Nuclear Instruments and Methods in Physics Research A 1063 (2024) 169320

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

An experimental application of machine learning algorithms to optimize the FEL lasing via beam trajectory tuning at Dalian Coherent Light Source

Jitao Sun [a,c], Xinmeng Li [a,c], Jiayue Yang [a], Li Zeng [b], Jiahang Shao [b,d], Yong Yu [b,d,*], Weiqing Zhang [a,**], Xueming Yang [a,d]

[a] Dalian Coherent Light Source and State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, 116023, China
[b] Institute of Advanced Science Facilities, Shenzhen, 518107, China
[c] University of Chinese Academy of Sciences, Beijing, 100049, China
[d] Center for Advanced Light Source, College of Science, Southern University of Science and Technology, Shenzhen, 518055, China

ARTICLE INFO

Keywords:
Free-electron laser
Genetic algorithm
Deep reinforcement learning
FEL lasing optimization
Dalian Coherent Light Source

ABSTRACT

The lasing optimization of Free-Electron Laser (FEL) facilities is a time-consuming and challenging task. Instead of operating manually by experienced operators, implementation of machine learning algorithms offers a rapid and adaptable approach for FEL lasing optimization. Recently, such an experiment has been conducted at the vacuum ultraviolet FEL facility - Dalian Coherent Light Source (DCLS). Four algorithms, namely the standard and the neural network-based genetic algorithms, the deep deterministic policy gradient and the soft actor critic reinforcement learning algorithms, have been employed to enhance the FEL intensity by optimizing the electron beam trajectory. These algorithms have shown notable efficacy in enhancing the FEL lasing, especially the reinforcement learning ones which achieved convergence within only approximately 400 iterations. This study demonstrates the validity of machine learning algorithms for FEL lasing optimization, providing a forward-looking perspective on the automatic operation of DCLS.

1. Introduction

Free-Electron Laser, characterized by its high peak power, narrow bandwidth, and ultrafast pulses, has been developed as an indispensable tool in the fields of physics, chemistry, material, and life sciences [1–3]. In the extreme ultraviolet and X-ray region, several high-gain FEL facilities have been constructed and operated as user facilities for the science community [4–12]. In these FEL facilities, beam tuning is frequently required by the FEL user experiments, including regular FEL wavelength switching and pulse energy optimization. It is crucial to swiftly optimize the FEL performance to guarantee the continuity of the user experiments. Traditionally, such optimization is manually conducted by FEL operators. This approach proves to be time-consuming due to the large number of tuning parameters. Furthermore, the optimization results are contingent upon operators' experience.

To overcome these issues, machine learning has been introduced into the field of accelerators and FELs in recent years [13–18], especially the genetic algorithms (GA) [19–21] and the deep reinforcement learning algorithms (RL) [22–27]. GA has been used to optimize electron beam dynamics in accelerators, such as the linac optimization for various application requirements in the proposed UK's New Light Source [19], the Pareto fronts of beam emittance and bunch length optimization in superconducting FEL injectors [20]. However, GA is not universally effective in addressing search problems characterized by a large number of variables. Consequently, a neural network-based genetic algorithm (NNGA) has been formulated and effectively utilized for the optimization of nonlinear beam dynamics in High Energy Photon Source (HEPS) [21]. RL has also been adopted in accelerator facilities, such as precise alignment of superconducting modules in CAFe-II [22], beam pointing adjustment and beam size matching in the DESY ARES particle accelerator [23]. Furthermore, RL has also been used to optimize FEL Lasing performance with theoretical simulation and experimental demonstration respectively in SXFEL and FERMI@Elettra [24–27].

In these optimizations, it is a core issue to balance the exploration and exploitation [28]. Exploration refers to trying unknown actions to discover more valuable strategies, while exploitation involves using

* Corresponding author. Institute of Advanced Science Facilities, Shenzhen, 518107, China.
** Corresponding author.
E-mail addresses: yuyong@mail.iasf.ac.cn (Y. Yu), weiqingzhang@dicp.ac.cn (W. Zhang).

https://doi.org/10.1016/j.nima.2024.169320
Received 3 January 2024; Received in revised form 3 April 2024; Accepted 3 April 2024
Available online 4 April 2024

known knowledge to gain the maximum reward. In machine learning algorithms, since the absolute optimal solution is unknowable in complex environments, the algorithms aim to find a solution that is as good as possible, one that approximates the global optimum in terms of performance. Genetic algorithms ensure exploration by randomly generating initial populations and through crossover and mutation operations. By employing selection processes to retain individuals with better fitness, they can maintain diversity in sufficiently large populations, enabling the exploration of various solution regions. This approach facilitates the discovery of and eventual convergence to a global optimum [29,30]. In reinforcement learning, the introduction of noise or entropy enhances the algorithm's exploration capabilities [31]. This method not only encourages the exploration of previously unencountered states but also prevents the algorithm from prematurely converging on suboptimal solutions. By systematically utilizing the information from explored areas, the algorithm balances the relationship between exploration and exploitation. This balance enables the agent to efficiently explore the solution space and gradually refine its policy based on the cumulative rewards received, until it converges on a solution that maximizes long-term rewards [32].

The machine learning shows different performance at different FEL facilities. To evaluate the performance of machine learning at Dalian Coherent Light Source facility, an exploratory experiment with GA/NNGA and RL has been recently conducted for FEL lasing optimization. First, both GA and NNGA have been employed to optimize the machine, as well as gather data for developing a surrogate model intended for RL training purposes. Then, both deep deterministic policy gradient (DDPG) and soft actor critic (SAC) RL algorithms [33,34] have been evaluated in the surrogate model and the machine optimization. The remainder of the manuscript is organized as follows. Sec.2 introduces the machine learning algorithms used in DCLS experiment. Sec.3 describes the experimental conditions and procedures. Sec.4 presents the optimization of FEL lasing performance with GA/NNGA and DDPG/SAC at DCLS, accompanied by a detailed discussion of these algorithms. Sec. 5 concludes this study.

