International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 4423-4434

https://doi.org/10.70517/ijhsa464288

Reinforcement learning algorithm-driven energy transfer loss suppression and efficiency optimization strategy for high-power laser systems

Juanhui Ren[1,*] and Qin Liu[2]
[1] Chengdu Aeronautic Polytechnic, Chengdu, Sichuan, 610100, China
[2] Chengdu Guangxunda Technology Co., LTD., Chengdu, Sichuan, 610100, China
Corresponding authors: (e-mail: 13880087132@163.com).

Abstract Conventional laser systems have disadvantages such as high energy loss and low transmission efficiency, which need to be optimized and improved by new methods. In this paper, a deep reinforcement learning (RL)-based energy transmission loss suppression and efficiency optimization method for high-power laser systems is proposed. First, the attenuation mechanism of laser transmission in the atmosphere is analyzed and the corresponding thermodynamic model is established. Then, the A-TD3 algorithm in deep reinforcement learning is used to optimize the energy transmission efficiency of the laser system. Simulation results show that the A-TD3 algorithm has better convergence under different learning rates, and the algorithm converges within 150 rounds at a learning rate of 0.0005 and improves the average energy transfer efficiency of the laser system to 9.7. Compared with the traditional algorithms (e.g., DQN, DDPG, and TD3), the A-TD3 algorithm has faster convergence speed and higher transfer efficiency (9.3 vs. 9.7). In addition, the energy transfer loss of the system is optimized to reduce up to 30%-70% compared to the unoptimized system. These results demonstrate the potential application of deep reinforcement learning in the optimization of high-power laser systems. By this method, not only the loss in the transmission process can be effectively reduced, but also the overall efficiency of laser energy transmission can be improved.

Index Terms High-power laser, Energy transmission, Deep reinforcement learning, A-TD3 algorithm, Optimization, Transmission efficiency

I. Introduction

Currently, high-power laser system has become one of the high points of national defense strategy, frontier science and technology and emerging industry, which plays a significant role in national security and national economic construction [1], [2]. Large-caliber high-performance reflective films, high-precision beam-splitting films and high-performance laser spectroscopic combining mirrors are three types of key components for controlling the optical path transmission of laser systems [3], [4]. With the technological breakthrough of high-power solid-state lasers and the increase of output power, laser systems have put forward more demanding requirements on the comprehensive performance of these three types of thin-film devices [5], [6].

Firstly, thin film devices need to have high spectral efficiency to ensure efficient transmission of the laser beam [7]. Secondly, thin film devices need to have good resistance to laser damage to ensure the long-term stability of the system [8]. Finally, thin-film devices also need to maintain a low temperature rise in service to ensure that the quality of the laser beam does not deteriorate [9]. The current research has some knowledge gaps and limitations in addressing the problem that the three devices cannot simultaneously achieve high spectral efficiency, low temperature rise and high damage threshold [10], [11]. At the same time, high-power laser systems have different power density requirements for the laser far-field according to their different uses, and the system energy transmission loss suppression and efficiency optimization is to establish the system performance optimization objective function under the premise of satisfying the performance requirements of the laser far-field while considering various design constraints, such as system development cost, volume, weight, power consumption, etc., and finally obtain the performance parameter set of the system through trade-off optimization [12]-[15].

This paper presents an optimization method based on deep reinforcement learning. Different from traditional optimization methods, deep reinforcement learning algorithms can automatically find the optimal strategy in the face of complex and dynamic environments through simulation and self-tuning. In this study, the A-TD3 algorithm, as an advanced algorithm in reinforcement learning, is applied to the optimization of a high-power laser system by virtue of its advantages in deep reinforcement learning. The A-TD3 algorithm adopts an asynchronous updating mechanism, which effectively solves the problems of over-estimation bias and slow convergence of traditional

algorithms in the training process. Through this method, the optimization of the laser system energy transmission process can be realized, significantly improving the transmission efficiency and reducing the energy loss. Specifically, the attenuation model of laser transmission in the atmosphere is first analyzed, and a deep reinforcement learning framework is designed based on this model. Next, through simulation experiments, the optimization effect of the A-TD3 algorithm is verified under different learning rates and different algorithm comparisons, and the best optimization strategy for laser energy transmission efficiency is finally derived. By comparing with traditional algorithms, this paper demonstrates the advantages of reinforcement learning algorithms in high-power laser energy transfer and proposes a new idea to improve the performance of existing laser systems.

II. Methodology

II. A.Beam Energy Transfer Losses in Laser Systems

The transmission of laser beams in the atmosphere is affected by both linear and nonlinear effects, leading to beam energy loss and beam quality degradation. The linear effects include atmospheric refraction, absorption, scattering and turbulence, while the nonlinear effects include thermal haloing, excited Raman scattering and breakdown.

II. A. 1) Laser atmospheric transmission transmittance

According to the Lambert-Beer law, the expression for the laser atmospheric transmission transmittance is:

$$T_{atm}(\lambda) = \exp\left(-\int_0^L \beta(\lambda) dl\right) \tag{1}$$

Here, $T_{atm}(\lambda)$ is the transmission transmittance of laser light of wavelength λ in the atmosphere.

L is the spatial transmission distance of the laser beam.

$\beta(\lambda)$ is the total attenuation coefficient of the laser beam, which consists of two processes: absorption and scattering:

$$\beta = \beta_a + \beta_m = A_a + S_a + A_m + S_m \tag{2}$$

Here, A_a and S_a are the absorption and scattering coefficients of aerosol particles.

A_m and S_m are the absorption and scattering coefficients of atmospheric molecules.

