IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 July 2023, accepted 13 August 2023, date of publication 15 August 2023, date of current version 23 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3305683

APPLIED RESEARCH

Energy Consumption Optimization for Heating, Ventilation and Air Conditioning Systems Based on Deep Reinforcement Learning

YI PENG[1], HAOJUN SHEN[1], XIAOCHANG TANG[2], SIZHE ZHANG[2], JINXIAO ZHAO[2], YURU LIU[2], AND YUMING NIE[2]

[1]College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China
[2]College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China

Corresponding author: Yi Peng (2016020219@s.upc.edu.cn)

ABSTRACT Heating, ventilation, and air conditioning (HVAC) energy consumption now accounts for a major portion of energy use for buildings. Therefore, finding the optimal energy-saving control strategy for HVAC systems to optimize energy consumption has become crucial in realizing energy savings, emission reductions, and green buildings. Traditional methods for HVAC parameter control require complex physical model calculation; continuous and coupled parameters are handled poorly. Developing deep reinforcement learning (DRL) methods provides new ideas for HVAC energy consumption optimization. Herein, a DRL-based energy consumption optimization framework for HVAC systems is proposed. First, an HVAC system energy consumption prediction model based on a convolutional neural network–long short-term memory (CNN-LSTM) network is suggested to approximate the real world. This model solves the efficiency problem of energy consumption prediction while also providing highly accurate predictions of HVAC energy consumption. We propose an enhanced deep deterministic policy gradient (E-DDPG) energy consumption optimization algorithm for HVAC systems based on an improved training strategy to obtain the best real-time energy consumption control strategy for HVAC systems. Finally, experiments using real-world building HVAC control data sets were conducted to evaluate our models. The experiments show that the CNN–LSTM model for HVAC system energy consumption prediction outperforms baseline models while reducing training time by 42.9%. Compared to the baseline algorithm, the E-DDPG algorithm using an improved training strategy requires 20% fewer iterations for convergence, has a 14.8% narrower fluctuation interval during the training process, and improves the energy efficiency ratio of HVAC systems by 49%.

INDEX TERMS HVAC systems, DDPG, deep reinforcement learning, energy consumption.

I. INTRODUCTION

At present, buildings are responsible for a major portion of global energy consumption. Indeed, heating, ventilation, and air conditioning (HVAC) systems account for about 50%–60% of total building energy consumption [1]. Therefore, energy-saving HVAC systems are key to reducing the energy consumption of buildings and realizing energy savings and emission reduction.

With the development of Internet of Things (IoT) technologies, most HVAC systems have been able to automatically collect system operation data and monitor system operation in real-time [2]. Among today's hot research topics is the use of real-time monitoring data to establish data-driven models for optimizing and controlling HVAC system energy consumption [3]. As artificial intelligence (AI) technologies continue to develop rapidly, new ideas and methods for managing and controlling HVAC systems continue to emerge, with two noteworthy trends. First, big data technologies are now being used to analyze the monitoring and operational data of HVAC systems to explore the essential characteristics

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

of HVAC system operation [4]. Second, deep reinforcement learning (DRL) technologies are now being used to provide improved energy consumption optimization strategies for the intelligent control of HVAC systems [5].

Technologies such as smart appliances and IoT have reduced the cost of data collection and the large amount of industrial data requires us to conduct data optimization [6] and data compression [7] to better meet the requirement of data mining. The processed industrial data often allows for more diverse analysis, such as state assessment [8] or damage detection [9],thus expanded the application areas of data mining [10]. Relying on a strong data basis, powerful data-driven models show promise in the field of energy-saving, optimized operation of HVAC systems. With the introduction of AlphaGo and AlphaZero, DRL has gradually become known to the public and has now become a hot research topic in the field of AI [11]. Reinforcement learning (RL), as a machine learning method emerging in recent years, features model-free learning, self-learning, and online learning. It is a data-driven control method that can achieve model-free adaptive optimization of controllers. RL-based controllers have been shown to be able to provide optimal control strategies in real time in the face of changes in the system environment. Since RL is model-free, the potentially complex system modeling process can be avoided under certain conditions [12]. Therefore, RL-based optimization of the control strategy for HVAC systems can not only improve the operational efficiency of HVAC systems but also make full use of the existing building operation data, exploit the value of the data, and achieve the ultimate purpose of energy consumption optimization [13]. Among the existing HVAC system energy consumption optimization models, the DDPG algorithm performs best [14]. In this paper, an optimization framework for HVAC system energy consumption based on the enhanced deep deterministic policy gradient (E-DDPG) algorithm is proposed. A convolutional neural network–long short-term memory (CNN-LSTM) neural network, combining a CNN and an LSTM model, is used to simulate and predict the energy consumption environment. The overall framework succeeds at providing stable, effective energy consumption optimization and control strategies at a rapid convergence rate.

The contributions of this paper are as follows.

1) **A CNN–LSTM–based energy consumption prediction algorithm** is proposed to predict the energy consumption of HVAC systems, provide a simulation environment for energy consumption optimization, and facilitate energy consumption optimization in the reducing energy consumption.

2) **The training strategy of the DDPG algorithm is improved**, including improvements to the DDPG algorithm's sampling method and its network update schedule. The adopted up and down sampling method solves, at each time point, the problem of an unbalanced distribution of agent learning experience. The Critic network is updated first, and then the Actor network is updated when the loss value of the Critic network is less than a preset threshold. By dynamiting the delayed update in such a way, the DDPG algorithm dynamically updates the network to solve the problem of wrong direction updating of the network. The training time is reduced with attendant improvements to training process stability and model robustness.

3) **An energy consumption optimization algorithm based on E-DDPG** is proposed for finding the optimal parameter values for the HVAC system control unit. The HVAC system control strategy is optimized and adjusted in a real-time manner in combination with the energy consumption prediction algorithm for environmental simulation, achieving energy savings and efficient HVAC system use.

The remainder of this paper is organized as follows. Section II describes previous work related to the prediction and optimization of HVAC system energy consumption. Section III presents the detailed architecture and the process of the energy consumption optimization framework for HVAC systems based on deep reinforcement learning. Section IV describes a series of comparative experiments on the proposed framework. Section V presents the conclusions and offers considerations regarding future work.

II. RELATED WORK

The traditional approach to improving HVAC system control consists of three steps. An equipment model is established first; then optimal parameters are found using an optimization algorithm, and finally, an optimized energy-saving control strategy is developed [15]. Li et al. [16] used a support vector machine to achieve short-term energy consumption prediction for HVAC systems, while Lu et al. [17] adopted a gray prediction method and multiple regression analysis to predic energy consumption in urban residential buildings. The difficulties encountered in applying these methods are due to three factors: i) the system operational parameters are mutually coupled, ii) the modeling process is complicated, and iii) the system parameters change slowly over time, so that models gradually fail to effectively achieve energy-saving optimization. Moreover, although some models can optimize the operating parameters online, they are not capable of self-learning and thus require optimization of control parameters at each step. As a result, the algorithms are computationally intensive, making it difficult to meet the demand for real-time control. In addition, due to the stochastic nature of the optimization algorithm, the optimization process can encounter other problems, such as slowness, hard convergence, and local optimum traps.

HVAC system operation parameter data are time-series data and vary significantly with time. Therefore, HVAC system operation prediction is a forecasting problem based on time series. Popular time-series prediction methods, including those based on time-series and neural network models, have been examined extensively [18]. Alireza et al. [19], [20] have utilized Multilayer Perceptron (MLP) to predict future stock movements. While in the field of HVAC system prediction. Kumar et al. [21] and Marino et al. [22] used LSTM models to predict the power consumption of residential

buildings, and compared with other models, it was verified that LSTM outperforms other models in solving nonlinear problems and memorizing historical data. Kim and Cho [23] combined CNN and LSTM networks to predict residential power consumption. However, power consumption prediction for HVAC systems alone has rarely been studied. The main reason for this is that the HVAC system is part of the electrical equipment, so its energy consumption is typically not separately measured or calculated. A typical HVAC system is relatively complex and consists of a refrigeration unit, recirculating cooling water circulation system, and chilled water system [24]. Therefore, it is typically quite difficult to predict the energy consumption of an HVAC system. Zhou et al. used the LSTM model to predict HVAC system power consumption [25], and Kim et al. used LSTM cells in recurrent neural network algorithms for dynamic simulation modeling of HVAC system power consumption [26]. In this paper, we propose to improve the accuracy of prediction by combining two neural networks, CNN and LSTM. Specifically, we use the CNN to extract data features to improve the stability of the model beyond the training set, while we use the LSTM network to extract historical features to improve model stability. The combined CNN–LSTM neural network is used to predict the ratio of cooling capacity to power, also known as the energy efficiency ratio (EER) for HVAC systems.

