and safety constraints. The problem is first posed as a constrained optimal control task and then reformulated as a Markov Decision Process (MDP) for reinforcement learning.

The quadrotor state is defined as: $\mathbf{x}_t = [\mathbf{p}_t^\top, \mathbf{q}_t^\top, \mathbf{v}_t^\top, \boldsymbol{\omega}_t^\top]^\top \in \mathbb{R}^{13}$, where $\mathbf{p}_t$, $\mathbf{q}_t$, $\mathbf{v}_t$, and $\boldsymbol{\omega}_t$ denote position, attitude quaternion, linear velocity, and angular velocity, respectively. The desired state $\mathbf{x}_{d,t} = [\mathbf{p}_{d,t}^\top, \mathbf{q}_{d,t}^\top, \mathbf{v}_{d,t}^\top, \boldsymbol{\omega}_{d,t}^\top]^\top \in \mathbb{R}^{13}$ is generated from the reference trajectory.

Wind-disturbed dynamics are modeled as:

$$\dot{\mathbf{x}}_t = f(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{d}_t, \tag{1}$$

where $f(\cdot)$ represents nominal Newton–Euler dynamics, $\mathbf{u}_t$ denotes motor inputs, and $\mathbf{d}_t$ captures stochastic wind-induced disturbances.

The tracking objective over horizon T is

$$\begin{aligned} \min_{\mathbf{u}(\cdot)} \quad & \int_0^T \left(\|\mathbf{p}_t - \mathbf{p}_{d,t}\|_2^2 + \lambda \|\mathbf{u}_t\|_2^2 \right) dt \\ \text{s.t.} \quad & \dot{\mathbf{x}}_t = f(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{d}_t, \ \mathbf{x} \in \mathcal{X}, \ \mathbf{u} \in \mathcal{U}. \end{aligned} \tag{2}$$

We reformulate the task as an MDP $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where the state includes quadrotor observations and reference information, the action space is $\mathcal{A} = [-1, 1]^4$, transitions follow wind-disturbed dynamics, and rewards penalize tracking errors and control effort. The learning objective is thus to find an optimal policy π^* that maximizes the expected cumulative discounted reward:

$$\pi^* = \arg\max_{\pi} \ \mathbb{E}_{\tau \sim \pi} \left[\sum_{t=0}^{T} \gamma^t r_t \right], \tag{3}$$

where $\tau = (s_0, a_0, r_0, s_1, \dots)$ denotes a trajectory generated under policy π, and the expectation is taken over stochastic transitions arising from wind disturbances.

4 Proposed Method: WA-TD3

Robust quadrotor control in spatially varying winds is difficult for standard RL due to dynamics mismatch. Instead of treating wind as noise, WA-TD3 exploits inconsistencies between nominal dynamics and onboard measurements as disturbance cues. It integrates acceleration residuals into policy learning without requiring wind sensors. Wind-induced forces appear as systematic deviations from thrust-driven dynamics; explicitly modeling these residuals enables proactive compensation and improved generalization. As shown in Fig. 2, WA-TD3 comprises: (i) residual computation from sensing and actuation, (ii) a residual perception network using short histories, and (iii) a residual-conditioned TD3 policy.

4.1 External Force Residual as a Wind-Sensor-Free Disturbance Proxy

WA-TD3 is based on the observation that wind appears as an unmodeled external force, causing systematic deviations from nominal thrust-driven dynamics. By comparing the force inferred from onboard inertial measurements with that predicted by the nominal model, we obtain a physically meaningful residual. Given known thrust commands and vehicle parameters, this discrepancy serves as a proxy without explicit wind sensing for wind-induced disturbances.

The translational dynamics of a quadrotor are given by

$$m\dot{\mathbf{v}}_t = \mathbf{R}(\mathbf{q}_t)\mathbf{f}_t - mg\mathbf{e}_3 + \mathbf{F}_{\text{wind},t}, \tag{4}$$

where m is the vehicle mass, $\mathbf{R}(\mathbf{q}_t)$ maps body to world frame, $\mathbf{f}_t \in \mathbb{R}^3$ is the total rotor thrust, and $\mathbf{F}_{\text{wind},t} \in \mathbb{R}^3$ is the unknown wind force.

Rather than explicitly modeling $\mathbf{F}_{\text{wind},t}$, we recover its effect indirectly from inertial information. Specifically, let $\mathbf{a}_t^w$ denote the quadrotor linear acceleration in the world frame, estimated from bias-corrected inertial measurements. The corresponding equivalent non-gravitational total force is given by

$$\mathbf{F}_{\text{meas},t} = m\left(\mathbf{a}_t^w + g\mathbf{e}_3\right). \tag{5}$$

In parallel, a nominal force is predicted from the thrust model with known inputs and parameters:

$$f_i = C_T \tilde{\omega}_i^2, \tag{6}$$

where $\tilde{\omega}_i$ is rotor speed and C_T the thrust coefficient. The total thrust $T_t = \sum_{i=1}^4 f_i$ yields the predicted world-frame force:

$$\hat{\mathbf{F}}_{\text{pred},t} = \mathbf{R}(\mathbf{q}_t) \begin{bmatrix} 0 \\ 0 \\ T_t \end{bmatrix}. \tag{7}$$

We define the external force residual as the difference between measured and nominal predicted forces:

$$\mathbf{r}_{f,t} = \mathbf{F}_{\text{meas},t} - \hat{\mathbf{F}}_{\text{pred},t}. \tag{8}$$

This residual captures wind-induced and other unmodeled aerodynamic effects using only onboard inertial data and known control inputs, removing the need for wind sensors or explicit aerodynamic models. The residual thus serves as a physically meaningful disturbance proxy for the perception module described next.

