IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 November 2023, accepted 6 December 2023, date of publication 8 December 2023,
date of current version 13 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3341212

RESEARCH ARTICLE

Optimizing Subway Train Operation With Hierarchical Adaptive Control Approach

GAOYUN CHENG[1], DIANYUAN WANG[1], MING SUN[1], ZHE FU[1], BINBIN YUAN[1], LEI ZHANG[1,2], AND XIAO XIAO[1]

[1]Traffic Control Technology Company Ltd., Beijing 100070, China
[2]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Gaoyun Cheng (gycheng96@163.com)

This work was supported by the National Key Research and Development Program of China under Grant SQ2020YFB1600702.

ABSTRACT The proportional integral derivative (PID) method is widely used in industrial control applications. However, when applied to complex and dynamic train operation control systems, real-time parameter adjustment becomes a formidable challenge. Moreover, the multifaceted nature of train operation control, encompassing safety, parking precision, passenger comfort, and energy efficiency, exacerbates the difficulty of parameter adjustment. To address this problem, this paper formulates train operation control as a Markov decision process (MDP) and introduces an innovative adaptive control approach. This approach features a hierarchical structure comprising an upper-level deep deterministic policy gradient (DDPG) controller and a lower-level PID controller, leveraging the learning capability of the DDPG algorithm, as well as the stability and interpretability of the PID method. The upper-level controller acquires train status information and autonomously fine-tunes the PID parameters, while the lower-level controller accepts these parameters and adjusts the percentage of traction or braking to achieve train operation control. Furthermore, the reward function has been meticulously designed to reconcile the diverse objectives of train operation. Extensive experiments conducted on a subway simulation platform substantiate the effectiveness and adaptability of the proposed approach in various operational scenarios.

INDEX TERMS Subway train operation, reinforcement learning, adaptive control, reward function.

I. INTRODUCTION

Efficient subway train operation is paramount for ensuring the high performance and reliability of urban transportation systems. In the face of ever-increasing public transportation demand, optimizing subway train system operations has become imperative in the aim to enhance overall performance [1].

The control of subway train operation mainly focuses on two aspects [2]: reference speed optimization and speed tracking control. Reference speed optimization entails the pre-calculation of an ideal speed profile for the train journey between stations. This profile is meticulously designed to encompass critical operational objectives encompassing safety, parking precision, passenger comfort, and energy efficiency. This optimized speed profile acts as a pivotal reference, thereby harmonizing with the train's operational dynamics and guiding the formulation of an effective speed control strategy. Speed tracking control, also known as subway train operation control, aims to ensure that the train operates as closely as possible to the reference speed profile, thus resulting in desired operational outcomes. Fig. 1 shows the speed-position/time curve and the applied train forces. During acceleration, the train experiences traction forces that propel it forward. Meanwhile, during braking, deceleration is achieved through braking force. Throughout the entire operational process, the train encounters resistance. Consequently, under the combined influence of these various forces, the train endeavors to closely adhere to the prescribed reference speed profile.

A highly efficient and reliable control system is essential for achieving effective train operation between stations. The most classical and commonly used method for train speed control is the proportional-integral-derivative (PID)

The associate editor coordinating the review of this manuscript and approving it for publication was Jesus Felez.

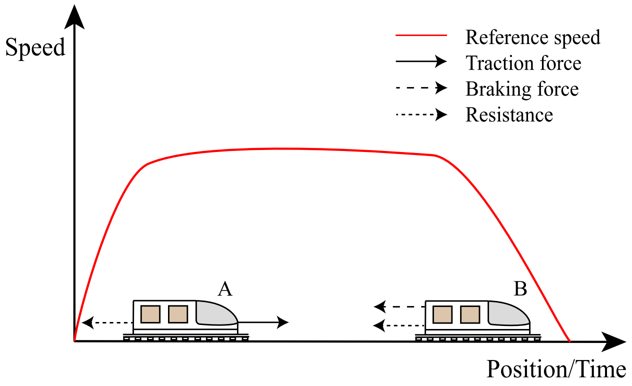


FIGURE 1. Speed-position/time curve and the applied train forces. Subway trains A and B simulate the force distribution during the traction and braking processes, respectively.

controller [3], [4], which is still applied in subway lines like the Yizhuang Line and Changping Line of Beijing Subway. The PID controller is widely used in the industry due to its simplicity and good robustness. However, conventional PID controller parameter tuning relies on manual experience and repetitive on-site adjustments, thus leading to high costs and difficulties in achieving dynamic adaptive parameter adjustments [5]. Moreover, during train operation, various internal and external factors, such as mechanical wear, normal aging, and weather conditions changes, continuously affect the train. The fixed-parameter PID controller inevitably experiences performance degradation. Additionally, the PID controller tends to frequently output speed adjustment commands to minimize tracking errors, thus possibly resulting in poor passenger comfort and an inability to balance multiple objectives [6]. Therefore, an advanced control method is needed to overcome these limitations and to enhance the efficiency and effectiveness of subway train operations.

Researchers have explored different methods to improve speed tracking performance. Fuzzy control is one of the most commonly used train operation control methods [7]. Fuzzy control involves fuzzifying the inputs of a control system, establishing a fuzzy rule base, and completing fuzzy inference. For instance, Pu et al. [8] designed a fuzzy PID controller to adaptively adjust PID gains, and they considered multiple objectives, including punctuality, energy consumption, parking accuracy, and comfort, to improve the tracking performance of the nonlinear train system. Similarly, for balancing multiple objectives, Zhu et al. [9] proposed a multi-objective model for urban railway train automatic operation, as well as designed a fuzzy controller to control train operations.

The fuzzy control methods heavily rely on the formulation of logical rules, which can be non-trivial [10]. Furthermore, researchers have recently incorporated neural networks through which to enhance train operation controllers. Sun et al. [11] investigated an adaptive neural fuzzy sliding mode controller to suppress the disturbance effects on the train model parameters, thus demonstrating its effectiveness in speed tracking control. Pu et al. [12] proposed an adaptive control method for subway trains, whereby the time-varying parameters of train motion were considered and a train model with dynamic parameters was established. They designed a model-free adaptive control system by combining neural networks and the PID algorithm to achieve adaptive control. However, the neural network methods based on supervised learning heavily rely on the quality and quantity of the samples.

Some of the above methods treat most parameters of the train system model as constants [13], [14], In actual subway operations, the traction/braking force and resistance of the train vary at different speeds. To overcome the limitations of previous methods, we propose a hierarchical adaptive control approach through which to optimize subway train operation. This approach leverages deep reinforcement learning technology to fine-tune PID parameters. Reinforcement learning enables agents to learn through interaction with the environment. It allows intelligent agents to autonomously learn and improve their behavioral strategies based on feedback and reward signals from the environment. Deep reinforcement learning combines deep neural networks with reinforcement learning algorithms to enable agents to learn complex representations and features from high-dimensional, unstructured input data, resulting in improved generalization and robustness when facing unseen states [15]. Deep reinforcement learning has achieved success in adaptive PID parameter tuning and found wide applications in various domains [16]. For example, in wind turbine control [17], [18], robot control [19], [20], [21], and unmanned aerial vehicle attitude control [22], [23], deep reinforcement learning has demonstrated its effectiveness. However, to the best of our knowledge, this strategy has not been applied in the field of train operation control. Thus, we explore this direction by adopting deep reinforcement learning to adaptively adjust PID parameters in train operation control, thereby expecting to achieve similar successful results in this domain.

