

# Soft Actor–Critic-Driven Adaptive Focusing under Obstacles

Huan Lu <sup>1</sup>, Rongrong Zhu <sup>1,2,\*</sup> , Chi Wang <sup>1</sup>, Tianze Hua <sup>1</sup>, Siqi Zhang <sup>1</sup> and Tianhang Chen <sup>3</sup> 

<sup>1</sup> Interdisciplinary Center for Quantum Information, State Key Laboratory of Modern Optical Instrumentation, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 310027, China

<sup>2</sup> School of Information and Electrical Engineering, Hangzhou City University, Hangzhou 310015, China

<sup>3</sup> China Aeronautical Establishment, Beijing 100029, China

\* Correspondence: rorozhu@zju.edu.cn

**Abstract:** Electromagnetic (EM) waves that bypass obstacles to achieve focus at arbitrary positions are of immense significance to communication and radar technologies. Small-sized and low-cost metasurfaces enable the accomplishment of this function. However, the magnitude-phase characteristics are challenging to analyze when there are obstacles between the metasurface and the EM wave. In this study, we creatively combined the deep reinforcement learning algorithm soft actor–critic (SAC) with a reconfigurable metasurface to construct an SAC-driven metasurface architecture that realizes focusing at any position under obstacles using real-time simulation data. The *agent* learns the optimal policy to achieve focus while interacting with a complex environment, and the framework proves to be effective even in complex scenes with multiple objects. Driven by real-time reinforcement learning, the knowledge learned from one environment can be flexibly transferred to another environment to maximize information utilization and save considerable iteration time. In the context of future 6G communications development, the proposed method may significantly reduce the path loss of users in an occluded state, thereby solving the open challenge of poor signal penetration. Our study may also inspire the implementation of other intelligent devices.

**Keywords:** deep reinforcement learning; reconfigurable metasurface; focusing; soft actor–critic



**Citation:** Lu, H.; Zhu, R.; Wang, C.; Hua, T.; Zhang, S.; Chen, T. Soft Actor–Critic-Driven Adaptive Focusing under Obstacles. *Materials* **2023**, *16*, 1366. <https://doi.org/10.3390/ma16041366>

Academic Editors: Theodore E. Matikas and George Kenanakis

Received: 28 November 2022

Revised: 31 January 2023

Accepted: 3 February 2023

Published: 6 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Metasurfaces, as two-dimensional metamaterials [1,2], have attracted extensive attention owing to their ability to generate arbitrary EM arrays by introducing corresponding field discontinuities at interfaces. Several interesting EM devices are based on metasurface technology including couplers [3,4], cloaking [5–7], focusing or imaging systems [8–11], and other devices [12–15]. A focusing metasurface is one of the most thought-provoking devices and are of great significance in promoting research in fields such as radar detection, imaging, and 6G communications. In particular, the development of tunable metasurfaces/metamaterials in recent years has led to a significant increase in the freedom of designing reconfigurable functions [16–27]. For example, flexible technologies can achieve tunable focusing by mechanically controlling the expansion and contraction of a structure [17]. Similarly, other smart tunable materials such as varactor diodes [18], electrolyte elastomers [19], and phase-change films [20] can also achieve focusing effects at different positions and focal lengths using reasonable voltage or light modulation. However, these traditional techniques for achieving focus are direct calculations of the compensated phase, and the traditional theoretical calculations fail when there is an obstacle to the incident wave source and metasurface.

An ideal metasurface-focusing system should quickly realize focusing tasks in different environments to adapt to different communication scenarios. In particular, in the presence of unknown obstacles, the fast realization of the focusing task is of great importance for user signal transmission and reception. However, this is difficult to achieve because the

amplitude-phase characteristics of the unit cell are difficult to analyze, resulting in the inability to analytically deduce the state of each meta-atom. Therefore, intelligent adaptive strategies are urgently required. Although adaptive optics have been extensively studied by enabling artificial intelligence (deep learning) on metasurfaces [28–30], the success of deep learning requires prior knowledge of the environment. Unlike traditional deep learning methods, which require a large number of training and test sets related to a specific environment [31–33], another branch of machine learning, reinforcement learning (RL), is used to describe and solve problems in which *agents* learn strategies to achieve specific goals in the process of interacting with the environment [34–36]. In particular, the development and improvement of deep RL (DRL) in the fields of Go and robotics has increased the demand for RL [37–40]. This makes it possible to solve focusing tasks in complex environments; however, it remains challenging.

In this study, we combine the DRL algorithm, soft actor–critic [41,42], with a metasurface to propose the SAC-M system for adaptive focusing design under an arbitrary obstacle. We first analyzed the adaptive focusing framework and implementation process of the SAC algorithm and used the designed metasurface to simulate and train the focusing task in the presence of different obstacles. Simulation results show that the proposed SAC-M system can operate stably in the presence of multiple obstacles of any shape and adaptively converge incident waves to user-defined locations. In addition, the knowledge learned by the *agent* based on the maximum entropy strategy in an environment can be used to initialize a new environment where the SAC-M system can learn a strong generalization ability. Both the simulation and network training results demonstrated that SAC-M exhibits effective and robust adaptive focusing capabilities when dealing with complex EM environments. Moreover, the SAC-M architecture is beneficial for the proposal of other EM wave smart devices and may be extended to other research fields, such as communication to solve more challenging problems.

