Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 250 (2024) 115–121

Procedia Computer Science

www.elsevier.com/locate/procedia

Proceedings of the 2024 International Conference on Biomimetic Intelligence and Robotics

A Spiking Neural Network Action Decision Method Inspired by Basal Ganglia

Tianyong Ao[a,b], Qiuping Liu[a,b], Le Fu[a,b,*], Yi Zhou[a,b]

[a] *School of Artificial Intelligence, Henan University, Zhengzhou 450046, China*
[b] *International Joint Research Laboratory for Cooperative Vehicular Networks of Henan, Zhengzhou 450046, China*

Abstract

Spiking Neural Networks (SNNs) have shown significant advantages in developing efficient artificial intelligence systems with cognitive characteristics, owing to their biological authenticity and temporal encoding characteristics. However, there is currently a lack of effective training algorithms for SNNs. In contrast, the brain can maintain accurate, efficient, and flexible decision-making in complex dynamic environments, mainly due to the joint function of multiple brain regions in the biological nervous system. Among them, Basal Ganglia (BG)-based reward learning mechanisms are crucial in action decision-making, but research on it is relatively scarce and most of it cannot cope with complex environments. Therefore, this paper explores the mechanism of biological reward learning and proposes an SNN-based action decision-making model with an Actor-Critic (AC) architecture. Toefficiently train the model and cope with complex situations, this paper proposes a novel cortico-striatal synaptic plasticity learning rule. Finally, the model is tested on the OpenAI Gym environment, and the experiment supported the effectiveness of the model. Compared with Deep Q-learning (DQN) and Proximal Policy Optimization (PPO) algorithms, this model demonstrates good learning performance and remains biologically plausible.

Keywords: Spiking Neural Network; Basal Ganglia; Reinforcement Learning; Cortico-Striatal synaptic plasticity ;

1. Introduction

SNNs use spiking signal encoding information and draw highly on bio-inspired neuron nodes and optimization methods. Therefore, it is expected to achieve complex cognitive functions at low power consumption. However, efficiently training SNNs is still a challenge. The non-differentiability of the activation functions of SNNs makes it difficult to train them using existing backpropagation methods. The local learning rules of synapses are biologically plausible, but they cannot help train large-scale networks. Although the ANN-SNN conversion approach has achieved

[*] Corresponding author. Tel.: +86037122822134.
E-mail address: lefu@henu.edu.cn

10.1016/j.procs.2024.11.017

good performance, it ignores the rich temporal dynamic behaviors of SNNs [8]. Therefore, compared with traditional Artificial Neural Networks (ANNs), the current field of SNNs still lacks recognized core training algorithms [13].

In contrast, the brain can achieve efficient decision-making and flexible motion control in situations of dynamic environmental changes and limited computing power, which cannot be achieved without the joint function of multiple brain regions such as the cortex, BG, and cerebellum[5, 16, 4]. Therefore, the joint function of multiple brain regions may provide a novel perspective for the study of SNN training algorithms. The BG, as the subcortical nuclei of the brain, are crucial for decision-making tasks in animals. At present, cerebellar models and some multi brain region joint models[1, 20]have been applied to robot tasks, but research on basal ganglia models is relatively scarce. Therefore, this paper focuses on the biological reinforcement learning mechanism in the BG and proposes a SNN decision-making model to provide a novel means for subsequent research.

The BG primarily achieves motor learning and reward action selection through the neurochemical dopamine(DA) driven reward learning mechanism. The classic basal ganglia model describes how DA regulates BG pathways to achieve normal control of movement. Existing computational models of BG can be broadly categorized into two types. One is mainly designed for in-depth exploration and exploitation of BG functions [17, 19, 10, 11, 3], and they are more biologically plausible, but most of them are mathematical models. The other is an ANN model inspired by the classic Cortico-Basal Ganglia-Thalamic (CBGT) circuit. Their tasks are more complex and often incorporate reinforcement learning (RL) theories to improve task completion efficiency. For example, articles [23, 21] mainly improves model performance by designing complex reward functions, but the construction of cortico-striatal synaptic plasticity learning rules is too simplified. The cortico-striatal synaptic is described as a mapping relationship from state to action, which is key to the reward learning mechanism and is mainly regulated by DA. The simulation of dopaminergic neuron activity in works [24, 7] is extremely simple and does not consider RL frameworks. The RL framework here refers to using the TD error of temporary difference learning to represent the reward prediction error in the BG model. Therefore, this paper focuses on the construction of cortico-striatal synaptic plasticity to develop a biologically plausible and efficient BG model. The main contributions of this article are as follows:

- To promote the research of SNN algorithms, a spiking basal ganglia decision-making model is established, inspired by the joint mechanism of multiple brain regions in the biological nervous system. Specifically, the BG and its reward learning mechanism were explored, and the decision-making function of CBGT was realized by synthesizing the artificial reinforcement learning theory as well as relevant data on the connectivity structure, electrophysiology and biochemistry of the BG.
- To efficiently train the BG model, a novel cortico-striatal synaptic plasticity learning rule is proposed to learn the correct mapping relationship between states and actions. Different from other BG-based computational models, our learning rule focuses on the exploration strategy, which can avoid falling into local optima when facing complex environments. In addition, real-time feedback rewards are utilized to directly regulate the synapses to accelerate the learning speed.
- To verify the validity of the model, the model performance is tested on FrozenLake-v1 environments in OpenAI Gym and the algorithm complexity of the model is analyzed. The results indicate that the proposed model attains high performance and retains low computational complexity when compared to DQN and PPO algorithms.

2. Neuroanatomy of Basal Ganglia Circuitry

BG is a group of subcortical nuclei in the brain of vertebrates, located at the bottom of the forebrain and the top of the midbrain. It receives inputs from the cerebral cortex and projects back to the frontal cortex through the thalamic nucleus. This anatomical structure implies that the BG are located in the main position that affects the executive function of the forebrain, and have important implications for behavior selection, motor control, working memory, and cognition [14]. The main anatomical structure and components of BG are shown in Fig. 1(a). In the nervous system, the CBGT circuit processes the main decision-making process, and Fig. 1(b) is a schematic diagram of the CBGT loop. D1-MSNs guide the direct pathway (Cx-Str-GPi) to inhibit the activity of the output nucleus GPi, thereby reducing the inhibition of projection to the thalamus and cortical neurons, promoting motor behavior. In contrast, D2-MSNs guide the indirect pathway (Cx-Str-GPe-STN-GPi) to inhibit GPe and STN, promoting the activity of GPi by stimulating synapses in the subthalamic nucleus, thereby enhancing the inhibition of thalamus and cortical neurons,

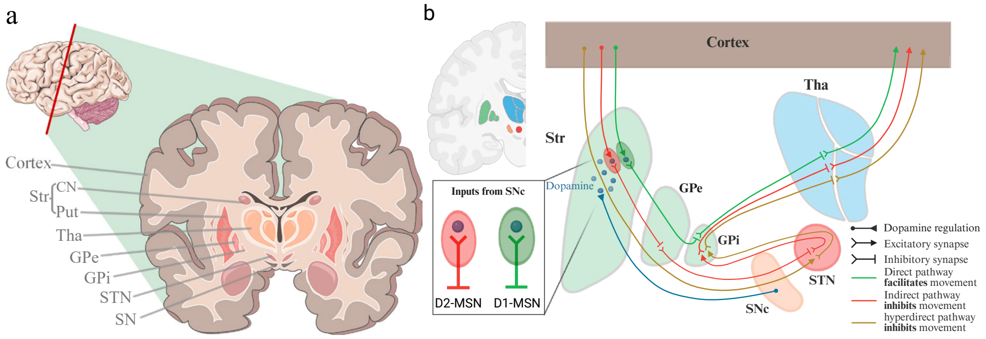

Fig. 1. Diagram of BG structure. (a) BG anatomical structure; (b) CBGT loop (D1/2-MSNs: Striatal medium spiny neurons expressing D1/2-type dopamine receptors, Str: Striatum, Tha: thalamus, GPe: globus pallidus external, GPi: globus pallidus internal, STN: subthalamic nucleus, SNc: Substantia Nigra), created with BioRender.com.

and preventing motor behavior. In addition, the cortex directly projects to the STN (hyperdirect pathway:Cx-STN-GPi) and then generates strong diffuse excitatory projections to the GPi to terminate any selection.

Based on connectivity patterns and functional roles, the striatum has been divided into two parts: the dorsal striatum (DS) and the ventral striatum (VS). The VS is located at the junction of the basal nuclei and the limbic system and supports the learning of state values [9]. It receives information about the current environment, stimuli and their emotional value, which is efferent to dopaminergic neurons, so that state value and reward information is carried by DA in the SNc and is used to update reward prediction error(RPE) [15]) to modulate the VS and DS.