RL algorithms perform well in the field of optimization and can learn by trial and error. They can obtain knowledge from the environment by receiving the highest reward to improve the control scheme. They then give the optimal control policy in real-time in the face of environmental changes, thus meeting the demand for real-time control [27]. For example, Yao et al. proposed a model-based reinforcement learning control for electro-hydraulic position servo systems [28].Building on the intensive study of RL in theory and practice, we make use of DRL, which has also been previously applied in this area [29]. The powerful fitting ability of neural networks in DRL can somewhat reduce the complexity of modeling.

Dounis et al. [30] used a fuzzy proportional differential method to optimize and control equipment in buildings to achieve energy savings. Azuatalam [31] developed a temperature set-point controller based on the proximal policy optimization (PPO) algorithm to control the temperature set-point of a room. Xing et al. [32] used an improved PSO based algorithm to model energy consumption optimization. Congradac and Kulic used genetic algorithms to optimize HVAC control systems to achieve energy savings [33]. This control optimization method requires experts to design and develop a model and then needs numerous analytical adjustments in order to find the appropriate parameter settings. Dalamagkidis [34] developed a linear controller based on a classical RL algorithm capable of monitoring energy consumption and policy decisions based on a temporal difference method. Although the controller performs well in energy consumption monitoring and control stability, its algorithm relies excessively on exploratory actions to determine the optimal policy, allowing incorrect control to occur in actual operation. Liu et al. [35] applied RL to optimize the operation of energy storage units for HVAC systems and pointed out that the classical Q-learning algorithm used might be inefficient in high-dimensional learning. In contrast, Hai et al. [36] significantly reduced the dimensionality of the action space and state space in reinforcement learning by integrating index selection with deep reinforcement learning using heuristic rules. Wei et al. [37] proposed a DRL-based control method for HVAC systems and verified the scalability of a DRL controller while noting that the required training time for a DRL controller might be long. TABLE 1 shows a summary of related algorithms.

Therefore, for a typically highly nonlinear, coupled, time-varying, uncertain, and complex multivariable system such as an HVAC system, a DRL [38] algorithm that can self-learn complex nonlinear relationships and adapt to environmental conditions to provide strategies in real-time is a suitable choice for energy consumption optimization and control strategy design. For example, Gao et al. [39] used the DDPG algorithm to control the set-point air temperature and humidity of HVAC system units. However, this method is not flexible enough to be adapted to the specific circumstances of the dwelling. Xia et al. [5] proposed a residential HVAC adaptive scheduling strategy based on the DRL method. The method assumes constant residential thermal parameters, but in reality residential thermal parameters vary with time and environment. Yu et al. [40] presented a multiagent DRL method involving an attention mechanism to minimize the energy cost in multizone buildings. However, the algorithm does not consider the dynamic characteristics of the HVAC system, which affects part of its robustness. At the scope of whole-building energy, Zou et al. [41] applied DDPG in a data-based LSTM environmental model. Ioannis et al, this method does not preprocess the data and the presence of noisy data affects the model performance. Reference [42] proposed a clustering-based DDPG training plan that optimized energy consumption more effectively than individual DDPG training. Ding et al. [43] proposed a branching dueling double Q-network to solve the high-dimensional action problem of four building subsystems, namely HVAC, lighting, blinds, and windows. However, this method network is updated in the wrong direction, leading to the risk of under-optimization of energy consumption.Yu et al. [44] used DDPG to minimize the energy cost of smart homes equipped with HVAC and energy storage systems. However, it takes substantial time for these methods to converge and obtain a stable control strategy.

In order to solve the above problems, the training strategy is optimized to find a better and more stable control strategy faster, this paper proposes an enhanced deep deterministic policy gradient (E-DDPG) algorithm that can fully learn the relationship between environment and state by outputting deterministic actions through a neural network

TABLE 1. Summary of algorithms.

Name	Stability	Training speed	Environmental adaptability
Heuristic algorithms	Very stable	Modeling required but no training	Poor
Model-based prediction and control	Stable	No training required	Positively correlated with system model accuracy
Linear controller	Stable	Slow	Fairly good
Q-learning	Relatively unstable	Slow	Ordinary
Traditional DRL	Relatively unstable	Slow	Fairly good
E-DDPG	Stable	Quick	Good

and incorporating mechanisms such as action exploration and experience playback. Combined with CNN–LSTM for environment modeling, our algorithm shows fast convergence and generates optimal control policies with continuous states and actions in a real-time manner to optimize the energy consumption of HVAC systems.

III. METHODS

In this section, we first introduce our framework for energy consumption optimization of HVAC systems and then describe the workflow of the framework in two modules: energy consumption prediction and energy consumption optimization.

A. DRL-BASED ENERGY CONSUMPTION OPTIMIZATION FRAMEWORK FOR HVAC SYSTEMS

To reduce the energy consumption of an HVAC system, it is necessary to find an appropriate optimization strategy. The traditional thermodynamic modeling approach is very difficult in complex systems such as HVAC, which requires the calculation of a large number of parameters and is hardly robust. The DDPG algorithm in the context of DRL has shown good performance in the optimized control of HVAC systems [9]. The DDPG algorithm for optimized control of an HVAC system can realize the learning process by exploring various state–control pairs, with no need to compute complex thermodynamic models to represent the HVAC system in the physical world. The DDPG algorithm's delaying of network updating can effectively deal with systems such as HVAC that feature time delays. The DDPG algorithm also works for continuous and coupled parameters in the sensors and controllers of HVAC systems. Therefore, we aim here to apply the DDPG algorithm to the optimization of HVAC system control.

While there are many benefits of using the DDPG algorithm to optimize the energy consumption of HVAC systems, the training process often reveals defects such as uneven distribution of learning experiences and a tendency to update the neural network in the wrong direction. For this reason, we improve the training strategy of the original DDPG algorithm in our proposed E-DDPG algorithm.

- Experience is collected from the experience replay pool using up and down sampling. The samples of the experience pool are divided into intervals based on the reward value. The oversampling technique is used for the data in the minority intervals, while the under sampling technique is used for the data in the majority intervals. In such a way, the problem of an uneven distribution of learning experiences can be solved at each time point to allow the E-DDPG algorithm to learn fully from the experiences.
- A dynamic delayed updating mode is adopted. The loss value of the Critic network is used as an indicator to compare the training results of different activation functions with the same loss function. The Critic network is updated first; the Actor network is updated only when the average variance of the Critic network's loss value falls below some threshold. Then the Actor network is dynamically updated according to the difference between the two loss values given by the Critic network. This technique provides for a training process of enhanced stability.

The training of the E-DDPG algorithm for the energy consumption optimization of an HVAC system should take place in a real-world HVAC system control space to ensure that it can receive real-time feedback on each parameter of the HVAC system's sensors and controllers. However, it is not possible for the E-DDPG algorithm to explore all the state–control pairs to obtain the reward values. On the other hand, the alternative control strategy, namely that obtained by using simulation tools to simulate the training for the environment, shows poor effectiveness in practical applications because the simulation tool contains only limited information. This limit makes it impossible for simulation tools to realistically restore all the details of the HVAC system in operation. Therefore, we propose to use a CNN–LSTM neural network to predict and simulate the operating environment of the HVAC system.