2. Machine learning algorithms used in DCLS facility

In this study, several machine learning algorithms have been adopted for the FEL lasing optimization, including genetic algorithms (GA, NNGA) and reinforcement learning algorithms (DDPG, SAC). This section gives a brief description of these algorithms.

2.1. Genetic algorithms – GA and NNGA

GA, known for its robustness in variable space exploration, exhibits strong global optimization capabilities [35,36]. Inspired by the principles of biological evolution, GA seeks the optimal objectives through several evolutionary steps including population initialization, fitness computation, selection, crossover and mutation. However, constraints in computational capacity and data acquisition efficiency limit the population size of GA, resulting unsatisfactory and time-consuming results in certain scenarios. NNGA has been proposed in order to magnify the population size and accelerate the convergence speed. Data generated by the GA is used to train the neural network, and the network is then employed to predict the fitness of a large number of individuals. According to the fitness, the most competitive individuals are selected for the next generation evolution. Benefiting from the outstanding performance of neural network in regression problems [37,38], NNGA shows faster convergence speed and richer diversity. In our study, a single-objective GA and NNGA have been applied for FEL lasing optimization, and the corresponding environments and results have been used to create a surrogate model for the testing of DDPG and SAC methods.

2.2. Reinforcement learning algorithms – DDPG and SAC

Neural networks that elucidate the implicit relationship between inputs and outputs are the core of deep RL algorithms, ensuring the optimization capabilities in complex, dynamic, and partially observable systems [32,37,39]. RL trains the neural networks by the interaction of an agent with the environment continuously. It aims to maximize the total return R_t by optimizing algorithm policy π in the interaction. After the agent selects action a_t following policy π at time step t, the environment staying at the state s_t transfers to state s_{t+1} and provides the agent with a present reward r_t as feedback. The agent iteratively optimizes its policy π through training the actor-critic network to maximize the total return $R_t = \sum_{k=0}^{\infty}\gamma^k r_{t+k}(s_{t+k}, a_{t+k})$, where the discount factor γ^k denotes the discount evaluation of the future k-step rewards reflected in the current state s_t.

DDPG and SAC are commonly employed in the realm of RL algorithm. The primary differences between them are embodied in the computation of the value function $Q(s_t, a_t)$ within the critic network and the introduction of policy entropy, where $Q(s_t, a_t)$ represents the expected return for choosing action a_t in state s_t. As shown in Fig. 1(a), DDPG employs one critic network and one actor network to respectively estimate the Q function and select the optimized actions. The critic training purpose is to minimize the error of the predicted Q function and the target Q_{target}^{DDPG} function, as expressed in Eq. (1) [33]. Meanwhile, the actor training purpose is to maximize the policy's objective function $J^{DDPG}(\pi)$ shown in Eq. (2).

$$Q_{\text{target}}^{DDPG}(s_t, a_t) = r_t + \gamma Q(s_{t+1}, \pi(s_{t+1})) \tag{1}$$

$$J^{DDPG}(\pi) = \mathbb{E}_{(s_t, a_t)\sim\pi}[Q(s_t, a_t)] \tag{2}$$

SAC, as an enhancement of DDPG, utilizes two critic networks ($k = 1, 2$) for Q function prediction to improve the training stability, as shown in Fig. 1(b). Additionally, policy entropy H is introduced to encourage exploration [34]. The target value function Q_{target}^{SAC} and the policy's objective function $J^{SAC}(\pi)$ are consequently converted into Eq. (3) and Eq. (4),

$$Q_{\text{target}}^{SAC}(s_t, a_t) = r_t + \gamma\left(min_{k=1,2}Q_k(s_{t+1}, \pi(s_{t+1}))\right) + \alpha H(\pi(\cdot|s_{t+1}))\big) \tag{3}$$

$$J^{SAC}(\pi) = \mathbb{E}_{(s_t, a_t)\sim\pi}[Q(s_t, a_t) + \alpha H(\pi(\cdot|s_t))] \tag{4}$$

where α is the temperature parameter to control the level of policy entropy. As a result, SAC exhibits more stable and superior performance in most complex environments.

3. The experimental conditions and procedures

This section provides a comprehensive description of the experiment, encompassing the experimental conditions and procedures in detail.

3.1. The experimental conditions

The experiment was carried out in the FEL amplifier section at DCLS, a high-gain FEL facility operating in the vacuum ultraviolet (VUV) region with wavelength of 50–150 nm [8,40], as shown in Fig. 2. The 300 MeV, 200 pC electron beam from the linac travels though the FEL amplifier and generates FEL pluses with a repetition rate of 10 Hz. To obtain fully-coherent FEL pulses, the facility was designed to operate in high-gain harmonic generation (HGHG) mode [41], in which the electron beam is modulated by a seed laser in the modulator (Mod) and generates FEL pulses at a higher harmonic of the seed laser in the radiator (Rad). Additionally, corrector and quadrupole magnets are implemented to control the trajectory and electron beam size to enhance FEL lasing amplification.

