

Article

Research on the Supporting Dynamics and Adaptive Intelligent Control Method for Hydraulic Support

Xuliang Lu ^{1,2}, Lin Zhang ³  and Dong Wei ^{4,*} 

¹ College of Tobacco Science, Henan Agricultural University, Zhengzhou 450046, China; xulianglu@henau.edu.cn

² College of Mechanical and Electrical Engineering, Henan Agricultural University, Zhengzhou 450046, China

³ School of Robotics Engineering, Yangtze Normal University, Chongqing 408100, China; linzhang@yznu.edu.cn

⁴ School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China

* Correspondence: weidongcmee@cumt.edu.cn

Abstract

The intelligent level of hydraulic support directly impacts the safe mining capacity and efficiency of the entire fully mechanized mining face. Adaptive intelligent control of its supporting pose, to adapt to the complex and ever-changing geological conditions of coal seams and variations in roof characteristics, is crucial for intelligent mining. This paper proposes and validates a novel adaptive adjustment method for the supporting pose, utilizing hydraulic support dynamics and transfer reinforcement learning. First, the supporting dynamics based on the coupling relationship between the hydraulic support and the surrounding rock of the coal seam are analyzed, and a supporting reinforcement learning model based on the Markov Decision Process is designed. Based on this model, a gradient-optimized Proximal Policy Optimization method is proposed, and a virtual dynamic simulator is built for training a supporting pose control policy. To transfer the virtually trained strategy to real hydraulic supports for practical application and to bridge the gap between support strategy simulation and reality, a progressive neural network architecture is introduced to mitigate the execution gap of support strategies under real-world conditions. Experimental results demonstrate that the proposed method can effectively and autonomously adjust the supporting pose to adapt to changes in the complex and variable coal seam roof. Furthermore, this work provides a theoretical foundation and practical engineering application for the development of intelligent support robots in coal mines.

Keywords: mining; hydraulic support dynamics; Markov decision process; transfer reinforcement learning; adaptive intelligent control



Academic Editor: Carmine Maria Pappalardo

Received: 30 March 2026

Revised: 23 April 2026

Accepted: 24 April 2026

Published: 29 April 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Currently, coal mine intelligentization has become an inevitable pathway for China to achieve safe, efficient, and green coal production. However, due to the complexity of coal seam distribution and harsh mining environments [1], the level of intelligence and reliability of mining equipment remains limited. With the development of artificial intelligence, the coal mining industry is increasingly integrating intelligence into coal production and operations to improve the automation and intelligence levels of mining operations [2]. As a core component of intelligent coal mining faces, hydraulic supports primarily function to stabilize the coal roof of the surrounding rock through its canopy,

thereby providing a safe working space for workers and other mining equipment [3]. In recent decades, the increasing mining depth has led to increasingly complex geological environments, which poses significant challenges to hydraulic supports in adapting to the constantly changing coal seam roofs [4,5].

In recent years, scholars and research institutions have been striving to address the challenge of supporting surrounding rock roofs by monitoring the support force and posture of hydraulic support, combined with hydraulic support kinematics [6]. Hydraulic support dynamics, including forward kinematics, inverse kinematics, and statics, are the foundation for its pose perception and regulation. Support mechanics monitoring mainly focuses on analyzing the roof state, roof movement patterns, and methods for determining reasonable hydraulic support resistance [7]. The WELSON estimation method described in [8] shows that the stability of the roof can be improved by adopting reasonable support techniques to strengthen the rock and redistribute high stresses. The empirical formula for support intensity in [9] lays the foundation for designing mechanisms to support surrounding rock. The Chinese theoretical method described in [10] qualitatively studies the coupling relationship between hydraulic supports and the surrounding rock roof of coal seams. Current posture monitoring research is divided into two areas: using contact sensors to measure the posture of hydraulic supports [11], and using virtual reality technology to model virtual prototypes [12,13]. This includes obtaining key sensor data from hydraulic supports through digital twin technology [14] to virtually monitor their motion state [15]. In hydraulic support posture regulation and prediction, ref. [16] proposes an LSTM-based method for predicting tail-beam inclination, addressing its control in adaptive top-coal caving with electro-hydraulic systems. In [3], refined algorithms using the Newton–Raphson, secant, and Broyden methods are developed to solve posture and height, supporting rapid control and intelligent management. For pose detection and model prediction, ref. [17] establishes models for both self and relative position and pose based on the spatial three-point plane principle. In practice, the canopy of the hydraulic support directly contacts and interacts with the surrounding rock roof to maintain dynamic equilibrium. Therefore, analyzing only support kinematics or mechanics cannot solve the problem of adaptive adjustment of the supporting pose. However, the interaction information between the hydraulic support and the surrounding rock roof can facilitate autonomous adjustment of the supporting pose to accommodate continuously changing roof conditions.

Reinforcement learning is mainly used to solve autonomous decision-making and control problems in the process of interacting with the environment [18], providing a new method for autonomously adjusting the supporting pose of hydraulic support when interacting with the coal seam roof. It has been widely applied in fields such as autonomous driving [19,20], robot control [21,22], game competitions [23,24], industrial automation [25,26], and drone collaboration [27]. At the same time, in order to solve the issue of insufficient data samples in the field of real robot control, transfer learning based on sim-to-real technology is adopted, enabling policies trained via reinforcement learning in simulation to be transferred to real-world scenarios [28]. For instance, Rabinowitz introduced Progressive Neural Networks, a novel architecture specifically designed to facilitate transfer across sequential tasks. This approach allows for rapid policy learning on real robots [29], providing practical support for sim-to-real autonomous adjustment of hydraulic support postures.

Existing studies on hydraulic support posture control largely rely on fixed strategies or models under single operating conditions, with limited adaptability to dynamic roof variations and insufficient attention to sim-to-real transfer. To address this, this paper proposes a support control method based on transfer reinforcement learning [30,31], which enables policy learning and deployment within a simulation-to-real framework. A

dynamics-based virtual environment is constructed for policy training, and the learned policy is then transferred to real hydraulic supports. Compared with existing methods, the proposed approach better adapts to complex conditions and improves the stability and generalization of posture control.

The remainder of this paper is organized as follows: Section 2 analyzes the kinematics of the hydraulic support and its supporting environment. Section 3 establishes a reinforcement learning model for the supporting pose based on a Markov decision process. Section 4 presents an autonomous decision-making and control method for the supporting pose. Section 5 develops a support virtual dynamic simulator for policy training and analyzes the experimental results. Finally, Section 6 concludes the paper.

2. Kinematic Analysis of Hydraulic Supports and the Supporting Environment

In an intelligent coal mining working face, the hydraulic support serves as the primary structure to ensure a safe production space by supporting the roof and protecting the coal wall while relying on the floor of the working face. However, the hydraulic support operates as part of a dynamic equilibrium system, characterized by interaction and mutual restriction with the surrounding rock. Furthermore, the supporting pose continuously adjusts to changes in the coal seam roof, ensuring adaptability to the surrounding rock environment. The support attitude can broadly be classified into three states: head-up posture, horizontal posture, and head-down posture, as illustrated in Figure 1.

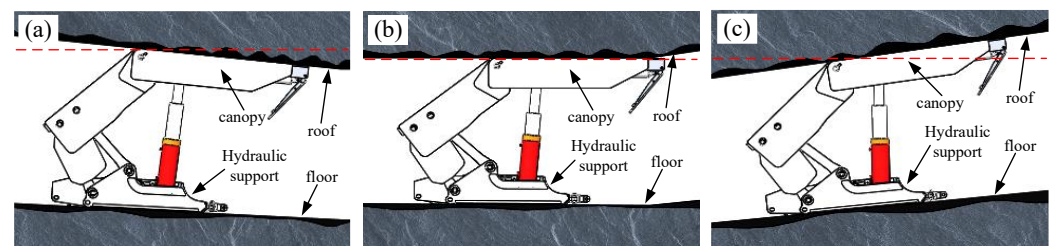


Figure 1. Support attitude of hydraulic support: (a) head down posture; (b) horizontal posture; (c) head up posture.

2.1. Kinematics Analysis of Hydraulic Support

The operating principle of the hydraulic support for maintaining the roof stability involves a fixed base, while the canopy's posture is adjusted via the leg and stabilizing ram. The motion analysis of the hydraulic support during the supporting process can be modeled as a planar multi-link mechanism. The kinematic diagram of this mechanism, as shown in Figure 2, exhibits two degrees of freedom. In this framework, the base serves as the frame, while the leg and stabilizing ram function as the drivers of the planar linkage, with the canopy acting as the follower. The solution model for the supporting pose can be represented using a functional, as follows:

$$y = \mathcal{F}(s) \quad (1)$$

where $y = [h \ \theta \ \varphi \ \beta]$ represents the support posture variables; $s = [s_1, s_2]$ represents the driving variables; and $\mathcal{F}: \mathbb{R}^2 \rightarrow \mathbb{R}^4$ is a nonlinear vector operator, which is the kinematic model of the hydraulic support.

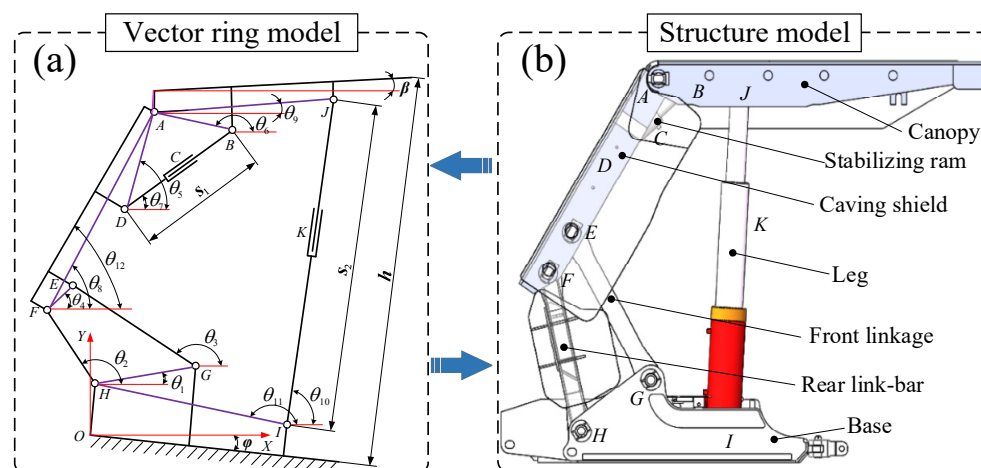


Figure 2. The kinematic diagram of mechanism of hydraulic support: (a) schematic of the closed-loop vector equations describing the positional relationships of the components in the hydraulic support mechanism; (b) schematic of the topology and connection relationships of the hydraulic support mechanism.