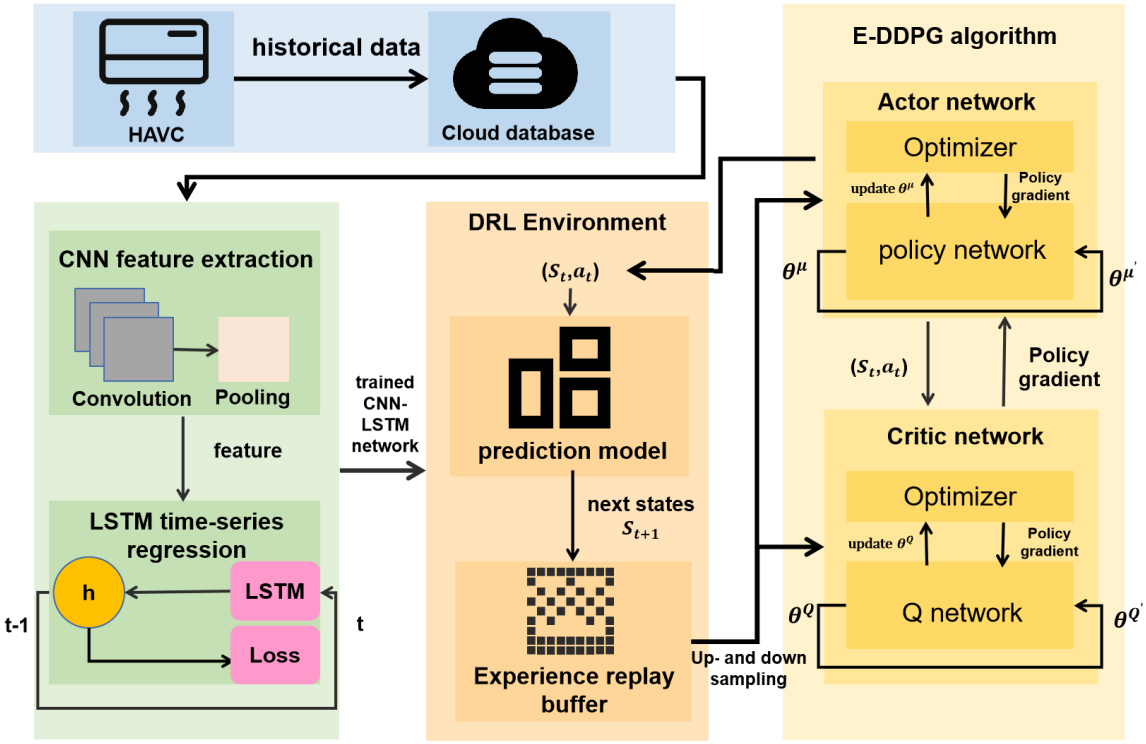

FIGURE 1. Flow chart for the proposed framework.

We first use a CNN to extract data features and then feed the extracted features into an LSTM network for learning the data time series. In contrast with a simple CNN or LSTM model, the CNN–LSTM network can both extract features and learn time-series trends. It can efficiently learn the time-series features of the HVAC system parameter data, thereby accelerating the training process. After training, the E-DDPG algorithm needs only to input the control parameter values of the HVAC system into the prediction model to achieve an accurate prediction of energy consumption at the next moment. This capability can assist in modeling using the E-DDPG algorithm and provide a basis for decisions regarding the optimization of HVAC system energy consumption.

Combining the above processes, we construct a complete framework for the energy optimization of HVAC systems based on DRL to achieve energy-efficient operation. FIGURE 1 shows a flow chart of the framework implementation.

As shown in the flow of Fig.1, the workflow of the framework is as follows.

1) Obtain the historical data uploaded by HVAC equipment to the cloud database and input it into the CNN network for feature extraction.
2) Input the feature extraction results into the LSTM network to learn the temporal order of the data, and obtain the trained CNN–LSTM energy consumption prediction model.
3) Use E–DDPG algorithm for the construction of energy consumption optimization model, Actor network obtains the energy consumption in various states by using the trained energy consumption prediction model.
4) Put it into the experience replay, the Critic network collects experience from the experience replay by up and down sampling to obtain its Loss value, when the Loss value is less than a certain threshold, the Actor network starts to update dynamically.

Finally, through this process, the HVAC energy consumption optimization model with robustness and good generalization ability is trained.

B. CNN–LSTM NEURAL NETWORK FOR HVAC SYSTEM ENERGY CONSUMPTION PREDICTION

The CNN–LSTM network can learn the long-term time dependence of the energy consumption of the HVAC system. However, the simulation of HVAC system energy consumption trends requires very long input sequences, which in turn require a lengthy training process. To reduce the training effort while ensuring prediction accuracy, we instead use the low-dimensional data obtained after feature engineering to construct our model.

FIGURE 2 shows the CNN–LSTM network structure for energy consumption prediction. At the top is a set of historical observations of the energy consumption of the HVAC system. The EER is used to measure the energy consumption of the

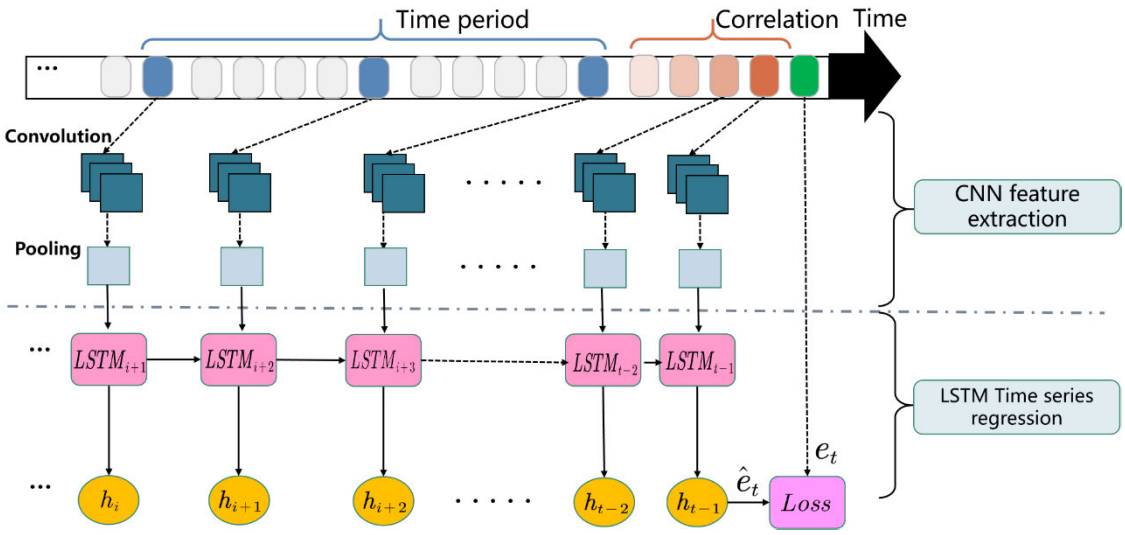

FIGURE 2. Schematic diagram of the convolutional neural network – long short-term memory (CNN–LSTM) algorithm.

HVAC system; the larger the EER, the lower the energy consumption of the HVAC system. Green dots indicate the EER values predicted for time points t. The blue hues of the dots reflect their time proximity; as the time distance widens, the temporal correlation becomes weaker, and the blue dots become lighter in color. Each orange dot indicates a period. Time series are first input to the convolutional and pooling layers for feature extraction, and then the CNN output is input to the LSTM network. The bottom half of FIGURE 2 shows the output values corresponding to each LSTM cell state. The input-to-output process can be expressed by the following four equations:

$$i_t = \sigma\left(W_{gi} * G_t^s + W_{hi} * H_{t-1}^s + W_{ci} C_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma\left(W_{gf} * G_t^s + W_{hf} * H_{t-1}^s + W_{cf} \circ C_{t-1} + b_f\right) \quad (2)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh\left(W_{gc} * G_t^s + W_{hc} * H_{t-1}^s + b_c\right) \quad (3)$$

$$o_t = \sigma\left(W_{go} * G_t^s + W_{ho} * H_{t-1}^s + W_{co} \circ C_{t-1} + b_o\right) \quad (4)$$

Here, i is the input gate, j is the forget gate, C is the update cell, o is the output gate, σ is the activation function, and G is the input to the LSTM layer at time t. The final LSTM output is as follows:

$$H_t^s = o_t \circ \tanh\left(C_t\right) \quad (5)$$

When the input for time $t-1$ is completed, the LSTM network gives the predicted value for time t. The predicted value is compared with the true value to obtain the loss value to optimize the Critic network.