4.2 Residual Dynamics Learning via Online Supervision

The external force residual $\mathbf{r}_{f,t}$ provides a physically grounded representation of unmodeled aerodynamic disturbances. To capture temporally correlated wind effects, we employ a residual dynamics network that predicts $\mathbf{r}_{f,t}$ from short-horizon onboard state histories. The network maps a history of k system states

$$\mathcal{H}_t = [\mathbf{s}_{t-k+1}, \dots, \mathbf{s}_t], \quad \mathbf{s}_t = \mathbf{x}_t, \tag{9}$$

to a 3D residual force estimate. This temporal input allows the network to learn structured disturbance patterns without explicit wind sensing or aerodynamic models.

The analytically computed residual $\mathbf{r}_f$ from Eq. (8) serves as a self-supervised target. Network parameters are optimized via

$$\mathcal{L}_{\text{res}} = \|\hat{\mathbf{r}}_f - \mathbf{r}_f\|_2^2. \tag{10}$$

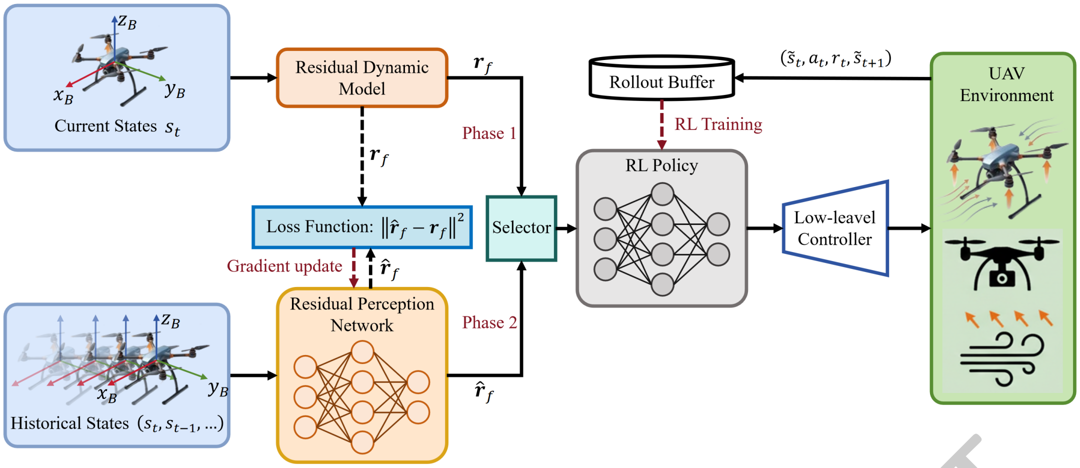

Figure 2: Overview of WA-TD3. The framework employs a two-phase training strategy for wind-robust quadrotor control. In phase 1, a residual dynamics module computes ground-truth disturbance residuals r_f to supervise the Residual Perception Network and to train the RL policy via the rollout buffer. In phase 2, historical states are used to predict residuals $\hat{r}_f$, which augment the policy input to generate control actions in windy environments.

Training samples are collected online in a dedicated residual replay buffer, decoupled from the policy buffer, improving data efficiency and stabilizing learning.

To handle non-stationary residuals and sparse wind excitation, we adopt a two-phase strategy that separates exploration, residual learning, and deployment. In phase I (Exploration & Bootstrapping): TD3 is trained under randomized winds, while the residual network is updated online from the buffer every 5 steps; In phase II (Residual Refinement): the actor and critics are frozen; the residual network is refined offline for 200 epochs to improve estimation accuracy. The trained network predicts residuals online, which are incorporated into translational acceleration:

$$\dot{\mathbf{v}}_t = \frac{1}{m}\mathbf{R}(\mathbf{q}_t)\mathbf{f}_t - g\mathbf{e}_3 + \frac{1}{m}\hat{\mathbf{r}}_f(\mathcal{H}_t) \tag{11}$$

Conditioning on short-horizon histories, the network encodes a compact latent representation of external disturbances—magnitude, direction, and temporal evolution—derived solely from onboard measurements. This residual representation provides a physically interpretable interface for high-level policy learning, enabling robust control under unseen wind conditions.

4.3 Residual-Conditioned Policy Learning via WA-TD3

Using the learned residual dynamics, we integrate the predicted external force into the reinforcement learning pipeline for adaptive quadrotor control under wind disturbances. TD3 serves as the backbone, extended with residual-conditioned state representations.

At each timestep, the policy observes the quadrotor state augmented with the predicted residual:

$$\tilde{\mathbf{s}}_t = [\mathbf{p}_t - \mathbf{p}_{d,t}, \mathbf{v}_t - \mathbf{v}_{d,t}, \mathbf{q}_t, \boldsymbol{\omega}_t, \hat{\mathbf{r}}_f(\mathcal{H}_t)]. \tag{12}$$

This provides explicit disturbance information for proactive compensation.