Based on the vector ring model of the hydraulic support illustrated in Figure 2, the vector ring equations are derived as follows:

$$\begin{cases} \overrightarrow{HG} + \overrightarrow{GE} = \overrightarrow{HF} + \overrightarrow{FE} \\ \overrightarrow{DA} + \overrightarrow{AB} = \overrightarrow{DB} \\ \overrightarrow{IH} + \overrightarrow{HF} + \overrightarrow{FA} + \overrightarrow{AJ} = \overrightarrow{IJ} \end{cases} \quad (2)$$

The corresponding complex equation form of the vector ring equation, as outlined in Equation (2) is presented as follows:

$$\begin{cases} L_{HG}e^{j\theta_1} + L_{GE}e^{j\theta_3} = L_{HF}e^{j\theta_2} + L_{FE}e^{j\theta_4} \\ L_{DA}e^{j\theta_5} - L_{AB}e^{j\theta_6} = L_{DB}e^{j\theta_7} \\ L_{IH}e^{j\theta_{11}} + L_{HF}e^{j\theta_2} + L_{FA}e^{j\theta_8} + L_{AJ}e^{j\theta_9} = L_{IJ}e^{j\theta_{10}} \end{cases} \quad (3)$$

Let $\theta_1 = \alpha_1 - \varphi$, $\theta_{11} = \alpha_2$, $\theta_5 = \alpha_3 + \theta_{12}$, $\theta_8 = \alpha_4 + \theta_{12}$, $\theta_4 = \alpha_5 + \theta_{12}$, $\theta_9 = \alpha_6 + \beta$, and $\theta_6 = \alpha_7 + \beta$, where α_1 , α_2 , α_3 , α_4 , α_5 , α_6 and α_7 represent known quantities that can be calculated based on the geometrical dimensions of the hydraulic support structure. Consequently, Equation (3) is expressed as follows:

$$\begin{cases} L_{HG}e^{j(\alpha_1 - \varphi)} + L_{GE}e^{j\theta_3} = L_{HF}e^{j\theta_2} + L_{FE}e^{j(\alpha_5 + \theta_{12})} \\ L_{DA}e^{j(\alpha_3 + \theta_{12})} - L_{AB}e^{j(\alpha_7 + \beta)} = L_{DB}e^{j\theta_7} \\ L_{IH}e^{j\alpha_2} + L_{HF}e^{j\theta_2} + L_{FA}e^{j(\alpha_4 + \theta_{12})} + L_{AJ}e^{j(\alpha_6 + \beta)} = L_{IJ}e^{j\theta_{10}} \end{cases} \quad (4)$$

The Euler equation $e^{j\phi} = \cos\phi + j\sin\phi$ is employed to solve Equation (4). Additionally, by separating the real and imaginary components, the positional relationships among various hydraulic support components are derived. This approach yields the kinematic

mathematical model for the position and orientation of the hydraulic support, formulated as follows:

$$\begin{cases} L_{HG} \cos \alpha_1 \cos \varphi + L_{HG} \sin \alpha_1 \sin \varphi + L_{GE} \cos \theta_3 = L_{HF} \cos \theta_2 \\ + L_{FE} \cos \alpha_5 \cos \theta_{12} - L_{FE} \sin \alpha_5 \sin \theta_{12} \\ L_{HG} \sin \alpha_1 \cos \varphi + L_{HG} \cos \alpha_1 \sin \varphi + L_{GE} \sin \theta_3 = L_{HF} \sin \theta_2 \\ + L_{FE} \cos \alpha_5 \sin \theta_{12} + L_{FE} \sin \alpha_5 \cos \theta_{12} \\ L_{DA} \cos \alpha_3 \cos \theta_{12} - L_{DA} \sin \alpha_3 \sin \theta_{12} - L_{AB} \cos \alpha_7 \cos \beta + L_{AB} \sin \alpha_7 \sin \beta = s_1 \cos \theta_7 \\ L_{DA} \sin \alpha_3 \cos \theta_{12} + L_{DA} \cos \alpha_3 \sin \theta_{12} - L_{AB} \sin \alpha_7 \cos \beta - L_{AB} \cos \alpha_7 \sin \beta = s_1 \sin \theta_7 \\ L_{IH} \cos \alpha_2 + L_{HF} \cos \theta_2 + L_{FA} \cos \alpha_4 \cos \theta_{12} - L_{FA} \sin \alpha_4 \sin \theta_{12} + L_{AJ} \cos \alpha_6 \cos \beta \\ - L_{AJ} \sin \alpha_6 \sin \beta = s_2 \cos \theta_{10} \\ L_{IH} \sin \alpha_2 + L_{HF} \sin \theta_2 + L_{FA} \sin \alpha_4 \cos \theta_{12} + L_{FA} \cos \alpha_4 \sin \theta_{12} + L_{AJ} \sin \alpha_6 \cos \beta \\ + L_{AJ} \cos \alpha_6 \sin \beta = s_2 \sin \theta_{10} \end{cases} \quad (5)$$

Finally, the support height of the hydraulic support is determined based on its geometric relationship, as presented in Equation (6).

$$h = l_b + l_r \sin(\theta_2 + \varphi) + l_c \sin(\theta_{12} + \varphi) + l_p \cos(\beta + \varphi) + l \sin(\beta + \varphi) \quad (6)$$

Assuming the lengths $L_{IJ} = s_2$ and $L_{DB} = s_1$ of the leg and stabilizing ram are considered independent motion variables, Equation (5) represents a non-linear transcendental equation with $(\theta_2, \theta_3, \theta_7, \theta_{10}, \theta_{12}, \beta, \varphi)$ as the variables to be determined. Thus, the support attitude β and support height h can be adjusted by varying the lengths s_2 of the leg and s_1 of the stabilizing ram; however, the relationship between these parameters is non-linear. However, the lengths of the leg and stabilizing ram cannot be explicitly calculated based on the support attitude angle and the support height of the hydraulic support.

2.2. Support Environment of Hydraulic Support

The support environment of a hydraulic support refers to the dynamic equilibrium system formed by the coupled interaction between the hydraulic support and the roof of the coal seam. This environment can be quantified as the mechanical coupling relationship representing the corresponding interactions and mutual constraints between the hydraulic support and the coal seam roof during the support process. Based on the coupling relationship between the hydraulic support and the surrounding rocks in the stopes, the coal wall, hydraulic support, and caved rock in the gob are treated as an elastomer. This allows the leg support force and roof subsidence to be determined using the unlimited bearing beam model. Then, by integrating the resultant force model of the hydraulic support and the surrounding rocks of the stopes proposed by Qian et al. [32], the coupling mechanics of the supporting pose and surrounding rock are formulated, as illustrated in Figure 3.

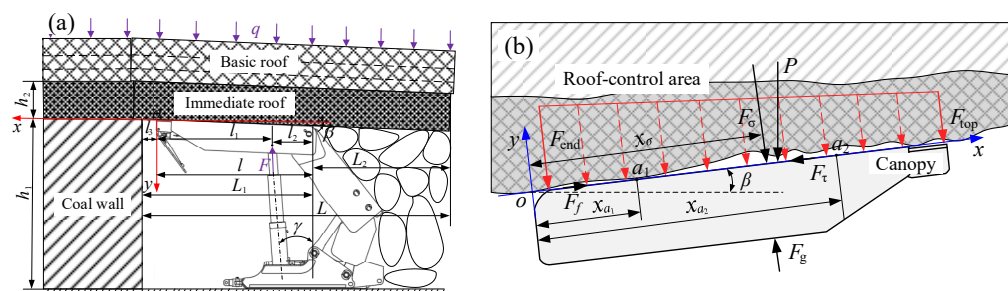


Figure 3. Coupling the mechanical model of hydraulic support and surrounding rock: (a) coupling the relationship between supporting pose and surrounding rock; (b) coupling the relationship between canopy and roof-control area.

In the mechanical system depicted in Figure 3, differential analysis is performed on the rock beam with unit width, resulting in the following mathematical model:

$$EI(dy/dx)^4 = q(x) - p(x) \quad (7)$$

where $p(x)$ represents the vertical force exerted by the hydraulic support on the roof; $q(x)$ denotes the load of the overlying strata acting on the roof rock beam; E corresponds to the elastic modulus of the overlying strata; I represents the moment of inertia of the overlying strata about the neutral plane; and y is a continuous function of x .

According to E. Winkler's elastic foundation beam hypothesis, the bearing pressure value in the elastic bearing pressure region is proportional to the settlement at each point [33]. Assuming the relative resistance coefficient of the hydraulic support is k and the roof rock beam subsidence is y , the supporting force of the hydraulic support on the roof, denoted as F , is expressed as follows:

$$F = ky \cos \beta \cos(\beta - \gamma) \quad (8)$$

where β represents the support attitude angle, and γ represents the angle between the leg and the vertical direction.

By substituting Equation (8) into Equation (7) and solving the differential equation based on the mechanical model presented in Figure 3, the deflection curve of the roof rock beam above the hydraulic support can be derived as follows:

$$y(x) = \frac{qL_2}{4\lambda^2 EI} e^{-\lambda(l-x)} \left[-L_2 \sin(\lambda(l-x)) + \left(1 + \frac{2}{\lambda}\right) \cos(\lambda(l-x)) \right] + \frac{q}{k \cos \beta \cos(\beta - \gamma)} \quad (9)$$

where $\lambda = \sqrt[4]{\frac{k_p}{4EI}}$, $k_p = k \cos \beta \cos(\beta - \gamma)$ denotes the combined stiffness of the hydraulic support and the immediate roof, and EI is the bending rigidity of the roof rock beam per unit width.