II. A. 2) Absorption and Scattering of Atmospheric Molecules

The gaseous composition of the atmosphere is very complex, and to simplify the arithmetic, only a few gas molecules with the largest share are considered. Only the absorption coefficients of seven gas molecules, namely, water vapor, carbon dioxide, ozone, nitrous oxide, carbon monoxide, methane and oxygen, are calculated. The total absorption coefficient of atmospheric molecules is the sum of the absorption coefficients of these seven molecules:

$$A_m = \sum_{i=1}^{7} \alpha_m^i \tag{3}$$

Similar to the absorption coefficient, the scattering coefficient of atmospheric molecules is likewise the sum of the scattering coefficients of the various molecules:

$$S_m = \sum_{i=1}^{K} s_m^i \tag{4}$$

The size of atmospheric molecules is on the order of $10^{-4}\mu m$ According to the law of scattering, Rayleigh scattering occurs when the laser is in the visible and infrared wavelength bands at $(>0.4\mu m)$, where the wavelength is much larger than the particle size. The scattering coefficient depends on the concentration of atmospheric molecules, the refractive index of the atmosphere, and the wavelength of the laser.

II. A. 3) Absorption and Scattering of Aerosol Particles

The size of aerosol particles is larger than atmospheric molecules, $0.1\mu m - 100\mu m$, and according to the Mie scattering theory, the laser attenuation coefficients β_a^i, the absorption coefficients α_a^i, and the scattering coefficients s_a^i of aerosol particles can be calculated:

$$\begin{cases} \beta_a^i = \int_{r_{\min}}^{r_{\max}} \pi r^2 Q_{ext}^i(m^i, r) n^i(r) dr \\ \alpha_a^i = \int_{r_{\min}}^{r_{\max}} \pi r^2 Q_{abs}^i(m^i, r) n^i(r) dr \\ s_a^i = \beta_a^i - \alpha_a^i \end{cases} \tag{5}$$

Here, Q_{ext}^i is the extinction factor, Q_{abs}^i is the absorption factor, $n^i(r)$ is the scale distribution of aerosol particles, m^i is the complex refraction index of the i th aerosol particle.

According to the formula, the total attenuation coefficient of aerosol particles in any weather is the sum of the aerosol attenuation coefficient in clear weather and the aerosol attenuation coefficient in special weather such as cloud and rain.

II. B.Thermodynamic modeling of laser energy converters

Thermodynamics studies heat, work, and temperature and how they relate to energy, entropy, matter, and radiation. Considering the laser energy process, the thermodynamic model of a laser energy converter is shown in Figure 1. The laser energy converter (LPC) receives the energy flux $\dot{E}_L$ and entropy flux $\dot{S}_L$ from the laser source, and emits the energy flux $\dot{E}_C$ and entropy flux $\dot{S}_C$ through radiation (Luminescence).The LPC transfers heat to the surroundings at a rate $\dot{Q}$. Both are at temperature T and the LPC outputs electrical energy at rate $\dot{W}$.

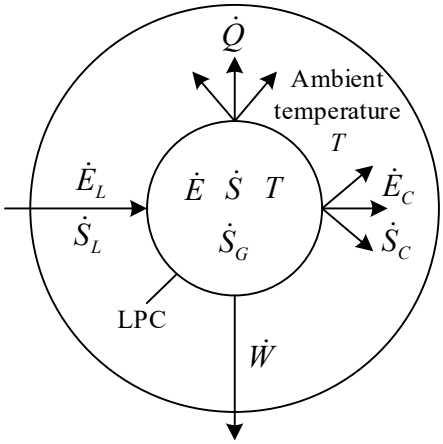

Figure 1: Thermodynamic model of laser energy converter

According to the first and second laws of thermodynamics, the flux balance equations for energy and entropy are given by the following two equations:

$$\dot{E} = \dot{E}_L - \dot{W} - \dot{Q} - \dot{E}_C \tag{6}$$

$$\dot{S} = \dot{S}_L - \dot{Q} / T - \dot{S}_C + \dot{S}_G \tag{7}$$

where $\dot{E}$, $\dot{S}$ and $\dot{S}_G$ are the energy rate of change, entropy rate of change and entropy generation rate in the LPC, respectively. Considering the steady state $\left(\dot{E}=\dot{S}=\dot{T}=0\right)$, eliminating $\dot{Q}$ in the flux balance equation gives the output power

$$\dot{W} = \dot{E}_L(1 - T / T_{FL}) - \dot{E}_C - T(\dot{S}_G - \dot{S}_C) \tag{8}$$

where $T_{FL} \equiv \dot{E}_L / \dot{S}_L$ is the effective flux temperature of the laser radiation. Depending on the arrangement of the terms, Eq. (8) can be rewritten in various forms, for example:

$$\dot{W} = (\dot{E}_L - T\dot{S}_L) - (\dot{E}_C - T\dot{S}_C) - T\dot{S}_G \tag{9}$$

According to the second law of thermodynamics, $\dot{S}_G$ is non-negative, so the available energy is at most $(\dot{E}_L - T\dot{S}_L) - (\dot{E}_C - T\dot{S}_C)$, which represents the difference in free energy between the absorbed and emitted radiation. The reason that the form of Eq. (8) is used in the analysis is that it distinguishes between loss terms that dissipate heat and those that do not require heat dissipation. The first term on the right-hand side of Eq. (8) contains the Carnot coefficient $(1-T/T_{FL})$, which represents the efficiency of the Carnot cycle constituted by two systems with temperatures T_{FL} and T. The second and third terms are the radiation loss and entropy loss, respectively. The entropy generation rates $\dot{S}_{Ga}$ during absorbed radiation and $\dot{S}_{Ge}$ during emitted radiation are:

$$\dot{S}_{Ga} = (\dot{E}_L - \mu\dot{N}_L)/T - \dot{S}_L \tag{10}$$

$$\dot{S}_{Ge} = \dot{S}_C - (\dot{E}_C - \mu\dot{N}_C)/T \tag{11}$$

where $\dot{N}_L$ is the particle number flux absorbed by the LPC from the laser, $\dot{N}_C$ is the particle number flux emitted by the LPC itself, and μ is the chemical potential of the photons interacting with the carriers in the LPC. Combining Eqs. (8) and (11) shows that the process of emitting radiation takes away the entropy flux of $\dot{S}_C$ at the same time, and the net effect is to reduce the entropy in the LPC, so the main source of entropy increase is the process of absorbing radiation.

