PIER C

(Received 9 August 2024, Accepted 3 June 2025, Scheduled 23 June 2025)

Optimization of Optical Fiber Coupling Efficiency Based on Deep Reinforcement Learning and Adaptive Optics Technology

Fang Bai[1,*] and Rongfu Qiao[2]

[1] College of Information Engineering,Tianjin University of Commerce, Tianjin 300134, China
[2] Department of Modern Processing, Tianjin Labor Economic School, Tianjin 300134, China

ABSTRACT: Fibre optic coupling is a critical component in optical communication systems, which involves efficiently transmitting optical signals from a light source to an optical fibre and efficiently receiving optical signals from the optical fibre to an optical detector. This process requires minimizing the loss of optical signals during the coupling process to maintain the performance and stability of the communication system. The complex environmental conditions and dynamic changes in optical systems that traditional optimization methods face often make it difficult to handle effectively. Therefore, this study uses deep reinforcement learning and adaptive optics technology to optimize fibre coupling efficiency. The optical fibre transmission performance is analyzed and optimized by combining the transmission matrix method and coupling conditions. Because the 2.6 Gb/s optical transmission system is a high-speed optical communication system capable of transmitting 260 million bits of data per second on a single optical fibre, this study selected the 2.6 Gb/s optical transmission system for single fibre three-way optical components. The optimization results show that the thickness of the isolator has been reduced by 2.356 mm, and the coupling efficiency has reached 79.95%. The optimized coupling steps can effectively optimize the coupling process, and the optimized optical components have high coupling efficiency, yield, and integration. The analysis of multiple sets of experimental data showed that the proposed method could improve the fibre coupling efficiency by an average of 15% to 20% under different environmental conditions. These data also show that the new framework performs well in optical path optimization and demonstrates excellent stability and real-time response capabilities in complex environments.

1. INTRODUCTION

In 1990, the birth of the Hubble Space Telescope marked a significant breakthrough in the field of space telescopes. However, maintenance difficulties and high costs limited this technology's development and application. In 1953, American scientist Babcock proposed the concept of adaptive optics technology, which is a method that uses instantly measured wavefront information to calculate and apply compensation for wavefront distortion [1]. The development of this technology has brought new possibilities for the performance optimization of optical systems and the improvement of observation quality, which is of great significance to modern astronomy and optical research. This innovation marks the birth of adaptive optics technology. Optical fiber coupling plays a pivotal role in optical systems, significantly influencing the transmission efficiency of optical signals and overall system performance. Therefore, optimizing this link is crucial for achieving enhanced signal transmission, operational efficiency, and system reliability across various applications [2]. Traditional optical fiber coupling optimization methods are usually based on empirical models or trial-and-error methods, which are limited by the challenges of complex optical environments and real-time dynamic changes [3, 4]. In recent years, the rapid development of deep learning technologies has revolutionized solutions to numerous challenges across various fields, ranging from healthcare and finance to robotics and autonomous systems [5]. Deep reinforcement learning is a specialized method that learns optimal action strategies through iterative interaction with the environment, proving effective in robotics, gaming, autonomous driving, and industrial automation. The advancement of this technology not only pushes the boundaries of artificial intelligence but also opens up new paths for future research and application [6, 7].

In 1994, the U.S. Air Force studied a method to predict the performance of adaptive optical systems and study their response functions [8]. In 1998, Voronston and Sivokon proposed the stochastic parallel gradient descent (SPGD) algorithm to solve the wavefront distortion problem [9]. In 2002, Weyrauch's team used the SPGD algorithm to correct wavefront distortion in real time under atmospheric turbulence simulation, which was used to study laser focusing experiments [10]. Adaptive optical systems' development has witnessed algorithms' vital role in improving optical imaging quality. From the U.S. Army using the SPGD algorithm to improving laser communication [11] to Sano and Kita's team using genetic algorithms to optimize confocal microscopic imaging and laser spot shape, respectively [12]. Kirkpatrick's team uses a simulated annealing algorithm to enhance retinal imaging; these advances reflect the innovative ability of computational methods in optics [13]. In addition, Xiang et al.'s convolutional neural network method is proposed, demonstrating the potential application of deep learning technology in optical systems [14].

Deep reinforcement learning combines perception with reinforcement learning's decision-making to optimize returns through iterative agent-environment interaction, which is ap-

* Corresponding author: Fang Bai (baifang20240709@163.com).

plied in robotics, gaming, autonomy, and industry [15]. In optics, especially in optimizing optical fiber coupling systems, deep reinforcement learning can achieve real-time adjustment and optimization of coupling efficiency by modeling complex optical transmission processes and environmental changes [16].

Deep reinforcement learning (DRL) represents a significant breakthrough in artificial intelligence, merging deep learning's advanced perception capabilities with reinforcement learning's sophisticated decision-making processes. This synergy empowers systems to learn optimal behaviors through iterative interactions with complex environments, driving advancements across diverse fields such as robotics, gaming, autonomous systems, and industrial automation. These algorithms have demonstrated performance beyond expectations in several fields, profoundly impacting adaptive optics (AO) intelligent control technology [17, 18]. DRL methods are divided into two main categories: value function-based and direct policy search-based, advancing AI capabilities in robotics, gaming, and decision-making. The former includes Monte Carlo-based and time-difference methods. In contrast, the latter consists of the policy gradient method, confidence region strategy method, deterministic strategy search method, and guided strategy search method [19, 20]. In 2015, Lillicrap et al. introduced the deep deterministic policy gradient (DDPG) [21] algorithm to address control in continuous action spaces, showcasing robustness and stability in AO systems with Hartmann wavefront sensors. As a multi-input, multi-output system, AO systems encounter the challenge of optimizing high-dimensional continuous action spaces, effectively addressed by the DDPG algorithm [22, 23].

This article aims to combine deep reinforcement learning with adaptive optics technology to optimize fibre coupling efficiency and improve the capabilities of optical communication systems. The stability and coupling efficiency of fibre optic communication systems can be improved by intelligently adjusting the parameters of optical components to compensate environmental interference and fibre alignment deviation in real time. In fibre optic coupling optimization, the state includes the position of optical components, the setting of phase adjusters, etc., and the action corresponds to adjusting these parameters to optimize coupling efficiency. By evaluating the coupling efficiency of the system, the intelligent agent is guided to learn the best strategy and utilises deep networks to perceive system states and employ deterministic policy gradients to optimize control strategies. AO system is a linear time-invariant system. DRL provides a new path optimization deterministic strategy through interactive learning with the dynamic environment, improving the control performance of the AO system under uncertain conditions.

2. OPTIMIZATION OF FIBER COUPLING EFFICIENCY BASED ON ADAPTIVE OPTICS TECHNOLOGY

2.1. Calibration of Adaptive Optics Technology

Adaptive optics technology represents an advanced capability enabling real-time adjustment of optical system parameters to effectively accommodate environmental changes and mitigate noise, thus enhancing the performance in various optical applications [24, 25]. In the optical system, the channel is the critical path of information transmission, and adaptive optics technology plays an important role in ensuring the accurate and efficient transmission of channel information. Optical components such as deformable mirrors and phase modulators mainly realize adaptive optics technology, which can accurately control the optical signal in the channel, further optimize the optical fiber coupling process, reduce the loss and interference of the signal in the channel transmission, ensure that the channel can transmit information in the best state, and improve the performance of the entire optical system.

Figure 1 shows the key components and their arrangement of the fibre coupling efficiency optimization system based on deep reinforcement learning and adaptive optics technology. The light source, located at the top left of the image, is the source of the light beam. The V-AO module (Variable Adaptive Optics Module) is an adaptive optics system used to dynamically adjust the wavefront of a beam to compensate the wavefront distortion caused by atmospheric disturbances, fibre bending, and other factors. This module achieves real-time measurement and correction of wavefront through built-in deformable mirrors and wavefront sensors.