Equation (9) reveals that when the support attitude angle aligns with or matches the change trend angle of the surrounding rock roof, the roof deflection decreases. This reduced deflection enhances the ability of the hydraulic support to stabilize the surrounding rock roof. Similarly, the deflection curve of the rock beam above the coal wall is expressed as follows:

$$h(x) = \frac{e^{-\kappa x}}{4\kappa^2 EI} \left\{ \left[F(l_1 + l_3) - qL^2 \right] \sin(\kappa x) - \frac{2\kappa^3 [qL^2 - 2F(l_1 + l_3)] + 4\kappa^2 EI}{k_c + 2\kappa^3} \cos(\kappa x) \right\} + \frac{q}{k_c} \quad (10)$$

where $\kappa = \sqrt[4]{\frac{k_c}{4EI}}$, k_c and corresponds to the stiffness of the coal wall.

An analysis of Equations (9) and (10) reveals that the deflection and stress of the coal wall and hydraulic support on the roof rock beam follow approximate exponential functions. To ensure safe and effective roof support, the two deflection curves must align, meaning that at $x = 0$, the deflection-stress curves transition smoothly. Otherwise, as illustrated in Figure 4b, the roof rock beam is prone to fracturing (represented by the yellow curve), which can cause significant impact damage to the hydraulic support. Thus, according to $y(0) = h(0)$, the coupling relationship between the hydraulic support leg and surrounding rock can be expressed as follows:

$$F = \frac{qL^2}{2l_1} + \frac{q}{\kappa l_1} + \frac{qL\lambda(k_c + 2EI\kappa^3)(\lambda + 2) \cos(\lambda l)}{4EI\kappa^5 l_1 e^{\lambda l}} \quad (11)$$

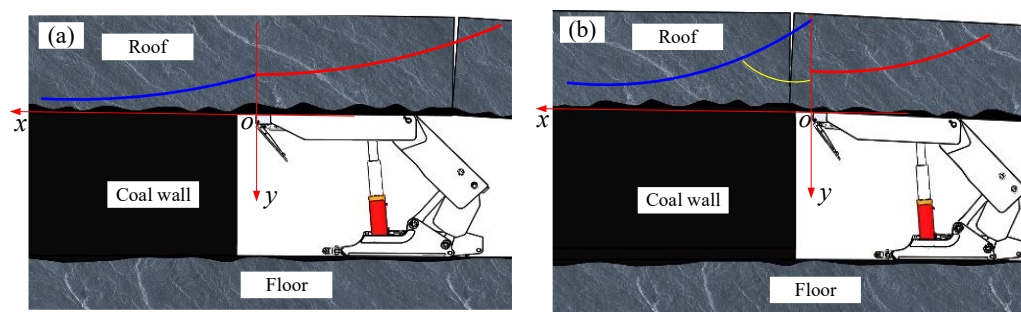


Figure 4. Roof deflection curve above hydraulic support and coal wall: (a) reasonable safety deflection curve; (b) non-ideal dangerous deflection curve.

Based on the theory of mechanics equilibrium region, roof rock beam instability places the hydraulic support in the tension or compression working range of the stabilizing ram, significantly reducing its bearing capacity. Assuming that torsional effects caused by uneven forces on the leg and linkage mechanism are neglected, the mechanical analysis model of the plane linkage system of the hydraulic support can be constructed as Figure 5.

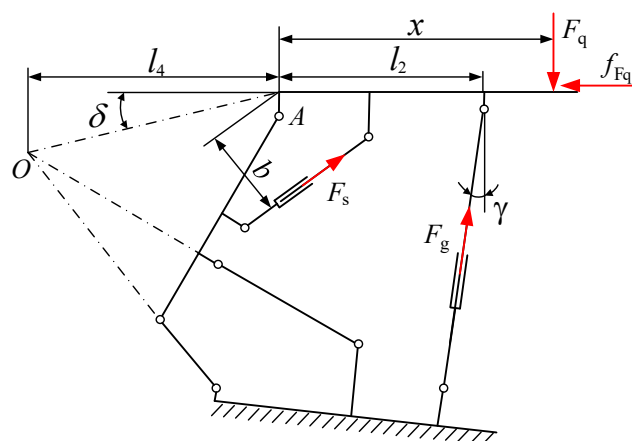


Figure 5. Mechanical model of a plane bar system of the hydraulic support.

The canopy, caving shield and canopy are each considered as the isolation units for force analysis. Subsequently, the moments about points O and A are analyzed to derive the torque equilibrium equation as follows:

$$\begin{cases} (l_4 + l_2)F_g \cos \gamma + (fF_q - F_g \sin \gamma)l_4 \tan \delta = F_q(x + l_4) \\ F_sb + F_g \cos \gamma l_2 = F_q x \end{cases} \quad (12)$$

where O denotes the instantaneous velocity center of the hydraulic support, and A is the hinge point between the canopy and the caving shield. l_2 represents the distance from the socket of the hydraulic support to point M, and l_4 is the horizontal distance from point O to M. b is the distance from point M to the stabilizing ram, which is its action position. F_q denotes the external load on the hydraulic support, and x represents the position where this load acts. f is the friction coefficient between the canopy and the roof, while F_g and F_s are the forces acting on the leg and stabilizing ram, respectively.

By substituting the maximum pressure T and maximum tension— T of the stabilizing ram into Equation (12), the coupling relationship between the stabilizing ram of the hydraulic support and the surrounding rock can be derived as follows:

$$F_q = \begin{cases} -\frac{l_4+l_2}{(x-l_2)l_4} Tb, 0 \leq x < x_{a_1} \\ \frac{l_4+l_2}{l_4+x} F_g \cos \gamma, x_{a_1} < x \leq x_{a_2} \\ \frac{l_4+l_2}{(x-l_2)l_4} Tb, x_{a_2} < x \leq l \end{cases} \quad (13)$$

where $x_{a_1} = (l_4 l_2 F_g \cos \gamma - l_4 b T) / (l_4 F_g \cos \gamma + b T)$ and $x_{a_2} = (l_4 l_2 F_g \cos \gamma + l_4 b T) / (l_4 F_g \cos \gamma - b T)$ represent the abscissa of points a_1 and a_2 on the canopy, respectively.

The coupling curve of the force balance between the stabilizing ram and the surrounding rock is then obtained from Equation (13), as illustrated in Figure 6.

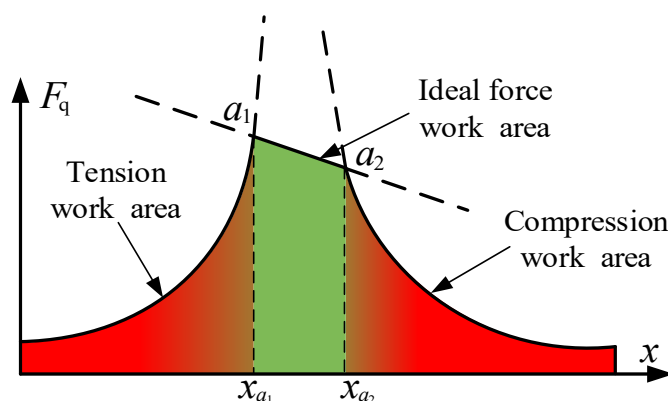


Figure 6. Coupling curve of force balance between the stabilizing ram and the surrounding rock. The three curves in the figure represent the three sections of the hydraulic support's force balance area, i.e., the back end of the canopy to point a_1 , point a_1 to point a_2 and point a_2 to the front end of the canopy.

3. Reinforcement Learning Model for the Supporting Pose Based on Markov Decision Process

The hydraulic support process is a complex, dynamic adjustment of its supporting pose. As analyzed in Section 2, the coupling interaction between the supporting pose and the surrounding rock forms the basis for training a self-control policy of supporting pose using reinforcement learning. This training can be viewed as a process of continuously optimizing the decision-making policy. This optimization is achieved through timely feedback during the ongoing trial-and-error interaction between the hydraulic support and the surrounding rock environment. Support actions derived from accumulated experience can achieve the optimal supporting pose to adapt to the coal seam roof. However, the state change process of the supporting pose follows the Markov property. This section establishes a reinforcement learning model for the supporting pose, along with a simulator for policy training, based on the fundamental elements of a Markov decision process. The sequential decision-making process for the supporting pose of the hydraulic support is shown in Figure 7. At each support posture regulation moment, the decision-making process involves four steps: obtaining the current support state information, making a support action decision, obtaining the immediate reward, and updating the support state.

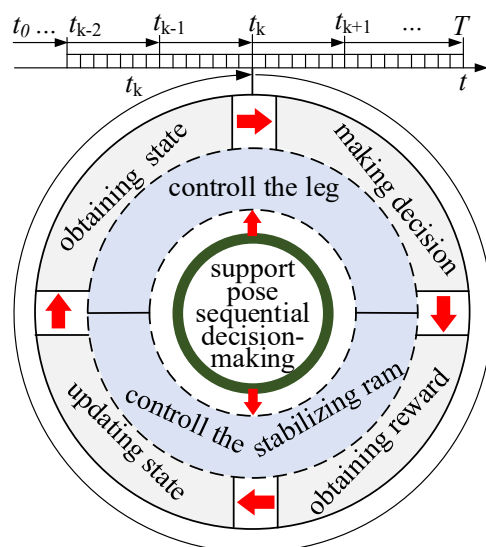


Figure 7. Sequential decision for the supporting pose of the hydraulic support.

3.1. Supporting Action Space

The adjustment of the hydraulic support's supporting pose is mainly controlled by the thrusts of the leg and stabilizing ram, with each offering three possible settings: constant, increase, and decrease. These combinations result in nine basic support actions, as shown in Table 1. These actions are labeled as $L_C S_C$, $L_C S_I$, $L_C S_D$, $L_I S_C$, $L_I S_I$, $L_I S_D$, $L_D S_C$, $L_D S_I$ and $L_D S_D$, and correspond to the nine dimensions of the action space.

Table 1. Basic actions of support.