As a whole, the proposed approach possesses a hierarchical structure comprising an upper-level deep deterministic policy gradient (DDPG) controller and a lower-level PID controller. The reinforcement learning DDPG algorithm dynamically adjusts PID parameters online, thus allowing for the learning of optimal control strategies for different operating conditions and effective management of the train's complex continuous state and action spaces. The integration of the DDPG algorithm enables the online adjustment of PID parameters through value function approximation, thus facilitating an adaptive control based on continuously changing operational requirements. In addition, in order to balance multiple optimization objectives, the reward function is meticulously designed.

By combining the learning capability of the DDPG algorithm with the stability and interpretability of the PID controller, stable and interpretable control signals can be provided under complex and time-varying conditions, thus achieving precise and efficient train operation.

In summary, the main contributions of this paper are as follows:

1) We propose an adaptive control approach through which to optimize subway train operation, thus addressing the limitations of other PID control strategies.
2) We developed a hierarchical structure that combines the learning capability of the DDPG algorithm with the stability and interpretability of the PID controller, thus enabling accurate tracking of the reference speed profile while considering multiple objectives.
3) We conducted extensive numerical experiments on a city subway simulation platform to evaluate the effectiveness and superiority of the proposed approach.

The rest of this paper is organized as follows. Section II provides a detailed analysis of the train dynamic model, traction/braking force, and resistance, thus laying the theoretical foundation for the proposed approach. In Section III, we elaborate on the adaptive control approach, including the design of the upper-level controller, the lower-level controller, and the reward function. Section IV describes the simulation settings and presents the numerical experiment results, thus demonstrating the adaptability of the proposed approach. Finally, Section V summarizes this paper and suggests potential future research directions.

II. TRAIN MODEL ANALYSIS

A. TRAIN DYNAMIC MODEL

The single-particle train model is the most commonly used model for solving train operation problems [24], [25]. In this paper, we adopt the single-particle model to simulate the train. According to Newton's laws of motion, the train's dynamics can be expressed as

$$\begin{cases} \dfrac{dx}{dt} = v \\ \dfrac{dv}{dt} = \dfrac{F-R}{M} \end{cases} \tag{1}$$

where x, t, and v represent the train's position, time, and speed during its operation, respectively. M represents the mass of the train, and F and R represent the traction/braking force and resistance of the train, respectively.

The proposed approach controls the train by sampling at the time interval Δt, thus enabling an iterative calculation of the train's position and speed using (2) and (3).

$$\begin{cases} \Delta x = v\Delta t + \dfrac{F-R}{2M}\Delta t^2 \\ x = x + \Delta x \end{cases} \tag{2}$$

$$\begin{cases} \Delta v = v + \dfrac{F-R}{M}\Delta t \\ v = v + \Delta v \end{cases} \tag{3}$$

B. TRACTION/BRAKING FORCE

The traction and braking forces of the train are provided by the traction system and braking system, respectively. The maximum traction/braking force that the train can provide at a specific moment depends on the train's speed and is usually represented by the traction/braking force characteristic curve [5], which is illustrated in Fig. 2.

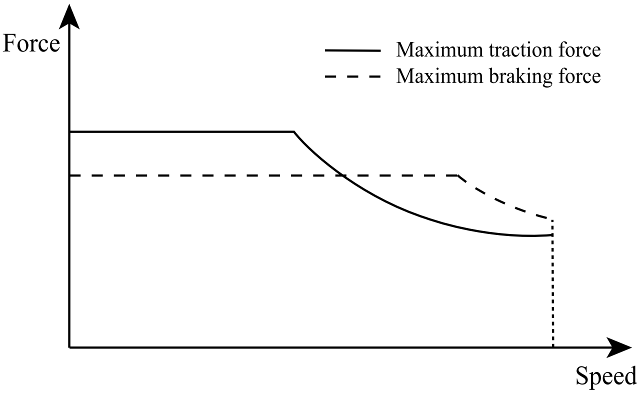


FIGURE 2. The traction and braking characteristic curves of the train.

From Fig. 2, it can be observed that the maximum traction and braking forces remain constant at lower speeds. However, as the speed approaches the critical value, these forces start to decrease. The maximum traction and braking forces at time t can be expressed as

$$\begin{cases} F_{\text{tmax}}(t) = f_{\text{tmax}}(v_t) \\ F_{\text{bmax}}(t) = f_{\text{bmax}}(v_t) \end{cases} \tag{4}$$

where $F_{\text{tmax}}(t)$ and $F_{\text{bmax}}(t)$ represent the maximum traction and braking forces at time t, respectively. $f_{\text{tmax}}(v_t)$ and $f_{\text{bmax}}(v_t)$ are functions describing the variation of the train's maximum traction and braking forces with respect to speed v_t at time t.

The train control system determines the percentage of traction or braking force output [12], which is denoted as u. Therefore, the traction/braking force $F(t)$ can be expressed as

$$\begin{aligned} F(t) &= u(t)F_{\max}(t) \\ &= \begin{cases} u(t)F_{\text{tmax}}(t) & u(t) > 0 \\ 0 & u(t) = 0 \\ u(t)F_{\text{bmax}}(t) & u(t) < 0 \end{cases} \end{aligned} \tag{5}$$

where $F_{\max}(t)$ is the maximum traction/braking force at time t, and $u(t)$ is the percentage of the maximum traction/braking force at time t. When $u(t) > 0$, the traction system applies traction force; when $u(t) < 0$, the braking system applies braking force; and when $u(t) = 0$, both the traction force and braking force are zero.

C. RESISTANCE

Resistance is an important factor that cannot be ignored when controlling a train. There are many influencing factors during a train's operation, thus making it difficult to precisely solve the resistance. Empirical formulas obtained through numerous experiments are typically used for calculation.

In this paper, we use the Davis equation to represent the resistance [26], which is formulated as

$$R(v) = D_1 + D_2 v + D_3 v^2 \tag{6}$$

where D_1, D_2, and D_3 are empirical coefficients, each of which is greater than or equal to zero. The specific values of these coefficients may vary depending on the different trains and track conditions.

From (6), it can be seen that the resistance increases with increasing speed. To evaluate the adaptability of our proposed approach, various combinations of these coefficients were considered.

III. HIERARCHICAL ADAPTIVE CONTROL APPROACH

A. PROBLEM DEFINITION

Subway train control is a complex task that can be effectively addressed by formulating it as an optimal control problem. The main objective is to determine the optimal strategy for adjusting traction and braking forces throughout the entire journey between stations. By carefully adjusting these forces, the train can achieve efficient and safe operation while maximizing performance.

Train control is composed of a series of control commands for the train, and its output depends solely on the current input state of the train, i.e., it is independent of historical states. Therefore, the optimal control of subway train operation can be formulated as a Markov decision process (MDP), which is a fundamental framework used to model decision-making problems involving sequential interactions. An MDP is typically composed of a state space $\mathcal{S}$, an action space $\mathcal{A}$, a state transition function $\mathbf{P}$, a reward function $\mathcal{R}$, and a discount factor γ, which are represented as a quintuple $< \mathcal{S}, \mathcal{A}, \mathbf{P}, \mathcal{R}, \gamma >$.