Compared with the LSTM model, the CNN–LSTM network both extracts features and learns time-series trends. It efficiently learns the time-series features of the HVAC system parameter data, accelerating the training process. In addition, it accurately predicts the energy consumption at the next moment to assist in the enhanced learning algorithm's modeling and provide a basis for the subsequent decisions on energy consumption optimization.

C. IMPROVING THE ORIGINAL DDPG ALGORITHM TO ACHIEVE ENERGY CONSUMPTION OPTIMIZATION

The basic framework of the RL algorithm is a Markov decision process (MDP). To design the RL algorithm, we must first mathematically model the MDP, including the essential elements of RL, such as states, actions, and rewards. After the essential framework is modeled, the training process for the DDPG algorithm is constructed and improved by replacing random sampling with up and down sampling methods. In addition, dynamic delayed network updating is adopted, and the E-DDPG algorithm is obtained. The following is a detailed description of the E-DDPG model framework.

- **States (s)**: The state space determines the content of the agent's environmental perception. In the E-DDPG algorithm, the parameters selected as the state space of the MDP are those with a high impact on the energy consumption of the HVAC system in the original data, i.e., feature importance. Before constructing each solution, the agent first obtains the current state, i.e., the current parameters related to the energy consumption of the HVAC system. Based on these parameters, the agent then makes energy consumption predictions and selects the optimized control strategy for energy consumption.
- **Actions (a)**: The decision action of the agent is the output of the algorithm, which changes the state of the environment with probability P. In most cases, the action is the result of the decision. To ensure that the HVAC system can meet the current working requirements and reduce energy consumption after optimization, only two relevant parameters are adjusted in this paper. Therefore, the action here is the adjustment of these two HVAC system parameters. The action output by the algorithm according to the current state at each time point is the result of the energy consumption optimization decision.

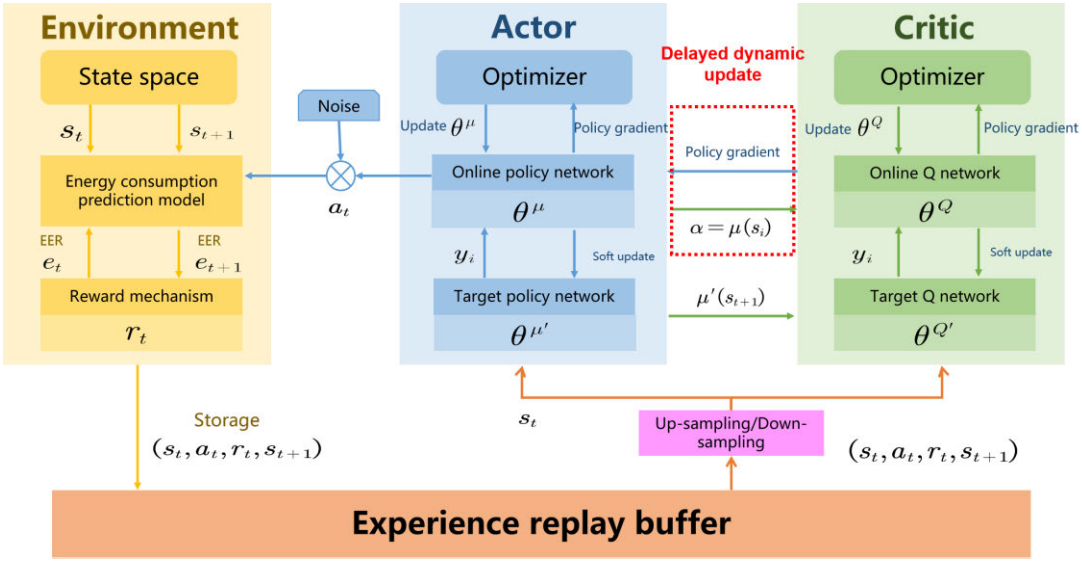

FIGURE 3. Flow chart for the enhanced deep deterministic policy gradient (E-DDPG) algorithm.

- **Reward mechanism**: Since the goal of this paper is to minimize the total energy consumption of the HVAC system, i.e., to maximize the EER, the reward feedback given by the RL model environment should guide the neural network to update in the direction of energy consumption reduction. Therefore, the reward setting in this model is to optimize the difference between the EER before and after the control parameters of the HVAC system. The greater the absolute value of the difference, the better the optimization performance, which leads to the greater the reward value.
- **Environment**: The agent's interactions with the environment provide feedback, i.e., a reward or penalty. In the energy consumption optimization task, the environmental feedback required by the agent after each state update is the energy consumption corresponding to the current state of the HVAC system and the predicted energy consumption of the HVAC system at the next moment. Therefore, the above energy consumption prediction model can be used to simulate the environment. That is, with corresponding control parameters of the HVAC system as input, the energy consumption prediction model gives the energy consumption corresponding to the next time point as feedback from the environment. The reward value is calculated based on environmental feedback.

The overall process of the E-DDPG algorithm is shown in FIGURE 3. The initial HVAC system control parameter value (s_0) is manually set or is read from the sensor. The agent takes an action (a) based on the current HVAC system parameter value (s), and then the state is transferred to obtain the next HVAC system control parameter value (s'), which is expected to correspond to the parameter value after energy consumption reduction. Subsequently, the energy consumption prediction model calculates the energy consumption of the HVAC system after 30 s according to the current control parameter value and the next control parameter value, respectively, as environmental feedback. The reward value is calculated based on the environmental feedback, and the generated state (s), action (a), reward value (r), and next state (s') in this calculation are stored in the experience playback pool until the set maximum number of steps is reached, thus completing an episode.

The above process is repeated until the maximum data storage capacity of the experience replay pool is reached. After that, a step is added at each iteration to update the network weights. The Actor network is responsible for learning the optimal policy for successive actions, while the Critic network maintains the fitting with the real actions.

The online Actor network observes the HVAC system control parameter state (s) and takes actions based on the policy. Then, the energy consumption prediction model simulates the environmental feedback and returns the next state (s') and the reward obtained from this action. The target Actor network also selects the virtual optimal action based on the state and sends it to the target Critic network to calculate the target Q-value. When the target Q-value is entered, the online Critic network calculates the temporal difference error (TD-error) and completes the gradient update of its network parameters. The parameter update equation is as follows:

$$\theta^{\mathrm{Q}} \leftarrow \theta^{\mathrm{Q}} + \alpha_{\mathrm{Q}} \delta \cdot \nabla_{\theta^{\mathrm{Q}}} \mathrm{Q}\left(s, a \mid \theta^{\mathrm{Q}}\right) \tag{6}$$

With feedback from the online Critic network, the online Actor network also completes the update of the policy, i.e., the gradient update of its parameters. Its update equation is as follows:

$$\theta^{\mu} \leftarrow \theta^{\mu} + \alpha_{\mu} \cdot \nabla_{a} Q\left(s, a \mid \theta^{Q}\right)\Big|_{a=\mu(\delta \mid \theta^{\mu})} \nabla_{\mu} Q\left(s \mid \theta^{\mu}\right) \tag{7}$$

TABLE 2. Compilation environment.

Tool	Version
Python	3.7
pandas	1.1.3
scikit_learn	1.1.1
tensorflow	2.6.2
Keras	2.6.0
tensorlayer	2.2.5
numpy	1.19.2

Finally, a soft update of the target Q-network and the target policy network is completed. The update equation is as follows:

$$\begin{cases} \theta^{Q} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'} \\ \theta^{\mu} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \end{cases} \tag{8}$$

It is worth noting that the E-DDPG algorithm uses an experience replay mechanism to break the temporal correlation of the training samples. At each update, a certain amount of data is taken out from the experience replay pool, using up and down sampling to calculate the gradient. The data, including s and a, generated by each iteration for network updates are also stored in the experience replay pool until the corresponding maximum number of iterations is reached and the model training is complete.