3. Methods

3.1. Spiking Neuron Model

Leaky Intergrate and Fired (LIF) neurons are currently the simplest and most commonly used model for describing the dynamics of spiking neurons, and their membrane potential updates are as follows:

$$\begin{cases} \tau\frac{dV_i(t)}{dt} = -V_i(t) - I_i(t) \\ V_i(t) = V_{reset}, O_i(t) = 1, \;\; if \;\; V_i(t) > V_{th} \\ I_i(t) = \sum_j^N W_{ij}O_j(t) \end{cases} \tag{1}$$

Among them, $V_i(t)$ is the neuronal membrane potential at time t, τ is the time constant, and $I_i(t)$ denotes the pre-synaptic input which is determined by the pre-neuronal activities or external injections and the synaptic weight (W_{ij}) from the *jth* neuron in pre-synaptic layer to the *ith* neuron in the post-synaptic layer. When the $V_i(t)$ exceeds a given threshold V_{th}, the neuron emits a spike ($O_i(t) = 1$) and resets its potential to V_{reset}.

3.2. Action Decision Model

Inspired by the function of the VS, the constructed basal ganglia decision-making model adopts a reinforcement learning AC network structure, and its overall architecture is shown in Fig. 2(a). The Critical network is a traditional ANN used to calculate state value, while the Actor is a SNN, mainly simulating the function of the CBGT loop.

Compared with previous studies on BG computational models, our model only utilizes 16 prefrontal cortex neurons and 8 MSNs to encode states and actions as Fig. 2(b), and proposes a cortico-striatal synaptic plasticity update rule incorporating exploration strategies. Therefore, our BG network is lightweight and can solve more complex tasks. In the proposed model, state information is conveyed from the cortex to the striatum, where the direct and indirect pathways guided by D1-MSNs and D2-MSNs compete with each other and converge at the BG output nucleus GPi to complete action selection. Finally, their choice is sent to the thalamo-motor cortex. SNc receives reward information and state value to calculate TD error, and then adjusts the synaptic weights between BG and PFC for learning.

Note that the Actor network in this paper is more lightweight than the general reinforcement learning AC algorithm. The specific network structure in the brain itself has certain prior knowledge, so by simulating the network structure

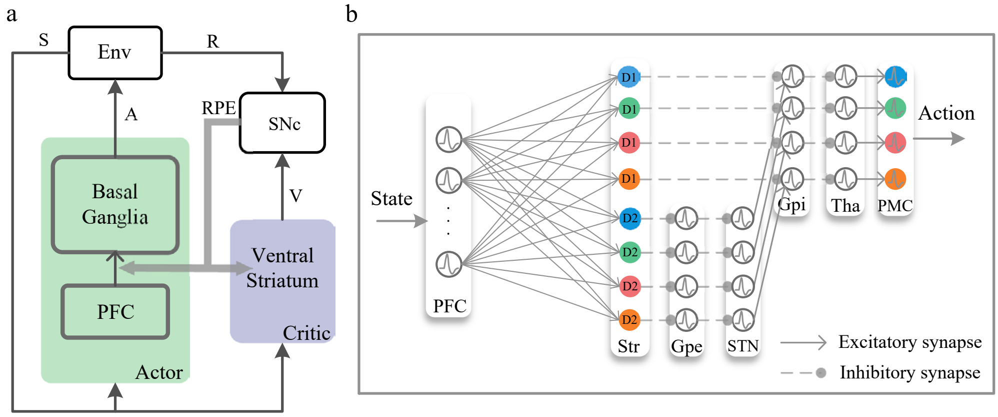

Fig. 2. The action decision model. (a) The spiking basal ganglia model with Actor-Critic architecture; (b) The spiking neural network structure based on BG(PFC:prefrontal cortex, PMC:premotor cortices)

of specific brain regions, the prior knowledge features can be quickly transferred and generalized to ANNs. In other words the simulation of the unique topology of the basal ganglia brings certain advantages to our model.