In thermodynamics, the chemical potential μ_i of some particle i is defined by the fundamental thermodynamic relation:

$$dU = \delta Q - \delta W + \sum_{i=1}^{N} \mu_i dN_i \tag{12}$$

where dU is the change in the internal energy U of the system in microelements and dN_i is the change in the number of particles of the i th class in microelements N_i. The physical meaning of the chemical potential is the amount of change in the internal energy of the system due to the addition or removal of a particle. For high power devices, the particles in the system are electrons and holes, and under illumination, electrons and holes have chemical potentials μ_e and μ_h (also known as quasi-Fermi energy levels), respectively, and when electrons flow through the external circuit and return to the system, the energy supplied to the load is the difference in the chemical potentials of the two $\mu = \mu_e - \mu_h$, which is the chemical potential of the photons interacting with the carriers. The chemical potential, the magnitude of which is related to the output voltage V. The main part $\left(\dot{E}_L - \mu\dot{N}_L\right)/T$ of the entropy generation rate $\dot{S}_{Ga}$ due to absorbed radiation can be written as $(E_{laser} - \mu)\dot{N}_L/T$, where E_{laser} is the photon energy of the laser. According to the concept of chemical potential, this part of the entropy generation rate comes from the difference between the photon energy received by the LPC and the energy provided to the load by the carriers therein. At different output voltages V, this energy difference is different and the entropy loss is therefore different. In the microscopic view, the physical process behind the entropy loss is that, under the combined effect of light and applied voltage, the internal carriers of the LPC and the crystal lattice are collided to reach the steady state, part of the energy is dissipated in the form of heat, and the remaining energy is reflected to the macroscopic level and is described by the chemical potential. The term $\left(\dot{E}_C - \mu\dot{N}_C\right)/T$ in the equation is similar to the above analysis, in the process of emitting the radiation, the radiation carries the energy of $\dot{E}_C$, while the energy provided by the carriers in the system $\mu\dot{N}_C$ is smaller than the energy of the radiation, so only the chemical potential is described. is smaller than the energy of the radiation, so the system generates negative heat when only the emitted radiation is considered, a phenomenon also known as radiative cooling.

Substituting Eq. (10) and Eq. (11) into Eq. (8) yields:

$$\dot{W} = \mu\left(\dot{N}_L - \dot{N}_C\right) \tag{13}$$

The implication is that the product of the energy (μ) carried by the carriers and the number of effective carriers $\left(\dot{N}_L - \dot{N}_C\right)$. Consider the ideal case where the difference between the chemical potentials of electrons and holes, $\mu = qV$, when the carrier mobility is infinity, contact losses at the electrodes are negligible, and there are no other non-intrinsic losses. Combined with this equation, Eq. (13) can be written:

$$\dot{W} = V \times q\left(\dot{N}_L - \dot{N}_C\right) = V \times I \tag{14}$$

Thus, the thermodynamic model portrays the ideal current-voltage relationship for high power devices.

II. C.Deep reinforcement learning algorithm

The Markov decision-making [16] process is the foundational framework for reinforcement learning [17], enabling reinforcement learning algorithms to analyze and solve sequential decision-making problems.

The Markov decision process contains five key elements, which are:

State: represents the environmental situation in which the intelligent body is located. The state is the basis of the intelligent body's decision-making, and the set composed of all states is called the state space.

Action: the intelligent body chooses an action based on the current state, and each action may lead to a change in the state of the environment and affect the reward obtained by the intelligent body. The set of all actions that an intelligent body can choose is called the action space.

Transfer probability: describes the probability that an intelligent body will transfer from the current state to the next state after taking an action. This probability is usually represented by a conditional probability distribution, i.e., $S \times A \rightarrow (A)$.

Reward: the value of the reward that an intelligent body receives from the environment after taking an action and transferring to the next state. The reward value can be positive, negative or zero and is used to assess how good the action is.

Discount Factor: a value between 0 and 1 used to calculate the long-term reward. The smaller the discount factor, the more the intelligent body focuses on immediate rewards. The larger the discount factor, the more the intelligence focuses on long-term rewards.

The commonly used algorithm in deep reinforcement learning is the DQN algorithm, but when faced with a scenario that requires the output of continuous actions, the DQN algorithm is difficult to achieve ideal results. Therefore, this subsection will be settled based on the Deep Deterministic Policy Gradient algorithm (DDPG) [18]. The DDPG algorithm belongs to the heterogeneous policy deep reinforcement learning algorithm, which also adopts the experience replay mechanism when selecting samples, so drawing on the success of prioritized experience replay in DQN, this experiment uses the prioritized experience replay instead of the original experience pool sampling method when training the DDPG algorithm.

The DDPG algorithm adopts an actor-critic structure and contains a policy network and a value function network with network parameters θ^{μ} and θ^{Q}, respectively. The update process of the strategy network is shown in Eq:

$$\nabla_{\theta^{\mu}} \mu \approx E_{\mu'}^{\pi}\left[\nabla_a Q(s,a \mid \theta^{Q})\Big|_{s=s_t, a=\mu(s_t)} \nabla_{\theta_{\mu}} \mu(s \mid \theta^{\mu})\mid_{s=s_t}\right] \tag{15}$$

$$\theta_{t+1}^{\mu} = \theta_t^{\mu} + \alpha_{\mu} \nabla_{\theta^{\mu}} \mu \tag{16}$$

Here, α_{μ} denotes the learning rate of the strategy network.