Leg Thrust \ Stabilizing Ram Thrust	Stabilizing Ram Thrust		
	Constant	Increase	Decrease
Constant	$L_C S_C$	$L_C S_I$	$L_C S_D$
Increase	$L_I S_C$	$L_I S_I$	$L_I S_D$
Decrease	$L_D S_C$	$L_D S_I$	$L_D S_D$

3.2. Supporting State Space

The supporting pose state information comprises the surrounding rock roof state and the hydraulic support's canopy state. The roof state information is derived from the pressure data of the hydraulic support's leg and stabilizing ram using an inverse approach. The pressure value of both the leg and the stabilizing ram must correspond to the pressure exerted by the roof on the hydraulic support during the support process. The canopy state information includes the pressure ratio between its front and rear ends, its attitude angle, and its support height. Figure 3b illustrates the coupling relationship between the hydraulic support's canopy and the roof-control area. Here, P represents the pressure exerted by the roof-control area on the canopy, while $F_\sigma = P \cos \beta$, $F_\tau = P \sin \beta$. To ensure the stability of the roof-control area, $F_\tau \leq F_f = F_\sigma f$ (where f denotes the friction coefficient) must hold true on the contact friction surface between the canopy and roof. This implies that $\tan \beta \leq f$, and consequently, $\beta \leq \arctan f$.

Thus, the state space for reinforcement learning in supporting pose adjustment is defined as $s = F_g \times F_s \times \beta \times h$, where F_g is the leg pressure of the hydraulic support, $0 \leq F_g \leq F_{max}$, and F_{max} represents the maximum working resistance of the hydraulic cylinder for the leg. F_s is the stabilizing ram pressure, where $-T \leq F_s \leq T$, and $-T$ and T are the maximum tension and maximum pressure of the stabilizing ram cylinder, respectively. β represents the attitude angle of the hydraulic support's canopy, where

$\beta \leq \arctan f$. h denotes the support height, where $H_{min} \leq h \leq H_{max}$. Here, H_{min} and H_{max} represent the minimum and maximum working heights of the hydraulic support, respectively. The support state of the hydraulic support can be expressed as a vector, and is denoted as follows:

$$s = [F_g \ F_s \ \beta \ h] \quad (14)$$

3.3. Supporting Reward Function

The fuzzy set membership function within the fuzzy comprehensive evaluation method is used to quantitatively assess the supporting pose. This function helps design a comprehensive reward function for the support process by considering factors such as support height, canopy pitch angle, support resistance, and the resultant force action point of the hydraulic support. The following reward function is expressed in matrix form using vectors and indicator functions.

(1) Support Height Reward Function

The support height must be maintained within a reasonable working range, as excessively low or high heights can result in transitional damage to the roof. Hence, the membership function for support height must conform to a triangular membership function. The center of this function is defined based on the average height of each hydraulic support measured along the intelligent coal mining working face. The closer the support height is to the average value, the higher the membership degree, resulting in a greater reward for the support height as depicted in Figure 8a. The reward function for the support height is defined as follows:

$$R_h(h) = \begin{bmatrix} I_1(h) & I_2(h) & I_3(h) & I_4(h) \end{bmatrix} \begin{bmatrix} 0 & \frac{h-H_{min}}{H_{ave}-H_{min}} & \frac{H_{max}-h}{H_{max}-H_{ave}} & 0 \end{bmatrix}^T \quad (15)$$

where H_{min} and H_{max} represent the minimum and maximum working heights of the hydraulic support, respectively; H_{ave} is the average height of the hydraulic support on the intelligent coal mining working face, and h is calculated using the kinematic model of the hydraulic support shown in Figure 5; $I_i(h)$ is the indicator function, $I_1(h) = \mathbf{1}_{(0,H_{min})}(h)$, $I_2(h) = \mathbf{1}_{[H_{min},H_{ave})}(h)$, $I_3(h) = \mathbf{1}_{[H_{ave},H_{max}]}(h)$, $I_4(h) = \mathbf{1}_{(H_{max},\infty)}(h)$.

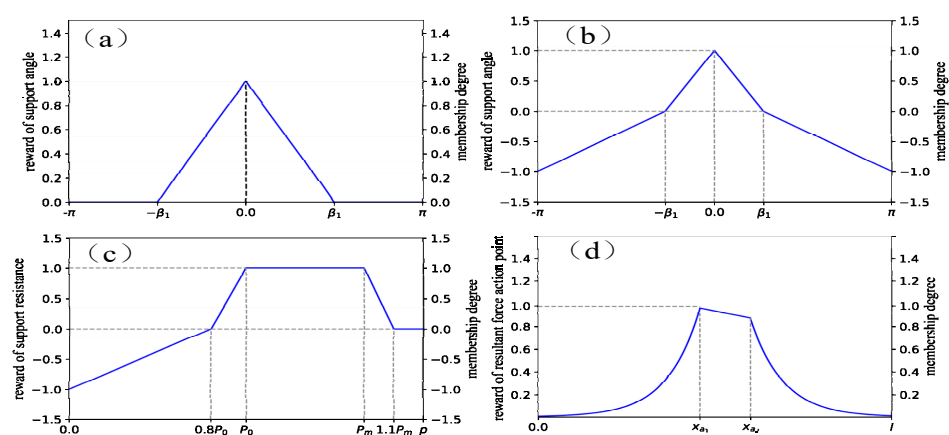


Figure 8. Reward function and membership degree of the hydraulic support process: (a) reward of the support height; (b) reward of the support angle; (c) reward of the support resistance; (d) reward of the resultant force action point.

(2) Support Angle Reward Function

The pitching angle of the hydraulic support's canopy typically fluctuates around 0° during normal operation. Therefore, the membership function for the support angle

also adopts a triangular membership function, and the support angle reward is expressed as follows:

$$r_a(\beta) = \begin{bmatrix} I_1(\beta) & I_2(\beta) \end{bmatrix} \begin{bmatrix} \frac{\beta+\beta_1}{\beta_1} & \frac{\beta_1-\beta}{\beta_1} \end{bmatrix}^T \quad (16)$$

where $\beta_1 = \arctan f$ represents the friction angle between the canopy and the roof-control area; $I_i(\beta)$ is the indicator function, $I_1(\beta) = \mathbf{1}_{[-\beta_1, 0)}(\beta)$, $I_2(\beta) = \mathbf{1}_{[0, \beta_1)}(\beta)$.

Additionally, considering the combined stiffness relationship between the hydraulic support and the surrounding rock, if the support angle is less than β_1 or greater than β_1 , the system applies a negative penalty (below zero), which disrupts the combined stiffness. The constraint-based reward for the support angle is defined as follows:

$$p_a(\beta) = \begin{bmatrix} I_3(\beta) & I_4(\beta) \end{bmatrix} \begin{bmatrix} \frac{\beta_1+\beta}{\pi-\beta_1} & \frac{\beta_1-\beta}{\pi-\beta_1} \end{bmatrix}^T \quad (17)$$

where $I_3(\beta) = \mathbf{1}_{[-\pi, -\beta_1)}(\beta)$, $I_4(\beta) = \mathbf{1}_{[\beta_1, \pi)}(\beta)$. With comprehensive consideration of rewards and penalties, the reward function for the support angle is depicted in Figure 8b and is defined as follows:

$$R_a(\beta) = \begin{bmatrix} I_1(\beta) & I_2(\beta) & I_3(\beta) & I_4(\beta) \end{bmatrix} \begin{bmatrix} \frac{\beta+\beta_1}{\beta_1} & \frac{\beta_1-\beta}{\beta_1} & \frac{\beta_1+\beta}{\pi-\beta_1} & \frac{\beta_1-\beta}{\pi-\beta_1} \end{bmatrix}^T \quad (18)$$

(3) Support Resistance Reward Function

Based on the research conducted on the working cycle process of hydraulic supports [34], the variation curve of support resistance is presented in Figure 9. During the same working cycle, the support resistance $p \in [P_0, P_m]$, where P_0 represents the initial setting load of the hydraulic support's leg, and P_m denotes the maximum working resistance at the end of the cycle. Consequently, a trapezoidal membership function is chosen as the membership function of support resistance, with boundary values defined as $[0.8P_0, 1.1P_m]$. The positive reward for support resistance is formulated as follows:

$$r_p(p) = \begin{bmatrix} I_1(p) & I_2(p) & I_3(p) & I_4(p) \end{bmatrix} \begin{bmatrix} \frac{p-0.8P_0}{0.2P_0} & 1 & \frac{1.1P_m-p}{0.1P_m} & 0 \end{bmatrix}^T \quad (19)$$

where $I_1(p) = \mathbf{1}_{(0.8P_0, P_0)}(p)$, $I_2(p) = \mathbf{1}_{[P_0, P_m)}(p)$, $I_3(p) = \mathbf{1}_{[P_m, 1.1P_m)}(p)$, $I_4(p) = \mathbf{1}_{(1.1P_m, \infty)}(p)$. Similarly, to maintain sufficient support strength during the hydraulic support working process and to prevent roof fracture, collapse, or strength instability of overlying surrounding rock due to support failure, the support resistance must not fall below a critical value. Thus, the penalty (negative reward) for inadequate support resistance is designed as follows:

$$p_p(p) = I_5(p) \frac{p - 0.8P_0}{0.8P_0} \quad (20)$$

where $I_5(p) = \mathbf{1}_{(0, 0.8P_0)}(p)$. By combining both positive and negative rewards, the reward function for support resistance is illustrated in Figure 8c and is defined as follows:

$$R_p(p) = \begin{bmatrix} I_1(p) & I_2(p) & I_3(p) & I_4(p) & I_5(p) \end{bmatrix} \begin{bmatrix} \frac{p-0.8P_0}{0.2P_0} & 1 & \frac{1.1P_m-p}{0.1P_m} & 0 & \frac{p-0.8P_0}{0.8P_0} \end{bmatrix}^T \quad (21)$$

(4) Resultant Force Action Point Reward Function

Based on the relationship between the action position of the external load's resultant force on the hydraulic support and its support capacity, as shown in Equation (13) and Figure 6, the membership function for the resultant force action point is derived by normal-

ization, as illustrated in Figure 8d. Thus, the reward function for the action point of the canopy's resultant force is defined as follows:

$$R_q = F_q / F_m \quad (22)$$

where F_q represents the resultant force acting on the hydraulic support at the current action point, and F_m denotes the maximum resultant force the hydraulic support can withstand.