IV. EXPERIMENTS

In this section, we first introduce the development environment and the data sets used. We then validate the performance of the proposed framework with experimental results.

A. DEVELOPMENT ENVIRONMENT

System development relies on a computing environment. Our system development environment, as shown in TABLE 2, integrates and comprehensively uses various resources.

B. BASELINE

We compare the neural network part of our proposed method with three common neural networks, and the performance of the DRL part with a model that uses the PPO algorithm to achieve energy optimization. In the experiments with heterogeneous data, we use the following benchmark model:

- LSTM [25], a commonly used neural network (RNN) structure, suitable for nonlinear data problems
- GRU [45], a commonly used neural network structure, simpler compared to LSTM structure
- CNN [23], a deep learning network structure, suitable for processing topological data
- PPO [31], a reinforcement learning algorithm with good performance in HVAC energy optimization

TABLE 3. Evaluation indicators for various models.

Model	MAE	MAPE	Training time (s)
LSTM	0.047	17.9548	214
GRU	0.0514	18.7705	116
CNN	0.0622	37.3155	57
CNN-LSTM	0.0458	16.9898	122

C. DATA SET

HVAC systems are commonly employed in plants and can offer typical data on energy consumption. Further, to improve the robustness of the model, geographical locations that may influence the performance of HVAC systems must be considered. Therefore, the data sets used were collected from five plants in different regions. Noise reduction and feature filtering were used to select the four most important features from among their uniform set of 15 features: compressor suction temperature, evaporator side water outlet temperature, condenser side water inlet temperature, and condenser side water inlet temperature. These four features formed a data set to support the energy consumption prediction and optimization models.

D. PERFORMANCE IN ENERGY CONSUMPTION PREDICTION

Four models, namely CNN–LSTM, gated recurrent unit (GRU), LSTM, and CNN, were used to independently predict the same segment of data, and their results were compared.

As can be seen from TABLE 3, under the same conditions for all models, the LSTM model exhibits a relatively small mean absolute error (MAE) and mean absolute percentage error (MAPE) but the longest training time and, therefore, relatively low efficiency. The GRU model has a simpler internal structure compared with the LSTM due to its reduced number of gates, and its relatively small MAE and MAPE values and relatively short training time lend it an efficiency higher than that of the LSTM. The CNN model requires the shortest training time but has the worst prediction accuracy. Finally, the CNN–LSTM model, constructed by adding the convolutional and pooling layers before inputting to the LSTM, is relatively fast at the reduced dimensionality of the input LSTM data due to the convolutional layer. Compared with the other three models, the CNN–LSTM model has the smallest MAE and MAPE values and the best prediction performance. At the same time, the CNN–LSTM model requires fully 42% less training time than the simple LSTM model, making the CNN–LSTM model the most efficient model.

The prediction performance of various models can be seen in FIGURE 4. The CNN model performs poorly, with significant differences between true values and predicted values. As expected, the other three models demonstrate no such significant gaps. However, many outliers appear in

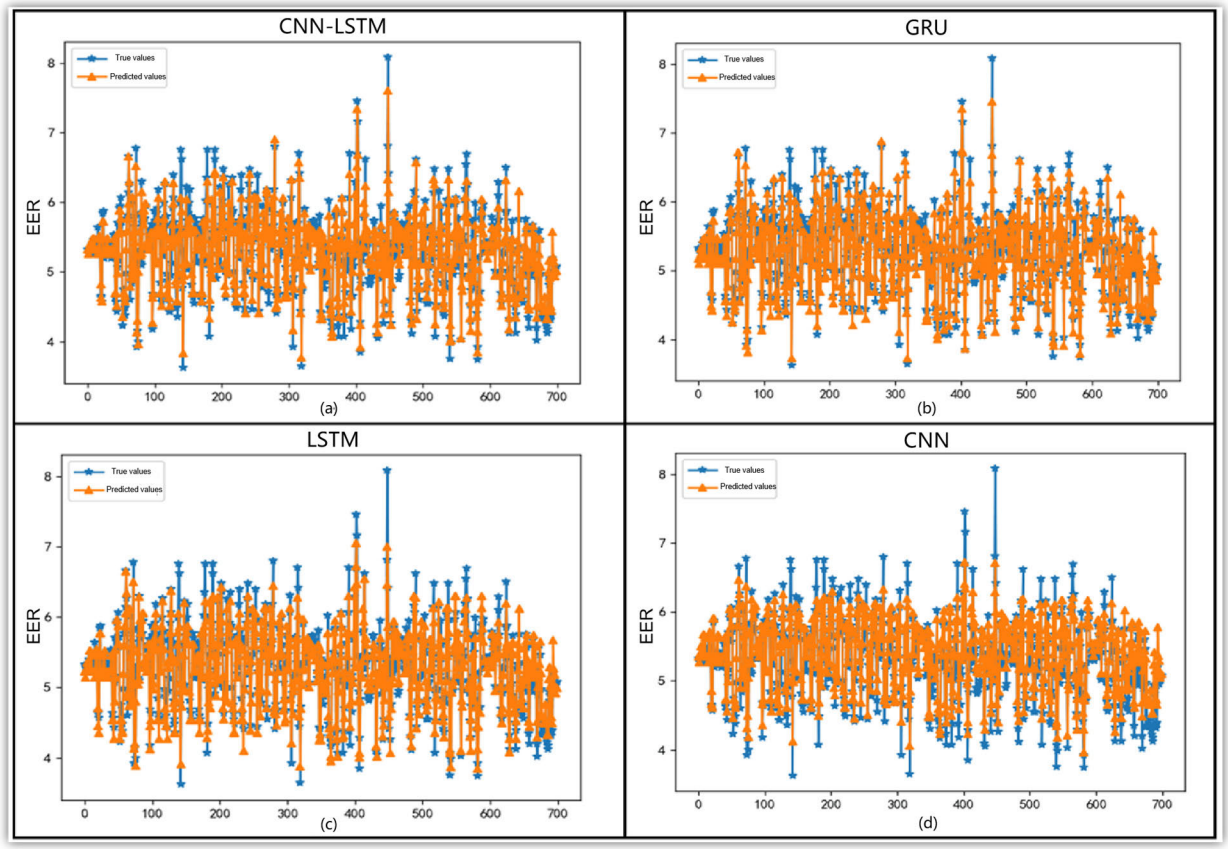

FIGURE 4. Comparison of prediction performance among four models.

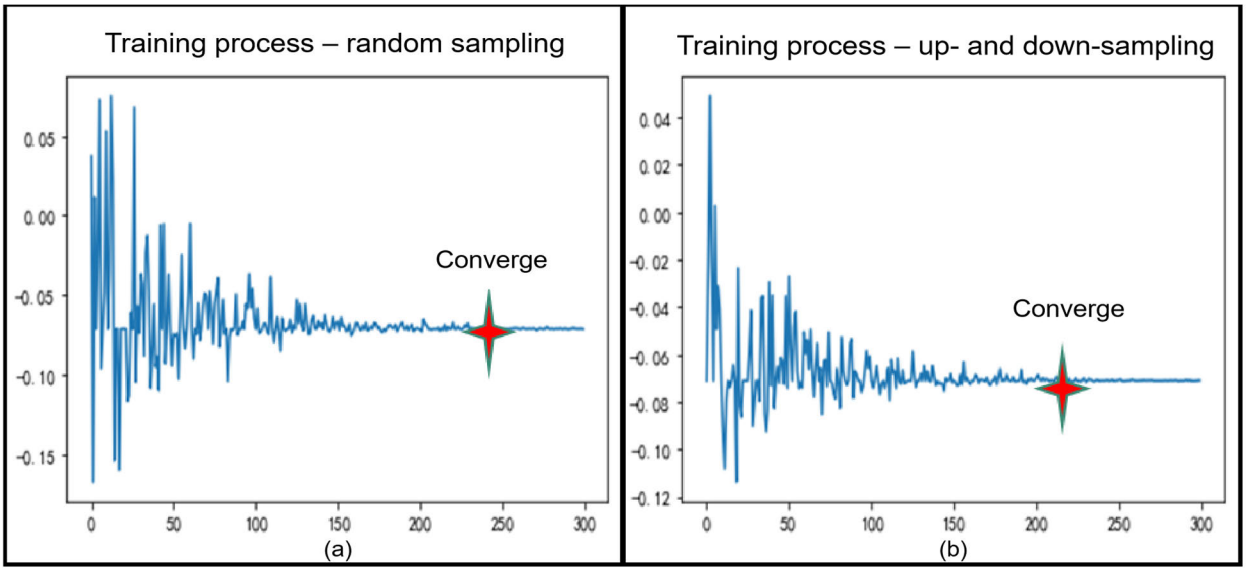

FIGURE 5. Effect of sampling method on model performance.

epochs 300–500. Outliers can reflect the stability of a model's prediction performance. In this framing, the CNN–LSTM model is more stable than the other two models. The reason is that the neural network with a convolutional layer not only improves the training efficiency compared with the simple LSTM model but also enhances the ability to capture data features. The noise of the data is further reduced through the convolutional layer, making the data more suitable for the model to learn and use in the improvement of prediction performance.