The actor $\pi_\theta(\tilde{\mathbf{s}}_t)$ outputs normalized control actions corresponding to total thrust and desired body angular rates, while two critics Q_{ϕ_1}, Q_{ϕ_2} evaluate residual-augmented states, mitigating overestimation bias. Actions $\mathbf{a}_t = [a_z, \omega_x, \omega_y, \omega_z]$ are mapped to thrust and body-rate references, tracked by a deterministic low-level controller to ensure stability and actuator compliance. This separation allows high-level policies to focus on disturbance-aware decision making.

The reward function encourages precise trajectory tracking by penalizing position and velocity tracking errors:

$$r_t = -\|\mathbf{p}_t - \mathbf{p}_{d,t}\|_2^2 - 0.1\|\mathbf{v}_t - \mathbf{v}_{d,t}\|_2^2. \tag{13}$$

The learning procedure follows the standard TD3 algorithm with delayed policy updates and target smoothing. Given a transition $(\tilde{\mathbf{s}}_t, \mathbf{a}_t, r_t, \tilde{\mathbf{s}}_{t+1})$, the target Q-value is computed as

$$y_t = r_t + \gamma \min_{i=1,2} Q_{\phi_i'}\left(\tilde{\mathbf{s}}_{t+1}, \pi_{\theta'}(\tilde{\mathbf{s}}_{t+1}) + \epsilon\right), \tag{14}$$

where ϵ is clipped Gaussian noise for target policy smoothing. The critics are updated by minimizing the Bellman error, while the actor is updated less frequently by maximizing the expected Q-value estimated by the first critic. Incorporating the residual-augmented state $\tilde{\mathbf{s}}_t$ ensures that learned policies proactively compensate for wind-induced disturbances.

5 Experiments and Results

We evaluate the robustness and generalization of WA-TD3 under both nominal and severe wind disturbances. Extensive experiments are conducted in high-fidelity simulation, with a particular focus on trajectory tracking accuracy under out-of-distribution (OoD) wind conditions. WA-TD3 is systematically compared against six representative baselines spanning classical control, model-free reinforcement learning, hybrid learning–control frameworks, and disturbance-aware methods, providing a comprehensive assessment across diverse control paradigms.

Specifically, PID [Lopez-Sanchez and Moreno-Valenzuela, 2023] serves as traditional model-based controllers relying on manual tuning and linearized dynamics. TD3 [Fujimoto *et al.*, 2018] is included as a model-free RL baseline without explicit disturbance modeling. Neural-Fly [O'Connell *et al.*, 2022], linear method (Linear) and OoD-Control [Ye *et al.*, 2024], and OMAC [Shi *et al.*, 2021] represent recent learning-based and adaptive control approaches with varying levels of disturbance awareness and online adaptation. To isolate the effect of model expressiveness, we also include a lightweight linear residual estimator that maps short-horizon state history to external force estimates.

All methods are trained and evaluated under identical state observations, actuator constraints, and in-distribution wind conditions, ensuring fair and controlled comparisons.

5.1 Experimental Setup

All experiments are conducted in *WindSim*, our in-house quadrotor simulation environment with an aerodynamic wind-disturbance model. Wind is modeled as a piecewise-constant 3D velocity vector updated every 2 s. During training, each Cartesian component is generated by sampling a nonnegative magnitude from a Gamma distribution and assigning it an independent random sign:

$$w_{t,i} = s_{t,i} m_{t,i}, \qquad m_{t,i} \sim \mathrm{Gamma}(1.0, \beta_e), \tag{15}$$

where $s_{t,i} \sim \mathrm{Unif}\{-1, +1\}$, $i \in \{x, y, z\}$, e is the episode index, and $\beta_e = 0.5((e \bmod 3) + 1)$. Thus, $\beta_e \in \{0.5, 1.0, 1.5\}$, corresponding to mean component-wise wind magnitudes of 0.5, 1.0, and 1.5 m/s. During testing, each component is independently sampled as

$$w_{t,i} \sim \mathcal{U}(-W_v, 0), \qquad i \in \{x, y, z\}, \tag{16}$$

where W_v denotes the prescribed test wind level. This setup enables evaluation of out-of-distribution generalization under distribution-shifted wind conditions.

Parameter	Value
State dimension (s_{dim})	16
Action dimension (a_{dim})	4
Hidden layer sizes	[256, 256, 128]
Replay buffer capacity	10^6
Batch size	256
Warm-up steps ($start\_learn$)	10,000
Learning rate (lr)	3×10^{-4}
Initial exploration variance (var_init)	1
Variance decay rate (var_decay)	0.9995
Minimum variance (var_min)	0.001
Discount factor (γ)	0.99
Soft update rate	5×10^{-3}
Policy noise	0.1
Noise clip	0.2
Policy update frequency	2

Table 1: Hyperparameters for WA-TD3 method.

Implementation Details and Hyperparameters. All methods are implemented in Python using PyTorch 1.13 and trained on a single NVIDIA RTX 3090 Ti GPU. Both the TD3 actor–critic networks and the residual estimator are implemented as multi-layer perceptrons with ReLU activations.