The state space $\mathcal{S}$ encompasses all possible states s within the environment. In order to comprehensively encapsulate the environment's state and ensure seamless coordination with the lower-level controller, this paper chooses the tracking speed v_t, the percentage of traction/braking u_t, the difference speed $\hat{v}_t$ between the reference speed and tracking speed, and the distance $\hat{d}_t$ between the reference position and tracking position to form the state at time t, which is defined as

$$s_t = [v_t, u_t, \hat{v}_t, \hat{d}_t] \tag{7}$$

The action space $\mathcal{A}$ refers to the set of all possible actions a. In this paper, it consists of the PID parameters $k_p(t)$, $k_i(t)$ and $k_d(t)$, as shown in (8), which will be discussed in Section III-D.

$$a_t = [k_p(t), k_i(t), k_d(t)] \tag{8}$$

The state transition function $\mathbf{P}$ is a conditional probability density function that is denoted as $p(s_{t+1}|s_t, a_t)$. It represents the probability of the state transitioning to s_{t+1} if the current state s_t and action a_t are observed. The state transition function is automatically completed by the train dynamic model.

Reward $\mathcal{R}$ is a numerical value returned by the environment to the agent after executing an action. The reward function design is discussed in Section III-E.

The reward discount factor $\gamma \in [0, 1]$ is used to reduce the weight value with increasing time steps. The goal of reinforcement learning is to find an optimal policy π^* that maximizes the cumulative reward R_t, which can be expressed as

$$R_t = \sum_{i=t}^{T} \gamma^{i-t} r_i \tag{9}$$

where r_i is the reward at time i.

B. APPROACH OVERVIEW

In this section, we provide a comprehensive overview of the proposed approach for optimizing subway train operations. The approach adopts a hierarchical control structure, whereby the advantages of DDPG and PID controllers are combined to improve train operation efficiency and performance, as shown in Fig. 3. The upper-level controller is responsible for adjusting the parameters of the lower-level controller based on environmental state information. The DDPG algorithm effectively addresses the problem of continuous action space in reinforcement learning by combining deterministic policy and the Actor-Critic architecture, thereby reducing exploration complexity and enabling efficient value function estimation. This provides an effective solution for reinforcement learning problems that involve continuous action spaces, such as PID parameter tuning. The lower-level controller dynamically adjusts the percentage of the traction and braking forces applied to the train, thereby allowing it to operate according to the desired control strategy. It is worth noting that the reward function was meticulously designed to balance multiple optimization objectives.

C. UPPER-LEVEL CONTROLLER DESIGN

The upper-level controller design in the proposed approach utilizes reinforcement learning techniques to learn the optimal control strategy for subway train operation. Reinforcement learning involves the interaction between an agent and its environment, with the agent making decisions based on feedback in the forms of states and rewards.

The DDPG is a deep reinforcement learning algorithm developed by Lillicrap et.al [27], and it is composed of two key components (where θ^μ and θ^Q are the parameters): an Actor (online) network $\mu(s|\theta^\mu)$ and a Critic (online) network $Q(s, a|\theta^Q)$.

The Actor network is shown in Fig. 4, and it consists of an input layer, two hidden layers, and an output layer. To enhance non-linear expression capacity, ReLU [28] is used as the activation function for the intermediate layers, and Tanh [29] is used for the output layer. It takes the state information as input and outputs the PID parameters. Note that, due to the use of the Tanh function, the output x_{out} range is limited to $[-1, 1]$, and it is scaled to a reasonable range

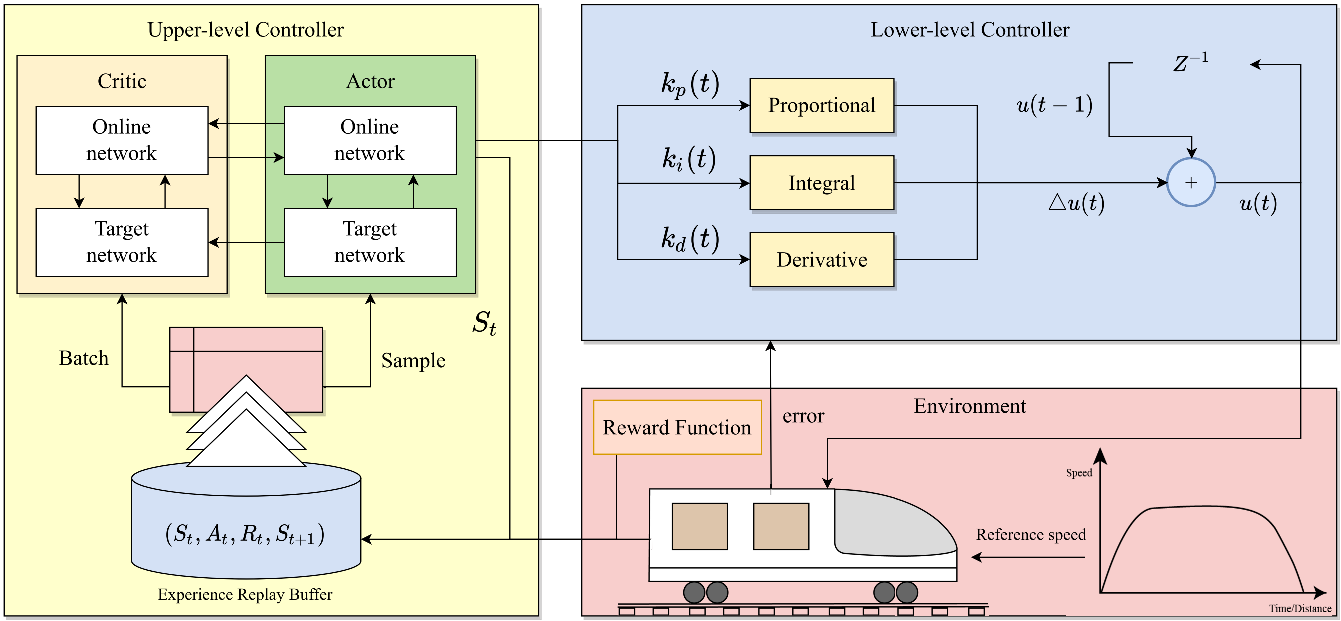


FIGURE 3. Structure of the proposed approach.

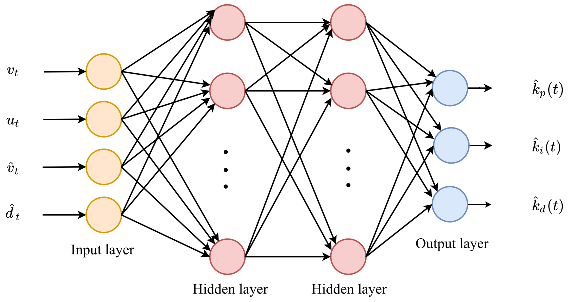


FIGURE 4. Actor network structure.

using (10).

$$\begin{cases} k_p(t) = \dfrac{(\hat{k}_p(t)+1)k_p^{\text{up}}}{2} \\ k_i(t) = \dfrac{(\hat{k}_i(t)+1)k_i^{\text{up}}}{2} \\ k_d(t) = \dfrac{(\hat{k}_d(t)+1)k_d^{\text{up}}}{2} \end{cases} \tag{10}$$

where k_p^{up}, k_i^{up}, and k_d^{up} denote the upper limit of the allowable PID parameter values.