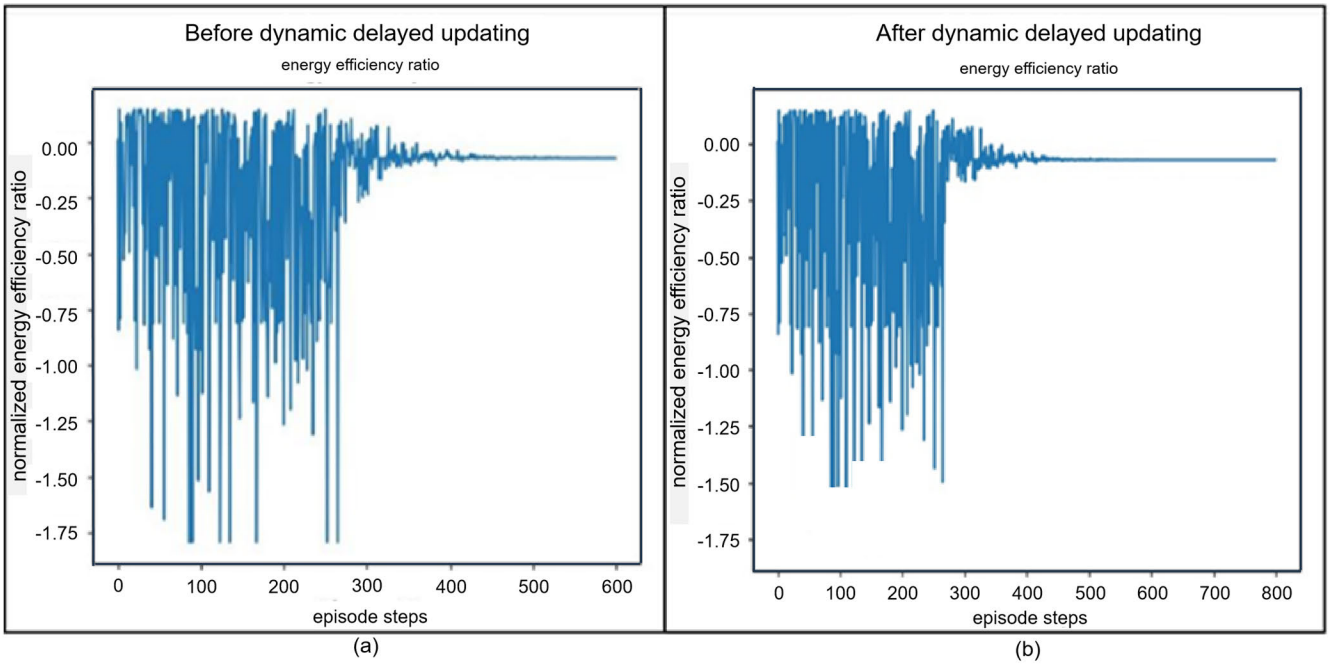

FIGURE 6. Effect of dynamic delayed updates on model results.

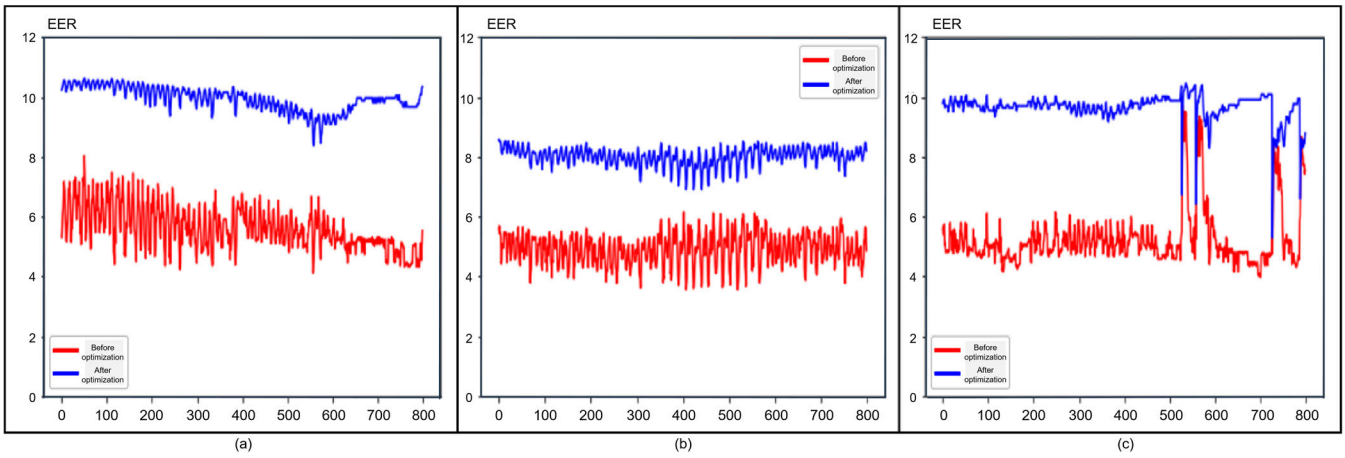

FIGURE 7. Comparison of energy efficiency ratio (EER) before optimization (red) and after optimization (blue) based on parameter data at various timescales.

In summary, the CNN–LSTM model provides the best prediction performance.

E. EFFECT OF IMPROVED TRAINING STRATEGIES

To determine the effects of our optimized up and down sampling method, the model was trained using each of two methods: a random sampling method and our optimized method, separately. FIGURE 5 shows the EER change after the neural network update once the experience pool was filled. Panel FIGURE 6(a) shows the training process corresponding to random sampling, which converges at the 250th iteration, while panel FIGURE 6(b) shows the training process corresponding to up and down sampling, which converges at the 200th iteration. The number of iterations for convergence was thus reduced by 20%. Therefore, the experimental results demonstrate that our sampling method improves the model training speed while reducing the training time. Under our sampling method, the training results converge, allowing a stable control strategy for the HVAC system to be obtained with fewer iterations.