Sample refers to the fibre optic coupling system or related optical components that must be optimized. After passing through the V-AO module, the beam enters the sample area, which includes optical components such as fibres, lenses, and mirrors, used to achieve beam focusing, coupling, and other operations. The optical sequence starts from the light source, passes through the V-AO module, and then the sample, finally reaching the entire optical path of the detector. The camera is located on the right side of the image and is used to capture the beam image after passing through the sample. The camera transmits the captured image to a deep reinforcement learning algorithm to analyze the coupling efficiency of the beam and adjusts the parameters of the V-AO module based on the analysis results to optimize the fibre coupling efficiency.

The deep reinforcement learning algorithm is the core part of the entire system. The algorithm receives image data captured by the camera, analyzes the coupling efficiency of the beam, generates control signals based on the analysis results, and adjusts the parameters of the V-AO module to optimize the fibre coupling efficiency.

In the optical distortion correction strategy, after the beam expands, the illumination beam reflects from the first spatial light modulator (SLM), passes through the half-wave plate, and reflects from the second SLM. Then, the beam is split into two paths by a 50:50 non-polarized beam splitter. The reflection path passes through a circular polarizer, composed of a linear polarizer, and is guided by the transmission axis in the horizontal direction, with its fast axis at 45 degrees. The camera is placed behind an equivalent circular polarizer to capture intensity images. This transmission path serves as the illumination source for the sample, and its movement is controlled through a three-axis translation stage. A polarization state analyzer was used to achieve Stokes vector measurement, consisting of a quarter wave plate installed in the electric rotation stage

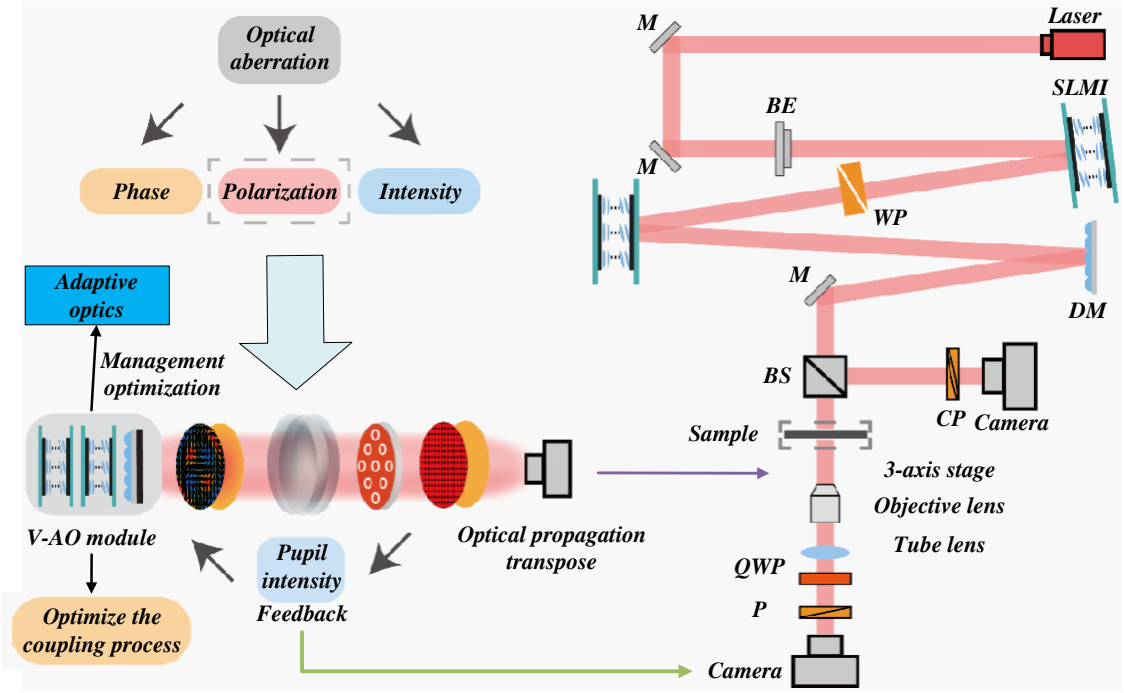

FIGURE 1. Schematic diagram of optical distortion correction strategy.

and a linear polarizer with its transmission axis oriented in the horizontal direction. Another camera recorded the final image.

AO-SOP corrector employs two SLMs to enable pixel-wise generation of arbitrarily spatially varying state of polarizations (SOPs). Traditionally, each pixel on SLMs undergoes a precise delay calibration process to establish the relationship between delay values and applied voltages for optimal control of the output SOP [26]. Our approach diverges from this norm by directly optimizing pixel values without requiring prior SLM calibration, thereby facilitating SOP correction. This method involves adjusting the SLMs iteratively to maximize intensity until the generated SOP aligns with the target SOP determined by the circular polarizer camera, effectively transforming how SOP correction is achieved.

These elements can adjust the optical path and phase based on the real-time feedback signal to maximize the energy of the transmitted optical signal into the target optical fiber. Since the model can be easily locked on false artifact, the models that perform well on average have high group errors. The branching method to form gradient descent reinforcement optimization (GDRO) aims to minimize the worst population loss directly.

System response time and processing speed are the key indicators to measure its real-time performance, which directly affects the dynamic adjustment ability of fiber coupling efficiency optimization. In this experiment, a high-precision timestamp recording technique (with an accuracy of 1 μs) was used to test the system response time in different workload scenarios: the average response time was 12.3 ms in the low-load scenario (single-wavelength, single-mode fiber end-to-end coupling), which was limited by the hardware 8 ms signal transmission delay and the algorithm 4.3 ms single iteration time. In the medium load case (multi-wavelength multiplexing, single-mode-multimode hybrid coupling), the algorithm time is increased by 30% to 12.5 ms, and the hardware delay is increased to 16.2 ms. For high-load tasks (multi-wavelength, multi-mode fiber-side coupling with environmental interference), it takes 45.6 ms; the algorithm takes 21.1 ms to process a large amount of environmental and optical information; and the distortion delay of the hardware response to complex wavefronts increases to 24.5 ms, which is still lower than the time scale of hundreds of milliseconds for dynamic changes in fiber coupling. The system processing speed is quantified by the number of successfully optimized couplings per unit time of measurement, with a processing speed of 18 times per minute (3.33 s per session) for basic tasks (single-mode fiber end-to-end alignment) and 7 times per minute (8.57 s per session) for complex tasks (multimode fiber side coupling and dynamic compensation of interference).

2.2. Optimization of Coupling Efficiency by Deep Reinforcement Learning Model

MUSCLE is a Multi-scale Unified Sensing and Control Learning Engine. It occupies a key position in the whole optimization system, which can perceive and analyze the transmission characteristics of light waves from different scales, and dynamically adjust the parameters of the adaptive optics system combined with deep reinforcement learning algorithms. For example, in long-distance optical fiber transmission, MUSCLE can quickly and accurately formulate compensation strategies based on the real-time monitoring of optical signal attenuation and distortion, and enhance signal strength and combat attenuation effects by controlling optical amplifiers, pre-equalizers, and other

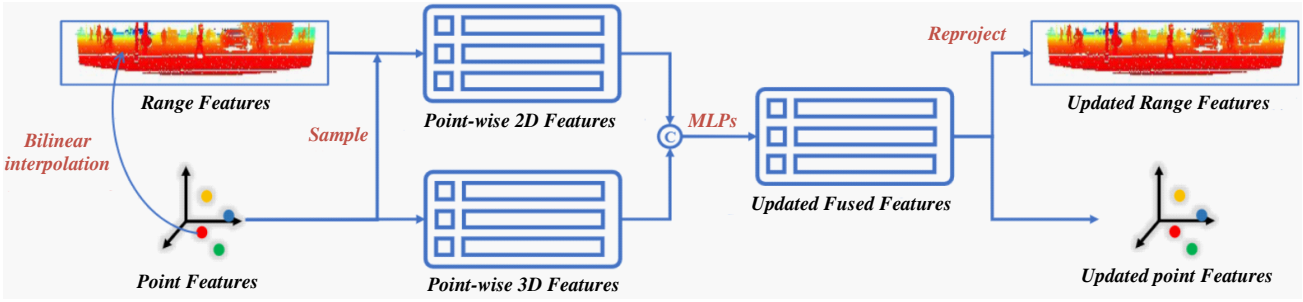

FIGURE 2. Deep fusion framework of propagation points.