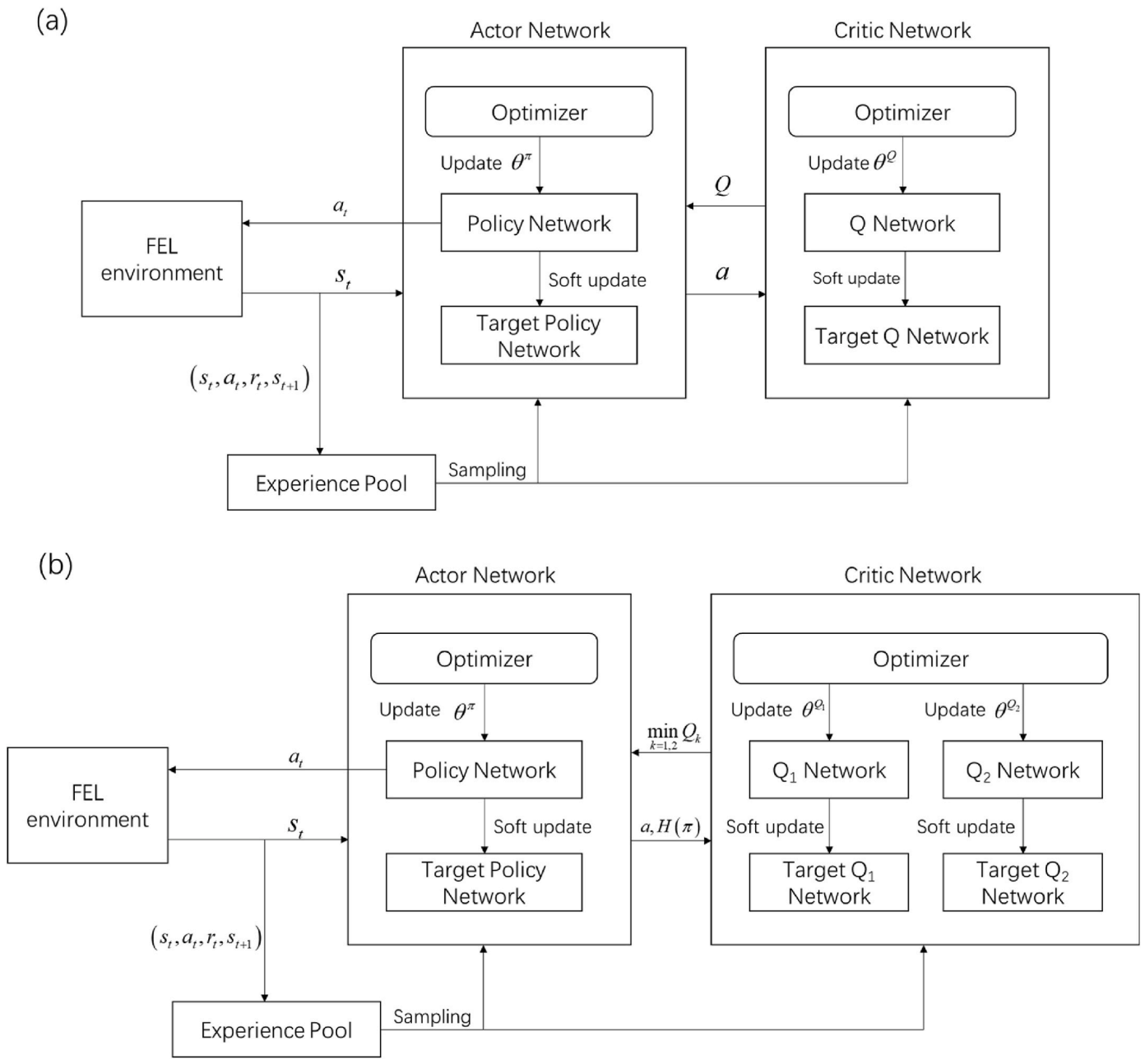

Fig. 1. The flow charts of the DDPG (a) and SAC (b) algorithms.

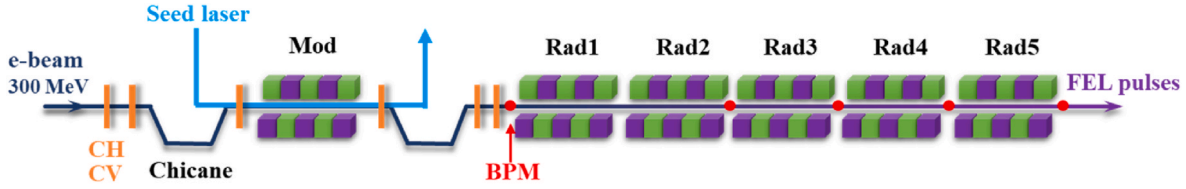

Fig. 2. The FEL amplifier layout of DCLS.

3.2. *The experimental procedures*

This study aims to evaluate the performance of GA/NNGA and DDPG/SAC algorithms for FEL lasing optimizations at DCLS facility. In the experiment, a total of 12 correctors (CH/CV) positioned between the linac and the radiator were employed as both actuators and optimized variables. They influence the FEL lasing performance by controlling the electron beam trajectory for laser modulation in the modulator and FEL amplification in the radiator. The correctors and quadrupoles along the radiator were not set as optimized variables. Their strengths were optimized by experienced operators in the initial stage and kept unchanged throughout the following stages. Although the beam trajectory was affected by the quadrupoles while the beam passed off-axis, it didn't affect the algorithms' optimizations as the radiator was treated as an

entire component. The global beam trajectory along the radiator was controlled solely by the aforementioned upstream 12 correctors and was checked by the beam position monitors (BPMs) along the radiator.

In our experiment, in order to mitigate the challenge of reward sparsity caused by the electron beam loss, the objective function encompasses both the FEL pulse energy measured by a photodiode [42] at the exit of radiator and the electron bunch charge obtained from the BPMs, shown as:

$$0.01\sum_i \frac{Q^{BPM_i}}{Q_0^{BPM_i}} + \frac{I}{I_0} \tag{5}$$

where $Q_0^{BPM_i}$ (I_0) and Q^{BPM_i} (I) represent the initial and currently measured bunch charge (FEL pulse energy), respectively. As the optimization starts, the objective function value is primarily influenced by the bunch charge, predominantly due to the beam loss issue, which is gradually resolved as the charge-driven reward increases. Subsequently, the optimization process shifts its focus towards maximizing the FEL pulse energy.

Before the experiment, the machine was manually optimized to about 60 μJ at the wavelength of 115 nm by experienced operators. Then, the aforementioned 12 correctors were randomly initialized, and the pulse energy decreased to a few microjoules. This condition was marked as the initial state of the subsequent algorithm optimizations. Both GA/NNGA and DDPG/SAC algorithms employ identical optimization variables, objective function, and initial state. The state space for the DDPG/SAC algorithms includes the currents of 12 correctors, as well as the beam positions and bunch charges obtained from the BPMs in the linac and FEL amplifier. In order to mitigate the impact of the jitters induced by the instabilities of RF power, magnet currents and other hardware, each data point obtained in the experiment was averaged over 10 shots. Taking into account the corrector response time, the duration for collecting each data point was approximately 3 seconds.

