

Article

Digital Model of Automatic Plate Turning for Plate Mills Based on Machine Vision and Reinforcement Learning Algorithm

Chunyu He *, Song Xue, Zhiqiang Wu, Zhong Zhao and Zhijie Jiao * 

State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, China; 2170552@stu.neu.edu.cn (S.X.); wuzq@ral.neu.edu.cn (Z.W.); zhaozhong@ral.neu.edu.cn (Z.Z.)

* Correspondence: hecy@ral.neu.edu.cn (C.H.); jiaozj@ral.neu.edu.cn (Z.J.); Tel.: +86-024-83686419 (C.H.)

Abstract: Plate turning is an essential step in the plate rolling process. The traditional control mode relies on the manual observation of billets and mainly manual operation. Manual plate turning becomes an external disturbance of the automatic control system of plate mills, which reduces the reproducibility and accuracy of the rolling process. The automatic plate turning function is urgently needed to improve the control level of the rolling line. In this paper, based on the improved image processing algorithm, the position and angle information of the billet conversion process are detected in real time, and the real-time processing of detection data in a complex production environment is realized. Based on the change in the billet rotation angle in the actual plate turning process, a mathematical model is constructed to simulate the plate turning process. On this basis, the digital model and optimization algorithm for automatic plate turning based on reinforcement learning are established, and the automatic optimization of plate turning speed and accuracy is completed. The field application of data-driven plate turning systems replaces manual plate turning control. The plate turning angle detection error of the system is $\leq 2^\circ$. The average plate turning time of each billet is greatly shortened compared with manual plate turning mode, and the fastest time can be shortened by more than 1 s, which greatly improves the production efficiency and is of great significance for improving the automatic control level and digital upgrade of plate mills.

Keywords: automatic plate turning; machine vision; image processing; data-driven; reinforcement learning



Citation: He, C.; Xue, S.; Wu, Z.; Zhao, Z.; Jiao, Z. Digital Model of Automatic Plate Turning for Plate Mills Based on Machine Vision and Reinforcement Learning Algorithm. *Metals* **2024**, *14*, 709. <https://doi.org/10.3390/met14060709>

Academic Editor: João Manuel R. S. Tavares

Received: 14 May 2024

Revised: 10 June 2024

Accepted: 12 June 2024

Published: 14 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plates are among the most important varieties of steel, which is extremely important for infrastructure construction, major engineering projects, and the national economy [1,2]. According to the international development trend of the plate industry and the major demand at present, it is of great significance for the technological progress and development of China's iron and steel industry to improve the rolling automation level and build intelligent rolling integration.

As shown in Figure 1, the plate rolling process generally includes the following three stages: sizing, broadening, and elongation [3,4]. In order to meet the requirements of the rolling process, it is necessary to carry out 1~2 plate turning operations in some passes. The specific operation is to rotate the billet by 90 degrees so that the length and width dimensions are reversed [5]. The traditional manual plate turning operation determines the speed and direction of a conical roller table by visually judging the position of a billet, manually setting the rotational speed of the variable frequency motor of the conical roller table, stopping the plate turning when the billet turns to a suitable angle, starting side guide clamping, moving the side guide to the billet from both sides of the conical roller table to carry out the centering operation, and starting the conical roller table again to transport the billet into a rolling mill for rolling.

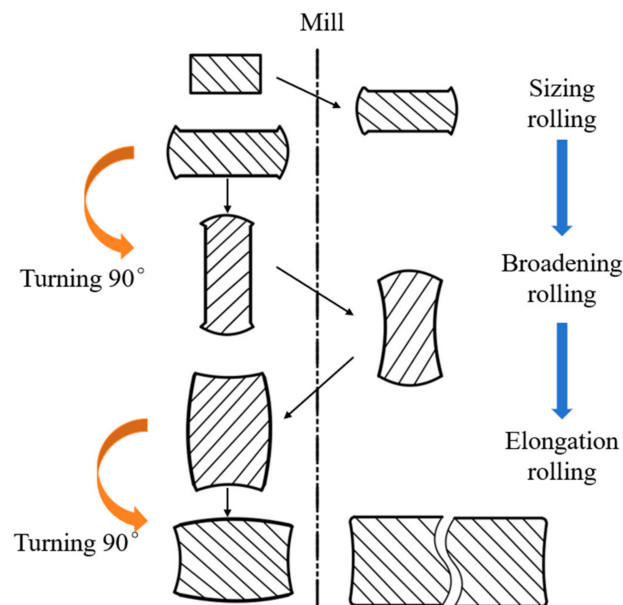


Figure 1. Schematic diagram of the plate rolling process.