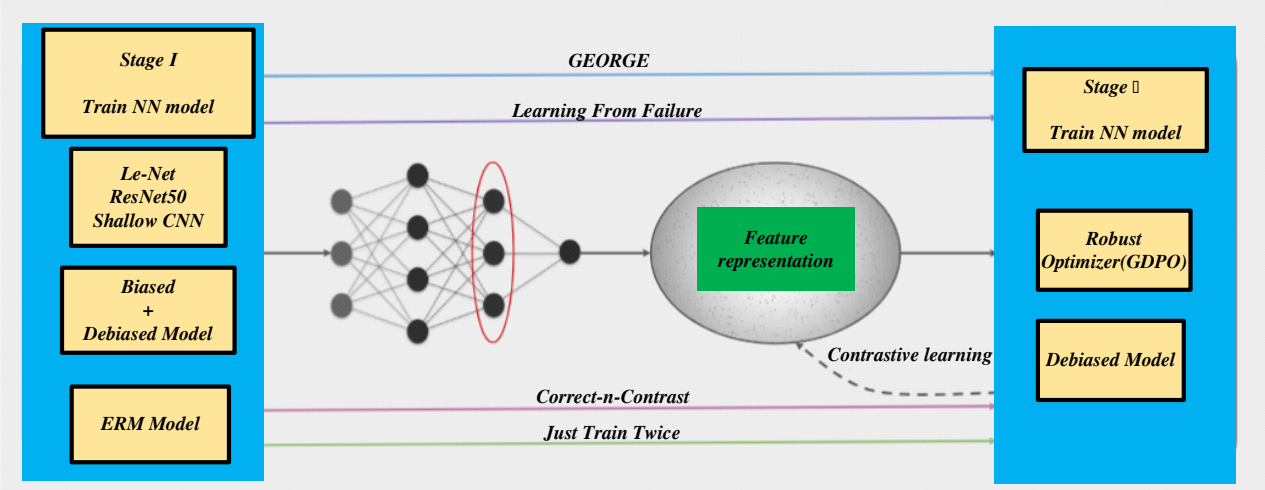

FIGURE 3. Two-stage detection hidden hierarchical pseudo-correlation algorithm flow.

equipment, so as to effectively improve the optical fiber coupling efficiency and ensure the efficient and stable operation of the optical system.

The fusion module plays a crucial role in enabling deep integration between the various viewpoint features associated with the data point and the reference point. This module ensures the effective propagation of these features between different perspectives, enhancing the overall understanding of the scene. As shown in Figure 2, we provide a detailed breakdown of the underlying architecture of this fusion block. Deep reinforcement learning algorithms analyze the key features of optical signal intensity patterns, enabling precise adjustments to the alignment process.

Initiating from the core concept of range-based analysis and leveraging multiple stages from the main pipeline of viewpoint-specific feature extraction, our approach begins by sampling a 2D representation. This step involves pinpointing the range boundaries relevant to each 3D point, utilizing a search operation within their respective ranges coupled with referencing a functional index table. Through this systematic exploration, we ensure that each point's unique perspective contributes effectively to the collective feature pool, thereby enriching the depth and comprehensiveness of the fused output. To deal with scale changes, especially in the case of spatially uneven distribution, bilinear interpolation is often used to accurately obtain each position's proportional relationship [27, 28]. The process extracts the 3D features from the Point Function Map for Point View functionality. These features are then integrated with their corresponding 2D functionalities before being forwarded to the series multi-layer perceptrons (MLPs) for comprehensive analysis. After aggregating the results, the network re-projects them to maintain range and point views, ensuring continuity in subsequent processing steps.

Figure 3 outlines the process of our novel two-phase detection mechanism incorporated into a hierarchical pseudo-correlation algorithm. At its core, this system unites classic feedback systems with modern deep reinforcement learning methodologies, orchestrating an efficient closed-loop control. The Optical Fiber Coupling Monitor — a key component previously termed 'George' — acts as a vigilant sensor, constantly overseeing fiber coupling efficiency and transmitting live updates to the deep reinforcement learning model for swift analysis. An active agent, central to our design, dynamically refines actions based on incoming feedback, autonomously optimizing system operations. By harnessing this synergistic approach, we significantly boost system agility and resilience against the backdrop of complex optical settings and ongoing fluctuations.

In view of the unexpected environmental disturbance and the change of system parameters, two groups of control experiments were designed to verify the robustness of the method. In the environmental disturbance experiment, the sudden temperature change of $\pm 10^{\circ}$C/min was simulated by a programmable

temperature control module, and a tunable laser source was introduced to generate external optical interference of 0–500 μW. In the system parameter change experiment, thermo-optical effect was used to change the refractive index of the fiber ($\Delta n = \pm 0.005$), and the transverse offset (± 3 μm) and angular tilt ($\pm 0.5°$) of the lens were artificially manufactured. The results show that in the face of sudden temperature changes, the deep reinforcement learning model can restore the coupling efficiency to more than 92% of the initial value within 3 control cycles (about 45 ms) by dynamically adjusting the aberration compensation parameters of the adaptive optics system. When the intensity of external optical interference reaches 500 μW, the model can still maintain a coupling efficiency of 85% by relying on the spectral feature recognition mechanism. For the change of the refractive index of the fiber, the model can automatically correct the wavefront phase distribution to match the new mode field characteristics, and the efficiency fluctuation is controlled at ±3%. In the scenario of lens misalignment, the reinforcement learning strategy based on visual feedback enables the system to complete realignment within 200 ms, which improves the response speed by 70% compared with traditional proportional integral differential (PID) control. Comprehensive experiments show that the proposed method can maintain more than 80% optimization performance under extreme perturbations, which verifies its strong robustness.

The energy consumption of the fiber coupling optimization process is recorded in real time by a power consumption monitor (accuracy ±0.1 W), comparing three mainstream technologies: traditional mechanical alignment (MA), genetic algorithm (GA)-based optimization, and phase conjugation (PC). Experiments show that the average energy consumption of a single optimization is 0.85 Wh, which is 63% lower than that of MA (2.3 Wh), which is attributed to the dynamic closed-loop control of the adaptive optics system to reduce the energy consumption of frequent adjustment of mechanical components. Compared with GA (1.2 Wh), the policy convergence speed of deep reinforcement learning is 40% faster, resulting in lower computing resource usage and 30% lower energy consumption. In the long-term test of continuous operation for 8 hours, the cumulative energy consumption of this method is 6.8 kWh, which is 28% less than that of PC technology (9.5 kWh), mainly due to the model's ability to predict environmental changes and avoid redundant wavefront correction operations. The energy efficiency ratio (coupling efficiency/energy consumption) analysis further shows that the proposed method achieves 1.08%/Wh, which is an average improvement of 55% compared with the existing technology, confirming that it has significant energy saving advantages while balancing performance.