To determine the effects of our model's dynamic delayed updating, we compared the normalized EERs before and after updating via a dynamic delayed update, with results shown in FIGURE 6. The EER interval of the model before the dynamic delayed update was [−1.80,0.16], but the narrower EER interval of the model after the dynamic delayed update was [−1.51,0.16], indicating that dynamic delayed updating can reduce the fluctuation interval of EER by 14.8%. Therefore, dynamic delayed updating can make the training

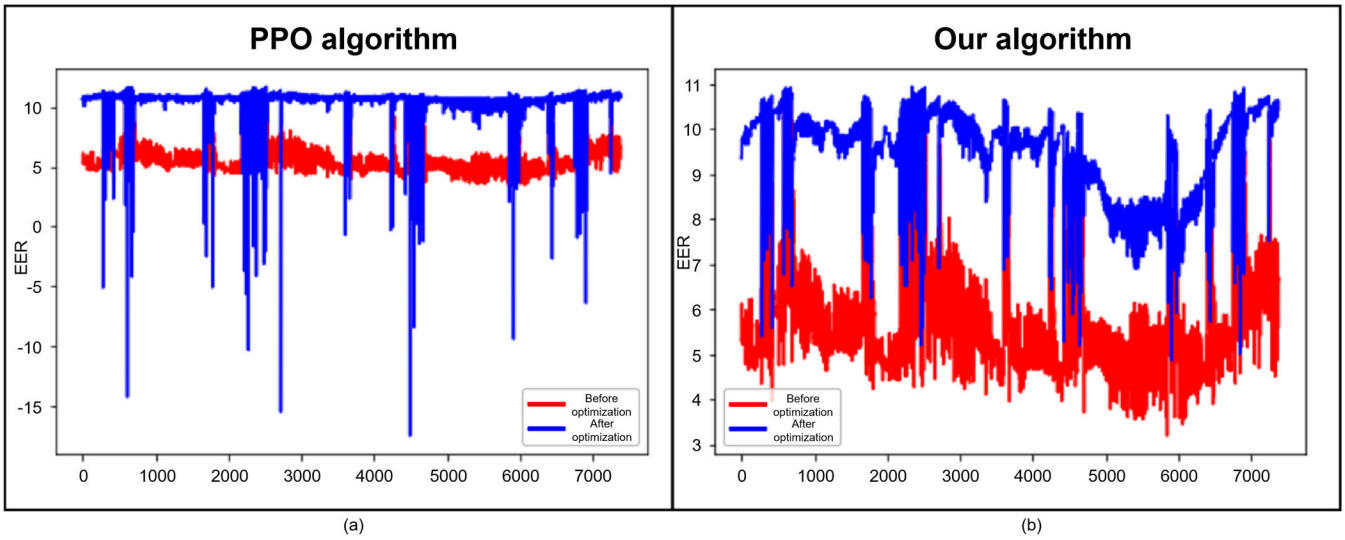

FIGURE 8. Energy consumption optimization performance: comparison of PPO and enhanced deep deterministic policy gradient (E-DDPG) algorithms.

process more robust by reducing the fluctuation interval of energy consumption in the training process.

In summary, the optimized training strategy can improve the training speed while making the training process more stable and the model more robust.

F. PERFORMANCE IN ENERGY CONSUMPTION OPTIMIZATION

To test the effectiveness and generalizability of the energy consumption optimization model based on the E-DDPG algorithm, experiments with different groups were conducted using the test data. We selected HVAC system parameter data at different timescales and on different days for energy consumption optimization and compared the test results.

FIGURE 7 compares the EERs before and after energy consumption optimization of 800 pieces of time-series continuous parameter data at different timescales using the E-DDPG algorithm. Clearly, optimization greatly improved the EER. As shown in panel (a), the EER of the original data began to decline around the 610th piece of data, but the model managed to adjust the strategy in real-time to improve the EER and reduce energy consumption accordingly, indicating that the model was able to make real-time policy adjustment for different environmental conditions, showing good generalizability.

To verify the particular effectiveness of the E-DDPG algorithm, an alternative algorithm was used to learn energy consumption optimization strategies, and its optimization results were compared with those of the E-DDPG algorithm. FIGURE 8 compares the energy consumption optimization results using the PPO and E-DDPG algorithms, with the red curve showing the EER before optimization and the blue curve showing the EER after optimization. The final EER after optimization using the PPO algorithm is maintained at a level greater than 10 but occasionally falls below 0. This behavior indicates that this algorithm cannot offer an optimized control strategy for the energy consumption of an HVAC system in all cases and thus provides only mediocre optimization performance. In contrast, the E-DDPG algorithm maintains the EER at a level of around 10 and improves to 11 at many points. This behavior shows that the E-DDPG algorithm outperformed the PPO algorithm. Indeed, the E-DDPG algorithm completed the energy consumption optimization task while improving the EER by 49% compared to not optimized.

V. CONCLUSION

In this paper, we proposed an energy consumption optimization system for HVAC systems that integrate CNN–LSTM–based energy consumption prediction and E-DDPG–based energy consumption optimization. Our key results are as follows.

1) In contrast with an LSTM model, the CNN–LSTM algorithm can both extract features and learn time-series trends. It efficiently learns the time-series features of the HVAC system parameter data, thus accelerating the training process by reducing the training time by 42.9% compared with an LSTM model. In addition, the CNN–LSTM algorithm accurately predicts the energy consumption at the next moment to simulate the interaction between the agent and environment in the RL algorithm and assist in the DRL algorithm's modeling;

2) During training with the E-DDPG algorithm using an improved training strategy, the up and down sampling methods can overcome the problem of uneven distribution of learning experience of the agents at each time point while improving model training speed, shortening training time, and reducing the number of iterations for convergence by 20%. Dynamic delayed updating can reduce the fluctuation interval of energy consumption during training by 12.5%, making the training process more stable.

3) The E-DDPG algorithm for energy consumption optimization of HVAC systems successfully makes real-time adjustments to its strategy in different environments, showing good generalizability. In our modeling experiments, the E-DDPG algorithm improved the EER of the HVAC system by 49%.

In summary, the CNN–LSTM–based energy consumption prediction model, with high prediction accuracy and high speed, successfully simulates the HVAC system environment and provides a basis for decisions on energy consumption optimization. The E-DDPG energy consumption optimization model finds stable control strategies for HVAC systems faster and is more robust than traditional models. It can help HVAC systems better meet the needs of daily operations and achieve efficient energy use.

Our work focuses mainly on the energy consumption optimization of HVAC systems. Future efforts may consider expanding the application of DRL in the optimization scenarios of other energy consumption subsystems in buildings, such as electric vehicles, electric water heaters, and washing machines. In this way, comprehensive energy consumption optimization can be realized for energy management systems in buildings.

Moreover, the present analysis of control and management of energy systems in buildings has focused mainly on optimizing energy consumption without considering privacy leakage during training. In future research, we will consider introducing federated learning techniques to protect the data privacy of HVAC systems during training.

ACKNOWLEDGMENT

(Yi Peng and HaoJun Shen contributed equally to this work.)

REFERENCES

[1] Building Energy Consumption Statistics Committee of China Building Energy Conservation Association, "China building energy consumption annual report 2020," *Building Energy Efficiency*, vol. 49, no. 2, pp. 1–6, 2021.

[2] M. A. A. Razali, M. Kassim, N. A. Sulaiman, and S. Saaidin, "A ThingSpeak IoT on real time room condition monitoring system," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS)*, Jun. 2020, pp. 206–211.

[3] M. Ala'raj, M. Radi, M. F. Abbod, M. Majdalawieh, and M. Parodi, "Data-driven based HVAC optimisation approaches: A systematic literature review," *J. Building Eng.*, vol. 46, Apr. 2022, Art. no. 103678.

[4] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, "CNN-LSTM architecture for predictive indoor temperature modeling," *Building Environ.*, vol. 206, Dec. 2021, Art. no. 108327.

[5] M. Xia, F. Chen, Q. Chen, S. Liu, Y. Song, and T. Wang, "Optimal scheduling of residential heating, ventilation and air conditioning based on deep reinforcement learning," *J. Modern Power Syst. Clean Energy*, early access, Nov. 28, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9965191

[6] L. Tang and Y. Meng, "Data analytics and optimization for smart industry," *Frontiers Eng. Manag.*, vol. 8, no. 2, pp. 157–171, 2021.

[7] D. Liu, H. Zhen, D. Kong, X. Chen, L. Zhang, M. Yuan, and H. Wang, "Sensors anomaly detection of industrial Internet of Things based on isolated forest algorithm and data compression," *Sci. Program.*, vol. 2021, pp. 1–9, Jan. 2021.

[8] K. Tao, Q. Wang, and D. Yue, "Data compression and damage evaluation of underground pipeline with musicalized sonar GMM," *IEEE Trans. Ind. Electron.*, early access, May 1, 2023, doi: 10.1109/TIE.2023.3270519.

[9] K. Tao, Q. Wang, Z. Yao, B. Jiang, and D. Yue, "Underground sedimentary rock moisture permeation damage assessment based on AE mutual information," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[10] E. B. Priyanka and S. Thangavel, "Influence of Internet of Things (IoT) in association of data mining towards the development smart cities—A review analysis," *J. Eng. Sci. Technol. Rev.*, vol. 13, no. 4, pp. 1–21, Aug. 2020.