## 2. Materials and Methods

### 2.1. Architecture of the SAC-M

The advantage of SAC-DRL lies in its ability to explore more ways to solve problems while learning the policy. For different focuses in unknown scenarios, the *agent* can rapidly complete new tasks and reduce unnecessary iterations such that it can be adapted to different types of environments. Figure 1 illustrates the proposed SAC-M mechanism. A plane incident wave, tunable metasurface, and arbitrarily shaped object (purple cube) simultaneously constitute a complex environment, as shown in Figure 1a. Two sets of one-dimensional data (crossed white-dotted lines) of the focal position in the focal plane (red dashed box in Figure 1a) were selected as the training data. When the state of the tunable metasurface changed, the state (training data) obtained at each moment changed accordingly. The agent collected the training data in the environment, compared it with the objective function (Figure 1b), and analyzed the contribution (termed rewards) of state variables to the final task in real time. Following the analysis, the output action initiated, which means that the voltage or capacitance change at the next moment is transmitted to the tunable metasurface, thereby altering the environment. The training data also changed in real time and were obtained by the agent. Finally, the SAC-M framework formed a closed-loop architecture of an environment-agent-action-environment and iterated continuously until the mission was accomplished.

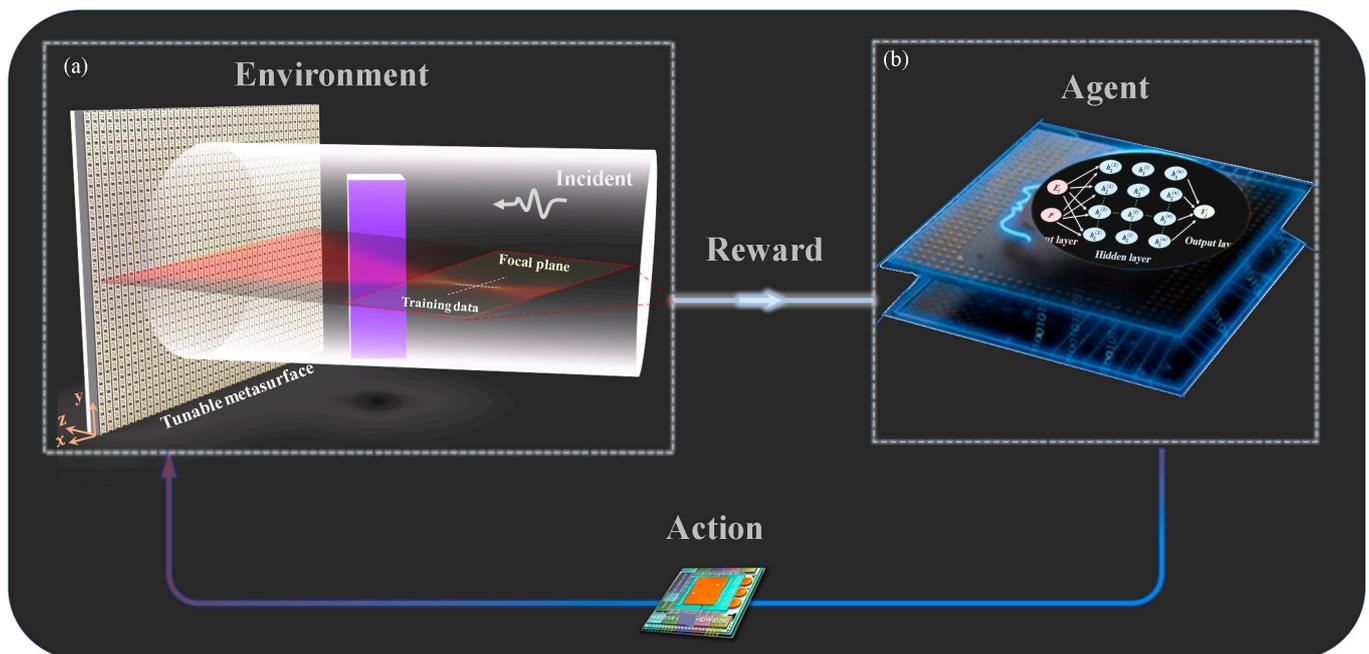
### 2.2. SAC Algorithm

In traditional DRL, the *agent's* goal is to learn a policy that maximizes the accumulated reward expectation [40]. It considers only one optimal action for a state and does not have the ability to cope with changing environments. However, the SAC-DRL is based on maximum entropy, and its core idea is to randomize the strategy and disperse the probability of each action output as much as possible, and not leave out any useful action and trajectory. The optimal strategy is defined as [41,42]:

$$\pi^* = \operatorname{argmax}_{\pi} E_{(s_t, a_t)} [\sum_t R(s_t, a_t) + \alpha H(\pi(\cdot | s_t))] \quad (1)$$

where  $\pi^*$  is the optimal strategy;  $s_t$  and  $a_t$  are the state and action at time  $t$ , respectively;  $R(s_t, a_t)$  represents the return;  $H(\pi(\cdot | s_t))$  is the entropy; and  $\alpha$  is the temperature parameter, which determines the randomness of the strategy.

The neural network explores all possible optimal ways for the learned policy to have stronger exploration and robustness and can be used to initialize more complex tasks. In other words, when the policy has completed the focusing task in an environment, the *agent* can update the policy faster when faced with a new environment.