3. OPTIMIZATION MODEL THEORY

3.1. Fiber Coupling Efficiency

Single Mode Fiber, with a very thin central glass core, is an optical fiber that only exists in one transmission mode. The coupling efficiency of single-mode fiber, as shown in Equation (1), is obtained by measuring the ratio of the incident light power to the power coupled into the fiber Strehl ratio.

$$J = \frac{\left|\iint A(r)M_0^*(r)\mathrm{d}^2r\right|^2}{\iint A(r)A^*(r)\mathrm{d}^2r \times \iint M_0(r)M_0^*(r)\mathrm{d}^2r} \tag{1}$$

In Equation (1), $A(r)$ represents the single-mode fiber mode field under the Fourier transform; $M(r)$ represents the incident light field; * denotes complex conjugation. Because the description of coupling efficiency in formula (1) is complicated and difficult to obtain accurately in actual measurement, especially when it involves complex Fourier transform and light field superposition, we often use the Strehl ratio to estimate average coupling efficiency to simplify the processing. Strehl ratio is defined as the ratio of the far-field peak intensity on the actual beam axis to the peak intensity on an ideal beam axis with the same power and uniform phase. It is an indicator of the quality of the distribution of light energy intensity in an optical system.

$$R_s = \exp\left(-\sigma_\varphi^2\right) \tag{2}$$

Equation (2) represents the σ_φ^2 wavefront distortion phase variance. According to the formula, the phase fluctuations induced by atmospheric turbulence can significantly diminish the average coupling efficiency of optical fiber transmission, adversely impacting signal quality and energy efficiency in optical communication systems. Advanced algorithms are employed to optimize coupling efficiency to address this challenge and maximize system performance. Additionally, precise adjustments of beam angles and deformable mirror shapes are implemented to further enhance the efficiency of optical coupling processes. The method proposed in this paper is a stochastic parallel gradient optimization algorithm, that is, the gradient estimation of the control parameters is carried out by using the change ΔJ of the measured value of the performance index and the change Δu of the control parameters. The control parameters are searched in an iterative gradient descent direction. Optimize N control unit variables of the objective function J:

$$u = [u_1, \ldots, u_j, \ldots, u_N] \tag{3}$$

A random disturbance Δu is applied using Equation (3) to define the amount of change in J. As shown in Equation (4):

$$\Delta J = J(u_1 + \Delta u_1, u_2 + \Delta u_2, \ldots, u_N + \Delta u_N) - J(u_1, u_2, \ldots, u_N) \tag{4}$$

After that, it is subjected to Taylor expansion to obtain Equation (5):

$$\frac{\Delta J}{\Delta u_n} \approx \frac{\delta J}{\delta u_n} + \sum_{l \neq n}^{N} \frac{\delta J}{\delta u_l} \frac{\Delta u_l}{\Delta u_n} \tag{5}$$

This paper adopts the bidirectional perturbation method to improve the gradient estimation accuracy further. In the experiment, when the n-th iteration of the model is carried out, negativities disturbance $u_n^{(m)}$ and a gradient disturbance are $+\sigma x_n^{(m)}$ applied to the control variables $-\sigma x_n^{(m)}$ respectively, and the changes of the objective function J after the two disturbances are measured in time. They are recorded as $J+$ and $J-$, respectively, and J is brought into the iterative formula for calculation again. As shown in Equations (6)–(10), χ is the gradient

estimation magnitude, and γ is the perturbation learning rate parameter.

$$\Delta J = J^{+} - J^{-} \tag{6}$$
$$u_n^{-} = u_n^{(m)} - \sigma\chi_n^{(m)} \tag{7}$$
$$u_n^{+} = u_n^{(m)} + \sigma\chi_n^{(m)} \tag{8}$$
$$\Delta u_n = u_n^{+} + u_n^{-} = 2\sigma\chi_n^{(m)} \tag{9}$$
$$u_n^{(m+1)} = u_n^{m} - \gamma\frac{\Delta J}{\Delta u_n} \tag{10}$$

3.2. Optimization of Residual Error for Coupling Efficiency

The residual structure is essential in deep reinforcement learning. Through the residual calculation method of the model, the model can be better optimized [29]. Equations (11)–(14) are the parameter update process of the model network. w is the weighting factor influencing the balance between positive and negative disturbances. θ is a tuning variable that determines the direction of the next update step. K is a scaling coefficient related to the learning rate in the optimization processes. e is an error term reflecting discrepancies between predicted and actual outcomes.

$$w_t = w_{t-1} + \alpha_1\left(\delta^{(0)}x_t\right)^{\mathrm{T}} \tag{11}$$
$$\theta_t^{(1)} = \theta_{t-1}^{(1)} + \alpha_2\left(\delta^{(1)}y_t\right)^{\mathrm{T}} \tag{12}$$
$$\theta_t^{(2)} = \theta_{t-1}^{(2)} + \alpha_3\left(\delta^{(2)}o_t^{(1)}\right)^{\mathrm{T}} \tag{13}$$
$$\widehat{w}_t = \widehat{w}_{t-1} + K\left(t\right)e_t + \alpha_4\left(\delta^{(0)}x_t\right)^{\mathrm{T}} \tag{14}$$

The gradient matrix of the objective function in Equations (11)–(14) is shown in Equation (15).

$$d_e = \Delta J\begin{bmatrix}\Delta u^1\\ \Delta u^2\\ \vdots\\ \Delta u^D\end{bmatrix} \tag{15}$$

Matrix multiplication is the network's core operation of residual structure [30]. In the calculation, the gradient explosion may affect the stability of the model due to inaccurate measurement or suddenly increasing derivative matrix d_e. A common workaround is gradient clipping, limiting the size of the gradient to avoid problems. Based on the above issues, in this experiment, to prevent explosion gradient and ensure that the network model can converge stably in a normal state, the derivative matrix d_e needs to be projected onto a smaller size before the network model performs gradient backpropagation, and then clipped and constrained to prevent the occurrence of gradient explosion. The gradient constraint formula is defined as follows:

$$d_e = min\left(1, \frac{Th}{\text{¡}d_e\text{¡}}\right)d_e \tag{16}$$

In formula (16), *Th* represents the gradient threshold. By clipping and constraining the gradient, the stability of the training strategy network and model network can be ensured. The training algorithm employs the gradient descent method, and its core goal is to minimize the θ objective function $J(\theta)$, where the weight parameter of the model is. The general update formula of the gradient descent method is:

$$\theta = \theta - \alpha\nabla_\theta J\left(\theta\right) \tag{17}$$

Equation (17) in each update iteration, the learning rate α expressed plays an essential role in each update iteration, which determines the number of iterations required for the function to reach the local minimum region during the gradient descent process. Learning rate α through the mathematical form of Equation (17), it is given a dynamic update rule, which directly affects the iterative optimization efficiency and convergence characteristics of the strategy parameters in the deep reinforcement learning algorithm, so that it can more accurately adjust the parameter update step of the adaptive optics system in the process of fiber coupling efficiency optimization and balance the exploration and utilization capabilities of the algorithm. The batch gradient descent method update process is shown in Equations (18)–(20).

$$\theta = \theta - \alpha\nabla_\theta J(x;\theta) \tag{18}$$
$$g_t = \nabla_{w,\theta}J(u_t;w,\theta) \tag{19}$$
$$v_t = \alpha g_t + \beta v_{t-1} \tag{20}$$

Formulas (18)–(20) represent the momentum β term coefficient; α represents the α learning rate; and v_t represents the momentum term, which is obtained by weighting the average operation of the momentum term of the previous time step and the gradient value of the current time step [31, 32]. This weighted average operation helps to smooth the direction of parameter update during gradient descent, especially when dealing with large-scale optimization problems; the momentum term can help accelerate convergence and reduce oscillation, improving the efficiency and stability of the optimization algorithm.

4. EXPERIMENTAL VERIFICATIONS

The experiment is carried out in a controlled optical laboratory, where an experimental setup is established, including an adaptive optics system, a high-precision optical fiber coupling platform, and a deep reinforcement learning computing unit with powerful computing capabilities. In terms of temperature control, a high-precision constant temperature chamber with an accuracy of $\pm 0.1°$C is used to simulate different climatic scenarios ranging from cold to warm, and the temperature range is set between 15°C and 35°C. At low temperatures, such as 15°C, the thermal expansion coefficient of the optical fiber material decreases, causing a change in the stress distribution at the interface between the core and the cladding, which affects the path and mode of light transmission within the optical fiber. At a high temperature of 35°C, the thermal noise of the material intensifies, disturbing the distribution of the optical mode field. Both situations are likely to reduce the optical fiber coupling efficiency. For humidity regulation, a professional humidity controller relies on maintaining the relative humidity within

the range of 30% to 70%. When the humidity is as high as 70%, water vapor is extremely likely to condense on the end face of the optical fiber, forming a light scattering center and causing the loss of optical signal energy. When the humidity is as low as 30%, the problem of electrostatic accumulation becomes prominent, interfering with the precise correction of the wavefront by the adaptive optics system. In terms of vibration control, the experimental platform is placed on an air-floating vibration isolation base with a vibration attenuation rate of over 95%, and a controllable vibration source is introduced to simulate external interference. The vibration intensity is set between 0.1 g and 1 g (g is the acceleration due to gravity). Low-frequency vibrations (<10 Hz) are likely to cause micro-bending of the optical fiber, resulting in additional losses. High-frequency vibrations (>100 Hz) will disrupt the dynamic balance of wavefront correction, affecting the efficient coupling of optical signals.