The update process of the value function network is shown in Eq:

$$\delta_t = r_t + \gamma Q'\left(s_{t+1}, \mu'(s_{t+1} \mid \theta^{\mu'}) \mid \theta^{Q'}\right) - Q\left(s_t, a_t \mid \theta^{Q}\right) \tag{17}$$

$$\theta_{t+1}^{Q} = \theta_t^{Q} + \alpha_Q \delta_t \nabla_{\theta^{Q}} Q\left(s_t, a_t \mid \theta^{Q}\right) \tag{18}$$

Here, α_Q denotes the learning rate of the value function network, $\theta^{\mu'}$ denotes the parameters of the target policy network, and $\theta^{Q'}$ denotes the parameters of the target value function network.

The update methods of the target network are shown in Eq. respectively:

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \tag{19}$$

$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'} \tag{20}$$

Here, τ denotes the update rate and has a value much less than 1.

II. D. Optimization scheme based on A-TD3 algorithm

II. D. 1) Markov decision-making process

In order to substitute the problem of energy transmission loss suppression and efficiency optimization of high power laser system into deep reinforcement learning algorithm, the above problem is transformed into tuple $\langle S, A, R \rangle$, i.e., state space S, action space A, and reward R.

(1) State space S

The state space represents the state of the environment that the decision maker can be in, including the position of the laser system and the position of the ground nodes:

$$S = \left\{ u^{(t)}, q_n^{(t)} \right\} \tag{21}$$

(2) Action space A

In each state, the intelligence has the option of choosing an action from the possible action space. The action space includes actions for energy transfer from the laser system and data decision actions:

$$A = \left\{ \eta^{(t)}, \omega^{(t)}, v^{(t)}, \lambda_n^{(t)}, f_n^{(t)}, \beta_n^{(t)} \right\} \tag{22}$$

(3) Reward R

The reward function R defines the immediate reward obtained after taking an action a in state s, representing the goodness of the action. The optimization objective of the system is to maximize the energy transfer efficiency of the laser system, so the reward value is defined as:

$$R^{(t)} = \frac{\sum_{n=1}^{N} \beta_n^{(t)} \bar{D}_n}{\zeta_f E_f^{(t)} + \sum_{n=1}^{N} \left(\zeta_C E_{C,n}^{(t)} + \zeta_D E_{D,n}^{(t)} \right)} \tag{23}$$

II. D. 2) A-TD3 algorithm

The A-TD3 algorithm is an asynchronous form of the TD3 algorithm. Although TD3 is an improved version of the DDPG algorithm to solve the problems of over-estimation bias and training instability, it adds to the complexity of the model, and thus requires longer execution time to learn a model with excellent performance. To solve this problem, this paper introduces an asynchronous updating mechanism A-TD3 in the TD3 algorithm to reduce the training time of the model and speed up the convergence.

Figure 2 shows the general framework of the proposed A-TD3 algorithm. In the asynchronous update architecture, multiple intelligences are used to interact with the environment separately and transmit the gradient information to the global network. The global network updates its own parameters based on the gradient information obtained from the intelligences and periodically transmits the parameters to each intelligence. This asynchronous update mechanism improves the convergence speed of the algorithm. The global network in the figure has the same structure as each Agent, a TD3 structure, but with different parameters.

The estimation network of the TD3 algorithm consists of an Actor network and two Critic networks with parameters denoted as ϕ', θ_1 and θ_1, respectively. In addition, each network has an associated target network with parameters denoted as ϕ', θ_1' and θ_2'.

For a given state s, the Actor network of the laser system generates the value $\pi_\phi(s)$ of its action a, and the action a of the laser system is shown in Eq:

$$a = \pi_\phi(s) + \varepsilon, \varepsilon \sim N(0, \sigma) \tag{24}$$

where ε is the noise, obeying a normal distribution with mean 0 and variance σ.

The Critic objective network has two targets to compute $Q_{\theta_i'}\left(s', \pi_{\phi'}(s') + \varepsilon\right)$, $i = 1, 2$, and in order to minimize over-estimation, the target Q values are the two target Critic network outputs are minimized:

$$y = r + \gamma \min_{i=1,2} Q_{\theta_i'}\left(s', \pi_{\phi'}(s') + \varepsilon\right) \tag{25}$$

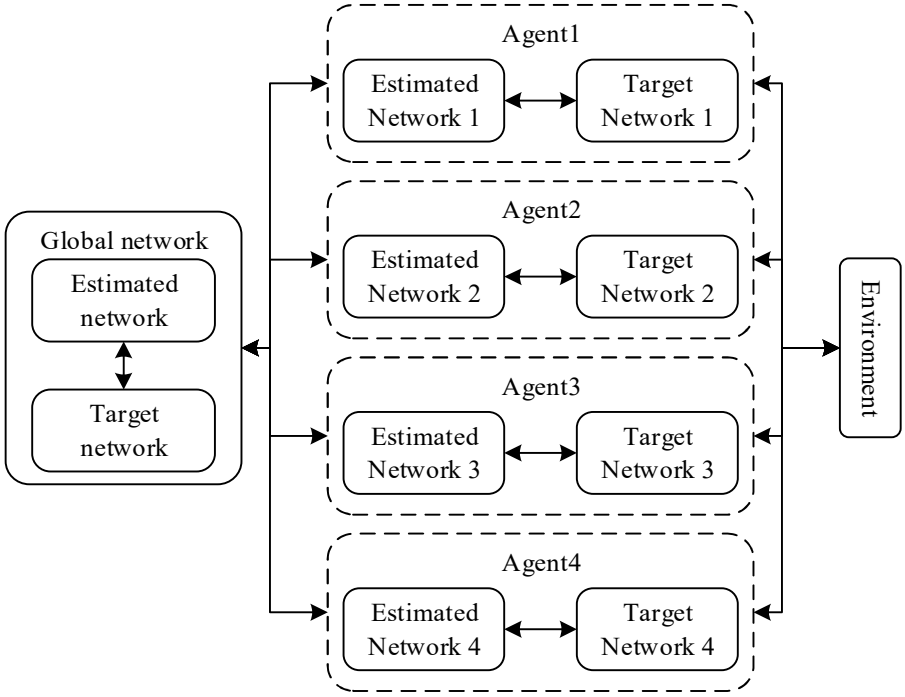

Figure 2: The structure diagram of the A-TD3 algorithm

Using the mean square error loss function, the parameters of the Critic network of the global network are updated asynchronously based on the target Q value and the output of the current Critic network, and then the updated parameters are passed to the corresponding Critic network:

$$L(\theta_i) = \mathrm{E}_{(s,a,r,s')\text{-}buffer}\left[\left(y - Q_{\theta_i}(s,a)\right)^2\right], \quad i = 1,2 \tag{26}$$

where $\mathrm{E}_{(s,a,r,s')\text{-}buffer}[*]$ denotes taking the mean of the data taken from the empirical playback buffer.