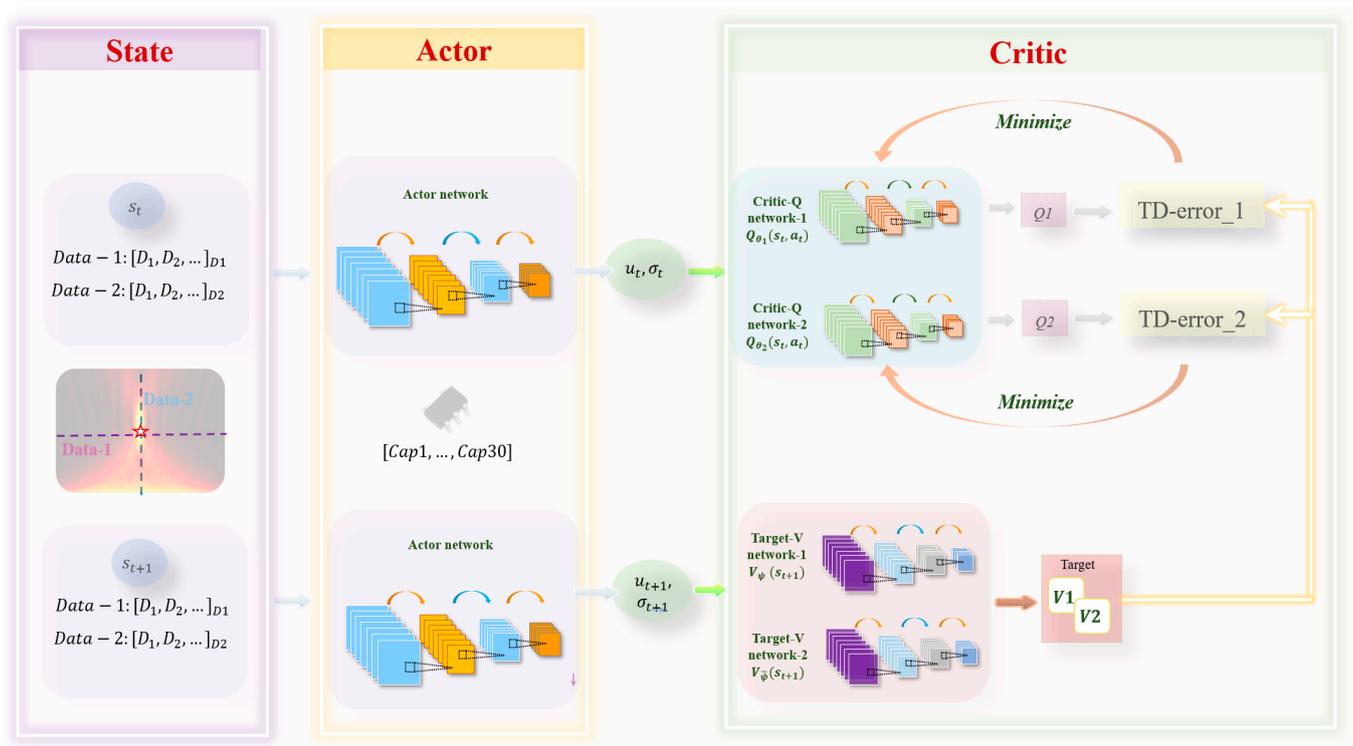
**Figure 1.** Proposed SAC-M Architecture. (a) Incident wave, metasurface, and object constitute complex environments simultaneously. (b) *Agent* continuously collects and analyzes data in real time.

Figure 2 shows the network framework of the SAC, including the state, actor, and critic. We used four-layer convolutional networks to build the actor and critic networks.  $s_t$  is the state (electric field) collected by the *agent* from the environment at time  $t$ . The actor network selects the appropriate action (capacitance matrix  $[Cap1, \dots, Cap30]$ ) based on the state, and the critic network evaluates the value of the state. For continuous actions, the actor outputs the mean ( $u_t$ ) and variance ( $\sigma_t$ ) of the action distribution.

To stabilize the training, “critic” uses two Q-value functions represented by  $\theta_1$  and  $\theta_2$ , and uses two value functions, represented by  $\bar{\psi}$  and  $\psi$ . The Target-V network represents the estimation of the state value, and the Critical-Q-network represents the estimation of the action value. In the SAC algorithm, the goal of the actor is to maximize the output action using Equation (1). The goal of the Critic-Q and Target-V networks is to make the output action value Q and state value V more accurate.

Similar to the traditional actor–critic algorithm, the network update iteration of SAC is divided into two steps: soft-policy evaluation and soft-policy improvement. In the soft-policy evaluation, the policy is fixed, and the Q value is updated using the Bellman equation until convergence occurs.

$$Q_{soft}^{\pi}(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1}, a_{t+1}} [Q_{soft}^{\pi}(s_{t+1}, a_{t+1}) - \alpha \log(\pi(a_{t+1} | s_{t+1}))] \quad (2)$$



**Figure 2.** Architecture of SAC. State ( $s_t$ ) is the data of the focal plane, and the action is the capacitance sequence:  $[Cap1, \dots, Cap30]$ . SAC network consists of one actor network and four critic networks (Q-network-1, Q-network-2, V-network-1, and V-network-2).