The layout of the plate turning roller table is shown in Figure 2.

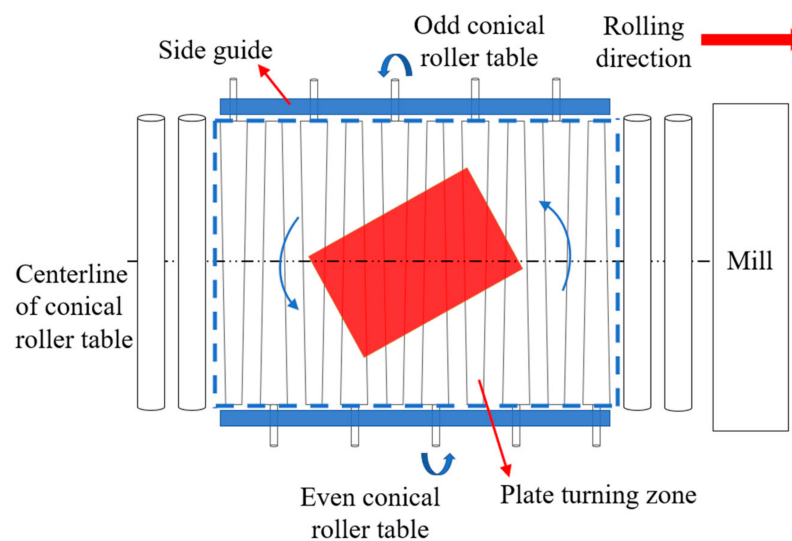


Figure 2. Schematic diagram of the conical roller table layout for plate turning.

Plate turning is an essential link in plate rolling production. Reducing the time of plate turning and improving the stability of the plate turning process are important means to improve the intelligent control of the rolling process. Nowadays, newly built modern plate production lines are well equipped and highly automated, especially in the rolling area, except for the plate turning operation, which has been automatically controlled. The plate turning operation has become the only bottleneck of automatic control in the plate rolling zone [6]. There are several main reasons that restrict automatic plate turning, as follows:

- (1) Because of the severe environment in the rough rolling area, the traditional visual inspection method cannot effectively perceive the position and angle information of the billet.
- (2) The control process of plate turning has characteristics of nonlinearity, strong coupling; it is multivariable; and there is no unified operation standard or control logic, so it is difficult to use a mechanism model for direct control.

At present, there are few studies in the field of automatic plate turning, both at home and abroad. Meng [7] proposed automatic plate turning technology in 2006 based on the image detection method. The original image is processed in advance through image grayscale, binarization, and opening an operation to remove the noise. The recognition program judges the coordinates of four vertices of ABCD of a rectangular billet in a binary image by setting effective values, predicting the position of the billet and the starting time of the side guide, and completing the plate turning process.

He [8] put forward a method of automatic plate turning for plate mills in 2009. The image processing includes histogram equalization, median filtering, optimal threshold image binary processing, Sobel operator billet image edge extraction, and sub-pixel edge location. Finally, high-precision billet edge information is obtained. In order to meet the requirements of image distortion and calibration, an eight-point correction method is used to establish the corresponding relationship between the billet size in the image and the actual size. Using the key points on the billet profile, the four boundaries of the billet are adjusted using the Hough transform method, and the rotation angle of the billet on the conical roller table is obtained. Finally, the conical roller table is driven to complete plate turning.

Wang [9] proposed full-automatic plate turning technology based on material tracking and positioning and image recognition technology. A canny edge detection algorithm is used to automatically extract billet edge contour information from field camera coordinates. The slope of a billet edge curve is fitted using the least square method. When the slope value changes by 90° (the error is 5%), the end of the plate turning is determined. In order to solve the water vapor problem caused by high-pressure primary descalers and laminar cooling, an axial flow fan is added to the purge, which improves the anti-interference ability of the image processing system.