Whenever the Critic network is updated N times, the Actor network of the global network is updated asynchronously by gradient ascent to maximize the value of the actions evaluated by the Critic network:

$$\nabla_\phi J(\phi) = \mathrm{E}_{s\text{-}buffer}\left[\nabla_a Q_{\theta_i}(s,a)\big|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s)\right] \tag{27}$$

Pass the updated parameters to the corresponding Actor network. Finally update the parameters of the target network using soft update:

$$\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta_i', \phi' \leftarrow \tau\phi + (1-\tau)\phi' \tag{28}$$

where τ is the soft update factor, which is usually much less than one.

II. D. 3) Actor network design

Since the variables of the objective optimization problem are coupled with each other and difficult to solve, the joint optimization of the coupled action variables is decomposed into several small sets of actions by designing the output layer of the Actor neural network so that the action variables are related to each other in order, and the structure of the Actor network is shown in Fig. 3.

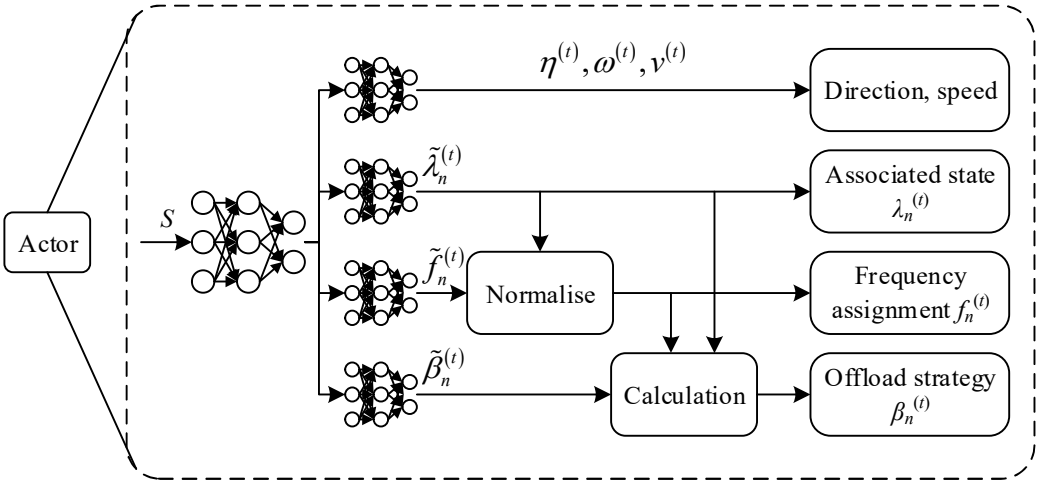

Figure 3: Actor network model

The action space $A=\left\{\eta^{(t)},\omega^{(t)},v^{(t)},\lambda_n^{(t)},f_n^{(t)},\beta_n^{(t)}\right\}$, defines the intermediate variables $\tilde{\lambda}_n^{(t)},\tilde{f}_n^{(t)},\tilde{\beta}_n^{(t)}$, and $\tilde{\lambda}_n^{(t)},\tilde{f}_n^{(t)},\tilde{\beta}_n^{(t)}\in[0,1]$.

To handle the binary association variable $\lambda_n^{(t)}$, the Actor network first outputs a variable $\tilde{\lambda}_n^{(t)}\in[0,1]$. To satisfy the binary association variable constraint, take:

$$\lambda_n^{(t)}=\left\lfloor\tilde{\lambda}_n^{(t)}+0.5\right\rfloor \tag{29}$$

For computational resource allocation, the Actor network generates $\tilde{f}_n^{(t)}\in[0,1]$, which denotes the proportion of CPU computational frequency allocated for ground mobile nodes n. The normalized $f_n^{(t)}$ is obtained by using the optimization done with the correlation variable $\lambda_n^{(t)}$ and keeping the constraints $\sum_{n=1}^{N}f_n^{(t)}\le f_{\max}$ at the optimum:

$$f_n^{(t)}=\frac{\lambda_n^{(t)}\tilde{f}_n^{(t)}}{\sum_{n=1}^{N}\lambda_n^{(t)}\tilde{f}_n^{(t)}}f_{\max} \tag{30}$$

For the offloading strategy $\beta_n^{(t)}$, the strategy is related to the correlation variable $\lambda_n^{(t)}$ and the allocation frequency $f_n^{(t)}$. The Actor network of the laser system generates variables $\tilde{\beta}_n^{(t)}\in[0,1]$. To satisfy the constraints of the offloading strategy, $\beta_n^{(t)}$ can be obtained as follows:

$$\beta_n^{(t)}=\frac{\tau\tilde{\beta}_n^{(t)}}{C\bar{D}_n}\left(\frac{1}{\lambda_n^{(t)}f_n^{(t)}}+\frac{\delta_D}{R_{D,n}^{(t)}}+\frac{\delta_U}{R_{U,n}^{(t)}}\right)^{-1} \tag{31}$$

III. Results and analysis

The software environment used for simulation is: Python 3.7, Tensorflow 1.14.0, Windows 10.