Phase I performs joint policy learning and online residual estimation under stochastic wind sampling, promoting coordinated adaptation of control and disturbance perception. Phase II freezes the policy and refines the residual estimator offline to enhance disturbance prediction accuracy. The history length is set to $k = 1$, corresponding to a 0.01 s temporal window at 100 Hz control frequency. Key hyperparameters are summarized in Table 1.

Dynamics Modeling. The quadrotor dynamics follow standard Newton–Euler equations:

$$\dot{\mathbf{p}}_t = \mathbf{v}_t, \tag{17}$$

$$\dot{\mathbf{q}}_t = \frac{1}{2}\,\Omega(\boldsymbol{\omega}_t)\mathbf{q}_t, \tag{18}$$

$$\dot{\mathbf{v}}_t = \frac{1}{m}\mathbf{R}(\mathbf{q}_t)\mathbf{f}_t - g\mathbf{e}_3 + \frac{1}{m}\mathbf{F}_{\text{wind},t}, \tag{19}$$

$$\dot{\boldsymbol{\omega}}_t = \mathbf{J}^{-1}\Big(\boldsymbol{\tau}_t - \boldsymbol{\omega}_t \times (\mathbf{J}\boldsymbol{\omega}_t) + \boldsymbol{\tau}_{\text{wind},t}\Big), \tag{20}$$

where $\mathbf{p}_t, \mathbf{v}_t \in \mathbb{R}^3$ are the position and linear velocity, $\mathbf{q}_t \in \mathbb{R}^4$ is the attitude quaternion, and $\boldsymbol{\omega}_t \in \mathbb{R}^3$ is the angular velocity. The matrices $\mathbf{R}(\mathbf{q}_t)$ and $\mathbf{J} \in \mathbb{R}^{3\times 3}$ denote the rotation matrix from body to world frame and the inertia matrix, respectively. Wind-induced effects are injected through $\mathbf{F}_{\text{wind},t}$ and $\boldsymbol{\tau}_{\text{wind},t}$. Motor dynamics are approximated using a first-order system.

Evaluation Metrics. Tracking performance is quantified using the translational position error, emphasizing disturbance rejection under wind perturbations. The primary metric is the Average Control Error (ACE):

$$\mathrm{ACE} = \frac{1}{T}\sum_{t=1}^{T} \left\|\mathbf{p}_t - \mathbf{p}_{d,t}\right\|_2, \tag{21}$$

where $\mathbf{p}_t$ and $\mathbf{p}_{d,t}$ denote the actual and reference positions at timestep t, respectively, and T is the total trajectory length. Lower ACE indicates better tracking accuracy and robustness against disturbances.

5.2 Overall Performance Comparison

Table 2 summarizes the ACE across four representative trajectories under increasing wind intensities. All learning-based methods are trained under Gamma-distributed wind and evaluated on unseen uniform headwinds, forming a strict OoD generalization scenario.

Across all trajectories and wind conditions, WA-TD3 consistently achieves the lowest ACE in all OoD test cases. On average, it reduces ACE by 62.4% compared to the strongest baseline (OoD-Control) and by 71.8% relative to PID. Under extreme Gale conditions (12 m/s), WA-TD3 maintains stable tracking with ACE values of 0.05–0.06 m, whereas PID and OoD-Control degrade to 0.31–0.35 m and 0.14–0.18 m, respectively. Moreover, WA-TD3 exhibits consistently lower variance, highlighting stable and repeatable behavior across diverse tasks.