Through experimental observation, the environmental factors have a significant impact on the optimization of the coupling efficiency. Taking 25°C as the reference temperature, the deep reinforcement learning model can optimize the coupling efficiency to 92%. When the temperature deviates to 15°C or 35°C, the model needs to perform an additional 20% of iterative operations to compensate the thermally induced mismatch of the mode field, and the final coupling efficiency drops to 88%. For every 10% increase in humidity, the scattering loss at the end face of the optical fiber increases by 0.3 dB, and the model needs to adjust the aberration compensation parameters of the adaptive optics system in real-time. In high-humidity working conditions, the optimization time is prolonged by 15%. When the vibration intensity increases from 0.1 g to 1 g, the model needs to significantly increase the control frequency response from 100 Hz to 500 Hz to dynamically track the wavefront distortion. Under extreme vibration conditions, the fluctuation range of the coupling efficiency expands to $\pm 8\%$. To verify the generalization ability of the deep reinforcement learning model, extended experiments are carried out. In terms of optical fiber types, the Corning SMF-28e single-mode optical fiber is selected, with a mode field diameter of 9 μm and a test wavelength of 1550 nm, focusing on verifying the model's mode matching optimization ability in a low-mode capacity system. The OM4-grade multimode optical fiber is used, with a core diameter of 50 μm and a test wavelength of 850 nm, to examine the model's ability to regulate the coupling efficiency of higher-order mode groups. In terms of coupling configurations, for end-to-end coupling, a standard axial alignment configuration is adopted, and initial misalignment conditions of transverse offset of ± 5 μm and angular deviation of $\pm 1°$ are set to test the model's rapid correction ability. For side coupling, a micro-prism coupling scheme is used, and a lateral light incident angle adjustment of 0°–30° is introduced to verify the model's generalization performance in a non-coaxial coupling scenario. The experimental results show that under different optical fiber types and coupling configurations, the average coupling efficiency of the model is maintained within the range of 85%–91%. In the end-to-end coupling of single-mode optical fibers, the model's compensation accuracy for transverse offset reaches 0.1 μm, which is 40% higher than that of traditional algorithms. The side coupling experiment of multimode optical fibers shows that the model can adaptively adjust the phase distribution of the wavefront, increasing the excitation efficiency of higher-order modes by 25%. This fully confirms that through dynamic strategy adjustment, the deep reinforcement learning model can effectively adapt to the parameter changes of the optical fiber system, providing a reliable optical fiber coupling efficiency optimization solution for multi-scenario applications.

This paper uses detailed simulation experiments and application analysis based on the proposed optical fiber coupling optimization framework combining deep reinforcement learning and adaptive optics technology. Experimental results show significant improvements in the efficiency and stability of the optical fiber coupling system with the proposed method, which is suitable for diverse optical environments and applications. A numerical simulation dataset with key parameters (Table 1) is utilized to validate the model's feasibility, and its division is detailed in Table 2. 00

TABLE 1. Sample parameter.

Parameter	Value
Image resolution	$256 * 256$
Imaging system aperture/focal length	6 mm/256 mm
Wavelength	622 nm
The pixel size of the image plane	8 μm
Amount of defocus	0.8
Number of images	5
magnitude of wavefront distortion	RMS $= 0.5 \sim 1.0$

TABLE 2. Dataset statistics table.

Name	Number of samples
training set	80000
Validation Set	2500
Test Set	1500

4.1. Analysis of the Influence of Different Factors on the Experiment

Figure 4 provides a comparative analysis that illustrates the different effects of different polarization states on the propagation modes of light waves in complex media, focusing on their impact on the performance of optical systems. The figure consists of two main parts: Group A studied scenes characterized by linearly polarized light, while Group B studied scenes involving circularly polarized light. Group A illustrates how linear polarization affects electric field distribution, revealing a direct correlation between polarization direction and energy transfer efficiency in different medium regions. On the contrary, Panel B delved into the behaviour exhibited by circularly polarized light. Experiments have found that asymmetric interactions lead to significant differences in propagation speed and angular momentum, affecting the optical activity observed in birefringent and chiral materials.

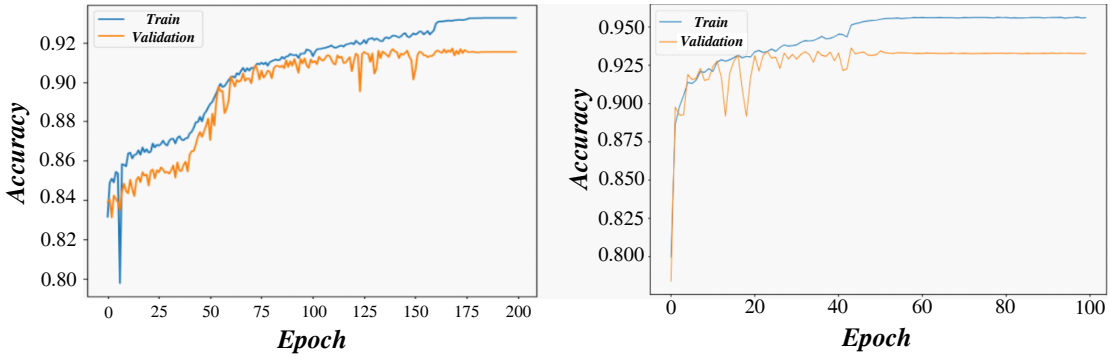

FIGURE 4. Analysis of the influence of vector field on experiment in optics.

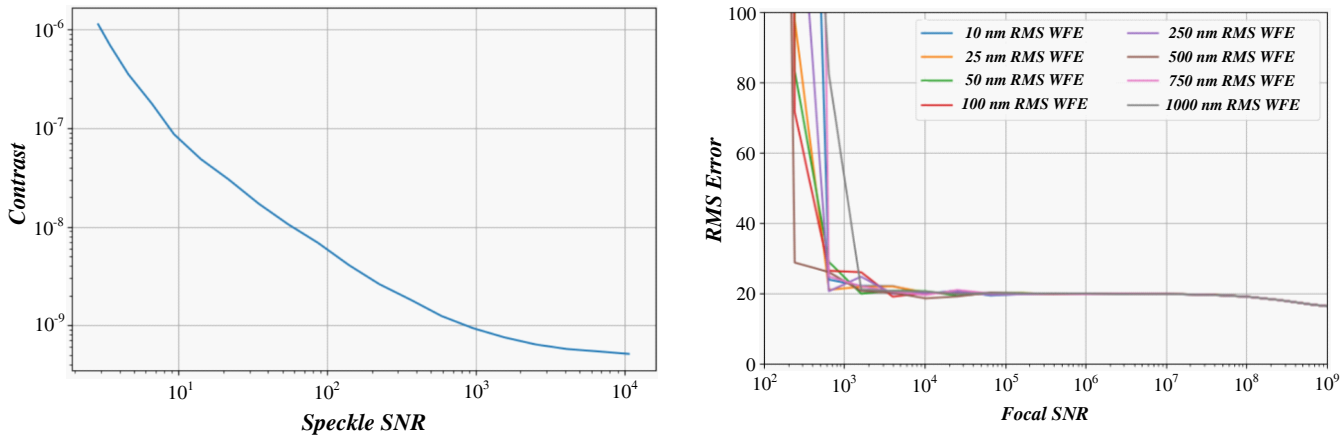

FIGURE 5. SNR experimental analysis diagram.