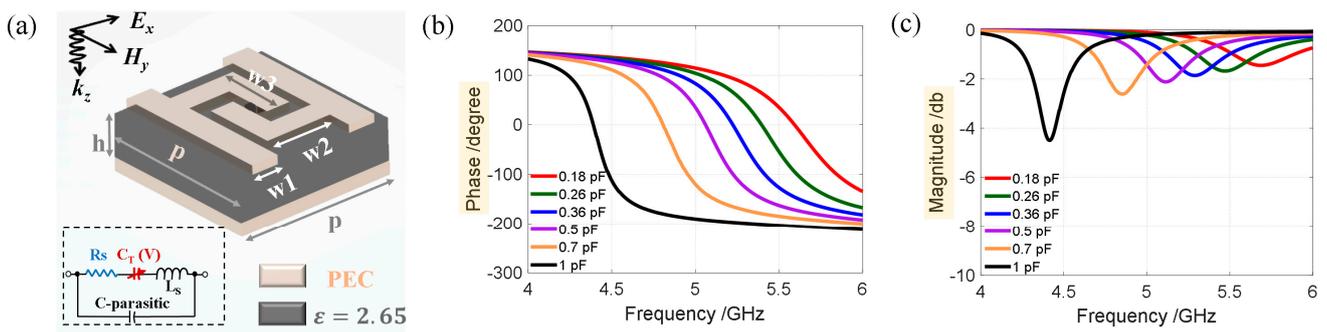
In soft-policy improvement, the policy is updated using Equation (3).

$$\pi' = \underset{\pi_k \in \Pi}{\operatorname{argmin}} D_{KL}(\pi_k(\cdot | s_t) || \frac{\exp(\frac{1}{\alpha} Q_{soft}^{\pi}(s_t, \cdot))}{Z_{soft}^{\pi}(s_t)}) \quad (3)$$

Soft-policy iteration algorithms alternate the soft-policy evaluation and soft-policy improvement steps. The detailed derivation process of the algorithm can be found in the references [41,42].

### 2.3. Element Configuration

The tunable reflective metasurface arrangement is composed of a group of unit cells with various capacitances. The physical dimensions of the proposed reconfigurable metasurface were  $300 \times 200 \text{ mm}^2$ , comprising  $30 \times 20$  unit cells along the  $x$  and  $y$  directions, where  $y$  direction was set as the periodic boundary. Figure 3a shows that the designed unit cell, which consists of a  $10 \times 10 \text{ mm}^2$  substrate ( $\epsilon_r = 2.65$ ) with 2.5 mm thickness, a fully reflective metal patch attached to the back, and a specially designed metal structure with a varactor diode. The varactor diode model is MAVR-000120-14110P, whose capacitance can be tuned between 0.14 and 1.1 pF with a parasitic resistance of approximately 2.5  $\Omega$ . Figure 3a shows an equivalent circuit diagram. The RC model is used as an equivalent diode, and the diode characteristics can be changed by adjusting capacitance. The metal sheet attached to the front is a center-symmetrical figure with a diode placed at the center. Using the commercial software CST2020 (CST Studio Suite 2020, Dassault aircraft company, France) to analyze the S-parameters of the unit cell, the frequency domain simulation mode is used, its  $x$  and  $y$  directions are set as the unit cell boundary, and the electric field is along the  $x$  direction. The S-parameters are shown in Figure 3b,c. Apparently, at the frequency  $f = 5 \text{ GHz}$ , the reflection phase changes continuously to  $320^\circ$  and the amplitude is greater than  $-2 \text{ dB}$ . Here, we show six kinds of capacitors as characteristic units, and the capacitance values are 0.18, 0.26, 0.36, 0.5, 0.7, and 1 pF.

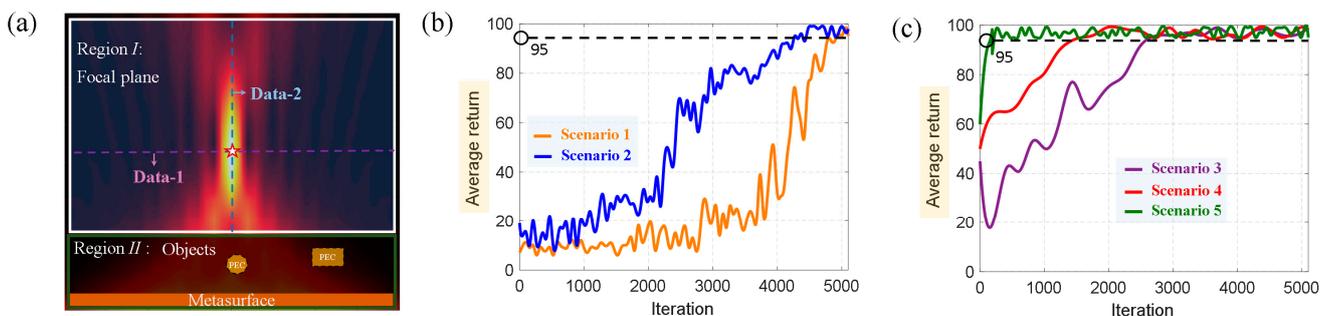


**Figure 3.** (a) 3D view of the designed unit cell.  $p = 10$  mm,  $w1 = 1.5$  mm,  $w2 = 4.2$  mm,  $w3 = 4.5$  mm,  $h = 2.5$  mm. (b,c): Reflection phase and amplitude at different capacitances and frequencies, respectively.