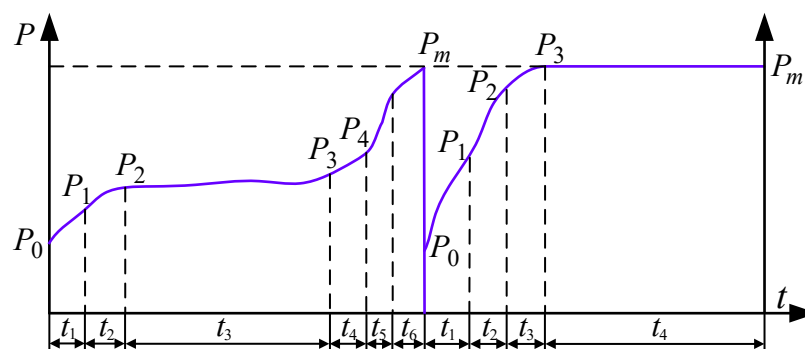


Figure 9. Support resistance curve in the support working cycle.

(5) Comprehensive Reward Function Design

Inspired by the concept of inverse reinforcement learning, the reward function for environmental feedback is derived from the agent's learning trajectory data [35] and is expressed as a linear function of state features. Consequently, the comprehensive reward function R is defined as a linear combination of the support height reward (R_h), support angle reward (R_a), support resistance reward (R_p) and resultant force action point reward (R_q). It is expressed as follows:

$$R = a_1 R_h + a_2 R_a + a_3 R_p + a_4 R_q \quad (23)$$

where a_1, a_2, a_3 and a_4 are non-negative weighting coefficients in the linear combination of the comprehensive reward function. However, their contributions are not the same; from an engineering perspective, the contribution of support height is the most critical, and they satisfy the following constraint:

$$a_1 + a_2 + a_3 + a_4 = 1 \quad (a_1, a_2, a_3, a_4 \geq 0) \quad (24)$$

where the four parameter values are determined based on the fuzzy consistency judgment matrix of the support pose, as shown in Table 2. Using the calculation for the fuzzy consistent judgment matrix, as in (24), specifically $a_1 = 0.35$, $a_2 = 0.26$, $a_3 = 0.14$, and $a_4 = 0.25$.

$$w_i = \frac{1}{n} - \frac{1}{2a} + \frac{1}{na} \sum_{k=1}^n j_{ik} \quad (a \geq (n-1)/2) \quad (25)$$

Table 2. Judgment matrix of the support posture.

State Variable	Support Height	Support Angle	Resultant Force	Support Resistance Action Point
support height	0.5	0.7	0.9	0.7
support angle	0.3	0.5	0.7	0.6
resultant force	0.1	0.3	0.5	0.2
support resistance action point	0.3	0.4	0.8	0.5

4. Autonomous Decision-Making and Control Method for the Supporting Pose

The proposed method for the supporting pose control strategy is based on the reinforcement learning model built on a Markov Decision Process described in Section 3. The hydraulic support is treated as an agent capable of perceiving changes in the support environment. Through trial-and-error learning via interaction with the environment, the agent enhances its decision-making ability to select appropriate support actions under uncertain support conditions. To address the limitations of trial-and-error interaction between the agent and the real support environment, a progressive neural network is used to transfer the support pose control policy from simulation to reality. Specifically, this paper proposes a Progressive Neural Network (PNN) transfer reinforcement learning algorithm based on Proximal Policy Optimization (PPO), referred to as PPO-PNN. This method effectively addresses the cross-task transfer problem in complex sequences of support regulation. Through lateral connections in the network, the support control policy trained in simulation is transferred to regulate the supporting pose in the real environment.

4.1. Policy Gradient-Based Proximal Policy Optimization

Proximal Policy Optimization is an off-policy reinforcement learning algorithm, which is well-suited for addressing control problems in both discrete and continuous action spaces. The network architecture of PPO under the Actor–Critic framework consists of a value network (Critic) and a policy network (Actor). The value network takes the environmental state as input, that is, the state information of the support pose, and outputs the corresponding state value. The policy network consists of two sub-networks: one responsible for interacting with the environment of the hydraulic support, and the other dedicated to updating the network parameters.

Assume that the agent follows a behavior policy π_θ with network parameter θ , also referred to as the Actor. Here, π_θ represents a mapping from the environmental state s to the action a taken by the agent. The goal of agents in reinforcement learning is to iteratively optimize their strategies to maximize the cumulative reward R . The expected value of R serves as a metric for evaluating the Actor's performance. Consequently, the policy objective function $J(\theta)$ is constructed to assess the performance of the behavior policy and is defined as follows:

$$J(\theta) = \bar{R}_\theta = \sum_{\tau} R(\tau) \pi_\theta(\tau) = E_{\tau \sim \pi_\theta(\tau)}[R(\tau)] \quad (26)$$

where $\tau = \{s_1, a_1, r_1, s_2, a_2, r_2, \dots, s_T, a_T, r_T\}$ represents a sequence of states, actions taken upon observing those states and rewards. The expectation $E_{\tau \sim \pi_\theta(\tau)}[\dots]$ denotes the empirical average over a finite batch of samples.

To enhance the data sampling efficiency of the PPO algorithm, importance sampling is employed to process the action distribution. Samples collected using the old policy $\pi_{\theta_{old}}$ are utilized to update the new policy π_θ , enabling the reuse of collected data. Additionally, PPO introduces the advantage function to replace the original reward in (26), allowing (26) to be rewritten as follows:

$$J^{CPI}(\theta) = E_{(s_t, a_t) \sim \pi_{\theta_{old}}} \left[\frac{\pi_\theta(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)} A(s_t, a_t) \right] \quad (27)$$

where θ represents the vector of new policy parameters, and θ_{old} represents the vector of old policy parameters. $A_\pi(s_t, a_t) = Q_\pi(s_t, a_t) - V_\pi(s_t)$ defines the advantage function, which evaluates the current action value relative to the average value. Here, $Q_\pi(s_t, a_t) = E_{s_{t+1}, a_{t+1}, \dots} [\sum_{k=0}^{\infty} \gamma^k r(s_{t+k})]$ represents the state-action value function,

$V_\pi(s_t) = E_{a_t, s_{t+1}, \dots} [\sum_{k=0}^{\infty} \gamma^k r(s_{t+k})]$ is the value function, $r: S \rightarrow \mathbb{R}$ denotes the reward function, and γ is the discount factor. Moreover $A_\pi(s_t, a_t) > 0$ indicates that the action is better than the average action, and the larger the value, the better the action is selected in state s_t .

In all cases, the divergence between the two distributions, $\pi_{\theta_{old}}$ and π_θ , must not be too large and must remain below a constraint value to ensure a larger objective function. To address this, the clip function is employed as a constraint to limit updates to the new policy. Based on (27), the latest surrogate objective function is redefined as follows:

$$J^{CLIP}(\theta) = E_{(s_t, a_t) \sim \pi_{\theta_{old}}} [\min(k_t(\theta) A(s_t, a_t), \text{clip}(k_t(\theta), 1 - \epsilon, 1 + \epsilon) A(s_t, a_t))] \quad (28)$$

where $k_t(\theta) = \pi_\theta(a_t|s_t) / \pi_{\theta_{old}}(a_t|s_t)$ and ϵ is a hyperparameter typically assigned a recommended value of 0.2. The first term inside the minimum function is J^{CPI} . The second term, $\text{clip}(k_t(\theta), 1 - \epsilon, 1 + \epsilon) A(s_t, a_t)$, modifies the surrogate objective by clipping the probability ratio, thereby removing the incentive for shifting k_t beyond the interval $[1 - \epsilon, 1 + \epsilon]$, effectively constraining the divergence between the old and new policies. Next, the gradient of the objective function is computed using the policy gradient method, wherein an estimator of the policy gradient is calculated and incorporated into a stochastic gradient ascent algorithm. The most commonly used form of the gradient estimator is expressed as follows:

$$\hat{g} = \hat{E}_{(s_t, a_t) \sim \pi_\theta} [\nabla_\theta \log \pi_\theta(a_t|s_t) \hat{A}(s_t, a_t)] \quad (29)$$

Finally, the optimal policy is obtained by iteratively updating the policy parameters using an optimizer to maximize the expected reward, as defined by the objection function in (28) and the gradient estimator in (29). The update rule for the optimizer, with step size, α is given as follows:

$$\theta \leftarrow \theta + \alpha \nabla J^{CLIP}(\theta) \quad (30)$$

4.2. Progressive Neural Network-Based Transfer Policy

Progressive Neural Networks are ideal for simulation-to-real policy transfer in robot control domains for several reasons. First, the features learned for one task can be transferred to multiple new tasks without being disrupted by fine-tuning. Second, the columns within the network can be heterogeneous, which is beneficial for solving diverse tasks and accelerating learning when transferred to real robots. Third, Progressive Neural Networks add new capacity when transferring to new tasks by incorporating new input connections. This mechanism helps bridge the reality gap by accommodating dissimilar inputs between simulated environments and real-world sensors.

Progressive Neural Networks leverage a series of pre-trained models trained from the source domain as the initialization. A progressive network begins with a single column, depicted as a deep neural network with L layers, hidden activations $h_i^{(1)} \in \mathbb{R}^{n_i}$, where n_i denotes the number of units at layer $i \leq L$, and parameters $\Theta^{(1)}$ trained to convergence. When adapting to a second task, the parameters $\Theta^{(1)}$ are frozen, and a new column with parameters $\Theta^{(2)}$ is instantiated using random initialization. In this setup, each layer $h_i^{(2)}$ receives input from both $h_{i-1}^{(2)}$ and $h_{i-1}^{(1)}$ through lateral connections. This connection mechanism generalizes to K tasks as follows:

$$h_i^{(k)} = f \left(W_i^{(k)} h_{i-1}^{(k)} + \sum_{j < k} U_i^{(k:j)} h_{i-1}^{(j)} \right) \quad (31)$$

where $W_i^{(k)} \in \mathbb{R}^{n_i \times n_{i-1}}$ represents the weight matrix of layer i in column k , $U_i^{(k:j)} \in \mathbb{R}^{n_i \times n_j}$ denotes the lateral connections from layer $i - 1$ of column j to layer i of column k and h_0 is the network input. f represents an element-wise non-linearity; we use $f(x) = \max(0, x)$ for

all intermediate layers. A progressive neural network with $k = 3$ is illustrated on the left in Figure 10d.