[11] X. Wang, S. Wang, X. Liang, D. Zhao, J. Huang, X. Xu, B. Dai, and Q. Miao, "Deep reinforcement learning: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 28, 2022, doi: 10.1109/TNNLS.2022.3207346.

[12] Z. Ding, Y. Pan, J. Xie, W. Wang, and Z. Huang, "Application of reinforcement learning in HVAC system operation optimization," *Building Energy Efficiency*, vol. 48, no. 7, pp. 14–20, 2020.

[13] J.-M. Liao, M.-J. Chang, and L.-M. Chang, "Prediction of air-conditioning energy consumption in R & D building using multiple machine learning techniques," *Energies*, vol. 13, no. 7, p. 1847, Apr. 2020.

[14] Y. Yao and D. K. Shekhar, "State of the art review on model predictive control (MPC) in heating ventilation and air-conditioning (HVAC) field," *Building Environ.*, vol. 200, Aug. 2021, Art. no. 107952.

[15] A. Afram and F. Janabi-Sharifi, "Theory and applications of HVAC control systems—A review of model predictive control (MPC)," *Building Environ.*, vol. 72, pp. 343–355, Feb. 2014.

[16] X. Li, S. Chen, H. Li, Y. Lou, and J. Li, "A behavior-orientated prediction method for short-term energy consumption of air-conditioning systems in buildings blocks," *Energy*, vol. 263, Jan. 2023, Art. no. 125940.

[17] S. Lu, Y. Huo, N. Su, M. Fan, and R. Wang, "Energy consumption forecasting of urban residential buildings in cold regions of China," *J. Energy Eng.*, vol. 149, no. 2, Apr. 2023, Art. no. 04023002.

[18] F. Divina, M. García Torres, F. A. Goméz Vela, and J. L. Vázquez Noguera, "A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings," *Energies*, vol. 12, no. 10, p. 1934, May 2019.

[19] A. Namdari and Z. S. Li, "Integrating fundamental and technical analysis of stock market through multi-layer perceptron," in *Proc. IEEE Technol. Eng. Manage. Conf. (TEMSCON)*, Jun. 2018, pp. 1–6.

[20] A. Namdari and T. S. Durrani, "A multilayer feedforward perceptron model in neural networks for predicting stock market short-term trends," *Oper. Res. Forum*, vol. 2, no. 3, p. 38, Sep. 2021.

[21] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, "Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster," in *Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. (EAIT)*, Jan. 2018, pp. 1–4.

[22] S. Frizzi, R. Kaabi, M. Bouchouicha, J. -M. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional neural network for video fire and smoke detection," in *Proc. 42nd Annual Conf. IEEE Ind. Electron. Soc.*, Dec. 2016, pp. 877–882.

[23] T. Y. Kim and S. B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.

[24] T. Hayashi and R. H. Howell, *Industrial Ventilation and Air Conditioning*. USA, 1985.

[25] C. Zhou, Z. Fang, X. Xu, X. Zhang, Y. Ding, X. Jiang, and Y. Ji, "Using long short-term memory networks to predict energy consumption of air-conditioning systems," *Sustain. Cities Soc.*, vol. 55, Apr. 2020, Art. no. 102000.

[26] C. H. Kim, M. Kim, and Y. J. Song, "Sequence-to-sequence deep learning model for building energy consumption prediction with dynamic simulation modeling," *J. Building Eng.*, vol. 43, Nov. 2021, Art. no. 102577.

[27] Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 115036.

[28] Z. Yao, X. Liang, G.-P. Jiang, and J. Yao, "Model-based reinforcement learning control of electrohydraulic position servo systems," *IEEE/ASME Trans. Mechatronics*, vol. 28, no. 3, pp. 1446–1455, Jun. 2023.

[29] W. Chen, X. Qiu, T. Cai, H.-N. Dai, Z. Zheng, and Y. Zhang, "Deep reinforcement learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1659–1692, 3rd Quart., 2021.

[30] A. I. Dounis, M. J. Santamouris, C. C. Lefas, and D. E. Manolakis, "Thermal-comfort degradation by a visual comfort fuzzy-reasoning machine under natural ventilation," *Appl. Energy*, vol. 48, no. 2, pp. 115–130, Jan. 1994.

[31] D. Azuatalam, W.-L. Lee, F. de Nijs, and A. Liebman, "Reinforcement learning for whole-building HVAC control and demand response," *Energy AI*, vol. 2, Nov. 2020, Art. no. 100020.

[32] Z. Xing, J. Zhu, Z. Zhang, Y. Qin, and L. Jia, "Energy consumption optimization of tramway operation based on improved PSO algorithm," *Energy*, vol. 258, Nov. 2022, Art. no. 124848.
[33] V. Congradac and F. Kulic, "HVAC system optimization with CO_2 concentration control using genetic algorithms," *Energy Buildings*, vol. 41, no. 5, pp. 571–577, May 2009.
[34] K. Dalamagkidis, D. Kolokotsa, K. Kalaitzakis, and G. S. Stavrakakis, "Reinforcement learning for energy conservation and comfort in buildings," *Building Environ.*, vol. 42, no. 7, pp. 2686–2698, Jul. 2007.
[35] S. Liu and G. P. Henze, "Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory," *Energy Buildings*, vol. 38, no. 2, pp. 148–161, Feb. 2006.
[36] H. Lan, Z. Bao, and Y. Peng, "An index advisor using deep reinforcement learning," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2105–2108.
[37] T. Wei, S. Ren, and Q. Zhu, "Deep reinforcement learning for joint datacenter and HVAC load control in distributed mixed-use buildings," *IEEE Trans. Sustain. Comput.*, vol. 6, no. 3, pp. 370–384, Jul. 2021.
[38] 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*.
[39] G. Gao, J. Li, and Y. Wen, "Energy-efficient thermal comfort control in smart buildings via deep reinforcement learning," 2019, *arXiv:1901.04693*.
[40] L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, and X. Guan, "Multi-agent deep reinforcement learning for HVAC control in commercial buildings," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 407–419, Jan. 2021.
[41] Z. Zou, X. Yu, and S. Ergan, "Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network," *Building Environ.*, vol. 168, Jan. 2020, Art. no. 106535.
[42] I. Zenginis, J. Vardakas, N. E. Koltsaklis, and C. Verikoukis, "Smart home's energy management through a clustering-based reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16363–16371, Sep. 2022.
[43] X. Ding, W. Du, and A. Cerpa, "OCTOPUS: Deep reinforcement learning for holistic smart building control," in *Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp.*, Nov. 2019, pp. 326–335.
[44] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020.
[45] X. He, Y. Wang, F. Guo, X. Zhang, X. Duan, and J. Pei, "Modeling for vehicle cabin temperature prediction based on graph spatial–temporal neural network in air conditioning system," *Energy Buildings*, vol. 272, Oct. 2022, Art. no. 112229.

YI PENG is currently pursuing the bachelor's degree in geographic information science with the China University of Petroleum. His research interests include reinforcement learning and data mining.

HAOJUN SHEN is currently pursuing the bachelor's degree in geomatics engineering with the China University of Petroleum. His research interests include reinforcement learning and data mining.

XIAOCHANG TANG is currently pursuing the bachelor's degree in Internet of Things engineering with the China University of Petroleum (East China). His research interests include edge computing and reinforcement learning.

SIZHE ZHANG is currently pursuing the bachelor's degree in intelligent science and technology with the China University of Petroleum. Her research interests include data mining and reinforcement learning.

JINXIAO ZHAO is currently pursuing the bachelor's degree in intelligent science and technology with the China University of Petroleum. Her research interests include data mining and AI security.

YURU LIU received the B.S. degree in communication engineering from Zhengzhou Shengda University, in 2021, and the M.S. degree from the China University of Petroleum. Her research interests include blockchain and federated learning.

YUMING NIE is currently pursuing the master's degree with the Department of Computer Technology, China University of Petroleum. Her research interests include time series data analysis and data mining.