### 3. Results and Discussion

#### 3.1. The Training Results and Unit Cell Design

The selection of the state data is crucial for the *agent* to quickly learn a good policy. To reduce the amount of data storage, we only selected two sets of data (dotted line in Figure 4a) at the focal position (the five-pointed star) in the focal plane as the input of the *agent*, denoted as *Data* – 1 :  $[D_1, D_2, \dots, D_{75}]_{D1}$ ; *Data* – 2 :  $[D_1, D_2, \dots, D_{36}]_{D2}$ . In Figure 4a, the solid white frame is the focal plane, and any focal position in this area can be defined as an *agent-learning* target. The green solid line frame represents the area where the object and metasurface are located. The object was a perfect electric conductor (PEC). Objects of any shape and number can be placed, and the *agent* does not need to be known in advance. Before the *agent* learns, we first define the target data: *Goal* – 1 :  $[D_1, D_2, \dots, D_{75}]_{G1}$ ; *Goal* – 2 :  $[D_1, D_2, \dots, D_{36}]_{G2}$ , which were obtained by the traditional focusing method [43]. The purpose of the *agent* is to learn a policy making the *Data* – 1 and *Data* – 2 rapidly approach *Goal* – 1 and *Goal* – 2 in any scenario. Here, we defined five scenarios, as listed in Table 1.

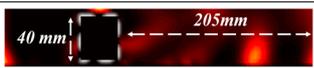


**Figure 4.** (a) Schematic illustration of focal plane and selection of training data. (b) Average returns for Scenarios 1 and 2. (c) Average returns for Scenarios 3, 4, and 5.

Figure 4b,c show the curves of the average return of the *agent* in the learning process with the number of iterations under different scenarios. A good focusing effect is generated when the average returns to 95. When the *agent* faces a new scene for the first time, large amounts of data are necessary, as the *agent* knows nothing about the environment. As shown in Figure 4b, the *agent* obtains a good policy after approximately 5000 iterations for Scenario 1 (orange curve). In Scenario 2, the *agent* continues to use the experience and policy learning in Scenario 1 to keep testing the new environment, and finally achieves a good focusing effect after approximately 4000 iterations. For the new task in Scenario 2 (blue curve), the *agent* already has a certain learning ability compared with Scenario 1. At this time, we changed the scene again, as shown in Figure 4c, and the new focusing task could be achieved after 2000 iterations in Scenario 3 (purple curve). The amount of data was twice as small as before because the *agent* combined the policies of Scenarios 1 and 2,

and the learning ability was further improved. Finally, we changed the environment to Scenarios 4 and 5. At this time, the speed of policy evaluation and policy improvement was faster. In particular, in Scenario 5, the *agent* only required a few dozen iterations to complete the focusing task. Combining the training data of different scenarios, it can be observed that the *agent* learned the solution to a set of problems and can quickly adapt to the new environment with the help of empirical knowledge.

**Table 1.** Define of different scenarios.

Scenario	Focal Position (x, z)/mm	Object
Scenario 1	(188, 130)	
Scenario 2	(150, 130)	
Scenario 3	(169, 100)	
Scenario 4	(225, 110)	
Scenario 5	(195, 150)	

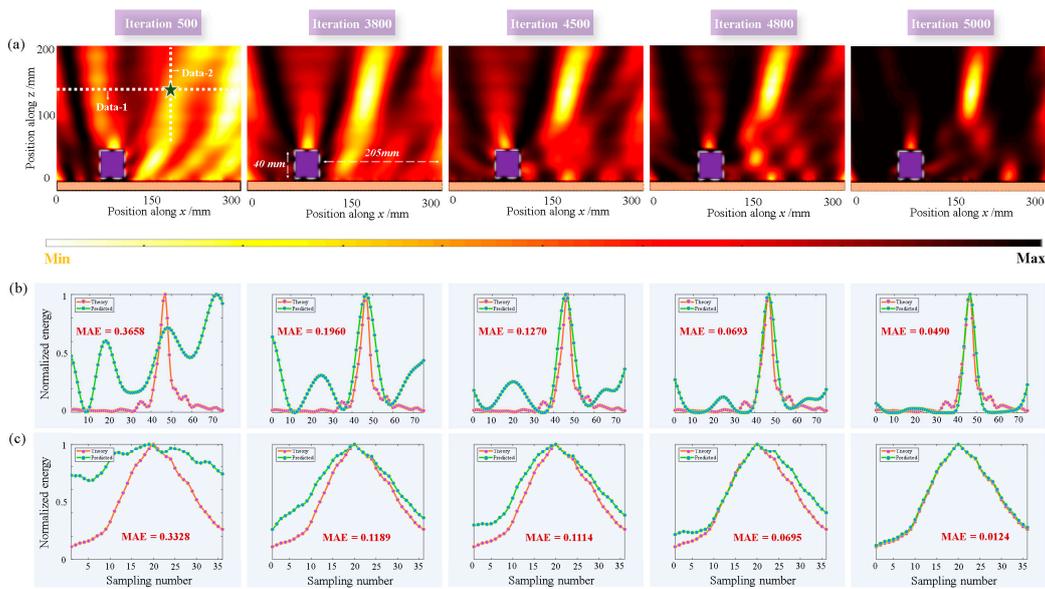
### 3.2. Adaptive Focusing Results at Different Positions

Figure 5 shows the focusing process in Scenario 1 under normal incidence. As shown in Figure 5a, the black five-pointed star marks the position of the target focus point, and the row and column data where the position is located are selected as training data (white-dotted line). The black square area represents an obstacle. In Scenario 1, an object with a side length of 40 mm was located 5 mm above the metasurface, and the object was 205 mm away from the edge of the metasurface. The five images in Figure 5a show the electric field energies at iterations of 500, 3800, 4500, 4800, and 5000 times. Scenario 1 is the unknown environment faced by the *agent* for the first time; therefore, the process of learning the policy was very long. It can be observed that when the iteration was 3800 times, a certain focusing effect was produced, but the energy was relatively scattered at this time. When the iteration reached approximately 5000 times, the energy almost converged to the position of the target focus, and the *agent* completed the task in Scenario 1. The plot in Figure 5b,c show the predicted (green line) and theoretical (red line) values for Data 1 and Data 2, respectively, and we calculated the mean absolute error (MAE) loss for both. With the continuous learning of the *agent*, the predicted value constantly approached the theoretical value, and the MAE constantly approached zero. The MAE is defined as:

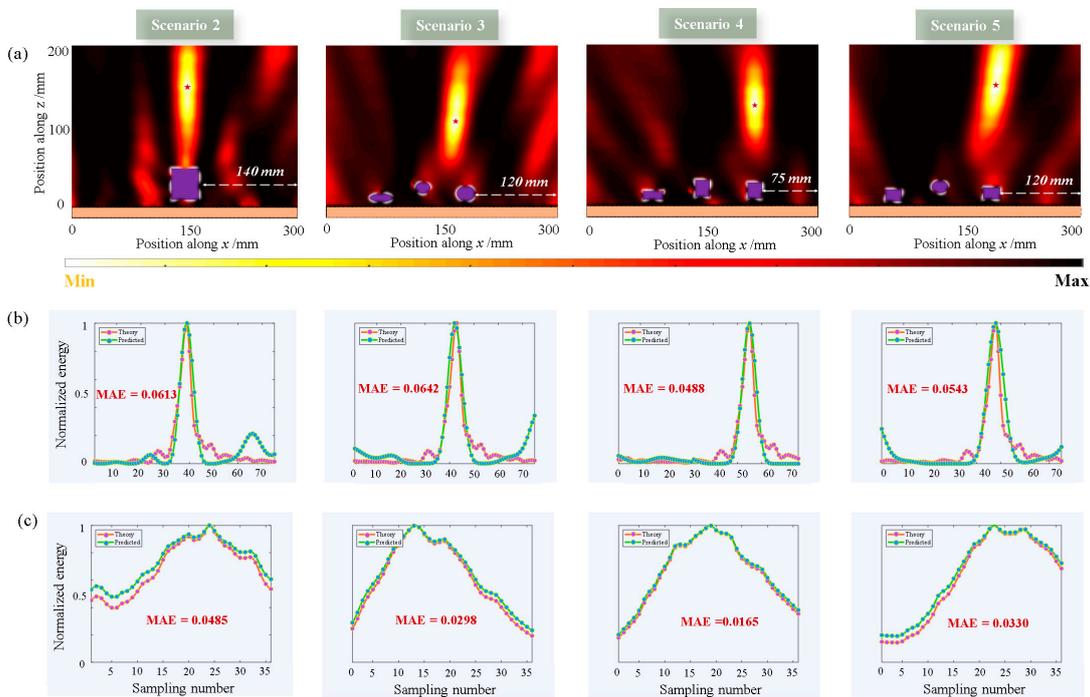
$$MAE = \frac{1}{m} \sum_{i=1}^m |E_{Theory} - E_{Predict}| \quad (4)$$

where  $m$  is the total number of iterations,  $E_{Theory}$  is the theoretical value, and  $E_{Predict}$  is the network prediction. The closer the MAE was to 0, a better the training result was realized.

Figure 6 shows the focusing results for Scenarios 2, 3, 4, and 5. The shape and number of objects are different in different scenes, and the red star is the focus position. Combining previous experiences and strategies, the *agent* can rapidly adapt to the new environment and achieve efficient focusing at the target location, as shown in Figure 6a. Figure 6b,c show the theoretical (red line) and predicted (green line) curves of Data 1 and Data 2, respectively, under different scenarios. It can be observed that the two have excellent fitting effects.



**Figure 5.** Focusing process in Scenario 1. (a) Dynamic focusing process with different iterations under normal incidence. Black five-pointed star represents the position of the target focus and the white-dotted line represents the data obtained by the *agent*. The length of the sides of the object was 40 mm. (b) Network prediction and theoretical results of Data 1 under different iterations and the MAE loss was calculated. (c) Network prediction and theoretical results for Data 2 under different iterations.



**Figure 6.** Focusing results for different scenarios. (a) Two-dimensional focusing results for different scenarios under normal incidence. Red five-point star represents the position of the target focus. (b) Network prediction and theoretical results of Data 1 under different scenarios, and the MAE loss was calculated. (c) Network prediction and theoretical results for Data 2 under different scenarios.

#### 4. Conclusions

In conclusion, combined with a soft actor–critic and reconfigurable metasurface, we proposed and designed an SAC-M-driven adaptive focusing system. The *agent* learns and improves policies in real time in changing environments, and the metasurface is guided by it and exhibits effective and robust adaptive focusing capabilities based on 1D electric-field data. The simulation and network results demonstrated that the proposed SAC-M system is highly adaptable for achieving focus at arbitrarily specified positions with multiple objects of any shape. Our novel combination of the classical EM theory and mainstream theories of RL uncovered the exciting potential of metasurfaces. The proposed SAC-M framework not only provided an idea for realizing adaptive focusing in unknown environments, but also provided a general architecture to solve more challenging problems. If similar designs can be decoded, smart metasurface systems offer incredible potential for communication and radar technologies.