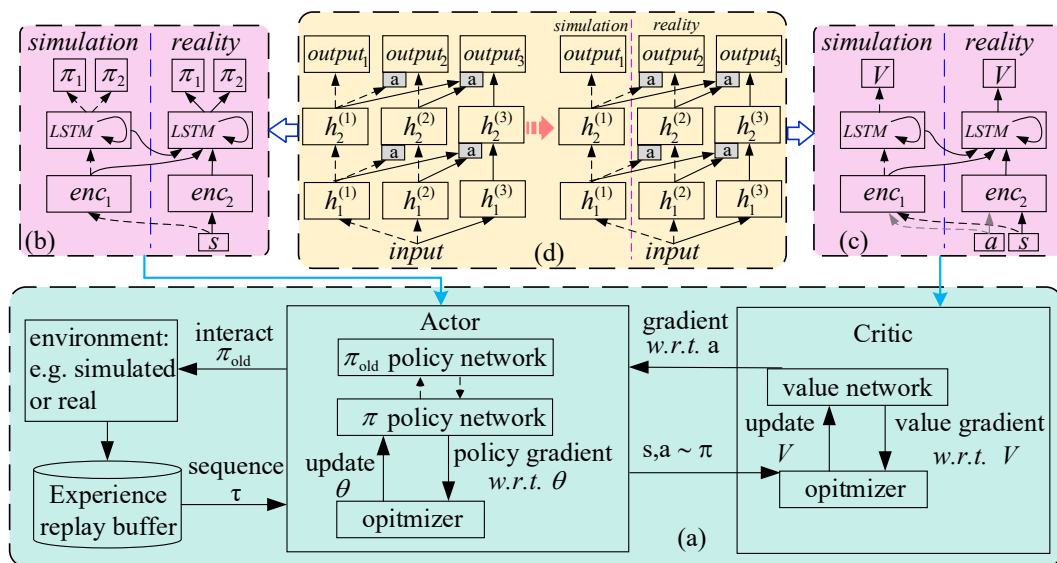


Figure 10. Detailed schematic of the proposed Progressive Transfer Network algorithm based on PPO: (a) the PPO algorithm base on the Actor–Critic architecture; (b) the architecture of the progressive network for policy transfer (c) the architecture of the progressive network for value transfer; (d) a three-column progressive network (shown on the left) and a modified progressive architecture for robot transfer learning (shown on the right).

In summary, Progressive Neural Networks can be described as follows: each column or layer can have an arbitrary network width to accommodate various levels of task complexity, and lateral connections from multiple independent networks can be combined in an ensemble setting.

4.3. Network Structure of Proposed Algorithm

This paper proposes an algorithm to simultaneously transfer both actor and critic networks in the PPO framework to ensure complete reuse of prior knowledge. The progressive network allows columns with different structures and capacities, with simulation-trained columns having sufficient depth to learn tasks from scratch in virtual environments, while reality-trained columns are lightweight to accelerate learning and limit parameter growth. The activations of the LSTM are connected as inputs to the progressive column to address the temporal dependencies in the sequential decision-making of supporting pose control. An LSTM is embedded into each column of the PNN, serving as the temporal feature extractor for each supporting pose regulation task. Lateral connections pass the LSTM representations from older columns to newer ones, helping new tasks handle similar temporal structures. To enhance exploration, reality-trained columns are initialized using parameters from simulation-trained columns, as illustrated in Figure 10. In Figure 10b, encoder1 and encoder2 represent a fully connected layer with two hidden layers, and the activations of the LSTM are connected as inputs to the progressive column. In Figure 10d, the first column is trained on Task1 in the simulated environment, the second column on Task1 in the real environment, and the third column on Task2 in the real environment. The columns may vary in capacity, with adapter functions (indicated by the gray boxes labeled ‘a’) used to reconcile differences in the scales of inputs.

5. Experiment

5.1. Experiment Platform Construction

An experimental platform was designed for autonomous control policy learning of the supporting pose with equal proportion reduction. This platform consists of both a virtual training environment and a real training environment, as shown in Figure 11. The virtual environment is established using the ROS system and Gazebo physics engine for supporting pose simulation training. As shown in Figure 11a, the experimental equipment in the real training environment includes a support attitude sensing system [10] for measuring support height and supporting pose, and a pressure transmitter for measuring cylinder pressure.

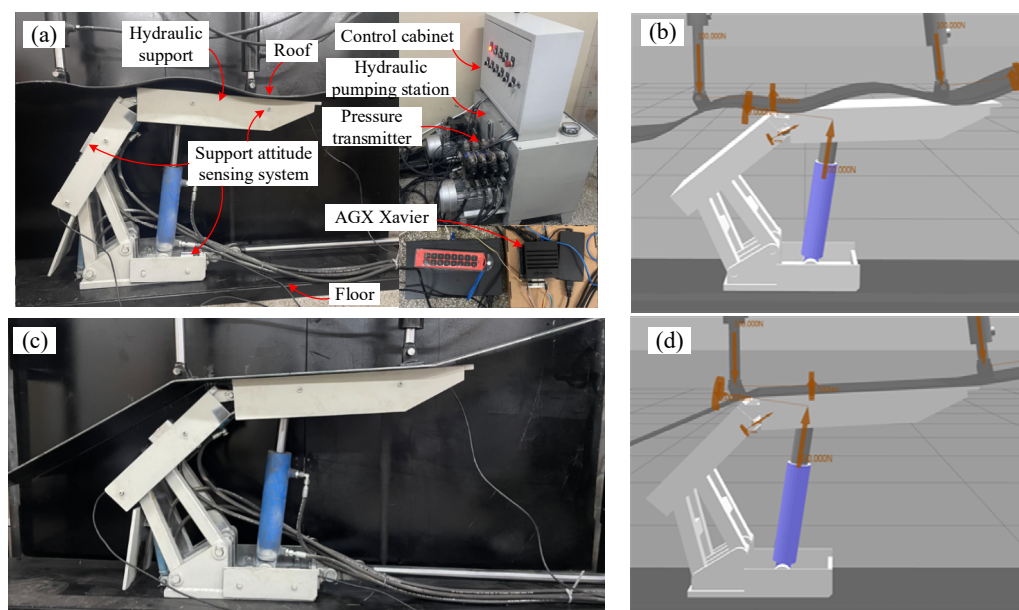


Figure 11. Experimental setup for self-adaptive adjustment of the supporting pose across simulation to reality under varying working conditions: (a) training the real hydraulic support interacting with a curved roof; (b) simulation training involving the virtual hydraulic support interacting with a curved roof; (c) training the real hydraulic support interacting with a plane roof; (d) simulation training involving the virtual hydraulic support interacting with a plane roof.

The experiments were conducted under two distinct working conditions. In these, the roof shape included an irregular curved surface and flat surfaces with varying inclinations, respectively.

5.2. Training in Simulation and Transfer to Reality

In the manipulation domain of the supporting pose for hydraulic support, the agent policy controls two degrees of freedom via hydraulic force commands, which act on two prismatic joints. The complete policy, denoted as $\Pi(A|s; \theta)$, consists of two prismatic joint policies learnt by the agent. Each prismatic joint policy, i , incorporates three discrete actions, denoted as $\pi_i(a_i|s; \theta_i)$: applying a fixed positive hydraulic force, a fixed negative hydraulic force, or a zero hydraulic force. Additionally, a softmax function and a single value function, $V(s; \theta_v)$, are employed. The value function is linearly connected to both the previous layer and lateral layers.

The simulation-trained columns are trained using a standard-sized network, while the reality-trained columns utilize a reduced-capacity network. The first column is trained in a hydraulic support physics simulator, with the ROS service providing observations, as illustrated in Figure 10a,b. In the real world, pressure transmitters serve as input providers. The experiment focuses on achieving the optimal supporting pose to effectively support

the coal seam roof, with comprehensive rewards serving as feedback. At the start of each episode, the supporting pose is initialized to a random starting position. The agent receives a reward, as defined in (23), if the roof pressure remains within its maximum allowable value. Each episode lasts up to 100 steps. The episode terminates if the agent destabilizes the coupling between the hydraulic support and the roof by exceeding the limits of the supporting pose, support resistance, or roof pressure.

In simulation training, the first column is trained using the PPO algorithm, as shown in Figure 10a. It is intuitive to train in simulation with a larger-capacity network to achieve maximum performance. This intuition is supported by comparing the wide network architecture (simulation-trained with two hidden layers of 64 units) and the narrow network architecture (reality-trained with two hidden layers of 16 units), as shown in Figure 12a,b. The narrow network exhibits poorer performance and slower learning.

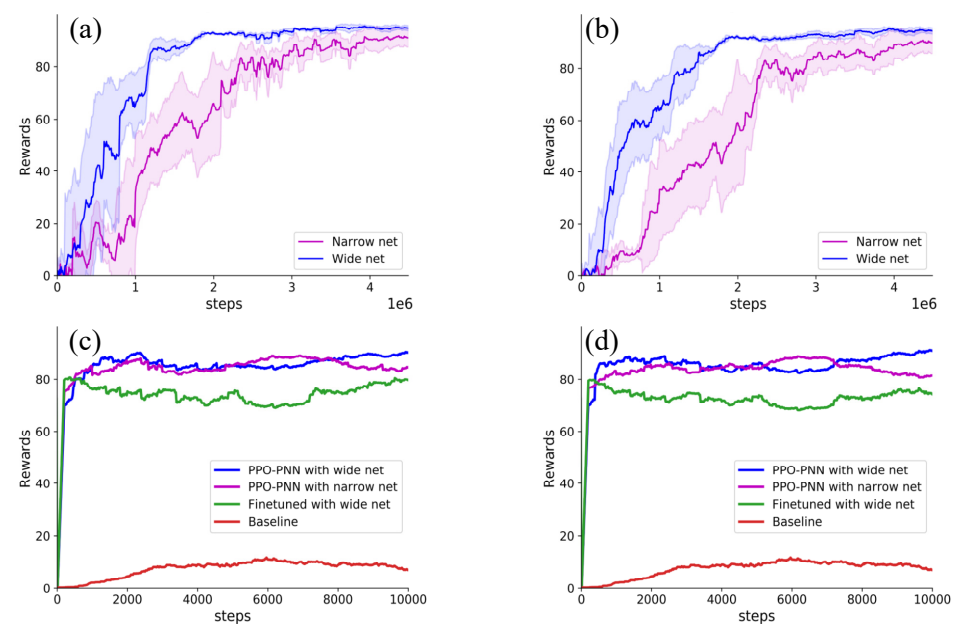


Figure 12. Learning curve for the supporting pose training under different working conditions: (a) training with wide and narrow networks in Experiment 1; (b) training with wide and narrow networks in Experiment 2; (c) training with three different network architectures in Experiment 1; (d) training with three different network architectures in Experiment 2.