The Critic network is similar to the Actor network, as shown in Fig. 5. The difference lies in the input, which consists of a joining of the state and action. The output layer has no activation function, and it only outputs a real value, which is represented by the Q-value $q(t)$.

The Actor network is updated by calculating the policy gradient, as shown in (11).

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_t \nabla_a Q(s, a \mid \theta^Q)|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^\mu} \mu(s \mid \theta^\mu)|_{s_t} \tag{11}$$

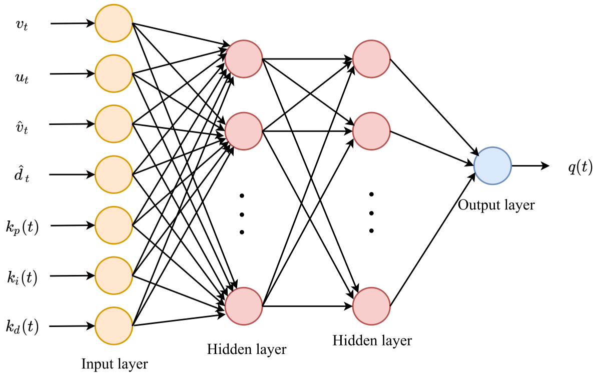


FIGURE 5. Critic network structure.

The parameters in the Critic network are updated by minimizing the value of the loss function L, which is expressed as

$$L = \frac{1}{N} \sum_t (y_t - Q\left(s_t, a_t \mid \theta^Q\right))^2 \tag{12}$$

where y_i is the estimate of the state-action value, and it is defined as

$$y_t = r_t + \gamma Q'\left(s_{t+1}, \mu'\left(s_{t+1} \mid \theta^{\mu'}\right) \mid \theta^{Q'}\right) \tag{13}$$

where $\theta^{\mu'}$ and $\theta^{Q'}$ are the parameters of the target networks.

In the DDPG algorithm, the incorporation of target networks constitutes a pivotal technique that is aimed at enhancing the stability and convergence of the training process. The Actor/Critic target network is a delayed copy of the Actor/Critic online network. During training, the target networks are updated periodically using a soft update approach, as shown in (14).

$$\begin{cases} \theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \\ \theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'} \end{cases} \tag{14}$$

where τ is the update rate of the target networks.

To enhance the learning process, the DDPG algorithm incorporates an experience replay buffer. This buffer stores past experiences, thereby enabling efficient training by reducing data correlation and preventing policy oscillation. During training, small batches of experiences are sampled from the replay buffer, and the DDPG network parameters are updated based on the calculated loss and policy gradient. This iterative process improves the control strategy over time, thereby enabling the upper-level controller to adapt to different operating conditions and to optimize the subway train's performance.

D. LOWER-LEVEL CONTROLLER DESIGN

The lower-level controller plays a crucial role in adjusting the percentage of the traction/braking force based on the parameters received from the upper-level controller.

The PID controller employs three types of items: proportional (P), integral (I), and derivative (D). The proportional item incorporates a suitable proportion of the error (difference between the desired value and the controlled object's output) into the control output. The integral item monitors the changing error variable over time and corrects the output by reducing the offset of the error variable. The derivative item control mode monitors the rate of change in the error variable, thus modifying the output in the presence of abnormal variations. By adjusting the parameters of the three items, the desired performance is obtained from the process.

The incremental PID is widely used in industrial applications as it overcomes the drawback of accumulating significant cumulative errors in the positional PID [19]. Therefore, the incremental PID control law is employed to design the lower-level controller, as shown in (15).

$$\begin{cases} \hat{u}(t) = u(t-1) + \Delta u(t) \\ \Delta u(t) = \begin{pmatrix} k_p(t)[e(t) - e(t-1)] + k_i(t)e(t) \\ +k_d(t)[e(t) - 2e(t-1) + e(t-2)] \end{pmatrix} \\ e(t) = v_r(t) - v(t) \end{cases} \tag{15}$$

where t represents the discrete sampling time, and the coefficients $k_p(t)$, $k_i(t)$, and $k_d(t)$ correspond to the proportional, integral, and derivative parameters of the incremental PID controller at time t, respectively. $\Delta u(t)$ represents the increment value at time t, and the output value $\hat{u}(t)$ at time t is obtained by adding $\Delta u(t)$ to the previous control output $u(t-1)$. The terms $e(t)$, $e(t-1)$, and $e(t-2)$ represent the system error at times t, $t-1$, and $t-2$, respectively, i.e., the difference between the reference speed v_r and the tracking speed v.

To ensure that the final output $u(t)$ of the PID controller stays within the desired range, we utilized the clip function. This function restricts the value of $u(t)$ to the range of -1 to 1, which is expressed as

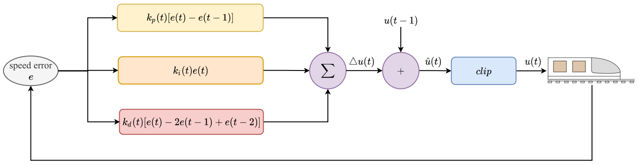


FIGURE 6. Design and integration of the lower-level controller.

$$u(t) = \text{clip}(\hat{u}(t), -1, 1) \tag{16}$$

As described above, The lower-level controller in the proposed approach employs the incremental PID as the PID controller, as well as applies constraints on the output to accurately adjust the percentage of the traction/braking force based on the received control signals. This enables the subway train to respond effectively to continuously changing conditions and helps to accurately track the reference speed curve. Fig. 6 provides an overview of the design and integration of the lower-level controller in the proposed approach.

E. REWARD FUNCTION DESIGN

The design of the reward function is crucial for incentivizing desired behaviors and penalizing undesirable ones. In this paper, the reward function is defined as a weighted combination of individual reward components that correspond to each control objective. The overall reward is represented as R, as shown in (17).

$$\begin{aligned} R = w_1 R_{\text{Safety}} + w_2 R_{\text{Tracking}} + w_3 R_{\text{Parking}} \\ + w_4 R_{\text{Comfort}} + w_5 R_{\text{Efficiency}} \end{aligned} \tag{17}$$

where $w_i (i = 1, 2, 3, 4, 5)$ determine the relative importance of each reward component.

The individual reward components are defined as follows.

1) Safety Reward (R_{Safety}): This component encourages the train to operate below the speed limit to ensure safety. It imposes a penalty for exceeding the speed limit, which is represented by a large negative constant value. It is worth noting that, in each simulation, the current iteration is terminated if the speed exceeds the speed limit.

$$R_{\text{Safety}} = \begin{cases} -C & v > v_{\text{limit}} \\ 0 & \text{else} \end{cases} \tag{18}$$

where C is a large number.