**Author Contributions:** R.Z. and H.L. conceived the idea and supervised the research. R.Z. guided the theory and simulations. H.L. and C.W. designed the algorithm framework. H.L. performed the simulation verifications. H.L., C.W., S.Z., T.H., T.C. and R.Z. analyzed the data. H.L. wrote the paper. All authors discussed the results and commented on the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Natural Science Foundation of Zhejiang Province (No. LQ21F050002).

**Institutional Review Board Statement:** The study did not require ethical approval.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data are available from the author.

**Conflicts of Interest:** The authors declare no competing financial interests.

#### References

1. Yu, N.; Genevet, P.; Kats, M.A.; Aieta, F.; Tetienne, J.P.; Capasso, F.; Gaburro, Z. Light propagation with phase discontinuities: Generalized laws of reflection and refraction. *Science* **2011**, *334*, 333–337. [[CrossRef](#)] [[PubMed](#)]
2. Yu, N.F.; Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **2014**, *13*, 139–150. [[CrossRef](#)] [[PubMed](#)]
3. Sun, W.; He, Q.; Sun, S.; Zhou, L. High-efficiency surface plasmon meta-couplers: Concept and microwave-regime realizations. *Light. Sci. Appl.* **2016**, *5*, e16003. [[CrossRef](#)]
4. Li, Z.; Kim, M.H.; Wang, C.; Han, Z.; Shrestha, S.; Overvig, A.C.; Yu, N. Controlling propagation and coupling of waveguide modes using phase-gradient metasurfaces. *Nat. Nanotechnol.* **2017**, *12*, 675–683. [[CrossRef](#)] [[PubMed](#)]
5. Yang, Y.; Jing, L.; Zheng, B.; Hao, R.; Yin, W.; Li, E.; Soukoulis, C.M.; Chen, H. Full-polarization 3D metasurface cloak with preserved amplitude and phase. *Adv. Mater.* **2016**, *28*, 6866–6871S. [[CrossRef](#)] [[PubMed](#)]
6. Cai, T.; Zheng, B.; Lou, J.; Shen, L.; Yang, Y.; Tang, S.; Li, E.; Qian, C.; Chen, H. Experimental realization of a superdispersion-enabled ultrabroadband terahertz cloak. *Adv. Mater.* **2022**, *34*, 2205053. [[CrossRef](#)]
7. Tan, Q.; Zheng, B.; Cai, T.; Qian, C.; Zhu, R.; Li, X.; Chen, H. Broadband spin-locked metasurface retroreflector. *Adv. Sci.* **2022**, *9*, 2201397. [[CrossRef](#)]
8. Cai, T.; Tang, S.; Zheng, B.; Wang, G.; Ji, W.; Qian, C.; Chen, H. Ultra-wideband chromatic aberration-free meta-mirrors. *Adv. Photonics.* **2021**, *3*, 016001.
9. Lu, H.; Zheng, B.; Cai, T.; Qian, C.; Wang, Z.; Yang, Y.; Chen, H. Frequency-controlled focusing using achromatic metasurface. *Adv. Opt. Mater.* **2021**, *9*, 2001311. [[CrossRef](#)]
10. Hao, H.; Ran, X.; Tang, Y.; Zheng, S.; Ruan, W. A Single-Layer focusing metasurface based on induced magnetism. *Prog. Electromagn. Res.* **2021**, *172*, 77–88. [[CrossRef](#)]
11. Zang, X.; Dong, F.; Yue, F.; Zhang, C.; Xu, L.; Song, Z.; Chen, X. Polarization encoded color image embedded in a dielectric metasurface. *Adv. Mater.* **2018**, *30*, 1707499. [[CrossRef](#)]
12. Wang, C.; Zhang, Z.; Zhang, Y.; Xie, X.; Yang, Y.; Han, J.; Gao, F. Enhancing directivity of terahertz photoconductive antennas using spoof surface plasmon structure. *New. J. Phys.* **2022**, *24*, 073046. [[CrossRef](#)]