The hardware environment is: NVIDIA RTX3050, Intel(R) Core(TM) i7-7700 CPU 3.60 GHz, 16GB RAM, 512GB hard disk.

III. A. Convergence analysis of laser energy transfer efficiency

In this section, the optimization of energy transfer efficiency of high power laser system by A-TD3 algorithm is demonstrated, and the convergence speed of A-TD3 algorithm at different learning rates is analyzed to prove the correctness of the algorithm. And this section compares the A-TD3 algorithm with the TD3 algorithm, DDPG algorithm and DQN algorithm in terms of average energy transfer efficiency.

The convergence performance comparison of A-TD3 algorithm at different learning rates is shown in Fig. 4. It can be seen that the algorithm reaches convergence at the 150th round when the learning rate is 0.0005. If the learning rate is reduced to 0.0001, the learning rate decreases significantly and convergence is achieved around the 330th round, which is due to the fact that the smaller learning rate results in a smaller update of the parameters of the A-

TD3 algorithm, and therefore the learning rate decreases. When the learning rate continues to decrease to 0.00001, the algorithm remains in a fluctuating state at the 500th round and does not complete convergence. And when the learning rate is 0.0005, the average energy transfer efficiency of the high-power laser is the highest, reaching about 9.7. Therefore, the learning rate is fixed at 0.0005 during the subsequent simulation.

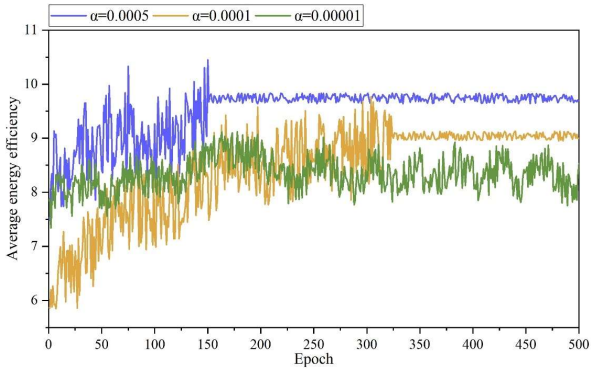

Figure 4: The convergence of A-TD3 algorithms under different learning rates

A comparison of the energy transfer efficiency of different algorithms is shown in Fig. 5, and the results of this simulation are analyzed as follows:

(1) The average energy transfer efficiency of TD3 algorithm and A-TD3 algorithm are close to each other, and the average energy transfer efficiency of the two algorithms is about 9.3 and 9.7 respectively, but the convergence speed of A-TD3 algorithm is faster.

(2) For the DDPG algorithm, the convergence speed is close to that of the TD3 algorithm, but the average energy transfer efficiency is lower than that of the TD3 algorithm.The DDPG algorithm converges to an average energy transfer efficiency of about 8.7 after 225 rounds of iterations.

(3) Compared with DQN, the A-TD3 algorithm achieves a significant improvement in energy transfer efficiency and convergence speed because DQN is unable to establish a mapping relationship between states and actions and can only randomly select actions. And the A-TD3 algorithm enhances the perception of the environment through the reward network and the cost network to maximize the energy transfer efficiency.

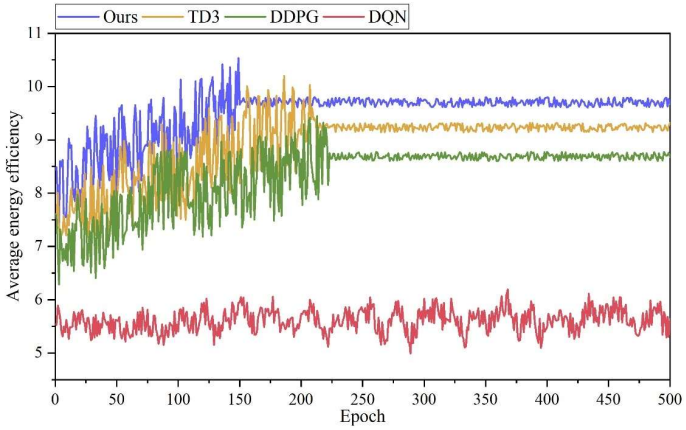

Figure 5: The efficiency of different algorithms is compared

III. B. Effect of transmission distance on laser energy loss

High-power laser has obvious advantages in the performance of atmospheric transmission, and the wavelength of 10.5 μm is exactly in the low-loss window of atmospheric channel transmission, so the far-infrared laser with a wavelength of 10.5 μm is selected as the light source of atmospheric laser communication. Computer simulation of the propagation characteristics of the Gaussian beam in free space is carried out, and the optical power distribution curves of the laser beam at different distances are obtained as shown in Fig. 6.

From the figure, it can be seen that the transmit power is certain, with the increase of the transmission distance, the optical power distribution of the Gaussian beam tends to be flat, and the optical energy in the unit area is less and less. When the laser energy transmission distance is 0, 1, 3, 5 and 10 km, the maximum optical power is 100%,

74.88%, 47.85%, 24.35% and 2%, respectively. For the receiving antenna with a certain aperture, the received optical power is reduced, which can be regarded as the relay loss of the laser beam in free space. The simulation results show that when the laser beam propagation distance reaches 10km, the energy peak of the spot at the receiving end is already very low, and the optical power distribution can be approximated as a uniform distribution.