3.3. Cortico-Striatal Synaptic Plasticity Learning Rule

The contribution of BG to learning behavior can be intuitively analyzed within the framework of RL. In this process, the striatum is an input area to the BG and learns to map information about the state of the world/body onto a variant of the action that generate rewards [9]. The output GPi of BG directly affects the control circuits of the midbrain, brainstem, and motor cortex. Biological reinforcement learning is believed to be a mechanism for establishing reward action responses, based on communication between midbrain dopaminergic neurons and the striatum. In other words, it is the regulatory process of DA on cortical-striatal synapses. Research has shown that the activity of SNc exhibits astonishing similarities with the TD error of temporary differential learning [12, 2]. Therefore, many DA signals in models are represented by TD errors as Eq. (2).

$$\delta(t) = r(t) - V(S_t) + \gamma V(S_{t+1}) \tag{2}$$

Where γ is the discount rate and $r(t)$ is the reward. The state value $V(S_t)$ is calculated by the Critical network.

This study is based on previous work in the field of neuroscience [11, 18]. Notably, only direct and indirect pathways are considered. This is because the hyperdirect pathway as a braking device for decision-making does not function in the FrozenLake-v1 environment. In recent years, a large number of BG models have been developed. However, although they can achieve the functions of BG, the specific coordination process of the three pathways of BG has not been interpreted in previous models. Therefore, combining neuroscience and RL theory may be a beneficial approach for exploring the mechanism of biological reward learning.

In mammals, instrumental behaviour includes goal-directed behaviour and habitual behaviour [6], where goal-directed behavioral systems enable animals to adapt flexibly to changes in the environment. However, the currently proposed BG model is a habitual behaviour system based on model-free RL, which is prone to falling into local optima when facing complex environments. Therefore this paper introduces an exploration strategy into the synaptic plasticity learning rule to improve the model's ability to cope with complex environments. In addition rewards are directly utilized to modulate cortico-striatal synapses to speed up learning.

The cortico-striatal synaptic plasticity learning rule proposed in this article mainly follows the Hebb rule regulated by dopamine, and the dynamic formula is as follows:

$$\begin{aligned} \Delta w_{PFC-D1m} &= \lambda_{D1} \cdot \delta(t) \cdot PFC \cdot D1_m + \alpha \cdot r \cdot D1_m + \beta \cdot \Phi \cdot D1_m \\ \Delta w_{PFC-D2m} &= \lambda_{D2} \cdot \delta(t) \cdot PFC \cdot D2_m + \alpha \cdot r \cdot D2_m + \beta \cdot \Phi \cdot D2_m \end{aligned}, \tag{3}$$

Where $\Delta w_{PFC-D1m}$ and $\Delta w_{PFC-D2m}$ are the amount of synaptic strength change from cortex to D1-MSNs and D2-MSNs, respectively. PFC, $D1_m$ and $D2_m$ represent the spiking activity of the prefrontal cortex and D1-MSNs and D2-MSNs. Reward r is negative or positive when the agent dies or reaches the target point. λ_{D1}, λ_{D2}, α, and β are