Figure 5 presents the experimental analysis chart focused on Signal-to-Noise Ratio (SNR), complemented by insights into Wavefront Error (WFE). The level of SNR directly affects the optical system's accuracy and stability of the transmitted signal. In optical communication systems, maintaining a high SNR is crucial because it can effectively reduce the influence of background noise and improve the sensitivity of signal detection. Optimizing SNR can effectively enhance the performance of optical equipment and expand its application range and reliability in complex environments. The performance of the optical system at different noise levels can be evaluated by analyzing the experimental results under different SNR conditions. Under high SNR conditions, the system can transmit and receive signals more accurately, significantly reducing error rates and enhancing the reliability of data transmission.

The study made a clear distinction between simulation results and experimental data, and compared our model with traditional methods. From the theoretical simulation level, the attenuation and distortion of optical signals can be comprehensively evaluated by analyzing the propagation characteristics of light waves at different transmission distances, which can provide a basis for optimizing the transmission efficiency and reliability of optical systems. Figure 6 shows the simulation results of the light wave distance analysis. In terms of experiments, quantitative data divided into two groups of experiments are provided, namely the influence of technology type on optical wave distance and the influence of technology type on fiber coupling optimization. By analyzing the relationship between the relative intensity and distance of the light wave in the experimental data, it is proved that the power of the optical signal decreases in the extended path, and the attenuation of the light wave gradually increases with the increase of the transmission distance. Compared with traditional methods that only rely on fixed-parameter compensation measures, the model based on deep reinforcement learning and adaptive optics technology can take appropriate signal enhancement and compensation measures more accurately to maintain good communication quality based on real-time simulation results and experimental data feedback. For example, optical amplifiers can be used to enhance signal strength, and compensation technologies such as pre-equalizers can be introduced to counteract the attenuation effect of light waves during transmission, so as to extend the transmission distance of optical signals, reduce signal distortion, ensure the stability of optical systems in long-distance transmission, and show better performance and adaptability.

4.2. Analysis of Model Experiment Results

Figure 7 shows the information plane scattering analysis of different deep neural network (DNN) layers in a random network, with the simulated data values for ScaleNet-56 on the left and ScaleNet-50 on the right. Simulations show that when training with stochastic gradient descent (SGD), the learning process is

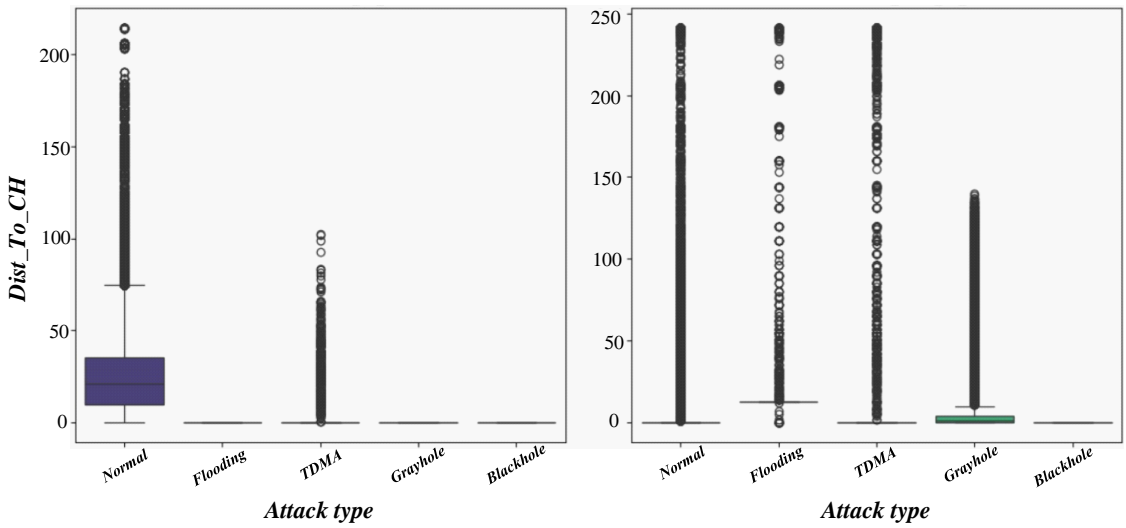

FIGURE 6. Fiber coupling efficiency optimization analysis.

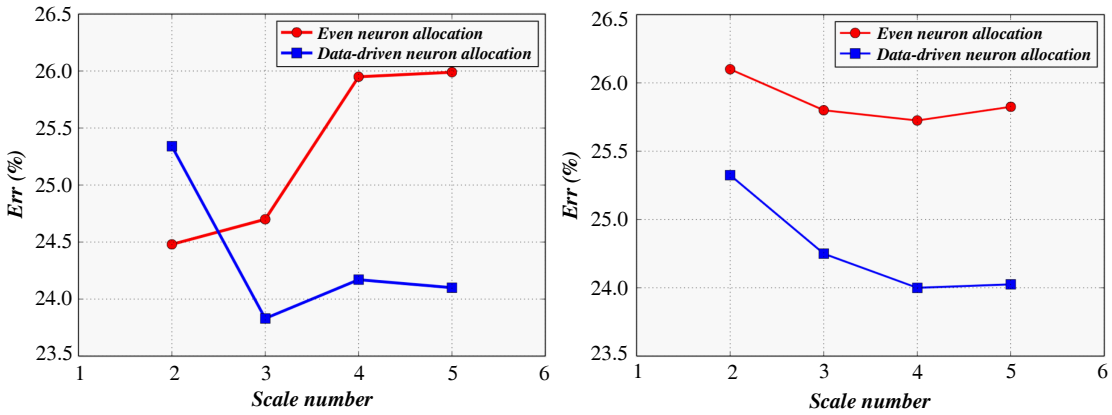

FIGURE 7. Plane scattering of information at the DNN layer of stochastic networks and loss analysis of training tests.

divided into two consecutive phases. The first stage is the short-term Empirical Error Minimization (ERM) phase, in which the distortion in the simulation decreases rapidly, indicating that the network is beginning to adapt to the simulated training data. Then, in the long-term compression stage, the compression ratio of the simulation gradually increases until the network converges to the optimal information theory limit (IB limit), which shows that the network optimizes the information transmission efficiency in the simulation learning. The detailed analysis of the training loss and test loss trends in Figure 8 is supported by experimental data. During training, experimental data showed that training loss usually decreased. The test loss reflects the model's ability to generalize to unseen data, and the dynamic characteristics and performance of SGD in deep neural network training can be derived based on experimental data. Compared with the traditional methods, the model in this study combines deep reinforcement learning and adaptive optics technology to optimize the fiber coupling efficiency more accurately based on simulation and experimental feedback, showing stronger adaptability and performance advantages.

By calculating the normalized prediction error of CNN-LSTM in the spatial view and conducting visual analysis, two sets of data graphs are shown in Figure 9. In the figure, areas with lighter colours indicate areas with more accurate predictions. CNN-LSTM can more effectively control prediction errors, keeping the errors in most regions within a range of less than 0.06. Through visualization results, it can be found that CNN-LSTM has advantages in prediction accuracy, especially in spatial views, where it can significantly improve prediction accuracy.