Therefore, the column policy trained in simulation is designed with sufficient capacity to capture and transfer rich feature information, thereby supporting pose training on the real hydraulic support. This performance is achieved after approximately five million interaction steps in the Gazebo simulator—an amount of experience that is impractical to obtain through training on a real hydraulic support. Compared with real-world training, simulation is substantially faster, benefiting from multithreaded algorithms and the ability to run continuously without human intervention. In contrast, learning this task on a real hydraulic support is estimated to require two months or more, even under continuous 24 h training.

To facilitate sim-to-real transfer and enable training on the real hydraulic support, the supporting pose is manually readjusted—within a predefined support range—every five episodes using the operation buttons on the electrical cabinet. Rewards are then assigned automatically based on the roof-support condition and the coupling relationship between the supporting pose and the roof. The diagram of the experimental architecture is presented in Figure 13, which illustrates three training scenarios: training from scratch in reality, fine-tuning the first column (starting with a simulation-trained column and continuing training

on the hydraulic support), and training a progressive second column. Two groups of experiments were conducted under different working conditions: Experiment 1 involved training the hydraulic support interacting with a curved roof, while Experiment 2 involved interaction with a plane roof.

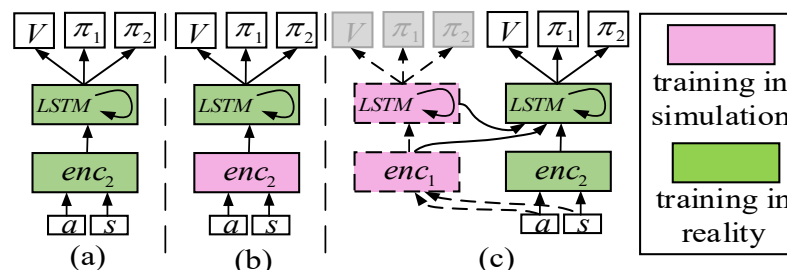


Figure 13. Model structures of the three approaches used in the experiment: (a) the baseline model, trained from scratch in the real environment; (b) the fine-tuning model, where the supporting pose is pretrained in simulation and further trained (fine-tuned) in the real environment; (c) the PPO-PNN model, employing a two-column progressive architecture with the parameters of the previous column deterministically initialized.

In Experiment 1, as illustrated in Figure 13c, the baseline (red curve) was trained from scratch with a randomly initialized narrow network in the real environment. The agent achieved very low rewards throughout the training process, indicating a failure to learn the adjustment policy for the supporting pose. The progressive second column achieved a score of 90.5 points, slightly underperforming the simulation column by an average of 5 points. However, the fine-tuned first column did not achieve the same performance as the progressive network. To more intuitively compare the differences among the three methods, a quantitative analysis was conducted from three aspects: performance, convergence speed, and sample efficiency. As shown in Figure 13c,d, PPO-PNN not only outperforms training from baseline and fine-tuning in terms of final return (improvements of 21.27 and 10.57, respectively), but also significantly reduces the number of training steps required to reach stable performance, demonstrating higher sample efficiency. In addition, from the perspective of early training performance, PPO-PNN exhibits better initial policy quality and shows a smoother convergence trend during training.

In Experiment 2, the results were broadly consistent with those of Experiment 1; however, learning progressed slightly faster, and the learning curve was smoother and more stable, as shown in Figure 12b,d. This difference is attributed to the greater complexity of the curved roof, which makes learning more difficult and slower than in the planar-roof case.

Compared to fine-tuning, the proposed simulation-to-reality transfer method can accommodate new input modalities and adapt to changes in network morphology, offering improved generalization performance. This advantage enables the method to transfer to new data sources while leveraging prior knowledge. To demonstrate this, a third progressive column was introduced to train the real hydraulic support.

As shown in Figure 14, a three-column architecture was implemented. State features sampled in simulation were used to train column one, features sampled in reality were used to train column two, and additional features sampled in reality were incorporated to train column three. To evaluate this architecture, two tasks were defined: the curved task, where the real hydraulic support interacts with a curved roof, and the plane task, where it interacts with a flat roof. The results of Experiment 1 were used to train the real hydraulic support in Experiment 2, and vice versa. Other aspects of the training tasks remained unchanged.

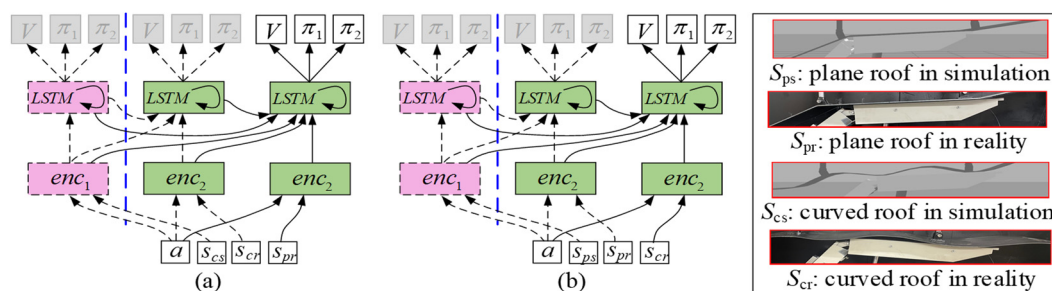


Figure 14. Task transfer architecture under two working conditions using a three-column progressive network: (a) transferring the supporting policy trained in Experiment 1 to the plane task; (b) transferring the supporting policy trained in Experiment 2 to the curved task.

In Experiment 1, when the second column was trained on the plane task, learning progressed relatively slowly (Figure 15a), reaching full performance after 8000 steps. By contrast, when the second column was trained on the curved task and the third column was subsequently trained on the plane task, transfer learning reached full performance approximately three times faster. In Experiment 2, the comparison between the plane and curved tasks trained with the second column was consistent with the findings of Experiment 1, as shown in Figure 15b.

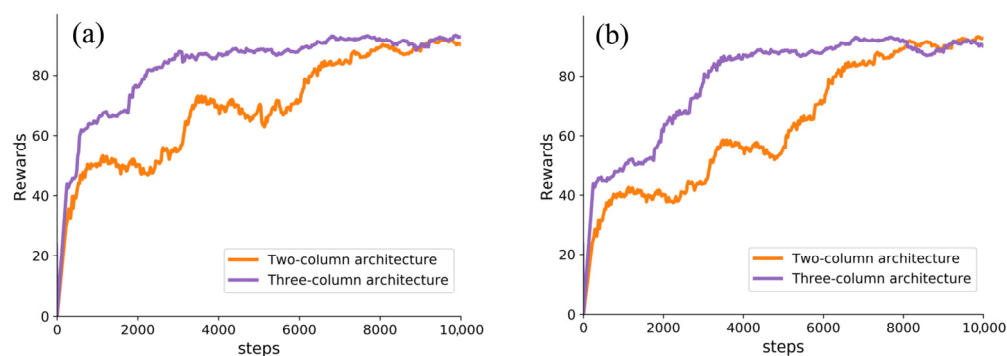


Figure 15. Transfer results of various network architectures for comparison: (a) plane task; (b) curved task.

These results demonstrate the capability of the proposed architecture to immediately reuse previously learned features and highlight its potential for application in real-world support of the robotic domains within coal mines.

5.3. Result Analysis and Discussion

By analyzing the learning curves of the support pose control strategy, it can be observed that the Markov decision process (MDP) model of the support pose based on kinematics can accurately reflect the process of a hydraulic support supporting the coal seam roof. Furthermore, by using a simulation-trained strategy and then transferring it to a real hydraulic support for further training, an optimal support pose control strategy can be formed and applied to the adaptive control of hydraulic supports. The experimental results show that the average reward obtained by the proposed algorithm is higher than that of other methods and is approximately equal to the reward in the simulation results. This confirms that the action space of the MDP established based on kinematics and the state space of the MDP established based on dynamics are accurate. At the same time, the simulation-to-real transfer of the support pose control strategy based on a progressive neural network to a real hydraulic support enables the support pose to adapt to changes in roof conditions; this method is feasible and achieves good application results. However, there is still a reality gap between the experimental environment and the simulation environment.

From the perspectives of theory, simulation, experiments, and algorithms, many factors affect the execution accuracy of the adaptive control strategy for support pose. The main factors are as follows:

- (1) The accuracy of the MDP for the support pose. This paper is the first in the field of hydraulic support research to define the MDP of the hydraulic support process and to propose a support pose control model based on an MDP. The supporting action space is defined based on the kinematic theory of hydraulic supports, the supporting state space is defined based on dynamic theory, and the reward function is defined according to the dynamic coupling relationship between the hydraulic support and the coal seam roof. As this is the first proposal, there remains substantial room for further research and optimization of the support pose MDP in the future.
- (2) Modeling errors in the Gazebo-based hydraulic support simulator. Modeling hydraulic supports in Gazebo requires simplifying key factors such as hydraulic dynamics (e.g., pressure–flow coupling and delays), structural flexibility, contact loads, and friction. Reinforcement learning strategies are optimized on this idealized model and often exploit these unrealistic assumptions. When transferred to the real environment, due to modeling errors such as nonlinearity, time delays, uncertain loads, and structural deformation in the actual system, the control performance of the strategy may degrade, manifesting as response lag, oscillation, insufficient accuracy, or even support instability, thereby affecting safety and reliability. To address this issue, this paper adopts randomization and robust training to cover uncertainties during training, including randomization of friction, time delay, load, and roof stiffness. However, experimental results still show a gap. Therefore, there remains room for further exploration and improvement in simulation and training methods.
- (3) PNN reuses existing features through lateral connections while learning to compensate for differences in the real system (such as hydraulic lag, friction, and load uncertainty), thereby largely mitigating modeling errors and improving the adaptability and stability of the strategy under complex real working conditions. However, the mining environment in which hydraulic supports operate is highly dynamic and complex, and PPO exhibits issues with exploration efficiency in such environments. To address this, this paper mainly mixes a batch of pre-collected stable-operation support data during training to reduce ineffective exploration. Therefore, from the perspective of algorithm optimization, there is still room to improve the final reward through other approaches.