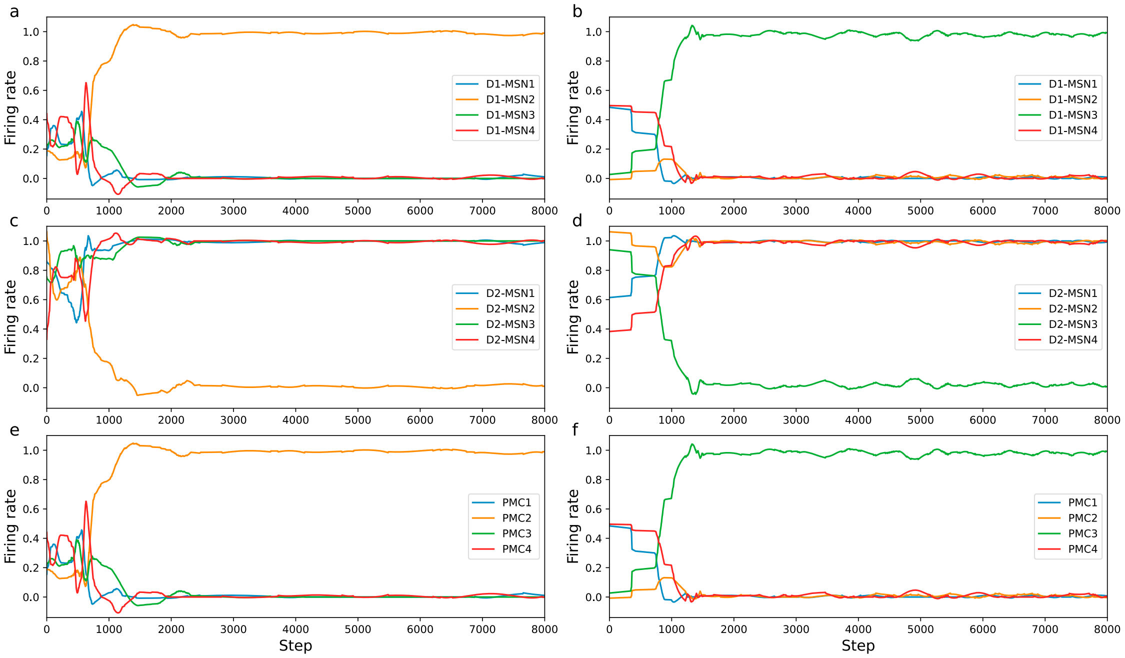

Fig. 3. Activity status of D1-MSNs, D2-MSNs and PMC under states 0(left) and states 8(right).

learning rates. The last term in the formula is the exploration strategy, and the exploration probability of the action will decrease with the increase of the number of attempts made, ensuring that the agent attempts all actions in the early stages of training. The formula for Φ is calculated as follows:

$$\Phi = \sqrt{2 * \frac{\ln(N)}{n}} \tag{4}$$

Where N is the number of attempts for all actions in a certain state; n is the number of times a certain action has been attempted.

4. Experiments

To evaluate the model, its performance on the FrozenLake-v1 environment from the OpenAI gym is measured and compared to the RL algorithm. We perform all experiments with BrainCog [22] library that supports accelerated and memory-efficient training on GPU(Quadro RTX 6000) machines.

4.1. Parameter initialization

The initialization of parameters, such as weights, thresholds, etc., is crucial for stabilizing the firing activity of the entire network. In the Actor network of our model, the two cortical-striatal weight matrices need to be initialized. One is randomly initialized, representing the model's random selection of actions at the beginning of training, while the other needs to meet certain constraint conditions as Eq. (5), mainly to cooperate with direct pathways to ensure stable firing of premotor cortical neurons. Unlike the biological nervous system, the striatum in this study only encodes four actions and is mutually exclusive, meaning that D1-MSNs and premotor cortical neurons can only activate one. Therefore, a threshold adaptive mechanism is introduced, where the threshold is mainly controlled by the product of presynaptic activity and weights.

$$\begin{cases} W_1 = (w_1^0, w_1^1, w_1^2, w_1^3) \\ W_2 = (w_1^3, w_1^2, w_1^1, w_1^0) \end{cases}, w_1^0 > w_1^1 > w_1^2 > w_1^3 \tag{5}$$

Other parameter settings are as follows: τ takes the value of 1; V_{reset} takes the value of 0; γ takes the value of 0.99; α takes the value of 0.19, and λ and β are function that decay with epoch.

4.2. Experimental results

In previous research on BG models, the tasks used for testing models are mostly binary action selection tasks or self built environments. The former is simplistic, while the latter improves the performance of the model mainly by designing complex reward functions. Therefore, to verify the superiority of the synaptic plasticity learning rule itself, the FrozenLake-v1 environment of OpenAI gym is chosen for model performance testing.

To demonstrate more intuitively the learning effect of the model, the activities of D1-MSN, D2-MSN, and PMC are visualized for states 0 and 8 as Fig. 3. The activities of PMC represent the final choices made by the agent as Fig. 3(e),(f). The higher activity of PMC0 represents the selection action 0, and so on. The learning of reward actions is reflected in the cortical-striatal synaptic intensity. In the early stages of training, the agent is in the exploratory stage, and the firing of D1-MSNs and D2-MSNs is chaotic. When the intelligent agent chooses reward actions, the DA signal emitted by SNc will enhance the synaptic strength of PFC with corresponding D1-MSNs and D2-MSNs, thereby increasing the activation of D1-MSNs responding to reward actions and D2-MSNs corresponding to non-reward actions as Fig. 3(a)-(d), and the activation of the two pathways will lock in the selection of reward actions.