Method	Trajectory	ACE for Different Wind Distributions (m) ↓			
		No Wind (0 m/s)	Breeze (4 m/s)	Strong Breeze (8 m/s)	Gale (12 m/s)
PID	Hover	$\underline{0.033 \pm 0.000}$	0.197 ± 0.016	0.260 ± 0.019	0.305 ± 0.028
Linear		0.133 ± 0.000	0.189 ± 0.035	0.212 ± 0.031	0.230 ± 0.034
OMAC		0.040 ± 0.000	0.089 ± 0.010	$\underline{0.114 \pm 0.006}$	0.157 ± 0.022
Neural-Fly		0.041 ± 0.002	0.116 ± 0.022	0.139 ± 0.010	0.172 ± 0.025
OoD-Control		0.040 ± 0.001	$\underline{0.089 \pm 0.008}$	0.118 ± 0.014	$\underline{0.138 \pm 0.006}$
WA-TD3 (Ours)		$\mathbf{0.021 \pm 0.000}$	$\mathbf{0.024 \pm 0.003}$	$\mathbf{0.040 \pm 0.004}$	$\mathbf{0.050 \pm 0.005}$
PID	Sin-forward	0.181 ± 0.000	0.227 ± 0.020	0.284 ± 0.025	0.315 ± 0.020
Linear		0.167 ± 0.037	0.176 ± 0.023	0.181 ± 0.018	0.200 ± 0.023
OMAC		0.112 ± 0.017	0.141 ± 0.006	0.167 ± 0.018	0.202 ± 0.022
Neural-Fly		0.102 ± 0.009	0.123 ± 0.007	0.152 ± 0.016	$\underline{0.166 \pm 0.015}$
OoD-Control		$\underline{0.098 \pm 0.015}$	$\underline{0.119 \pm 0.015}$	$\underline{0.143 \pm 0.022}$	0.171 ± 0.023
WA-TD3 (Ours)		$\mathbf{0.033 \pm 0.000}$	$\mathbf{0.038 \pm 0.002}$	$\mathbf{0.051 \pm 0.004}$	$\mathbf{0.062 \pm 0.006}$
PID	Figure-8	0.236 ± 0.000	0.259 ± 0.024	0.302 ± 0.020	0.353 ± 0.040
Linear		0.187 ± 0.015	0.213 ± 0.026	0.221 ± 0.026	0.253 ± 0.033
OMAC		$\underline{0.120 \pm 0.006}$	$\underline{0.138 \pm 0.010}$	0.160 ± 0.011	0.199 ± 0.021
Neural-Fly		0.141 ± 0.012	0.172 ± 0.019	0.196 ± 0.027	0.210 ± 0.036
OoD-Control		0.140 ± 0.013	0.139 ± 0.015	$\underline{0.146 \pm 0.006}$	$\underline{0.169 \pm 0.012}$
WA-TD3 (Ours)		$\mathbf{0.025 \pm 0.000}$	$\mathbf{0.035 \pm 0.003}$	$\mathbf{0.050 \pm 0.006}$	$\mathbf{0.064 \pm 0.010}$
PID	Spiral-up	0.170 ± 0.000	0.198 ± 0.016	0.260 ± 0.011	0.312 ± 0.016
Linear		0.146 ± 0.028	0.139 ± 0.021	0.172 ± 0.029	0.177 ± 0.029
OMAC		$\underline{0.094 \pm 0.014}$	0.139 ± 0.032	$\underline{0.139 \pm 0.020}$	0.164 ± 0.023
Neural-Fly		0.121 ± 0.014	0.129 ± 0.019	0.140 ± 0.026	0.170 ± 0.023
OoD-Control		0.104 ± 0.012	$\underline{0.124 \pm 0.015}$	0.146 ± 0.017	$\underline{0.158 \pm 0.014}$
WA-TD3 (Ours)		$\mathbf{0.032 \pm 0.000}$	$\mathbf{0.034 \pm 0.002}$	$\mathbf{0.047 \pm 0.003}$	$\mathbf{0.059 \pm 0.005}$

Table 2: Average Control Error (ACE) for different wind intensities. **Bold** indicates the best performer, underline indicates the second-best.

Qualitative trajectory results in Fig. 3 further corroborate these findings: baseline methods suffer from drift and trajectory distortion under strong wind, whereas WA-TD3 closely follows reference paths with smooth and stable motion, demonstrating effective real-time wind compensation without the need for dedicated wind sensors.

5.3 Ablation Studies

We perform ablation studies to evaluate the contributions of two key design choices in WA-TD3: (i) the residual-based wind compensation, and (ii) the length of temporal state history used for residual estimation. These experiments quantify how each component affects trajectory tracking accuracy and robustness under varying wind conditions.

Effect of Wind Compensation Residual. We evaluate the impact of the proposed residual $\hat{\mathbf{r}}_f$ by comparing WA-TD3 against vanilla TD3 across four trajectories (Hover, Sin-forward, Figure-8, Spiral-up) under wind velocities from 0 to 12 m/s. Performance is quantified using ACE with standard deviation over ten random seeds.

As shown in Fig. 4, WA-TD3 consistently outperforms TD3. At 0 m/s, both methods achieve low ACE (TD3: $\sim$0.08 m, WA-TD3: $\sim$0.02 m). Under stronger winds (12 m/s), TD3's ACE rises sharply to 0.19–0.21 m, while WA-TD3 maintains 0.05–0.06 m, representing a $\sim$70% reduction. The advantage is particularly pronounced in complex trajectories (e.g., Spiral-up), where TD3 suffers from unmodeled drift, whereas WA-TD3 leverages $\hat{\mathbf{r}}_f$ as a predictive disturbance proxy, enabling proactive compensation and robust OoD performance without additional sensors or tuning. Moreover, WA-TD3 exhibits lower variance (std $<$ 0.010 m vs. TD3's up to 0.036 m at 12 m/s), indicating more stable and predictable behavior. This improvement stems from explicitly disentangling disturbance dynamics from the control policy, avoiding overcompensation and high-frequency jitter common in implicit adaptation schemes.

These results demonstrate that incorporating explicit, sensor-free disturbance observables is critical for fast and reliable adaptation under extreme OoD wind conditions. By encoding unmodeled aerodynamics as a learnable residual, WA-TD3 achieves superior tracking precision, smoother control, and enhanced robustness without relying on online estimation, external sensors, or environment-specific tuning.

Impact of Temporal State History Length. We investigate the effect of state history length k on the residual-based wind compensation by ablating $k \in \{1, 2, 3, 4, 5\}$ on the Sin-forward trajectory under wind speeds of 0, 4, 8, and 12 m/s. The residual $\hat{\mathbf{r}}_f$ is computed as the difference between the predicted and nominal external forces over the selected horizon, while ACE is evaluated as the mean position tracking error.