The experiments have been carried out as follows. First, the GA/NNGA algorithms were implemented in the facility. Then, leveraging the robust variable space exploration capability of GA/NNGA, experimental data obtained from the first step has been utilized to develop a surrogate model for virtual testing of DDPG/SAC performance. Finally, the DDPG/SAC algorithms were applied in the surrogate model and the real facility to optimize FEL lasing.

4. The performance of GA/NNGA and DDPG/SAC in DCLS

This section mainly presents the FEL lasing performance with GA/NNGA and DDPG/SAC algorithms in DCLS facility, where the algorithm libraries in Python codes have been used [43–45]. This section also describes the development and application of the surrogate model.

4.1. The experimental performance of GA/NNGA and the corresponding surrogate model

Initially, the GA algorithm was employed with a population size of 200 individuals per generation. Considering the limitations of time, this optimization was terminated after 17 generations. As illustrated in Fig. 3 with blue curves, the FEL pulse energy of the optimal individual reached 69.13 μJ, whereas that of the optimal generation was merely around 9.86 μJ. This notable disparity between the two outcomes indicates non-convergence. To improve the convergence speed, NNGA algorithm was employed. The evolutionary process for the initial five generations remained consistent with that of the GA algorithm. Starting from the sixth generation, a multi-layer perceptron (MLP) [46] was introduced to increase the population size and identify superior individuals using accumulated data from previous experimental generation. A population of 2000 individuals, ten times the size of a single generation's population in GA, was generated with their fitness values evaluated using the MLP model. The top 200 individuals in the evaluation were selected to

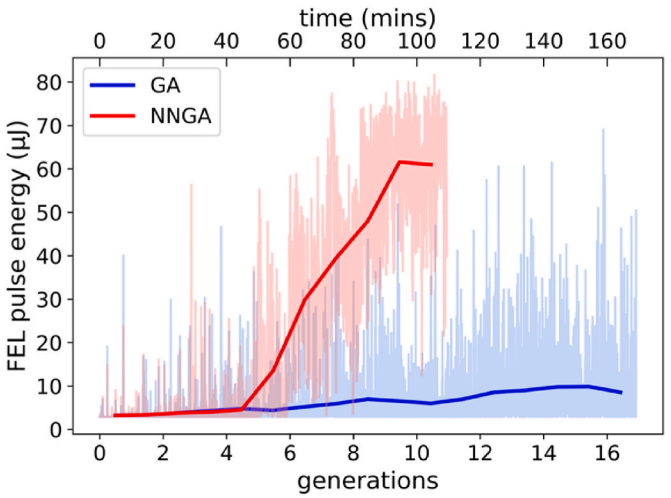

Fig. 3. The experimental results with GA (blue) and NNGA (red) at DCLS facility. The light color and dark color represent the individual and generation results, respectively.

form the next generation's population. These individuals were then utilized in the real facility to obtain actual measured fitness values, expressed in Eq. (5). As depicted by the red colors in Fig. 3, FEL pulse energies of both the optimal individual and the generation reached flat-top after the ninth generation. The converged energies were measured at 81.65 μJ and 61.57 μJ, respectively. Compared to GA, NNGA demonstrates a much faster convergence speed, considerably decreasing the convergence step required (approximately 1800 vs. much larger than 3400), as presented in Table 1.

As a user facility, DCLS has limited available time for this study. To address this constraint, a surrogate model was created using MLP based on the experimental data collected from the GA and NNGA optimizations. This surrogate model was then utilized for offline testing and preliminary hyperparameter selection of DDPG and SAC algorithms. The model takes the currents of the 12 correctors as its inputs and produces the predicted FEL pulse energy, beam positions and charges of all BPMs. As an illustration, the predicted FEL pulse energies (orange) and the original experiment data (green) are compared in Fig. 4. To evaluate the prediction performance of this model, the R^2 parameter is calculated as [27],

$$R^2 = 1 - \frac{\sum_i (I_i - P_i)^2}{\sum_i (I_i - \bar{I})^2} \tag{6}$$

where I_i and P_i denote the measured and predicted FEL pulse energies, respectively, $\bar{I}$ is the average value of I_i. A R^2-value of 0.92 was attained, signifying a high level of agreement between the surrogate model and the real facility. Although a slight deviation is observed in the high-energy region, which was caused by the sparsity of the corresponding experimental data (as depicted in Fig. 3), it does not significantly impact the training process and the surrogate model was deemed adequate for evaluating the performance of DDPG and SAC algorithms.

Table 1
The convergence performance of GA/NNGA and DDPG/SAC algorithms in DCLS FEL lasing optimization.

Algorithm	Converg. step (#)	Converg. time (mins)	FEL pulse energy (μJ)
GA	≫3400	≫164	9.86
NNGA	1800	90	61.57
DDPG-exp.	540	27	70.45
SAC-exp.	420	21	64.26
DDPG-sim.	443	–	57.60
SAC-sim.	333	–	66.20

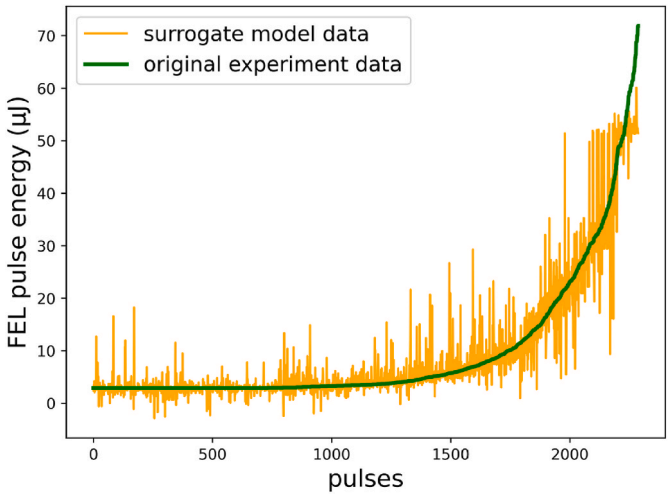

Fig. 4. Comparison of the predicted results in surrogate model (orange) and the experimental data (green) originated from GA/NNGA optimizations.