13. Cai, T.; Wang, G.; Tang, S.; Xu, H.; Duan, J.; Guo, H.; Zhou, L. High-efficiency and full-space manipulation of electromagnetic wave-fronts with metasurfaces. *Phys. Rev. Appl.* **2017**, *8*, 034033. [[CrossRef](#)]
14. Ding, F. A Review of Multifunctional Optical Gap-Surface Plasmon Metasurfaces. *Prog. Electromagn. Res.* **2022**, *174*, 55–73. [[CrossRef](#)]
15. Hu, Z.; He, N.; Sun, Y.; Jin, Y.; He, S. Wideband high-reflection chiral dielectric metasurface. *Prog. Electromagn. Res.* **2021**, *172*, 51–60. [[CrossRef](#)]
16. Wu, N.; Zhang, Y.; Ma, H.; Chen, H.; Qian, H. Tunable High-Q plasmonic metasurface with multiple surface lattice resonances. *Prog. Electromagn. Res.* **2021**, *172*, 23–32. [[CrossRef](#)]
17. Ee, H.-S.; Agarwal, R. Tunable metasurface and flat optical zoom lens on a stretchable substrate. *Nano. Lett.* **2016**, *16*, 2818–2823. [[CrossRef](#)]
18. Chen, K.; Feng, Y.; Monticone, F.; Zhao, J.; Zhu, B.; Jiang, T.; Qiu, C. A reconfigurable active Huygens' metalens. *Adv. Mater.* **2017**, *29*, 1606422. [[CrossRef](#)]
19. She, A.; Zhang, S.; Shian, S.; Clarke, D.R.; Capasso, F. Adaptive metalenses with simultaneous electrical control of focal length, astigmatism, and shift. *Sci. Adv.* **2018**, *2*, eaap9957. [[CrossRef](#)]
20. Wang, Q.; Rogers, E.; Gholipour, B. Optically reconfigurable metasurfaces and photonic devices based on phase change materials. *Nat. Photonics* **2016**, *10*, 60–65. [[CrossRef](#)]
21. Li, X.; Yang, H.; Shao, W.; Zhai, F.; Liu, B.; Wang, X.; Cui, T. Low-Cost and High-Performance 5-Bit Programmable Phased Array at Ku-Band. *Prog. Electromagn. Res.* **2022**, *175*, 29–43. [[CrossRef](#)]
22. Colburn, S.; Zhan, A.; Majumdar, A. Metasurface optics for full-color computational imaging. *Sci. Adv.* **2018**, *4*, eaar2114. [[CrossRef](#)] [[PubMed](#)]
23. Arbabi, E.; Arbabi, A.; Kamali, S.M.; Horie, Y.; Faraji-Dana, M.; Faraon, A. MEMS-tunable dielectric metasurface lens. *Nat. Commun.* **2018**, *9*, 812. [[CrossRef](#)]
24. Ren, Z.; Chang, Y.; Ma, Y.; Shih, K.; Dong, B.; Lee, C. Leveraging of MEMS technologies for optical metamaterials applications. *Adv. Opt. Mater.* **2020**, *8*, 1900653. [[CrossRef](#)]
25. Zhu, W.; Song, Q.; Yan, L.; Zhang, W.; Wu, P.; Chin, L.K.; Zheludev, N. Optofluidics: A flat lens with tunable phase gradient by using random access reconfigurable metamaterial. *Adv. Mater.* **2015**, *27*, 4739–4743. [[CrossRef](#)] [[PubMed](#)]
26. Kamali, S.M.; Arbabi, E.; Arbabi, A.; Horie, Y.; Faraon, A. Highly tunable elastic dielectric metasurface lenses. *Laser Photonics Rev.* **2016**, *10*, 1002–1008. [[CrossRef](#)]
27. He, Q.; Sun, S.; Zhou, L. Tunable/reconfigurable metasurfaces: Physics and applications. *Research* **2019**, *2019*, 1849272. [[CrossRef](#)]
28. Fan, Z.; Qian, C.; Jia, Y.; Wang, Z.; Ding, Y.; Wang, D.; Tian, L.; Li, E.; Cai, T.; Zheng, B.; et al. Homeostatic neuro-metasurfaces for dynamic wireless channel management. *Sci. Adv.* **2022**, *8*, eabn7905. [[CrossRef](#)] [[PubMed](#)]
29. Huang, M.; Zheng, B.; Cai, T.; Li, X.; Liu, J.; Qian, C.; Chen, H. Machine-learning-enabled metasurface for direction of arrival estimation. *Nanophotonics* **2022**, *11*, 2001–2010. [[CrossRef](#)]
30. Qian, C.; Zheng, B.; Shen, Y.; Jing, L.; Li, E.; Shen, L.; Chen, H. Deep-learning-enabled self-adaptive microwave cloak without human intervention. *Nat. Photonics* **2020**, *14*, 383–390. [[CrossRef](#)]
31. Jabir, B.; Falih, N. Deep learning-based decision support system for weeds detection in wheat fields. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 816. [[CrossRef](#)]
32. Succetti, F.; Rosato, A.; Di Luzio, F.; Ceschini, A.; Panella, M. A Fast Deep Learning Technique for Wi-Fi-Based Human Activity Recognition. *Prog. Electromagn. Res.* **2022**, *174*, 127–141. [[CrossRef](#)]
33. Brahim, J.; Loubna, R.; Nouredine, F. RNN-and CNN-based weed detection for crop improvement: An overview. *Foods Raw Mater.* **2021**, *9*, 387–396.
34. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Hassabis, D. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [[CrossRef](#)] [[PubMed](#)]
35. Ecoffet, A.; Huizinga, J.; Lehman, J.; Stanley, K.O.; Clune, J. First return, then explore. *Nature* **2021**, *590*, 580–586. [[CrossRef](#)]
36. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hassabis, D. Mastering the game of go without human knowledge. *Nature* **2017**, *550*, 354–359. [[CrossRef](#)]
37. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep reinforcement learning: A brief survey. *IEEE Signal. Process. Mag.* **2017**, *34*, 26–38. [[CrossRef](#)]
38. Fujimoto, S.; Hoof, H.; Meger, D. Addressing function approximation error in actor-critic methods. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1587–1596.
39. Chen, B.; Wang, D.; Li, P.; Wang, S.; Lu, H. Real-time actor-critic tracking. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 318–334.
40. Henderson, P.; Islam, R.; Bachman, P.; Pineau, J.; Precup, D.; Meger, D. Deep reinforcement learning that matters. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32, p. 1.
41. Haarnoja, T.; Zhou, A.; Hartikainen, K.; Tucker, G.; Ha, S.; Tan, J.; Levine, S. Soft actor-critic algorithms and applications. *arXiv* **2018**, arXiv:1812.05905.

42. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1861–1870.
43. Chen, W.T.; Zhu, A.Y.; Sanjeev, V.; Khorasaninejad, M.; Shi, Z.; Lee, E.; Capasso, F. A broadband achromatic metalens for focusing and imaging in the visible. *Nat. Nanotechnol.* **2018**, *13*, 220–226. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.