As shown in Fig. 5, the state history length $k = 1$ performs best under the tested wind models and control frequency, reaching 0.062 m at 12 m/s—15–30% lower than longer horizons. At low wind speeds (0–4 m/s), ACE is similar across k (0.038–0.065 m), indicating that a single-step residual suffices for effective disturbance observability. For higher wind intensities, longer histories ($k \geq 3$) introduce delayed responses and noise accumulation, increasing ACE (e.g., $k = 5$: 0.088 m at 12 m/s) due to over-smoothing of

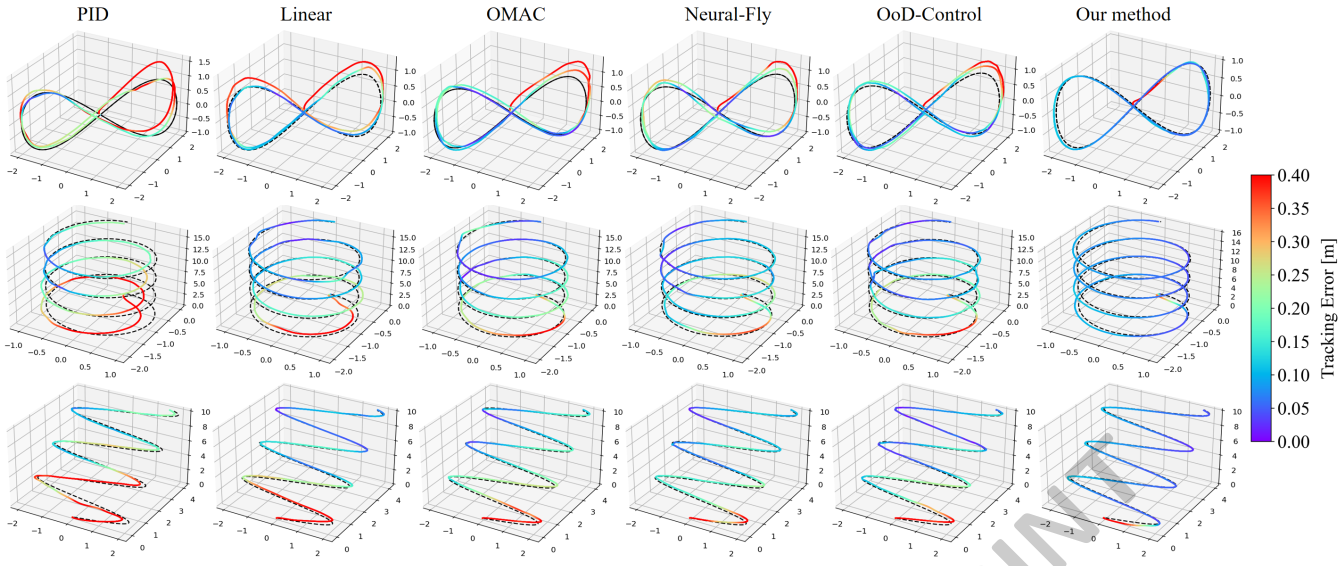

Figure 3: Qualitative trajectory tracking under 12 m/s Gale wind (from left to right: figure-8, spiral-up, and sin-forward trajectories). WA-TD3 maintains smooth and accurate tracking, while baseline controllers exhibit drift and oscillations.

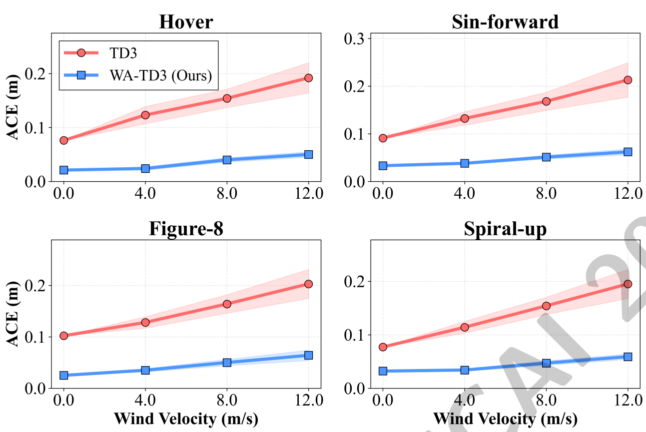

Figure 4: Comparison of TD3 and WA-TD3 across different trajectories and wind intensities. Shaded regions indicate standard deviation.

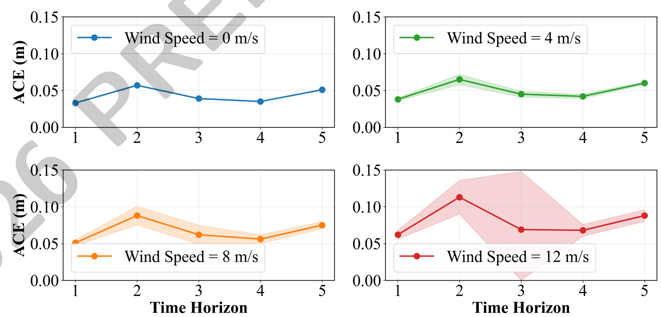

Figure 5: Average Control Error (ACE) on the Sin-forward trajectory across varying wind velocities.

transient gusts.

The optimal state history length $k = 1$ matches the control frequency and the temporal correlation of Dryden gusts: a single-step residual captures instantaneous model mismatch, enabling precise, low-latency compensation. Longer histories, though richer, dilute acute disturbances and increase sensitivity to velocity estimation errors. Overall, it demonstrates that the optimal state history length provides a lightweight, highly effective disturbance observable, maximizing robustness and responsiveness in severe, time-varying wind conditions without additional computational overhead.