4.2. DDPG and SAC performance in the surrogate model and the experiment

The performance of DDPG and SAC algorithms have been evaluated both in the surrogate model and the actual facility with the same procedures. At the beginning of each algorithm, a warm-up phase was conducted with the initial 128 steps executed with random actions. Subsequently, a training process consisting of 50 steps, followed by an evaluation process of 10 steps, was iterated until convergence. To enhance the exploration capability, a Gaussian noise was added to the actions in the training process, while in the evaluation process the performance of the algorithms was evaluated without noise.

The algorithms were firstly tested in the surrogate model before being implemented at the actual facility, with further hyperparameter optimization. Both the experimental and simulation results are shown in Fig. 5. Fig. 5(a) and (b) present the experimental results of DDPG and SAC respectively, illustrating the training process, evaluation process and corresponding tendency. Each tendency curve, as shown in Fig. 5(c), is generated by averaging the values obtained during each 10-step evaluation, effectively representing the convergence performance of the algorithm. The solid curves with light color represent the results for 100 random warm-up phases in the simulation, while the dark curves represent the corresponding average of these light-colored curves. For the experiment, the DDPG and SAC results are represented by the dashed curves with triangular markers in red and blue, respectively. The convergence criteria for the tendency curves in both the simulation and experiment were defined as the intensity increment between two consecutive evaluations being less than 5%. In the simulation (experiment), the average convergence steps for DDPG and SAC were recorded as 443 (540) and 333 (420), respectively, with corresponding FEL pulse energies of 57.60 (70.45) μJ and 66.20 (64.26) μJ, the experiment agrees well with the simulation. These results are also presented in Table 1. The FEL pulse energies yielded by both algorithms are comparable, with SAC demonstrating slightly better performance than DDPG in terms of convergence steps. Additionally, the RMS intensity fluctuations at the convergence step for DDPG and SAC in the simulation were 24.7% and 10.4% respectively. The SAC exhibits greater stability, whereas the DDPG, constrained by its limited exploration capabilities, is more susceptible to premature convergence.

4.3. Discussion about GA/NNGA and DDPG/SAC

The previous sections have extensively discussed the FEL lasing performance at DCLS with GA/NNGA and DDPG/SAC algorithms where the results are summarized in Fig. 6 and Table 1. In the experiments, NNGA, DDPG and SAC have demonstrated the ability to attain optimal FEL lasing performance, while GA has failed considering the time limitations. In addition, the convergence time of GA/NNGA is much longer than that of DDPG/SAC, the former is about a few hours or more while the latter is only about 20 minutes.

The trajectory of the electron beam in the radiator, which was controlled by the aforementioned 12 correctors, determined the FEL

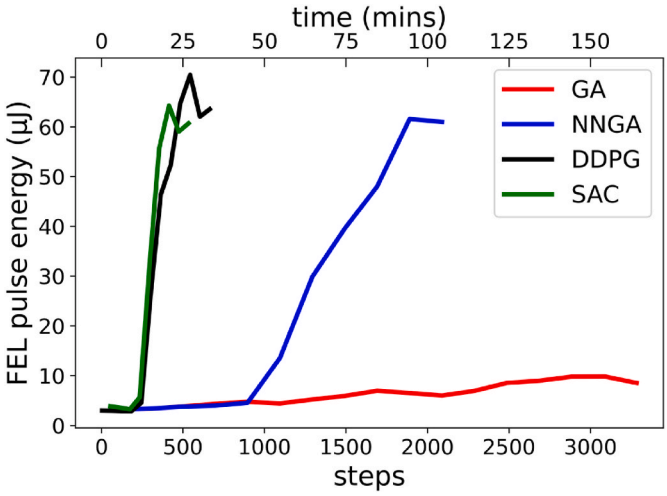

Fig. 6. The FEL lasing performance of GA/NNGA and DDPG/SAC in DCLS facility. The tendency curves of GA (red), NNGA (blue), DDPG (black) and SAC (green).

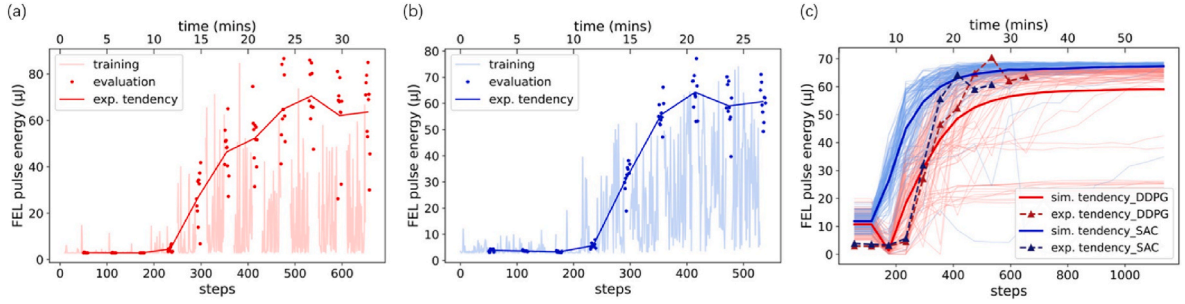

Fig. 5. The results of DDPG and SAC algorithms at DCLS. The experimental results with DDPG (a) and SAC (b), the light-colored curves and the dotted makers represent the training and evaluation processes respectively, the dark-colored curves indicate the average value in each evaluation. (c) The comparison of the simulation and experiment results with DDPG (red) and SAC (blue), the solid curves with light and dark colors represent the evaluated results with different initial random warm-up phases in the surrogate model and the corresponding average results respectively, the dashed curves marked with triangular markers represent the average evaluation results shown in (a) and (b).

lasing performance. The trajectory was measured by the BPMs at either the entrance or the exit of the radiator undulators (Rad), except for the one at the Rad2 entrance due to its faultiness. The trajectory indicated by the measured beam positions (x and y) is shown in Fig. 7 via heat maps. The region with positive values in each map represents the exploration area of each algorithm. The area explored by GA is the widest, followed by NNGA, SAC, and DDPG. This observation agrees well with the expected characteristics of each algorithm. Specifically, DDPG/SAC exhibit weaker exploration capabilities than GA/NNGA, and the integration of SAC with policy entropy provides it a wider exploration range than DDPG. As a consequence, the convergence speed of GA/NNGA is slower. Despite considerable variations in convergence speeds, all four algorithms ultimately converge to similar electron beam trajectory, as illustrated in Fig. 7(a–e). Compared to GA, NNGA/DDPG/SAC employ the MLP to distinguish the effective exploration area and obtain the optimal results quickly.