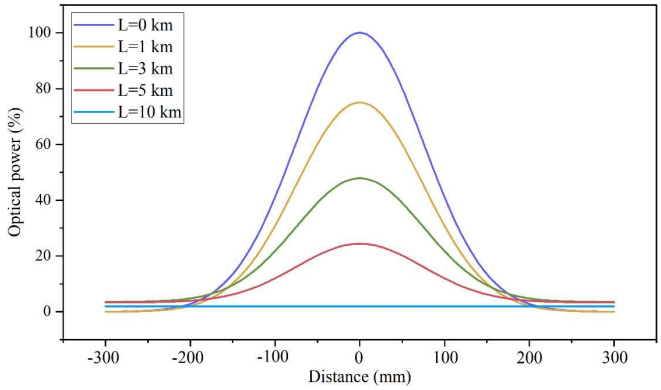

Figure 6: The light power distribution curve of different transmission distance

III. C. Study of beam convergence energy distribution in laser systems

In order to compare the high-power laser system before and after optimization, the energy distribution of the laser beam after convergence, five large fiber cores were used for combination, and the outside of the large fiber cores were tightly bundled together with a plastic sleeve. The structural design of the large fiber core double cladding, the duty cycle of the air holes in the inner cladding is 0.80, the outer diameter of the fiber is 200 μm, and the core diameter is 80 μm. Fig. 7 shows the laser beam energy distribution of the output of the unoptimized laser system. As can be seen from the figure, the laser energy distribution is not uniform, forming multiple energy centers. If it is focused with a lens, it cannot be focused to 1 point, and the same multiple energy centers will appear, which is not conducive to laser energy transmission.

For this reason, a high-power laser system driven by a reinforcement-based learning algorithm is used, and the optimized laser system output laser beam energy distribution is shown in Figure 8. Experiment with five 100W high-power laser as the input, hollow-core energy transfer fiber bundle length of 120cm, measured through the tapered hollow-core fiber after the output power of 440W, the output spot diameter of 0.1mm, the power density of 1120kW/cm². It can be seen that the optimized laser system can be output laser beams into a 0.1mm spot, and the spot laser energy density distribution to the center of the concentration. Density distribution to the center of the concentration.

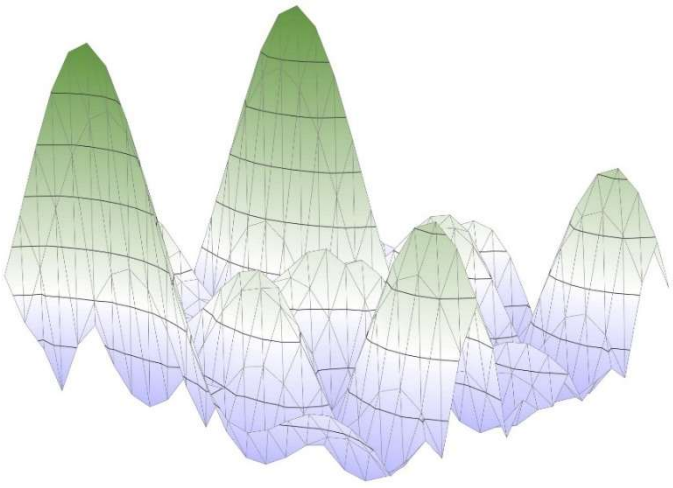

Figure 7: Unoptimized laser system output laser beam energy profile

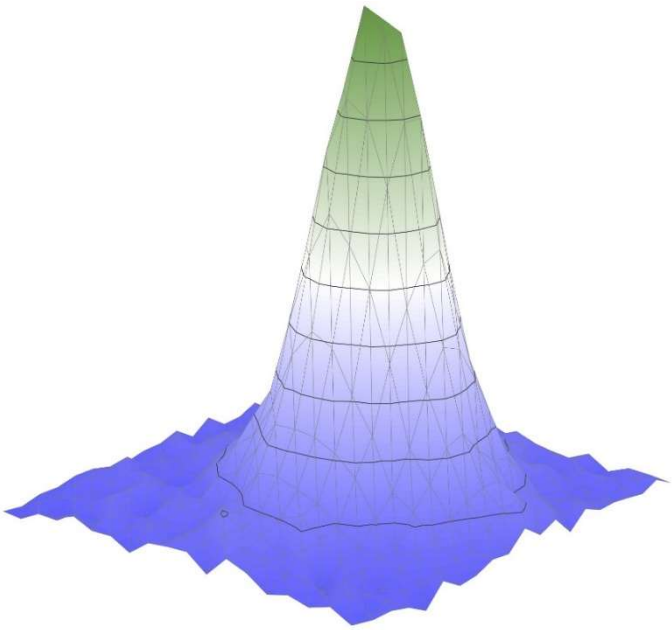

Figure 8: The optimized laser system outputs laser beam energy distribution

III. D. Laser System Energy Transfer Loss Suppression Experiment

In this section, a large fiber core with the same specifications as above is used to confine the high-power laser energy in the core and to reduce the nonlinear effects of the fiber. Meanwhile, the laser energy transmission loss of the optimized high-power laser system driven by reinforcement learning algorithm is compared with that of the unoptimized high-power laser system, and the loss characteristics of the fiber are tested by the truncation method. An ordinary lamp with a bandwidth of 400-1000 nm was used as the light source, and a spectrometer was used to record the data. The length of the test fiber is 120 m, and the truncation length is 40 m. The experimental results of the high-power laser energy transmission loss are shown in Figure 9.

As can be seen from the figure, the energy loss of the optimized high power laser system based on the algorithm of this paper is lower than that of the unoptimized one, with a reduction between 30% and 70%. Taking the optimized high power laser system energy loss as an example for analysis, the transmission loss for 446nm wavelength is 23dB/km, and the loss for 520nm and 545nm wavelengths is 27dB/km and 24dB/km, respectively. Due to the large difference in the refractive index between the inner and outer cladding and the core, resulting in the fiber's numerical aperture is large, which is beneficial to the coupling of the laser into the fiber, improving the coupling efficiency. The cladding with high air-filling rate is conducive to the propagation of the laser beam in the fiber core. The algorithm in this paper effectively suppresses the loss of high power laser system energy transmission into the core.