6 Conclusion

We presented WA-TD3, a RL framework that incorporates wind disturbance awareness for robust quadrotor control under strong and out-of-distribution wind conditions. By leveraging dynamics residuals from onboard state evolution, WA-TD3 enables implicit wind perception without dedicated sensors or precise aerodynamic models. Integrating this residual into a perception-augmented TD3 policy allows proactive compensation of wind effects rather than treating them as unstructured noise. Experiments show that WA-TD3 greatly improves trajectory tracking accuracy and stability under varied wind strengths and novel disturbances, proving the value of residual-based wind-aware RL for adaptive UAV control. Future work will explore more turbulent conditions and multi-agent systems to advance wind-aware UAV operations.

Acknowledgments

This paper is supported by the National Key Research and Development Program for Young Scientists (No. 2025YFB4711500), National Natural Science Foundation of China (No. 42474027, No. 62403085), Xiaomi Young Talents Program and Graduate Research and Innovation Foundation of Chongqing, China (No. CYB240066).

References

[Abdellatif *et al.*, 2025] Alaa Awad Abdellatif, Ali Elmancy, Amr Mohamed, Ahmed Massoud, Wadha Lebda, and Khalid K Naji. Pdsr: Efficient uav deployment for swift and accurate post-disaster search and rescue. *IEEE Internet of Things Magazine*, 8(3):149–156, 2025.

[Ahmed and Woolsey, 2025] Zakia Ahmed and Craig A Woolsey. An invariant extended kalman filter for wind estimation using a small, fixed-wing uncrewed aerial vehicle. *IEEE Transactions on Control Systems Technology*, 2025.

[Andres *et al.*, 2025] Alain Andres, Aritz D Martinez, Sümer Tunçay, and Ignacio Carlucho. Evaluating reinforcement learning-based neural controllers for quadcopter navigation in windy conditions. *Engineering Applications of Artificial Intelligence*, 161:112090, 2025.

[Cioffi *et al.*, 2025] Giovanni Cioffi, Leonard Bauersfeld, and Davide Scaramuzza. Hdvio2.0: Wind and disturbance estimation with hybrid dynamics vio. *IEEE Transactions on Robotics*, 41:5396–5410, 2025.

[Fujimoto *et al.*, 2018] Scott Fujimoto, Herke Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *International conference on machine learning*, pages 1587–1596. PMLR, 2018.

[Gao *et al.*, 2025] Feng Gao, Chao Yu, Yu Wang, and Yi Wu. Neural internal model control: Learning a robust control policy via predictive error feedback. *IEEE Robotics and Automation Letters*, 2025.

[Gu *et al.*, 2018] Jingjing Gu, Tao Su, Qiuhong Wang, Xiaojiang Du, and Mohsen Guizani. Multiple moving targets surveillance based on a cooperative network for multi-uav. *IEEE Communications Magazine*, 56(4):82–89, 2018.

[Gu *et al.*, 2025] Fuqiang Gu, Jiangshan Ai, Xu Lu, Xianlei Long, Yan Li, Tao Jiang, Chao Chen, and Huidong Liu. Heteroscedastic bayesian optimization-based dynamic pid tuning for accurate and robust uav trajectory tracking. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 17428–17435. IEEE, 2025.

[Hanover *et al.*, 2021] Drew Hanover, Philipp Foehn, Sihao Sun, Elia Kaufmann, and Davide Scaramuzza. Performance, precision, and payloads: Adaptive nonlinear mpc for quadrotors. *IEEE Robotics and Automation Letters*, 7(2):690–697, 2021.

[Hollenbeck *et al.*, 2018] Derek Hollenbeck, Gregorio Nunez, Lance E Christensen, and YangQuan Chen. Wind measurement and estimation with small unmanned aerial systems (suas) using on-board mini ultrasonic anemometers. In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 285–292. IEEE, 2018.

[Hovakimyan and Cao, 2010] Naira Hovakimyan and Chengyu Cao. *1 adaptive control theory: Guaranteed robustness with fast adaptation*. SIAM, 2010.

[Huang *et al.*, 2023] Kevin Huang, Rwik Rana, Alexander Spitzer, Guanya Shi, and Byron Boots. Datt: Deep adaptive trajectory tracking for quadrotor control. In *Conference on Robot Learning*, pages 326–340. PMLR, 2023.

[Ilyas *et al.*, 2025] Muhammad Ilyas, Bernd R Noack, Gang Hu, Hao Xiong, Nan Gao, Awais Khan, and Yang Yang. Drone anemometry of atmospheric winds—a review. *Physics of Fluids*, 37(6), 2025.

[Kong *et al.*, 2025] Chuixu Kong, Bin Wang, Liang He, Qinglong Jiang, and Chang Liu. Ann-mpc: An adaptive neural network-based nonlinear model predictive control method for robust trajectory tracking of quadrotors. *IEEE Robotics and Automation Letters*, 2025.

[Lee *et al.*, 2025] Easop Lee, Samuel A Moore, and Boyuan Chen. Sym2real: Symbolic dynamics with residual learning for data-efficient adaptive control. *arXiv preprint arXiv:2509.15412*, 2025.