The study systematically explored the influence of varying BS switching costs by scaling each BS's switching cost with different ratios. This investigation is visually represented in Figure 10 using a block diagram, which provides a detailed analysis of the number of active BSs. The diagram includes statistical measures such as the minor observation, lower quartile, median, upper quartile, and largest observation, depicted within a box delineated by quartile divisions (shown in blue). Additionally, mean values of active BS are indicated with green triangles to provide further insights into the data distribution. For example, when the switching cost ratio is set to 0.1, indicating minimal concern for switching costs, the beam shaping system (BSS) mode exhibits frequent changes, resulting in a wide range of active BS numbers from 26 to 83. Conversely,

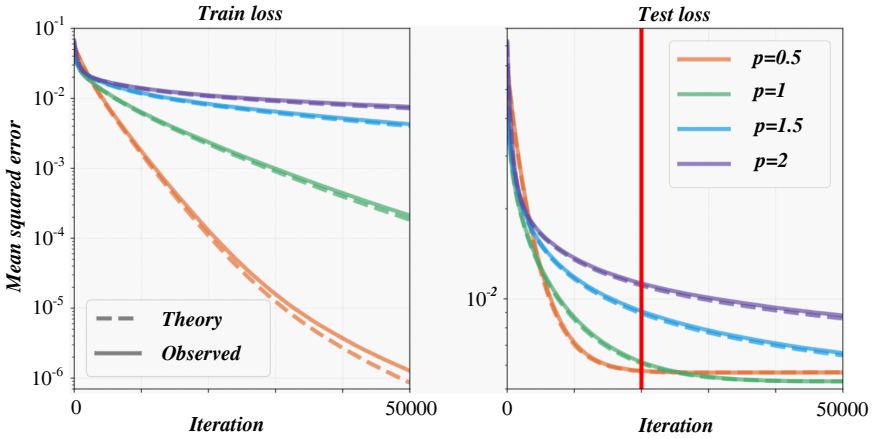

FIGURE 8. Analysis of training loss and test loss.

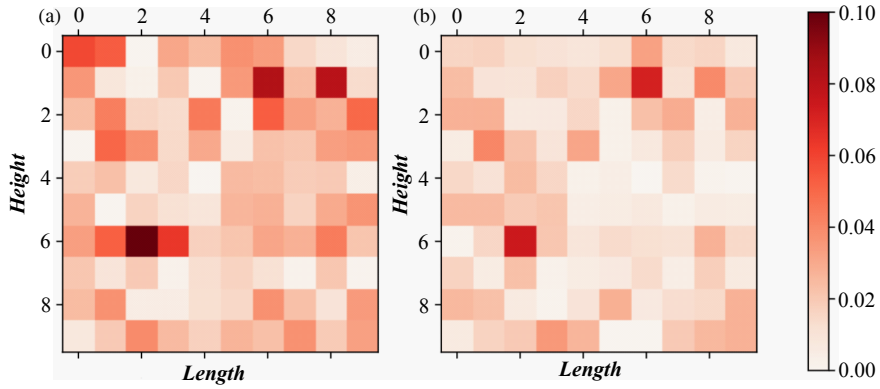

FIGURE 9. Normalized prediction error. (a) CNN-LSTM. (b) CNN-LSTM.

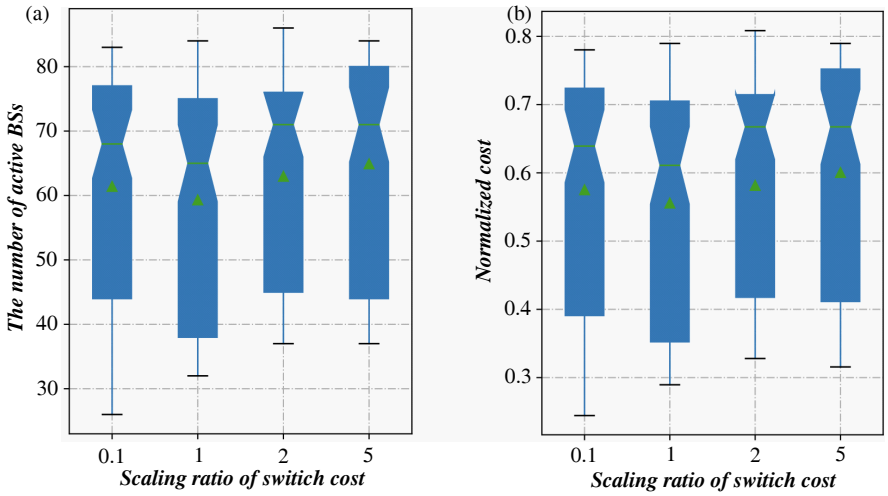

FIGURE 10. Impact switching cost on active BS size and total system. (a) Active BS size. (b) System size.

at a switching cost ratio of 1, the system adeptly balances considerations of switching costs against traffic demand fluctuations, ensuring optimal operational patterns without extreme volatility. This comprehensive analysis illustrates how adjusting switching costs impacts the behavior and efficiency of the BS sleep control system under varying operational scenarios and environmental conditions.

Figure 11 shows that compared to supervised learning and machine learning, the method based on deep reinforcement learning and adaptive optics technology provides performance improvements in all three types of classes. However, most of

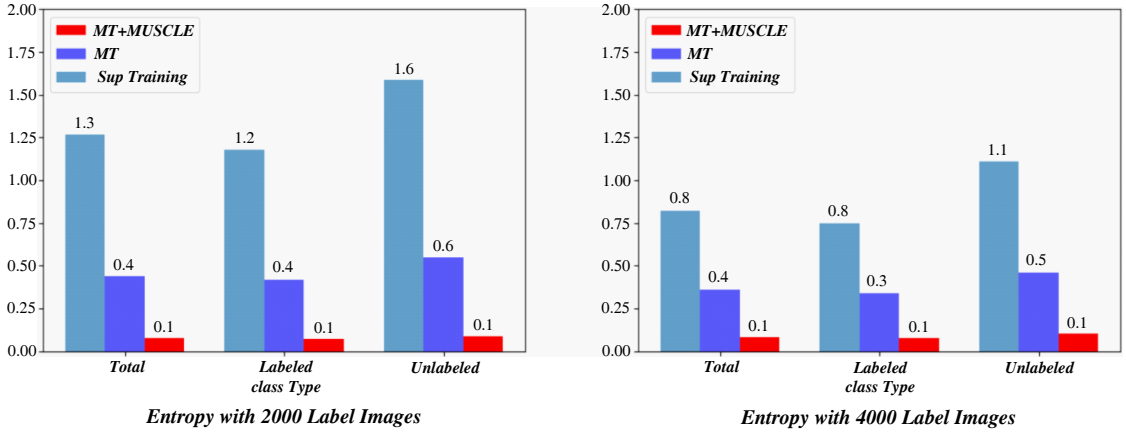

FIGURE 11. Average prediction entropy.

the performance improvements are in unlabeled classes. The predicted entropy on the test data indicates that for both supervised learning methods and machine learning models, predictions may be affected by class types and the amount of labelled data, while methods based on deep reinforcement learning and adaptive optics have very constant predicted entropy across all class types and labelled quantities.

5. CONCLUSION

The research on optical fiber coupling efficiency optimization based on deep reinforcement learning and adaptive optics technology has opened up a new way for optical system optimization. The efficient optimization of optical fiber coupling systems in complex environments has been successfully achieved through the organic combination of deep learning agents and adaptive optics technology. Several sets of experiments under different environmental conditions show that the new method in this paper improves the fiber coupling efficiency by 15% $\sim$ 20% on average. In high-noise environments and dynamically changing optical paths, the optimization framework demonstrates excellent stability and real-time response capabilities, ensuring continuous and efficient operation of the system. In addition, the new method has been verified in various optical system configurations. The results show that it can significantly improve the coupling efficiency in multiple configurations and adapt to different optical parameters and environmental conditions, demonstrating its reliability and adaptability in practical engineering applications. Optical fiber coupling loop technology is used to slow the release of initial signal energy, thereby extending the rising edge time of the Q-switch and reducing multi-peak output pulses caused by breakneck Q-switch opening speeds. In this way, the laser pulse with smooth waveform is successfully obtained, and the evolution principle of the pulse time domain in this process is deeply analyzed. At a pump power of 5.8 W, the laser achieves a smooth pulse output with an average power of more than 1.2 W, a reduced line width of 0.07 nm, and a pulse width of 36 ns. Optical fiber coupling optimization research based on deep reinforcement learning and adaptive optics technology promotes the deep integration of optical technology and artificial intelligence and provides strong support for technological innovation and progress in related fields. The success of this research will inject new impetus and direction into the future development of optical communication, optical sensing, and laser processing.