5. Conclusion

This study demonstrates the validity of machine learning algorithms for FEL lasing optimization at DCLS. Among the four algorithms applied in this study, GA is the most time-consuming one as it didn't reach convergence after several hours of experiment time. With the integration of neural networks, NNGA reached convergence within approximately 1.5 hours. DDPG and SAC have demonstrated much better performance where convergences were achieved about 20 minutes. Furthermore, a surrogate model was created based on the data generated in the GA/NNGA experiment. It reconstructed the facility environment virtually enabling the offline evaluation and initial hyperparameter selection of DDPG and SAC. This study was conducted in a normal-conducting FEL facility with a low pulse repetition rate of 10 Hz. For higher-repetition-rate facilities, such as superconducting continuous-wave FEL facilities with 1 MHz beam [47–50], machine learning algorithms possess a significant advantage in terms of convergence time.

This study demonstrated the outstanding performance of machine learning in the short-term FEL lasing optimization at DCLS. Another concerning point is about the long-term capability. Due to the influence of slow drifts from various devices (e.g. drive/seed lasers, radio-frequency powers, magnets, synchronizing system), the FEL facility environment changes over time, making the previously optimized populations/networks no longer fully effective. For genetic algorithms, they can adapt to small changes by proceeding the optimization based on previous populations [51]. For reinforcement learning, the understanding of the agent concerning the environment and targets is embedded within the actor-critic neural network. When the environment changes, the original network might not directly provide the best solution. However, the new network can be trained with the original ones. This approach, known as transfer learning [52], allows for the inheritance of previous experiences and can significantly expedite the training process. In a word, the reusability and robustness of these algorithms in a long-term could also be guaranteed. In future, additional optimization variables will be incorporated into the algorithms with the aim of achieving automated optimization and operation of DCLS in a long term.

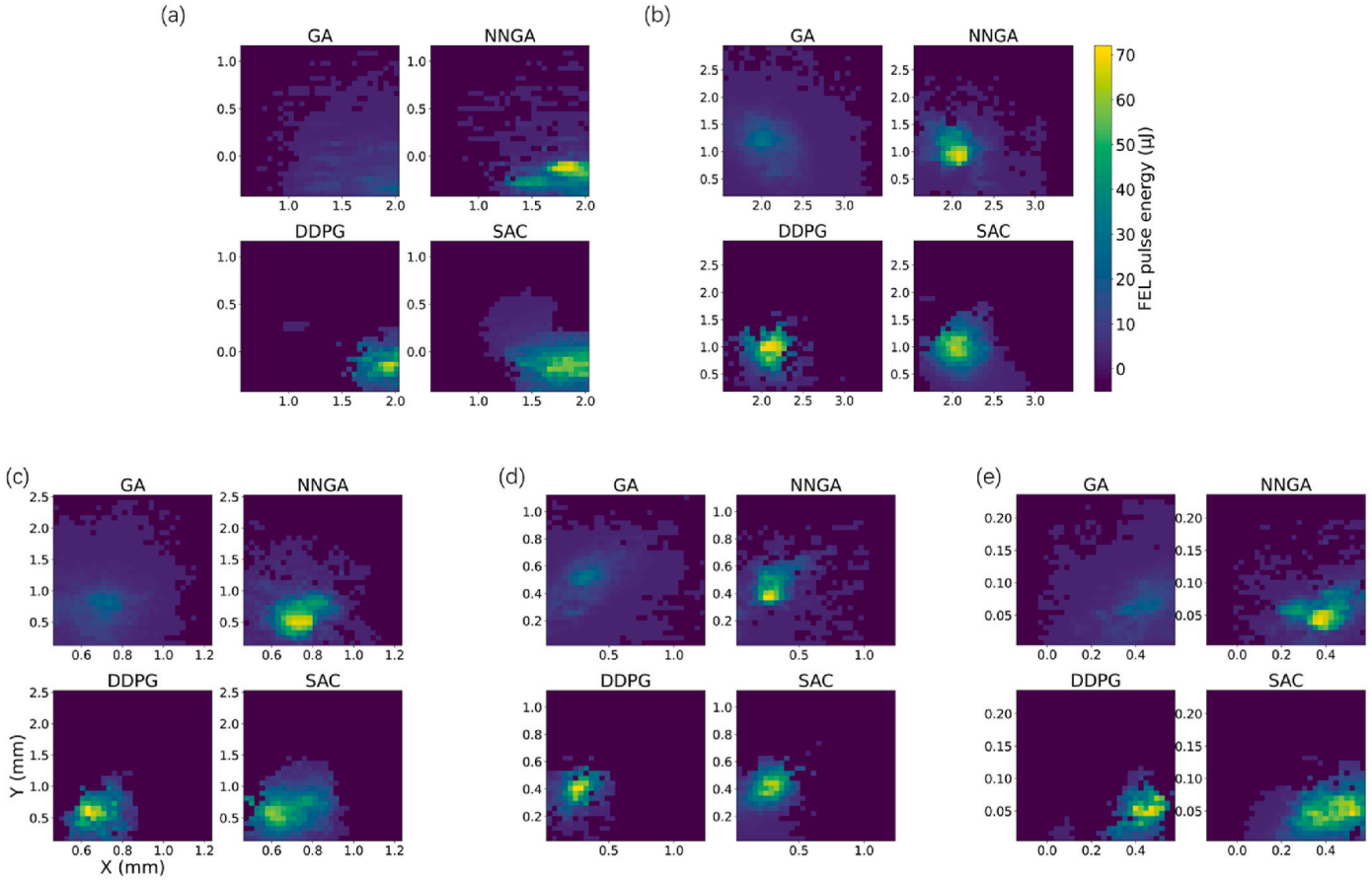

Fig. 7. The exploration capacity of GA/NNGA and DDPG/SAC in DCLS experiments. The exploration capacity and convergence status expressed in electron beam positions measured by the undulator BPMs, the alphabetical subgraphs correspond the BPMs located at the entrance of Rad1/3/4/5 and the exit of Rad5 respectively, the algorithm names are labeled at the top of the corresponding insets.