6. Conclusions

To achieve adaptive support of the coal seam roof by hydraulic supports, this paper proposes a novel autonomous decision-making and control method for support pose. The method analyzes the kinematics and dynamics of hydraulic supports, as well as their dynamically balanced coupling relationship with the coal seam roof. On this basis, for the first time in the field of hydraulic support research, the support pose is formulated as a Markov decision process. The action space, state space, and reward function are defined respectively based on the kinematics and dynamics of the hydraulic support and its coupling with the coal seam roof, thereby establishing a theoretical foundation for both single-agent and multi-agent Markov decision processes of hydraulic supports.

Meanwhile, using the open-source 3D dynamic simulator Gazebo combined with a physics engine, a virtual dynamic simulator for hydraulic support is developed to simulate and train autonomous adjustment strategies for support pose. In addition, a transfer reinforcement learning algorithm based on PPO-PNN is proposed to address cross-task transfer in complex continuous control sequences of support posture, ensuring

high sampling efficiency and improved robustness. Compared with traditional fine-tuning-based transfer learning methods, this approach is less sensitive to hyperparameters, thereby providing higher stability and better overall performance.

Finally, experimental results of reinforcement learning based on dynamic analysis show that the proposed method demonstrates good feasibility and effectiveness in a simulation environment, improves the stability of the support process to a certain extent, and exhibits potential for applications in ensuring safety and reliability. The results indicate that this method provides a theoretical basis for the intelligent and adaptive adjustment of hydraulic support pose and promotes its practical application toward fully intelligent control of hydraulic supports in coal mining environments. In practical applications, this method also has limitations: it can only be transferred under the same type of mining equipment and identical geological conditions. Although it shows good scalability to different mining equipment or geological conditions, it requires more training data and may even necessitate adjusting simulator parameters and retraining for transfer.

Author Contributions: Conceptualization, L.Z.; Methodology, X.L.; Software, X.L.; Validation, D.W.; Formal analysis, X.L. and L.Z.; Writing—original draft, X.L.; Writing—review & editing, D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Foundation of Henan, grant number 242300421454.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to acknowledge the support of the Natural Science Foundation of Henan (242300421454) in carrying out this research. The authors also greatly appreciate the reviewers' suggestions and corrections to the original manuscript and the editor's encouragement.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, H.; Cao, Y.; Wang, H. Research and practice on key technologies for intelligentization of coal mine. *Coal Geol. Explor.* **2023**, *51*, 44–54.
2. Peng, S.S.; Du, F.; Cheng, J.; Li, Y. Automation in U.S. longwall coal mining: A state-of-the-art review. *Int. J. Min. Sci. Technol.* **2019**, *29*, 151–159. [\[CrossRef\]](#)
3. Pang, Y.; Shi, Y. Intelligent control algorithms for posture and height control of four-leg hydraulic supports. *Sci. Rep.* **2025**, *15*, 3010. [\[CrossRef\]](#)
4. Lindsay, J.; Hall, A.; Cai, M.; Simser, B. Mitigating geotechnical challenges in deep mining: Lessons learned from shaft station excavations at extreme depths. *Can. Geotech. J.* **2025**, *62*, 1–23. [\[CrossRef\]](#)
5. Shakeri, J.; Ghorbani, E.; Taheri, A. A review of deep mining challenges, hazard mitigation, design approaches, and support methods. *Undergr. Space* **2026**, *28*, 86–136. [\[CrossRef\]](#)
6. Peng, S.; Cheng, Z.; Che, L.; Zheng, Y.; Cao, S. Kinematic performance analysis of a parallel mechanism for loading test of hydraulic support. *Mech. Mach. Theory* **2022**, *168*, 104592. [\[CrossRef\]](#)
7. Ren, H.; Zhao, G.; Zhou, J.; Wen, Z.; Ding, Y.; Li, S. Key technologies of all position and orientation monitoring and virtual simulation and control for smart mining equipment. *J. China Coal Soc.* **2020**, *45*, 956–971.
8. Bai, J.; Hou, C. Control principle of surrounding rocks in deep roadway and its application. *J. China Univ. Min. Technol.* **2006**, *35*, 145–148.
9. Gwiazda, A. Design of the roof support with strait-Line mechanism. *Adv. Mater. Res.* **2014**, *1036*, 553–558. [\[CrossRef\]](#)
10. Xu, Y.; Wang, G.; Ren, H. Theory of coupling relationship between surrounding rocks and powered support. *J. China Coal Soc.* **2015**, *40*, 2528–2533.
11. Lu, X.; Wang, Z.; Tan, C.; Yan, H.; Si, L.; Wei, D. A portable support attitude sensing system for accurate attitude estimation of hydraulic support based on unscented kalman filter. *Sensors* **2020**, *20*, 5459. [\[CrossRef\]](#)
12. Ma, K.; Xie, J.; Guo, X.; Wang, X.W.; Wang, X.S.; Wang, L.J. A pose monitoring method for floating connection mechanism of a hydraulic support based on virtual-real integration. *Gong-Kuang Zidonghua* **2025**, *51*, 158–163.

13. Mei, Z.; Wang, X.; Xie, J.; Li, S.; Liu, J. A sensing system and solving method for dynamic detection of relative pose of hydraulic support group. *Measurement* **2025**, *243*, 116145. [[CrossRef](#)]
14. You, X.; Ge, S. Research on decision-making and control technology for hydraulic supports based on digital twins. *Symmetry* **2025**, *16*, 1316. [[CrossRef](#)]
15. Feng, Z.; Xie, J.; Yan, Z.; Mei, Z.; Zheng, Z.; Li, T. An information processing method of software and hardware coupling for VR monitoring of hydraulic support groups. *Multimed. Tools Appl.* **2023**, *82*, 19067–19089. [[CrossRef](#)]
16. Yao, Y.; Zhang, J.; Xiong, W. Method of tail beam posture prediction of top coal caving hydraulic support based on LSTM. *Coal Sci. Technol.* **2025**, *53*, 362–371.
17. Zhang, Y.; Zhang, H.; Gao, K.; Zeng, Q.; Meng, F.; Cheng, J. Research on intelligent control system of hydraulic support based on position and posture detection. *Machines* **2023**, *11*, 33. [[CrossRef](#)]
18. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 1998.
19. Yang, S.; Zhang, M.; Feng, X.; Hua, Y.; Cao, Y. Deep reinforcement learning-based knowledge graph reasoning for autonomous driving systems. *IEEE Trans. Ind. Inform.* **2026**, *22*, 3341–3351. [[CrossRef](#)]
20. Zhang, H.; Wang, J.-Q.; Zhang, S.; Jiang, Y.; Li, M.; Yong, B.; Zhou, Q.; Zhou, X. Generative policy-driven HAC reinforcement learning for autonomous driving incident response. *Futur. Gener. Comput. Syst.* **2026**, *175*, 108106. [[CrossRef](#)]
21. Leonardo, L.R.; Leticia, B.; Paula, P.C.; Marcos, V.S.A.; Luiz, M.G.G. Dual or unified: Optimizing drive-based reinforcement learning for cognitive autonomous robots. *Cogn. Syst. Res.* **2026**, *95*, 101422.
22. Kim, T.; Choi, M.; Choi, S.; Yoon, T.; Choi, D. Deep reinforcement learning control of a hexapod robot. *Actuators* **2026**, *15*, 33. [[CrossRef](#)]
23. Barros, P.; Yalçın, Ö.N.; Tanevska, A.; Sciutti, A. Incorporating rivalry in reinforcement learning for a competitive game. *Neural Comput. Appl.* **2022**, *35*, 16739–16752. [[CrossRef](#)]
24. Florian, R.; Manuel, E.; Kai, E. Simulation-driven balancing of competitive game levels with reinforcement learning. *IEEE Trans. Games* **2024**, *16*, 903–913. [[CrossRef](#)]
25. Parnada, A.; Qu, M.; Castellani, M.; Chang, H.J.; Wang, Y. Towards cost-effective and safe contact-rich robotic manipulation with reinforcement learning: A review of techniques for future industrial automation. *Proc. Inst. Mech. Eng. Part I J. Syst. Control. Eng.* **2026**, *240*, 3–35. [[CrossRef](#)]
26. ElMenshawhy, M.; Wu, L.; Gue, B.; AbouRizk, S. Automating pipe spool fabrication shop scheduling for modularized industrial construction projects using reinforcement learning. *J. Comput. Civ. Eng.* **2025**, *39*, 04025013. [[CrossRef](#)]
27. Wang, G.; Peng, J.; Guan, C.; Chen, J.; Guo, B. Multi-drone collaborative shepherding through multi-task reinforcement learning. *IEEE Robot. Autom. Lett.* **2024**, *9*, 10311–10318. [[CrossRef](#)]
28. Rusu, A.; Vecerik, M.; Rothörl, T.; Pascanu, R.; Hadsell, R. Sim-to-Real Robot Learning from Pixels with Progressive Nets. *arXiv* **2016**. [[CrossRef](#)]
29. Rusu, A.; Rabinowitz, N.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Hadsell, R. Progressive Neural Networks. *arXiv* **2016**, arXiv:1606.04671. [[CrossRef](#)]
30. Zhu, Z.; Lin, K.; Jain, A.; Zhou, J. Transfer learning in deep reinforcement learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 13344–13362. [[CrossRef](#)] [[PubMed](#)]
31. Tiwari, R.; Khapre, S.; Singh, A. Reinforcement learning in robotic systems: A review on sim-to-real transfer. *Robot. Auton. Syst.* **2026**, *198*, 105327. [[CrossRef](#)]
32. Qian, M.; Miao, X.; He, F. Mechanism of coupling effect between supports in the workings and the rocks. *J. China Coal Soc.* **1996**, *21*, 21.
33. Ma, Q.; Zhao, X.; Song, Z. Break of main roof ahead of workface and ground pressure. *J. China Coal Soc.* **2001**, *26*, 473–477.
34. Tan, Y.; Wu, S.; Yin, Z. *Mining Pressure and Strata Control*; China Coal Industry Publishing House: Beijing, China, 2011.
35. Deshpande, S.; Walambe, R.; Kotecha, K.; Selvachandran, G.; Abraham, A. Advances and applications in inverse reinforcement learning: A comprehensive review. *Neural Comput. Appl.* **2025**, *37*, 11071–11123. [[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.