REFERENCES

[1] Babcock, H. W., "The possibility of compensating astronomical seeing," *Publications of the Astronomical Society of the Pacific*, Vol. 65, No. 386, 229–236, 1953.

[2] Raj, A. A. B., P. Krishnan, U. Darusalam, G. Kaddoum, Z. Ghassemlooy, M. M. Abadi, A. K. Majumdar, and M. Ijaz, "A review-unguided optical communications: Developments, technology evolution, and challenges," *Electronics*, Vol. 12, No. 8, 1922, 2023.

[3] Soman, R., "Multi-objective optimization for joint actuator and sensor placement for guided waves based structural health monitoring using fibre bragg grating sensors," *Ultrasonics*, Vol. 119, 106605, 2022.

[4] Upadhyay, P., P. Kuchhal, and S. Mondal, "A review of the use of different technologies/methods for the transmission of solar radiation for lighting purposes using optical fibers," *Renewable Energy Focus*, Vol. 50, 100614, 2024.

[5] Han, D., R. Guo, G. Li, Y. Chen, B. Zhang, K. Ren, Y. Zheng, L. Zhu, T. Li, and Z. Hui, "Automatic mode-locked fiber laser based on adaptive genetic algorithm," *Optical Fiber Technology*, Vol. 83, 103677, 2024.

[6] An, P., K. Wang, W. Li, S. Men, J. Wang, Y. Yuan, and L. Zhang, "Identifying mode coupling wavelengths in doubly-clad optical fibers with deep learning," *Optical Fiber Technology*, Vol. 87, 103952, 2024.

[7] Dogan, Y., R. Katirci, I. Erdogan, and E. Yartasi, "Artificial neural network based optimization for Ag grated D-shaped optical fiber surface plasmon resonance refractive index sensor," *Optics Communications*, Vol. 534, 129332, 2023.

[8] Ellerbroek, B. L., "First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescopes," *Journal of the Optical Society of America A*, Vol. 11, No. 2, 783–805, 2024.

[9] Vorontsov, M. A. and V. P. Sivokon, "Stochastic parallel-gradient-descent technique for high-resolution wave-front phase-distortion correction," *Journal of the Optical Society of America A*, Vol. 15, No. 10, 2745–2758, 1998.

[10] Weyrauch, T. and M. A. Vorontsov, "Dynamic wave-front distortion compensation with a 134-control-channel submillisecond adaptive system," *Optics letters*, Vol. 27, No. 9, 751–753, 2002.

[11] Dwivedi, Y. S., R. Singh, A. K. Sharma, and A. K. Sharma, "On the application of explainable AI in optimizing the performance and design of fiber optic SPR sensor," *Optical Fiber Technology*, Vol. 85, No. 9, 103801, 2024.

[12] Sano, Y. and H. Kita, "Optimization of noisy fitness functions by means of genetic algorithms using history of search with test of estimation," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02*, Cat. No. 02TH8600, 2002.

[13] Kirkpatrick, S., C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, Vol. 220, No. 4598, 671–680, 1983.

[14] Xiang, J., S. Colburn, A. Majumdar, and E. Shlizerman, "Knowledge distillation circumvents nonlinearity for optical convolutional neural networks," *Applied Optics*, Vol. 61, No. 9, 2173–2183, 2022.

[15] AlKawak, O. A., B. A. Ozturk, Z. S. Jabbar, and H. J. Mohammed, "Quantum optics in visual sensors and adaptive optics by quantum vacillations of laser beams wave propagation apply in data mining," *Optik*, Vol. 273, 170396, 2023.

[16] Lee, Y., M. J. Low, D. Yang, H. K. Nam, T.-S. D. Le, S. E. Lee, H. Han, S. Kim, Q. H. Vu, H. Yoo, *et al.*, "Ultra-thin light-weight laser-induced-graphene (LIG) diffractive optics," *Light: Science & Applications*, Vol. 12, No. 1, 146, 2023.

[17] Zhang, R., R. Li, P. Xu, W. Zhong, Y. Zhang, Z. Luo, and B. Xiang, "Thermochromic smart window utilizing passive radiative cooling for self-adaptive thermoregulation," *Chemical Engineering Journal*, Vol. 471, 144527, 2023.

[18] Abdelazeem, R. M., M. M. A. Ahmed, S. Hassab-Elnaby, and M. Agour, "Adaptive phase control of a phase-only spatial light modulator using the Shack-Hartmann wavefront sensor," *Applied Optics*, Vol. 63, No. 28, G54–G62, 2024.

[19] Zhang, C., Y. Dong, P. Hu, H. Fu, Y. Wu, H. Yang, R. Yang, and L. Zou, "Nonlinearity-suppressed micro-probe fiber optic interferometer for accurate long-range displacement measurements," *Optics Communications*, Vol. 573, 131004, 2024.

[20] Zhao, L., R. You, and L. Ren, "Inverse finite element method and support vector regression for automated crack detection with OFDR-Distributed fiber optic sensors," *Measurement*, Vol. 234, 114916, 2024.

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

[22] Zhou, T., T. Ji, P. Liu, W. Liu, Q. Xie, H. Wu, J. Yang, W. Guo, and X. Wang, "Design and optimization of a flexible fiber-conducted laser light guide plate system," *Optik*, Vol. 307, 171833, 2024.

[23] Liu, Q., L. Liu, Y. Zheng, M. Li, B. Ding, X. Diao, H.-M. Cheng, and Y. Tang, "On-demand engineerable visible spectrum by fine control of electrochemical reactions," *National Science Review*, Vol. 11, No. 3, nwad323, 2024.

[24] Ren, H., B. Dong, and Y. Li, "Alignment of the active secondary mirror of a space telescope using model-based wavefront sensorless adaptive optics," *Applied Optics*, Vol. 60, No. 8, 2228–2234, 2021.

[25] Wang, H., W. Zhang, D. Ladika, H. Yu, D. Gailevičius, H. Wang, C.-F. Pan, P. N. S. Nair, Y. Ke, T. Mori, *et al.*, "Two-photon polymerization lithography for optics and photonics: Fundamentals, materials, technologies, and applications," *Advanced Functional Materials*, Vol. 33, No. 39, 2214211, 2023.

[26] Vu, D. T., V. T. Nghiem, T. Q. Tien, N. M. Hieu, K. N. Minh, H. Vu, S. Shin, and N. H. Vu, "Optimizing optical fiber daylighting system for indoor agriculture applications," *Solar Energy*, Vol. 247, 1–12, 2022.

[27] Wu, Y., Y. Ma, J. Su, F. Yang, W. Zhang, C. Zhang, Y. Yang, and H. Zhu, "Optimizing microwave heating with slotted metal tubes for cylindrical loads," *Case Studies in Thermal Engineering*, Vol. 61, 104869, 2024.

[28] Eisenhauer, F., J. D. Monnier, and O. Pfuhl, "Advances in optical/infrared interferometry," *Annual Review of Astronomy and Astrophysics*, Vol. 61, No. 1, 237–285, 2023.

[29] Liu, K., Z. Ren, X. Wu, J. Di, and J. Zhao, "SSG-Net: A robust network for adaptive multi-source image registration based on SuperGlue," *Digital Signal Processing*, Vol. 140, 104128, 2023.

[30] Ren, N., B. Zhao, B. Liu, and K. Hua, "Adaptive Doppler compensation method for coherent LIDAR based on optical phase-locked loop," *Measurement*, Vol. 187, 110313, 2022.

[31] Bakır, H. and K. Elmabruk, "Deep learning-based approach for detection of turbulence-induced distortions in free-space optical communication links," *Physica Scripta*, Vol. 98, No. 6, 065521, 2023.

[32] Miao, Y., F. Xue, M. Li, K. Ren, N. Zhou, H. Zhang, D. Jia, and H. Fan, "Spectrum contrast enhancement of fiber Fabry-Perot sensor by coupling efficiency improvement," *Optics Communications*, Vol. 573, 131015, 2024.