In this game, our Critic network consists of three fully connected layers, and the specific structure of the Actor network is shown in Fig. 2. The latest performance of DQN, PPO, and the algorithms in this paper on the FrozenLake-v1 game is illustrated in Fig. 4. As shown in the figure, the proposed model has a faster learning speed and more stable learning performance. This is mainly due to our improvements in the learning mechanism and the network structure of the BG, which enables the model to learn state action mapping relationships more quickly. The PPO algorithm, although achieving high reward values, is not stable. We analyzed the algorithm complexity of each epoch in the Actor network during the learning process of the PPO algorithm and the model proposed in this paper as Fig. 5. The results indicate that the Actor network in this article has lower complexity. On the one hand, due to our improvement of the network structure of the BG , our model has fewer neurons; On the other hand, The proposed model uses Hebb learning rules without feedback propagation process.

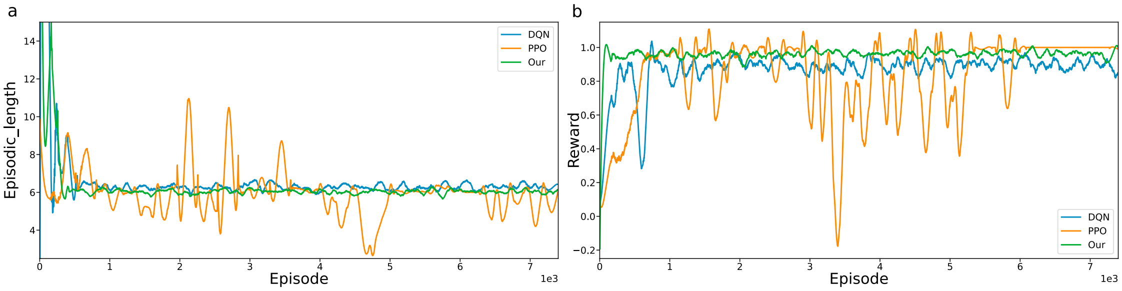

Fig. 4. Training curves of DQN, PPO, and our model in FrozenLake-v1 environment. (a) episodic length; (b) reward. Each curve is smoothed using a Savitzky Golay filter to improve readability.

Algorithms	FF	FB
PPO	$O(mg + gk + kn)$	$O(nk + kg + gm)$
Our	$O(2mn + 6n^2)$	----

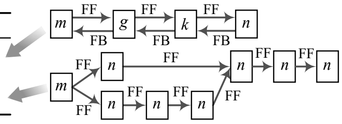

Fig. 5. Algorithmic complexity O(·) in each epoch during learning. It includes feedforward propagation (FF) and feedback propagation (FB). m, n, g and k represents the number of neurons in each layer of the network($m = 16$, $n = 4$, $g = 64$, $k = 64$).

5. Conclusion

Inspired by the learning mechanism of the brain, this paper explored the reward learning mechanism based on the BG and proposed a SNN decision-making model with a reinforcement learning AC architecture. To address a complex environment, a novel cortical-striatal synaptic plasticity learning rule is proposed. Firstly, an exploration strategy is introduced into the cortical-striatal synaptic plasticity to prevent the model from falling into local optima when facing complex environments. Secondly, rewards from the environmental feedback are utilized to directly modulate the synapses, thereby accelerating the learning speed. Finally, testing is performed in the OpenAI Gym environment.

The proposed model demonstrates good performance compared with DQN and PPO algorithms, particularly in terms of learning speed, and exhibited stabler in subsequent decision-making. This proved the validity and competitiveness of the spiking basal ganglia model and highlighted the significant enlightening value of the multi-scale network architecture and learning mechanisms in the biological nervous system for the investigation of SNN models. In future work, we will consider model testing in small robotics situations with energy constraints, providing opportunities for rapid and efficient inference on the pervasive low-power devices.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China(62176088, 62303160), the Program for Science & Technology Development of Henan Province(232102211011, 242102220111), the International Strategic Innovative Project of National Key Research & Development Program of China(2023YFE0112500) and the Key Project of Science and Technology Research of the Education Department of Henan Province(23A520007).

References

[1] Abadia, I., Naveros, F., Garrido, J.A., Ros, E., Luque, N.R., 2019. On robot compliance: A cerebellar control approach. IEEE transactions on cybernetics 51, 2476–2489.

[2] Averbeck, B.B., Costa, V.D., 2017. Motivational neural circuits underlying reinforcement learning. Nature Neuroscience 20, 505–512.

[3] Baladron, J., Nambu, A., Hamker, F.H., 2019. The subthalamic nucleus-external globus pallidus loop biases exploratory decisions towards known alternatives: a neuro-computational study. European Journal of Neuroscience 49, 754–767.