[Liu *et al.*, 2025a] Huidong Liu, Xianlei Long, Yong Li, Jinjin Yan, Mingyan Li, Chao Chen, Fuqiang Gu, Huayan Pu, and Jun Luo. Adaptive multi-uav cooperative path planning based on novel rotation artificial potential fields. *Knowledge-Based Systems*, 317:113429, 2025.

[Liu *et al.*, 2025b] Yan-Jun Liu, Benke Gao, Dengxiu Yu, Dapeng Li, and Lei Liu. Neuro-adaptive fault-tolerant attitude control of a quadrotor uav with flight envelope limitation and feedforward compensation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025.

[Lopez-Sanchez and Moreno-Valenzuela, 2023] Ivan Lopez-Sanchez and Javier Moreno-Valenzuela. Pid control of quadrotor uavs: A survey. *Annual Reviews in Control*, 56:100900, 2023.

[Ma *et al.*, 2023] Bodi Ma, Zhenbao Liu, Qingqing Dang, Wen Zhao, Jingyan Wang, Yao Cheng, and Zhirong Yuan. Deep reinforcement learning of uav tracking control under wind disturbances environments. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2023.

[O'Connell *et al.*, 2022] Michael O'Connell, Guanya Shi, Xichen Shi, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, and Soon-Jo Chung. Neural-fly enables rapid learning for agile flight in strong winds. *Science Robotics*, 7(66):eabm6597, 2022.

[Pereida and Schoellig, 2018] Karime Pereida and Angela P Schoellig. Adaptive model predictive control for high-accuracy trajectory tracking in changing conditions. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7831–7837. IEEE, 2018.

[Pravitra *et al.*, 2020] Jintasit Pravitra, Kasey A Ackerman, Chengyu Cao, Naira Hovakimyan, and Evangelos A Theodorou. 1-adaptive mppi architecture for robust and agile control of multirotors. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7661–7666. IEEE, 2020.

[Rosales *et al.*, 2018] C Rosales, S Tosetti, C Soria, and F Rossomando. Neural adaptive pid control of a quadrotor using efk. *IEEE Latin America Transactions*, 16(11):2722–2730, 2018.

[Shao *et al.*, 2025] Xingling Shao, Jun Du, Yi Xia, Zekai Zhang, Xiangwang Hou, and Merouane Debbah. Efficient path-following for urban logistics: A fuzzy control strategy for consumer uavs under disturbance constraints. *IEEE Transactions on Consumer Electronics*, 2025.

[Shi *et al.*, 2021] Guanya Shi, Kamyar Azizzadenesheli, Michael O'Connell, Soon-Jo Chung, and Yisong Yue. Meta-adaptive nonlinear control: Theory and algorithms. *Advances in Neural Information Processing Systems*, 34:10013–10025, 2021.

[Sikkel *et al.*, 2016] LNC Sikkel, Guido CHE de Croon, Christophe De Wagter, and QP Chu. A novel online model-based wind estimation approach for quadrotor micro air vehicles using low cost mems imus. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2141–2146. IEEE, 2016.

[Sun *et al.*, 2025] Sihao Sun, Xuerui Wang, Dario Sanalitro, Antonio Franchi, Marco Tognon, and Javier Alonso-Mora. Agile and cooperative aerial manipulation of a cable-suspended load. *Science Robotics*, 10(107):eadu8015, 2025.

[Vaidya and Keshavan, 2025] Varad Vaidya and Jishnu Keshavan. Dynamics-invariant quadrotor control using scale-aware deep reinforcement learning. *arXiv preprint arXiv:2503.09622*, 2025.

[Wang *et al.*, 2025] Shuai Wang, Xinquan Chen, Haoyu Wei, and Quan Quan. Heading adjustment by admittance control for lifting-wing quadcopters in strong winds. *IEEE Robotics and Automation Letters*, 2025.

[Xue *et al.*, 2023] Jun Xue, Ziniu Liu, Guanjun Liu, Ziyuan Zhou, Kaiwen Zhang, Ying Tang, and Jiacun Wang. Robust wind-resistant hovering control of quadrotor uavs using deep reinforcement learning. *IEEE Transactions on Intelligent Vehicles*, 2023.

[Yang *et al.*, 2025] Tao Yang, Huai-Ning Wu, and Jun-Wei Wang. cc-drl: A convex combined deep reinforcement learning flight control design of a morphing quadrotor. *IEEE Transactions on Cybernetics*, 2025.

[Ye *et al.*, 2024] Nanyang Ye, Zhaoyu Zeng, Jundong Zhou, Lin Zhu, Yuxiao Duan, Yifei Wu, Junqi Wu, Haoqi Zeng, Qinying Gu, Xinbing Wang, et al. Ood-control: generalizing control in unseen environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(11):7421–7433, 2024.

[Zhang *et al.*, 2025a] Dingqi Zhang, Antonio Loquercio, Jerry Tang, Ting-Hao Wang, Jitendra Malik, and Mark W Mueller. A learning-based quadcopter controller with extreme adaptation. *IEEE Transactions on Robotics*, 2025.

[Zhang *et al.*, 2025b] Ruiqi Zhang, Dingqi Zhang, and Mark W Mueller. Proxfly: Robust control for close proximity quadcopter flight via residual reinforcement learning. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13683–13689. IEEE, 2025.