2) Tracking Reward (R_{Tracking}): This component measures the deviation of the train speed and position from the reference values. It aims to provide a positive reward for accurate tracking.

$$R_{\text{Tracking}} = \frac{1}{1 + |v_r - v|} + \frac{1}{1 + |p_r - p|} \tag{19}$$

where v_r and v represent the reference speed and tracking speed, respectively, and p_r and p represent the reference position and tracking position, respectively.

3) Parking Reward (R_{Parking}): This component evaluates the accuracy of parking at the target position. It provides a reward for precise stops within the required tolerance, and it penalizes large position errors. Typical requirements dictate that the error in the parking position should be within ±0.3 m. It is worth noting that this reward only takes effect when the simulation time reaches the pre-planned time. Otherwise, its value is 0. Furthermore, when the simulation time reaches the pre-planned time and its speed is not 0, the iteration is terminated and considered a failure.

$$R_{\text{Parking}} = \begin{cases} \dfrac{1}{1+|p_{\text{target}}-p|} & |p_{\text{target}}-p| < 0.3 \\ -|p_{\text{target}}-p| & \text{else} \end{cases} \tag{20}$$

4) Comfort Reward (R_{Comfort}): Passenger comfort is closely related to the rate of acceleration change called jerk. A lower jerk value indicates higher comfort. Incorporating this component helps achieve smooth acceleration and deceleration, thus enhancing passenger comfort.

$$R_{\text{Comfort}} = \frac{1}{1+|\text{jerk}|} \tag{21}$$

5) Efficiency Reward ($R_{\text{Efficiency}}$): This component encourages energy-efficient operation by penalizing energy-consuming behavior, and it is the time integral of the product of train speed $v(t)$ and traction force $F(t)$.

$$R_{\text{Efficiency}} = \begin{cases} -\displaystyle\int_{\Delta t} v(t) \times F(t)dt & F(t) > 0 \\ 0 & \text{else} \end{cases} \tag{22}$$

Through the integration of multiple rewards, the proposed approach can learn to make decisions that balance multiple objectives, as well as optimize the operation of the subway train accordingly.

F. ALGORITHM STATEMENT

In this section, we present the algorithm of the entire process, as shown in Algorithm 1. The proposed approach first initializes the parameters of the DDPG. It then iterates in a loop that interacts with the environment. During each iteration, the current state of the subway train is obtained, and an appropriate action is chosen based on the current policy determined by the DDPG controller. This action parameterizes the PID controller, which, in turn, determines the traction/braking percentage of the train based on the current state. The environment executes the train dynamics based on this traction/braking ratio, and it then observes the resulting next state, reward, and termination signal. These transition experiences are stored in a replay buffer for experience replay. Then, the network parameters are updated using batch sampling from the replay buffer, thereby calculating the loss and policy gradient.

Algorithm 1 Hierarchical Adaptive Control Algorithm

```
Initialize actor (online) network $\mu(s|\theta^{\mu})$ and critic (online) network $Q(s, a|\theta^{Q})$
Initialize target network $\mu'$ and $Q'$ with weights $\theta^{\mu'} \leftarrow \theta^{\mu}$,$\theta^{Q'} \leftarrow \theta^{Q}$
Initialize the replay buffer $Rbf$ for experience replay
for $epsiode = 1$ to $M$ do
    $done \leftarrow False$
    Receive the initial current state $s_0$ of the subway train
    $t \leftarrow 0$
    while not done do
        Select action $a_t$ according to the current policy and exploration noise
        Calculate the percentage of traction/braking force $u$ according to $s_t$ and $a_t$
        Execute $u$ and observe next state $s_{t+1}$, reward $r_t$ and termination signal $done$
        Store transition $(s_t, a_t, r_t, s_{t+1})$ into replay buffer $Rbf$
        Sample a random minibatch transitions $T_b$ from $Rbf$
        Calculate the policy gradient $\nabla_{\theta^{\mu}} J$ and loss $L$ based on the $T_b$
        Update actor online network parameters using (11)
        Update critic online network parameters using (12)
        Update target networks using (14)
        $s_t \leftarrow s_{t+1}$
        $t \leftarrow t + 1$
    end while
end for
```

The integration of DDPG and PID controllers achieves adaptability and flexibility in the control process. The DDPG controller learns from experience and provides guidance to the PID controller, thus enabling it to adjust its parameters based on different environmental conditions. This adaptive interaction between controllers and the environment contributes to the continuous improvement and optimization of subway train operations.

IV. NUMERICAL EXPERIMENTS

In this section, we establish a train operation simulation environment to conduct experiments to validate the effectiveness of the proposed approach and compare its performance with other popular control methods. Additionally, various reference speeds and resistance parameters are adopted to explore the adaptability of the proposed approach.

A. SIMULATION SETUP

We simulated a subway line with a total length of 1280 m and a speed limit of 80 km/h. The total weight of the train was set to 300 tons. The parameters of the Davis equation were set as $D_1 = 0.6841$, $D_2 = 0.0229$, and $D_3 = 0.000345$.

The characteristics of the traction and braking forces varying with speed were modeled using (23) and (24),

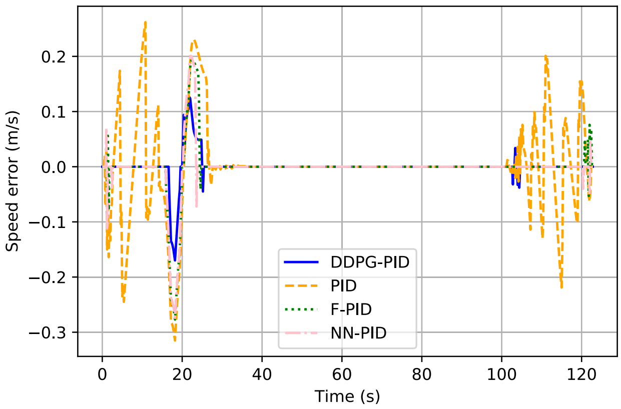


FIGURE 7. Speed tracking error curves.

respectively.

$$F_f(v) = \begin{cases} 345.19 & 0 \leq v \leq 40 \\ \begin{pmatrix} 714.0427 - 6.8018v \\ -0.1069v^2 + 0.0012v^3 \end{pmatrix} & 40 < v \leq 80 \end{cases} \tag{23}$$

$$F_b(v) = \begin{cases} 342.2 & 0 \leq v \leq 49 \\ \begin{pmatrix} 2561.2853 - 87.9215v \\ +1.1091v^2 - 0.0049v^3 \end{pmatrix} & 49 < v \leq 80 \end{cases} \tag{24}$$

where $F_f(v)$ and $F_b(v)$ represent the train's traction force and braking force, respectively (in units of kN, where v denotes the train's running speed in km/h).

For comparison, we considered three other control methods: the PID controller, F-PID controller [8], [9], and NN-PID controller [12]. The F-PID controller uses fuzzy logic and rules to optimize the PID parameters in real time. The NN-PID controller is an adaptive controller that combines neural networks with the PID algorithm for train operation control. In the NN-PID controller, PID parameters are controlled by a neural network, and the squared error between the reference speed and the tracking speed is used as a loss function for supervised training.