CRediT authorship contribution statement

Jitao Sun: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Xinmeng Li:** Visualization, Validation, Investigation, Data curation. **Jiayue Yang:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Li Zeng:** Writing – review & editing, Visualization, Validation, Funding acquisition, Formal analysis. **Jiahang Shao:** Writing – review & editing, Visualization, Validation, Formal analysis. **Yong Yu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Weiqing Zhang:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Xueming Yang:** Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank Xiaofan Wang (IASF) and Dengjie Xiao (IASF) for their helpful discussions. This work was supported by the National Key R&D Program of China (Grant No. 2018YFE0203000), the Pre-study Project of Dalian Advanced Light Source from city of Dalian, the National Natural Science Foundation of China (Grant Nos. 22288201 and 12305359), the Scientific Instrument Developing Project of Chinese Academy of Science (Grant No. GJJSTD20220001), DICP (Grant: DICP I202304), and the China Postdoctoral Science Foundation (Grant No. 2022M713081). Y. Yu also acknowledges the support of the Specific Research Assistant Funding Program from Chinese Academy of Sciences.

References

[1] J. Arthur, G. Materlik, R. Tatchyn, H. Winick, The LCLS: a fourth generation light source using the SLAC linac, Rev. Sci. Instrum. 66 (1995) 1987–1989.
[2] R. Neutze, R. Wouts, D. Van der Spoel, E. Weckert, J. Hajdu, Potential for biomolecular imaging with femtosecond X-ray pulses, Nature 406 (2000) 752–757.
[3] Y. Ding, A. Brachmann, F.-J. Decker, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, P. Hering, Z. Huang, Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source, Phys. Rev. Lett. 102 (2009) 254801.
[4] W.a. Ackermann, G. Asova, V. Ayvazyan, A. Azima, N. Baboi, J. Bähr, V. Balandin, B. Beutner, A. Brandt, A. Bolzmann, Operation of a free-electron laser from the extreme ultraviolet to the water window, Nat. Photonics 1 (2007) 336–342.
[5] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, First lasing and operation of an ångstrom-wavelength free-electron laser, Nat. Photonics 4 (2010) 641–647.
[6] D. Pile, First light from SACLA, Nat. Photonics 5 (2011) 456–457.
[7] E. Allaria, R. Appio, L. Badano, W. Barletta, S. Bassanese, S. Biedron, A. Borga, E. Busetto, D. Castronovo, P. Cinquegrana, Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet, Nat. Photonics 6 (2012) 699–704.
[8] G. Wang, Commissioning status of the dalian cohernet light source, in: Proceedings of the 8th International Particle Accelerator Conference, 2017, pp. 2709–2712. Copenhagen, Denmark.
[9] H.-S. Kang, C.-K. Min, H. Heo, C. Kim, H. Yang, G. Kim, I. Nam, S.Y. Baek, H.-J. Choi, G. Mun, Hard X-ray free-electron laser with femtosecond-scale timing jitter, Nat. Photonics 11 (2017) 708–713.
[10] W. Decking, S. Abeghyan, P. Abramian, A. Abramsky, A. Aguirre, C. Albrecht, P. Alou, M. Altarelli, P. Altmann, K. Amyan, A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator, Nat. Photonics 14 (2020) 391–397.
[11] C.J. Milne, T. Schietinger, M. Aiba, A. Alarcon, J. Alex, A. Anghel, V. Arsov, C. Beard, P. Beaud, S. Bettoni, SwissFEL: the Swiss X-ray free electron laser, Appl. Sci. 7 (2017) 720.
[12] Z. Zhao, S. Chen, L. Yu, C. Tang, L. Yin, D. Wang, Q. Gu, Shanghai soft X-ray free electron laser test facility, in: Proceedings of the 11th International Particle Accelerator Conference, 2011, pp. 3011–3013. San Sebastián, Spain.
[13] M. Aiba, M. Böge, N. Milas, A. Streun, Random Walk Optimization in Accelerators-Vertical Emittance Tuning at SLS, 2012.
[14] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, Bayesian optimization of a free-electron laser, Phys. Rev. Lett. 124 (2020) 124801.
[15] C. Emma, A. Edelen, M. Hogan, B. O'Shea, G. White, V. Yakimenko, Machine learning-based longitudinal phase space prediction of particle accelerators, Phys. Rev. Accelerators Beams 21 (2018) 112802.
[16] G. Gaio, M. Lonza, N. Bruchon, L. Saule, Advances in automatic performance optimization at FERMI, in: Proceedings of the 16th International Conference on Accelerator & Large Experimental Physics Control Systems (ICALEPCS), 2017, pp. 8–13. Barcelona, Spain.
[17] X. Huang, Robust simplex algorithm for online optimization, Phys. Rev. Accelerators Beams 21 (2018) 104601.
[18] A. Scheinker, X. Pang, L. Rybarcyk, Model-independent particle accelerator tuning, Phys. Rev. Spec. Top. Accel. Beams 16 (2013) 102803.
[19] R. Bartolini, M. Apollonio, I. Martin, Multiobjective genetic algorithm optimization of the beam dynamics in linac drivers for free electron lasers, Phys. Rev. Spec. Top. Accel. Beams 15 (2012) 030701.
[20] C. Papadopoulos, P. Emma, D. Filippetto, T. Raubenheimer, F. Sannibale, J. Schmerge, L. Wang, F. Zhou, RF injector beam dynamics optimization for LCLS-II, in: Proceedings of the 36th International Free Electron Laser Conference, 2014, pp. 1974–1976. Basel, Switzerland.
[21] J. Wan, P. Chu, Y. Jiao, Neural network-based multiobjective optimization algorithm for nonlinear beam dynamics, Phys. Rev. Accelerators Beams 23 (2020) 081601.
[22] X. Chen, X. Qi, C. Su, Y. He, Z. Wang, Trend-Based SAC Beam Control Method with Zero-Shot in Superconducting Linear Accelerator, 2023 arXiv preprint arXiv: 2305.13869.
[23] J. Kaiser, C. Xu, A. Eichler, A.S. Garcia, O. Stein, E. Bründermann, W. Kuropka, H. Dinter, F. Mayet, T. Vinatier, Learning to Do or Learning while Doing: Reinforcement Learning and Bayesian Optimisation for Online Continuous Tuning, 2023 arXiv preprint arXiv:2306.03739.
[24] N. Bruchon, G. Fenu, G. Gaio, M. Lonza, F.H. O'Shea, F.A. Pellegrino, E. Salvato, Basic reinforcement learning techniques to control the intensity of a seeded free-electron laser, Electronics 9 (2020) 781.
[25] M. Cai, Z. Zhu, K. Zhang, C. Feng, L. Tu, D. Gu, Z. Zhao, Twin delayed deep deterministic policy gradient for free-electron laser online optimization, J. Phys. Conf. 2420 (2023) 012027.
[26] S. Hirlaender, N. Bruchon, Model-free and Bayesian Ensembling Model-Based Deep Reinforcement Learning for Particle Accelerator Control Demonstrated on the Fermi Fel, arXiv Preprint arXiv:2012.09737, 2020.
[27] F. O'Shea, N. Bruchon, G. Gaio, Policy gradient methods for free-electron laser and terahertz source optimization and stabilization at the FERMI free-electron laser at Elettra, Phys. Rev. Accelerators Beams 23 (2020) 122802.
[28] M. Yogeswaran, S. Ponnambalam, Reinforcement learning: exploration–exploitation dilemma in multi-agent foraging task, Opsearch 49 (2012) 223–236.
[29] D.E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, 1988.
[30] T.V. Mathew, Genetic Algorithm, 2012, p. 53. Report submitted at IIT Bombay.
[31] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, Soft Actor-Critic Algorithms and Applications, 2018 arXiv preprint arXiv:1812.05905.
[32] R.S. Sutton, A.G. Barto, Reinforcement Learning: an Introduction, 2 ed., MIT press, Cambridge, Massachusetts, 2018.
[33] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning. arXiv Preprint arXiv:1509.02971, 2015.
[34] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a Stochastic actor, in: Proceedings of the 35th International Conference on Machine Learning, PMLR, 2018, pp. 1861–1870.
[35] J.H. Holland, Genetic algorithms, Sci. Am. 267 (1992) 66–73.
[36] M. Mitchell, An Introduction to Genetic Algorithms, 1 ed., MIT press, Cambridge, Massachusetts, 1998.
[37] J.A. Anderson, An Introduction to Neural Networks, 3 ed., MIT press, Cambridge, Massachusetts, 1995.
[38] B. Müller, J. Reinhardt, M.T. Strickland, Neural Networks: an Introduction, 2 ed., Springer Science & Business Media, Berlin, 1995.
[39] B. Recht, A tour of reinforcement learning: the view from continuous control, Annu. Rev. Control Robotics Autonomous Syst. 2 (2019) 253–279.
[40] Y. Yong, L. Qinming, Y. Jiayue, W. Guanglei, S. Lei, D. Hongli, T. Kai, T. Zhenxing, H. Zhigang, C. Zhichao, Dalian extreme ultraviolet coherent light source, Chin. J. Lasers 46 (2019) 0100005.
[41] L.-H. Yu, M. Babzien, I. Ben-Zvi, L. DiMauro, A. Doyuran, W. Graves, E. Johnson, S. Krinsky, R. Malone, I. Pogorelsky, High-gain harmonic-generation free-electron laser, Science 289 (2000) 932–934.
[42] Opto Diode Corporation, SXUV100: large EUV photodetector (1nm-190nm), available at, https://optodiode.com/photodiodes-sxuv-detectors.html. (Accessed 26 November 2019).