[4] Baladron, J., Vitay, J., Fietzek, T., Hamker, F.H., 2023. The contribution of the basal ganglia and cerebellum to motor learning: A neuro-computational approach. PLoS computational biology 19, e1011024.

[5] Caligiore, D., Arbib, M.A., Miall, R.C., Baldassarre, G., 2019. The super-learning hypothesis: Integrating learning processes across cortex, cerebellum and basal ganglia. Neuroscience & Biobehavioral Reviews 100, 19–34.

[6] Dickinson, A., Balleine, B., 2002. The role of learning in the operation of motivational systems. Stevens handbook of experimental psychology 3, 497–533.

[7] González-Redondo, Á., Garrido, J., Arrabal, F.N., Kotaleski, J.H., Grillner, S., Ros, E., 2023. Reinforcement learning in a spiking neural model of striatum plasticity. Neurocomputing 548, 126377.

[8] Guo, Y., Huang, X., Ma, Z., 2023. Direct learning-based deep spiking neural networks: a review. Frontiers in Neuroscience 17, 1209795.

[9] Kim, H.F., Hikosaka, O., 2015. Parallel basal ganglia circuits for voluntary and automatic behaviour to reach rewards. Brain 138, 1776–1800.

[10] Maith, O., Baladron, J., Einhäuser, W., Hamker, F.H., 2023. Exploration behavior after reversals is predicted by stn-gpe synaptic plasticity in a basal ganglia model. Iscience 26.

[11] Mulcahy, G., Atwood, B., Kuznetsov, A., 2020. Basal ganglia role in learning rewarded actions and executing previously learned choices: Healthy and diseased states. PLoS One 15, e0228081.

[12] Neftci, E.O., Averbeck, B.B., 2019. Reinforcement learning in artificial and biological systems. Nature Machine Intelligence 1, 133–143.

[13] Niu, L.Y., Wei, Y., Liu, W.B., Long, J.Y., Xue, T.h., 2023. Research progress of spiking neural network in image classification: a review. Applied intelligence 53, 19466–19490.

[14] Packard, M.G., Knowlton, B.J., 2002. Learning and memory functions of the basal ganglia. Annual review of neuroscience 25, 563–593.

[15] Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593–1599.

[16] Silkis, I., 2022. Mechanisms of functioning of the connectome including the neocortex, hippocampus, basal ganglia, cerebellum, and thalamus. Neuroscience and Behavioral Physiology 52, 1017–1029.

[17] Song, J., Lin, H., Liu, S., 2023. Basal ganglia network dynamics and function: Role of direct, indirect and hyper-direct pathways in action selection. Network: Computation in Neural Systems 34, 84–121.

[18] Topalidou, M., Kase, D., Boraud, T., Rougier, N.P., 2018. A computational model of dual competition between the basal ganglia and the cortex. eneuro 5.

[19] Vich, C., Dunovan, K., Verstynen, T., Rubin, J., 2020. Corticostriatal synaptic weight evolution in a two-alternative forced choice task: a computational study. Communications in Nonlinear Science and Numerical Simulation 82, 105048.

[20] Xing, D., Yang, Y., Zhang, T., Xu, B., 2023. A brain-inspired approach for probabilistic estimation and efficient planning in precision physical interaction. IEEE Transactions on Cybernetics 53, 6248–6262.

[21] Zeng, Y., Wang, G., Xu, B., 2017. A basal ganglia network centric reinforcement learning model and its application in unmanned aerial vehicle. IEEE Transactions on cognitive and developmental systems 10, 290–303.

[22] Zeng, Y., Zhao, D., Zhao, F., Shen, G., Dong, Y., Lu, E., Zhang, Q., Sun, Y., Liang, Q., Zhao, Y., Zhao, Z., Fang, H., Wang, Y., Li, Y., Liu, X., Du, C., Kong, Q., Ruan, Z., Bi, W., 2023. BrainCog: A spiking neural network based, brain-inspired cognitive intelligence engine for brain-inspired AI and brain simulation. Patterns , 100789.

[23] Zhao, F., Zeng, Y., Wang, G., Bai, J., Xu, B., 2018a. A brain-inspired decision making model based on top-down biasing of prefrontal cortex to basal ganglia and its application in autonomous uav explorations. Cognitive Computation 10, 296–306.

[24] Zhao, F., Zeng, Y., Xu, B., 2018b. A brain-inspired decision-making spiking neural network and its application in unmanned aerial vehicle. Frontiers in neurorobotics 12, 56.