During the experiments, we assigned a weight of $w_i = 0.2$ to each reward component, and the constant C in R_{safety} was defined as 100. Additionally, the upper limits for the PID parameters were set as $k_p^{\text{up}} = 5$, $k_d^{\text{up}} = 1.5$, and $k_i^{\text{up}} = 1.5$. The initial PID parameters were specified as $k_p = 3.4$, $k_i = 0.3$, and $k_d = 0.2$ with a sampling interval of 0.2 seconds.

B. MAIN EXPERIMENTS

In this section, we present the experiments conducted to evaluate the performance and effectiveness of the proposed approach. These experiments compare the proposed approach, named the DDPG-PID controller, with other control methods in subway train operation, namely the PID controller, F-PID controller, and NN-PID controller, Fig. 7 and 8 show the speed tracking error and position tracking error, respectively. Table 1 provides additional statistics of the tracking performance.

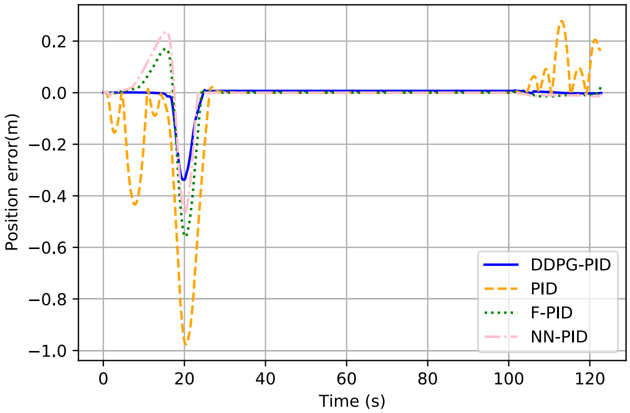


FIGURE 8. Position tracking error curves.

TABLE 1. Statistics of the tracking performance.

	Speed tracking (m/s)		Position tracking (m)			Other indicator	
	Total error	Max. error	Total error	Max. error	Parking error	Energy (kWh)	Jerk (m/s^3)
PID	24.29	0.31	48.96	0.97	0.17	17.55	0.93
F-PID	6.95	0.28	17.49	0.56	0.02	17.40	0.49
NN-PID	6.32	0.27	16.78	0.47	0.00	17.16	0.65
DDPG-PID	3.51	0.17	10.32	0.34	0.00	17.04	0.39

From Fig. 7, it can be observed that in terms of speed tracking around 18s (although all methods have speeds below the reference speed), the DDPG-PID controller produced a smaller error compared to the other methods. At around 23s, while all methods exhibited overshoot, the DDPG-PID controller demonstrated the least overshoot. It was also evident that inaccurate speed tracking often occurs during acceleration and deceleration periods, while the tracking performance was found to be excellent during stable operation. Fig.8 reveals that the pattern of position tracking errors was similar to that of speed tracking errors. At around 20s, all methods exhibited varying degrees of position tracking errors. However, the DDPG-PID controller proposed in this paper consistently maintains optimal qualitative performance, with errors consistently below 0.4m.

By combining the statistical data in Table 1, it is evident that the DDPG-PID controller outperforms other control methods in terms of tracking accuracy. Compared to the PID, F-PID, and NN-PID controllers, the DDPG-PID controller significantly reduced both the total error and maximum error in speed tracking. Specifically, when compared to the state-of-the-art NN-PID, the DDPG-PID reduces the total and maximum speed tracking error by 44.5% and 37%, respectively, and the total and maximum distance tracking error by 38.5% and 27.7%, respectively. Furthermore, from Table 1, it can be noted that, except for PID, all other methods almost perfectly achieved precise parking. The jerk indicator, which is calculated as the maximum acceleration rate over the time span of 2s during the entire tracking process, reflected that the DDPG-PID controller can provide superior passenger comfort compared to other methods.

The above experimental results indicate that the DDPG-PID controller outperforms other comparative

TABLE 2. Resistance coefficients of the different scenarios.

Scenario	D_1	D_2	D_3
1	a_{base}	b_{base}	c_{base}
2	$1.5\times a_{\text{base}}$	$1.5\times b_{\text{base}}$	$1.5\times c_{\text{base}}$
3	$0.5\times a_{\text{base}}$	$0.5\times b_{\text{base}}$	$0.5\times c_{\text{base}}$

methods. The potential reasons leading to this phenomenon could be attributed to the fact that the PID controller relies on fixed control parameters that are manually tuned for specific systems, thus lacking adaptability and learning capabilities, thereby leading to subpar tracking performance. The F-PID controller introduces some degree of adaptability by employing fuzzy rules to adjust PID parameters based on predefined conditions. Although it presents a certain degree of adaptability compared to the conventional PID controller, it still relies on rule-based methods, thus potentially failing to fully capture the complex dynamics of subway train systems. The NN-PID controller combines neural networks with the PID algorithm to adjust control parameters. However, it relies on supervised training, thereby using the squared error between reference speed and tracking speed as the loss function. This approach may not effectively balance multiple control objectives, thus leading to suboptimal performance.

In contrast, the DDPG-PID controller boasts an adaptive nature that is achieved through the fusion of reinforcement learning and the PID control mechanism. This innovative approach enables the controller to learn from previous experiences, and it helps it to optimize the control strategy via the signals received from the subway train, thus facilitating precise adjustments to the traction/braking force distribution. Furthermore, this controller operates independently of predefined rules or training datasets, thereby allowing it to adeptly capture the complex dynamics of train operations and deliver a significant enhancement to the adaptability of the underlying PID control mechanism.

C. EXPERIMENTS ON DIFFERENT RESISTANCE PARAMETERS

In this section, we investigate the performance of the DDPG-PID controller under different resistance conditions. Resistance is a critical factor that affects subway train operation and control. By exploring the controller's performance under different resistance scenarios without retraining the model, we can assess its adaptability under varying resistance conditions.

To simulate different resistance conditions, we adjusted the coefficients (D_1, D_2, and D_3) in the Davis equation. The baseline resistance parameters (Scenario 1) were set as $D_1 = a_{\text{base}} = 0.6841$, $D_2 = b_{\text{base}} = 0.0229$, and $D_3 = c_{\text{base}} = 0.000345$. Two additional resistance scenarios, denoted as Scenario 2 and Scenario 3, were created based on the baseline parameters. The adjusted resistance coefficients are shown in Table 2. The speed tracking curves under different resistance scenarios are shown in Fig. 9. The statistical data are presented in Table 3.

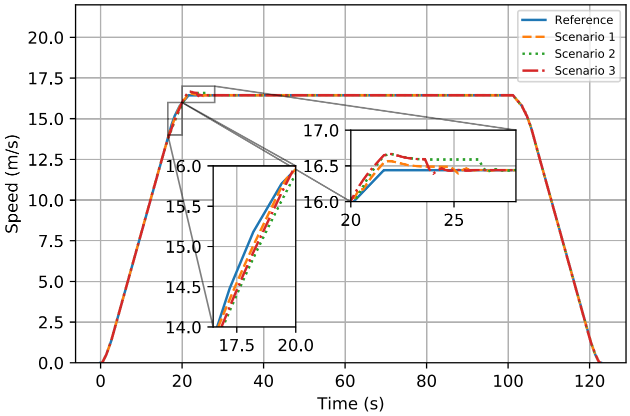


FIGURE 9. Speed tracking curves under different resistance scenarios.