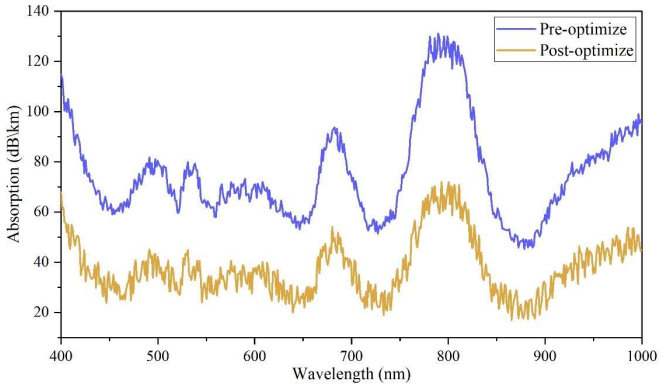

Figure 9: Experimental results of high efficiency laser energy transmission loss

IV. Conclusion

In this study, the A-TD3 algorithm was used to significantly improve the efficiency of the system in optimizing the energy transfer efficiency of a high-power laser system. The experimental results show that the energy transfer

efficiency of the system can reach up to 9.7 when using the A-TD3 algorithm, while the traditional TD3 algorithm and the DDPG algorithm have efficiencies of 9.3 and 8.7, respectively.In addition, the A-TD3 algorithm has a much faster convergence rate, which is achieved in 150 rounds, while the DDPG algorithm requires 225 rounds to achieve a similar effect.

When further analyzing the effect of transmission distance on laser energy loss, it is found that the laser energy loss increases significantly with the increase of transmission distance. At a transmission distance of 10km, the laser power has been reduced to 2% of the original power, a phenomenon that indicates that the energy loss of the laser in long-distance transmission is not negligible. Through the optimized laser system, the energy loss of the system is reduced between 30% and 70%, which is significantly lower than the energy loss of the unoptimized system.

In summary, the optimization strategy of A-TD3 algorithm based on deep reinforcement learning proposed in this paper can effectively improve the transmission efficiency and reduce the energy loss in the process of high-power laser energy transmission, which has a better application prospect, especially in the field of long-distance laser energy transmission.

References

[1] Zuo, J., & Lin, X. (2022). High-power laser systems. Laser & Photonics Reviews, 16(5), 2100741.

[2] Zhu, J., Zhu, J., Li, X., Zhu, B., Ma, W., Lu, X., ... & Lin, Z. (2018). Status and development of high-power laser facilities at the NLHPLP. High power laser Science and Engineering, 6, e55.

[3] Kwee, P., Bogan, C., Danzmann, K., Frede, M., Kim, H., King, P., ... & Willke, B. (2012). Stabilized high-power laser system for the gravitational wave detector advanced LIGO. Optics express, 20(10), 10617-10634.

[4] Bradford, P., Woolsey, N. C., Scott, G. G., Liao, G., Liu, H., Zhang, Y., ... & Neely, D. (2018). EMP control and characterization on high-power laser systems. High Power Laser Science and Engineering, 6, e21.

[5] Apollonov, V. V. (2014). High Power Lasers for New Applications. In High-Power Optics: Lasers and Applications (pp. 167-193). Cham: Springer International Publishing.

[6] Xu, M., Liu, B., Zhang, L., Ren, H., Gu, Q., Sun, X., ... & Xu, X. (2022). Progress on deuterated potassium dihydrogen phosphate (DKDP) crystals for high power laser system application. Light: Science & Applications, 11(1), 241.

[7] Eslamian, M. (2017). Inorganic and organic solution-processed thin film devices. Nano-micro letters, 9(1), 3.

[8] Wen, X., Wu, W., Ding, Y., & Wang, Z. L. (2013). Piezotronic effect in flexible thin-film based devices. Advanced Materials, 25(24), 3371-3379.

[9] Talin, A. A., Centrone, A., Ford, A. C., Foster, M. E., Stavila, V., Haney, P., ... & Allendorf, M. D. (2014). Tunable electrical conductivity in metal-organic framework thin-film devices. Science, 343(6166), 66-69.

[10] Kimura, M. (2019). Emerging applications using metal-oxide semiconductor thin-film devices. Japanese Journal of Applied Physics, 58(9), 090503.

[11] Winzer, P. J. (2012). High-spectral-efficiency optical modulation formats. Journal of lightwave technology, 30(24), 3824-3835.

[12] Zhao, Z., Zhang, G., Huang, Y., Zhou, J., Shi, T., Lin, Z., ... & Long, Y. (2024). Water jet guided high-power laser energy transmission loss analysis. The International Journal of Advanced Manufacturing Technology, 130(11), 5379-5389.

[13] Liu, H., Zhang, Y., Hu, Y., Tse, Z., & Wu, J. (2021). Laser power transmission and its application in laser-powered electrical motor drive: A review. Power Electronics and Drives, 6(41), 167-184.

[14] Fernández, E. F., García-Loureiro, A., Seoane, N., & Almonacid, F. (2022). Band-gap material selection for remote high-power laser transmission. Solar Energy Materials and Solar Cells, 235, 111483.

[15] Ludewigt, K., Liem, A., Stuhr, U., & Jung, M. (2019, October). High-power laser development for laser weapons. In High Power Lasers: Technology and Systems, Platforms, Effects III (Vol. 11162, pp. 46-53). SPIE.

[16] Zhehua Zhou,Xuan Xie,Jiayang Song,Zhan Shu & Lei Ma. (2024). GenSafe: A Generalizable Safety Enhancer for Safe Reinforcement Learning Algorithms Based on Reduced Order Markov Decision Process Model.. IEEE transactions on neural networks and learning systems,PP,

[17] Anupama Mampage,Shanika Karunasekera & Rajkumar Buyya. (2025). A deep reinforcement learning based algorithm for time and cost optimized scaling of serverless applications. Future Generation Computer Systems,173,107873-107873.

[18] Yan Wan,Yujia Zhai,Can Cui & Dexuan Song. (2024). Indoor energy-saving strategy optimization based on deep reinforcement learning and DDPG algorithm. Computing,107(1),26-26.