[43] geatpy Jazzbin, The genetic and evolutionary algorithm toolbox with high performance in python, available at, http://www.geatpy.com/. (Accessed 20 July 2023).
[44] evaZQR, TD3-master, available at, https://github.com/evaZQR/TD3-master. (Accessed 26 July 2023).
[45] zhaohaojie1998, DRL-for-Path-Planning, available at, https://github.com/zhaohaojie1998/DRL-for-Path-Planning. (Accessed 23 July 2023).
[46] R. Kruse, S. Mostaghim, C. Borgelt, C. Braune, M. Steinbrecher, Multi-layer perceptrons, in: Computational Intelligence: a Methodological Introduction, Springer, Cham, 2022, pp. 53–124.
[47] J.N. Galayda, The LCLS-II: A High Power Upgrade to the LCLS, SLAC National Accelerator Lab, 2018. Menlo Park, CA (United States).
[48] B. Liu, D. Wang, L. Yin, Progress of Shanghai HIgh Repetition Rate XFEL and Extreme Light Facility (SHINE), 2022.
[49] Z. Zhu, Z. Zhao, D. Wang, Z. Liu, R. Li, L. Yin, Z. Yang, SCLF: an 8-GeV CW SCRF linac-based X-ray FEL facility in Shanghai, in: Proceedings of the 38th International Free-Electron Laser Conference, 2017, pp. 20–25. Santa Fe, NM, USA.
[50] X. Wang, L. Zeng, J. Shao, Y. Liang, H. Yi, Y. Yu, J. Sun, X. Li, C. Feng, Z. Wang, Physical design for shenzhen superconducting soft x-ray free-electron laser (s3fel), in: Proceedings of the 14th International Particle Accelerator Conference, 2023. Venice, Italy.
[51] J.J. Grefenstette, Genetic algorithms for changing environments, in: Ppsn, 1992, pp. 137–144.
[52] M.E. Taylor, P. Stone, Transfer learning for reinforcement learning domains: a survey, J. Mach. Learn. Res. 10 (2009).