TABLE 3. Statistics of the different resistance scenarios.

	Speed track (m/s)		Position track (m)			Other indicator	
	Total error	Max. error	Total error	Max. error	Parking error	Energy (kWh)	Jerk (m/s^3)
Scenario 1	3.51	0.17	10.32	0.34	0.00	17.04	0.39
Scenario 2	9.48	0.32	25.80	0.94	0.00	19.07	0.40
Scenario 3	6.57	0.25	13.41	0.65	0.01	14.98	0.39

In Scenario 1, representing the baseline resistance parameters, the DDPG-PID controller achieves excellent speed and position tracking accuracy, as indicated by the low total and maximum tracking errors. This is because the model was trained using Scenario 1's parameters, and was then tested in Scenarios 1, 2, and 3.

In Scenarios 2 and 3, where resistance parameters are respectively increased by 50% and decreased by 50%, the DDPG-PID controller exhibited slightly higher speed and position tracking errors compared to Scenario 1. This is because the model was not trained under these resistance parameters. Despite the higher tracking errors, the controller achieved precise parking in almost all cases. From the above table, it can be observed that Scenario 2 saw the highest energy consumption, which can be attributed to the increased resistance. Thus, it required the controller to exert more effort in maintaining the desired speed and position. Regarding the jerk indicator, there was little difference among the scenarios, thus indicating that all three scenarios ensured passenger comfort.

The above experimental results demonstrated that the DDPG-PID controller dynamically adjusts the control actions in the face of different resistance conditions, and this is performed without the need for retraining the model, thus significantly reducing tuning costs. This reflects the DDPG-PID controller's good adaptability in subway train operation control, thereby further establishing its potential for practical applications in subway train control.

D. EXPERIMENTS ON DIFFERENT REFERENCE SPEEDS

In this section, we investigate the performance of the DDPG-PID controller under different reference speed scenarios. By examining the control performance under different reference speeds, we can assess the controller's ability to handle varying operational requirements and can accurately track the desired speed profile.

TABLE 4. Statistics of the different reference speed scenarios.

	Speed track (m/s)		Position track (m)			Other indicator	
	Average error	Max. error	Average error	Max. error	Parking error	Energy (kWh)	Jerk (m/s^3)
Scenario 1	0.01	0.17	0.02	0.34	0.00	17.04	0.39
Scenario 4	0.05	0.60	0.34	2.92	0.07	19.36	0.50
Scenario 5	0.01	0.05	0.00	0.01	0.02	15.22	0.44

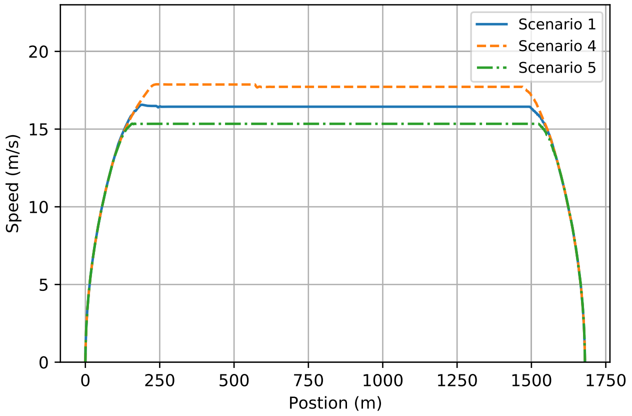


FIGURE 10. Speed tracking curves under different reference speed scenarios.

To evaluate the controller's performance, we constructed two additional reference speed profiles, reference-fast (Scenario 4) and reference-slow (Scenario 5). These profiles were based on the original reference speed profile (Scenario 1) used in previous experiments. These new speed profiles involved increasing or decreasing the running time, respectively. Fig. 10 displays the tracking curve of different scenarios. The statistics are summarized in Table 4. As shown in Fig. 10, we plotted the speed position curves due to the different total times for the different reference speed profiles. Similarly, as shown in Table 4, we replaced the total error with the average error.

In Scenario 4, where the reference speed increases compared to Scenario 1, the DDPG-PID controller exhibits slightly higher tracking errors. The increased speed introduced a more challenging control task, thus resulting in larger deviations in speed and position tracking. Correspondingly, the energy consumption increased. Compared to other scenarios, although Scenario 4 exhibited a relatively high error, it remained within an acceptable range. In Scenario 5, which had a decreased reference speed, the DDPG-PID controller demonstrated excellent accuracy in speed and position tracking. The reduced speed allowed for the controller to make precise adjustments, thus leading to significantly reduced tracking errors compared to other scenarios.

The experimental results demonstrate that the DDPG-PID controller can generate control parameters without retraining the model under different reference speed profiles. This allows for accurate speed tracking and reduces the resource waste that is caused by retraining the model or by adjusting the parameters when changing the reference speed profile due to the re-formulation of travel plans. These results highlight the adaptability of the DDPG-PID controller in accurately tracking the desired speed profile, as well as its superior performance, minimal deviation, and efficient control, which all contribute to the optimization of subway train operation.

V. CONCLUSION

In this paper, an adaptive control approach for optimizing subway train operation is proposed. The proposed approach utilizes a hierarchical structure consisting of an upper DDPG controller and a lower PID controller, which work together to improve the efficiency and performance of train operations. By conducting comparative experiments with other control methods, the superiority of the proposed approach is demonstrated. Moreover, the adaptability of the proposed approach is established through the manipulation of varying resistance parameters and desired speed profiles.

However, it is worth noting that our proposed approach relies on a reference speed profile, which, to some extent, limits the flexibility of train operation. In future work, we intend to focus on developing techniques for achieving the online control of trains without the need for predefined speed profiles. Furthermore, another area of future research is the coordination of multiple trains during operation. While our approach has demonstrated promising results in optimizing individual train operation, exploring other methods to coordinate the movements and interactions of multiple trains will be a crucial area of interest.

REFERENCES

[1] L. Dong, L. Qin, X. Xie, L. Zhang, and X. Qin, "Collaborative optimization method for multi-train energy-saving control with urban rail transit based on DRLDA algorithm," *Appl. Sci.*, vol. 13, no. 4, p. 2454, Feb. 2023.

[2] Y. Cao, Z.-C. Wang, F. Liu, P. Li, and G. Xie, "Bio-inspired speed curve optimization and sliding mode tracking control for subway trains," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6331–6342, Jul. 2019.

[3] P. Fu, S. Gao, H. Dong, B. Ning, and Q. Zhang, "Speed tracking error and rate driven event-triggered PID control design method for automatic train operation system," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2018, pp. 2889–2894.

[4] Y. Zhu and Z. Hou, "Data-driven MFAC for a class of discrete-time nonlinear systems with RBFNN," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 1013–1020, May 2014.

[5] Y. Sun, H. Qiang, J. Xu, and G. Lin, "Internet of Things-based online condition monitor and improved adaptive fuzzy control for a medium-low-speed Maglev train system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2629–2639, Apr. 2020.

[6] U. Akram, M. Khalid, and S. Shafiq, "An advanced control strategy for magnetic levitation train system based on an online adaptive PID controller," in *Proc. 9th IEEE-GCC Conf. Exhib. (GCCCE)*, May 2017, pp. 1–9.

[7] S. Yasunobu, S. Miyamoto, and H. Ihara, "Fuzzy control for automatic train operation system," *IFAC Proc. Volumes*, vol. 16, no. 4, pp. 33–39, Apr. 1983.

[8] Q. Pu, X. Zhu, J. Liu, D. Cai, G. Fu, D. Wei, J. Sun, and R. Zhang, "Integrated optimal design of speed profile and fuzzy PID controller for train with multifactor consideration," *IEEE Access*, vol. 8, pp. 152146–152160, 2020.

[9] X. Zhu, Q. Pu, Q. Zhang, and R. Zhang, "Automatic train operation speed profile optimization and tracking with multi-objective in urban railway," *Periodica Polytechnica Transp. Eng.*, vol. 48, no. 1, pp. 57–64, Jun. 2019.

[10] Y. Feng, M. Wu, X. Chen, L. Chen, and S. Du, "A fuzzy PID controller with nonlinear compensation term for mold level of continuous casting process," *Inf. Sci.*, vol. 539, pp. 487–503, Oct. 2020.

[11] Y. Sun, J. Xu, H. Qiang, and G. Lin, "Adaptive neural-fuzzy robust position control scheme for Maglev train systems with experimental verification," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8589–8599, Nov. 2019.

[12] Q. Pu, X. Zhu, R. Zhang, J. Liu, D. Cai, and G. Fu, "Speed profile tracking by an adaptive controller for subway train based on neural network and PID algorithm," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 10656–10667, Oct. 2020.

[13] W. Carvajal-Carreño, A. P. Cucala, and A. Fernández-Cardador, "Fuzzy train tracking algorithm for the energy efficient operation of CBTC equipped metro lines," *Eng. Appl. Artif. Intell.*, vol. 53, pp. 19–31, Aug. 2016.

[14] X. Li and H. K. Lo, "An energy-efficient scheduling and speed control approach for metro rail operations," *Transp. Res. B, Methodol.*, vol. 64, pp. 73–89, Jun. 2014.

[15] J. Arents and M. Greitans, "Smart industrial robot control trends, challenges and opportunities within manufacturing," *Appl. Sci.*, vol. 12, no. 2, p. 937, Jan. 2022.

[16] Q. Sun, C. Du, Y. Duan, H. Ren, and H. Li, "Design and application of adaptive PID controller based on asynchronous advantage actor-critic learning method," *Wireless Netw.*, vol. 27, no. 5, pp. 3537–3547, Jul. 2021.

[17] J. E. Sierra-Garcia, M. Santos, and R. Pandit, "Wind turbine pitch reinforcement learning control improved by PID regulator and learning observer," *Eng. Appl. Artif. Intell.*, vol. 111, May 2022, Art. no. 104769.

[18] J. E. Sierra-García and M. Santos, "Exploring reward strategies for wind turbine pitch control by reinforcement learning," *Appl. Sci.*, vol. 10, no. 21, p. 7462, Oct. 2020.

[19] X. Yu, Y. Fan, S. Xu, and L. Ou, "A self-adaptive SAC-PID control approach based on reinforcement learning for mobile robots," *Int. J. Robust Nonlinear Control*, vol. 32, no. 18, pp. 9625–9643, Dec. 2022.

[20] X. Yu, S. Xu, Y. Fan, and L. Ou, "A self-adaptive LSAC-PID approach based on Lyapunov reward shaping for mobile robots," 2021, *arXiv:2111.02283*.

[21] W. Ma, B. Li, Y. Cao, P. Wang, M. Liu, C. Chang, and S. Peng, "Velocity control of a multi-motion mode spherical probe robot based on reinforcement learning," *Appl. Sci.*, vol. 13, no. 14, p. 8218, Jul. 2023.

[22] Z. Zhang, X. Li, J. An, W. Man, and G. Zhang, "Model-free attitude control of spacecraft based on PID-guide TD3 algorithm," *Int. J. Aerosp. Eng.*, vol. 2020, pp. 1–13, Dec. 2020.

[23] A. Kim, "Deep reinforcement learning of position and velocity PID control for rotational wing unmanned aerial vehicles," in *Proc. 5th Int. Congr. Human-Computer Interact., Optim. Robotic Appl. (HORA)*, Jun. 2023, pp. 1–5.

[24] S. Su, X. Li, T. Tang, and Z. Gao, "A subway train timetable optimization approach based on energy-efficient operation strategy," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 883–893, Jun. 2013.

[25] J. Yin, T. Tang, L. Yang, J. Xun, Y. Huang, and Z. Gao, "Research and development of automatic train operation for railway transportation systems: A survey," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 548–572, Dec. 2017.

[26] Y. Cheng, J. Yin, and L. Yang, "Robust energy-efficient train speed profile optimization in a scenario-based position—Time—Speed network," *Frontiers Eng. Manage.*, vol. 8, no. 4, pp. 595–614, Dec. 2021.

[27] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

[28] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[29] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2146–2153.

GAOYUN CHENG received the M.S. degree from the College of Computer Science and Technology, Northeastern University, Shenyang, China, in 2021. He is currently with Traffic Control Technology Company Ltd., Beijing, China. His research interests include advanced train control methods, the optimization problem in rail transport, traffic flow theory, and reinforcement learning.

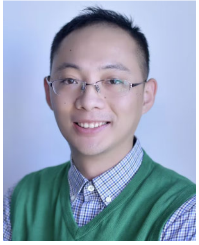
DIANYUAN WANG received the M.S. degree from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2021. He is currently with Traffic Control Technology Company Ltd., Beijing. His research interest includes the innovative application of artificial intelligence methods in the traditional rail transit industry, including passenger flow scheduling strategy, train scheduling strategy, train control strategy, and other fields.

MING SUN received the M.S. degree from the School of Mathematics, Harbin Institute of Technology, in 2021. She is currently with Traffic Control Technology Company Ltd., Beijing, China. In a professional capacity, she focuses on research areas, such as passenger flow scheduling, train control methods, and train braking systems. During academic pursuits, her research interests include machine learning, image processing, and federated learning.

ZHE FU received the M.S. degree from Beijing Jiaotong University, Beijing, China, in 2016. He is currently with Traffic Control Technology Company Ltd., Beijing. He is also with the Deputy Director of the Data Twin Research Laboratory. His research interests include train digital twins, switch state detection, and train state perception.

BINBIN YUAN received the M.S. degree from Beijing Jiaotong University, Beijing, China, in 2006. He is currently with Traffic Control Technology Company Ltd., Beijing. He is also with the Director of the Data Twin Research Laboratory. His research interests include train dynamics simulation, virtual-reality interaction, and train operation perception.

LEI ZHANG received the M.S. degree in computer science and technology from Beijing Jiaotong University, Beijing, China, in 2009, where she is currently pursuing the Ph.D. degree. She is with Traffic Control Technology Company Ltd., Beijing. She is also the Project Manager of the Safety Operation for the Virtual Coupling Project. Her research interests include train model parameter identification and analysis of train operational safety.



XIAO XIAO was born in 1987. Since July 2010, he has been with Traffic Control Technology Company Ltd., Beijing, China. In May 2020, he was the Director of the Research Institute. His research interests include train operation control, passenger flow prediction, and equipment predictive maintenance.