The above scholars have applied machine vision technology to realize automatic plate turning. However, most of the above research methods only use simple image detection technology, and it is difficult to achieve high-speed, accurate, and automatic measurements in the harsh environment near the rolling mill in the actual production process to avoid the influence of complex factors such as water vapor and pollution. Moreover, the control strategy of a conical roller table is realized directly in the L1 basic automation control system, which is relatively simple and has a low success rate of plates turning in place at one time.

Based on machine vision technology, this paper designs a billet angle detection scheme for an automatic plate turning control system and puts forward the image defogging algorithm, the improved image adaptive enhancement algorithm, the external rectangle fitting algorithm based on Tukey weight, and the angle smoothing algorithm, which realizes the automatic elimination of water vapor interference and abnormal angle situations and improves the stability of billet angle detection. Combining the reinforcement learning theory with the motion model of plate turning, the reinforcement learning model of the intelligent control of plate turning is constructed by defining reinforcement learning elements and defining state space and action space, which can learn the optimal roller speed setting rules independently. The automatic plate turning control system greatly improves productivity, reduces the labor intensity of workers, can realize unmanned and intelligent operation of the plate turning process, effectively improves the automatic production level of plate plants, and has important practical significance for the future development of the intelligent steel industry.

2. Billet Angle Measurement Method Based on Machine Vision

The detection algorithm of the billet angle is the precondition for realizing automatic control in the process of plate turning. The traditional manual plate turning operation mainly inspects billets visually, which depends too much on the working experience of the operators, and it also has low efficiency and accuracy levels. With the development of intelligent technology, machine vision recognition technology has been widely used in

industrial fields, with advantages of fast processing speed, high detection accuracy, and easy integration. It has also been widely used in steelmaking, continuous casting, and rolling processes in the domestic steel industry [10,11].

In the production process, due to the limitations of the actual environment, the collected billet images are partially blocked or disturbed by water vapor and other factors. Traditional image processing algorithms have few factors to consider, and it is difficult to obtain continuous and stable billet angle changes, which affect the input rate of automatic plate turning. In order to realize billet angle measurement in a complex production environment, the image defogging algorithm, the improved image adaptive enhancement algorithm, the external rectangle fitting algorithm based on Tukey weight, and the angle adaptive smoothing algorithm are proposed, which can accurately detect the rotation angle information of the plate turning process and provide all-weather real-time measurement results for the plate turning control process.

2.1. Detection Device

In order to realize the real-time extraction of a billet angle, an industrial camera is installed in the entry and exit parts of the plate mill, as shown in Figure 3. The camera adopts Basler acA640-120gm; the pixel size is $5.60 \times 5.60 \mu\text{m}$; and the resolution is 659×494 pixels. The billet image data are collected by connecting with the industrial camera through the special image network; the position and rotation angle information of the billet are obtained using machine vision technology; and the recognition results are stored on the local disk [12].

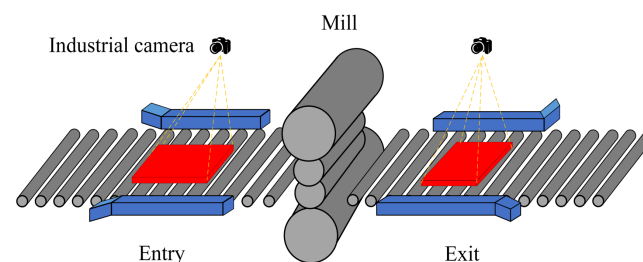


Figure 3. Schema of camera installation position.

2.2. Image Defogging Processing

In order to remove the scale on the surface of a rolled piece, a high-pressure primary descaler is needed during rough rolling. During this process, water comes into contact with the surface of the billet, and a large amount of water mist is produced due to temperature differences. This water mist shrouds around the billet, which causes the billet in the image to be partially or completely blocked and, at the same time, makes the billet image inevitably blurred along with other defects, which seriously affects the image quality and visibility. In this paper, the defogging algorithm based on the dark channel prior theory [13] is applied to restore a clear billet image, so as to solve the influence of water mist occlusion on billet recognition.

In the field of machine vision, the atmospheric scattering model is widely used in the field of image defogging, and the definition of the formula is as follows:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

In the formula, $I(x)$ denotes the foggy image; $J(x)$ denotes the fogless image to be restored; $t(x)$ is transmittance; and A represents the global atmospheric light value.

Firstly, for each pixel point in the billet image, the smallest pixel value in the three RGB channels is selected as the gray value, and the dark channel image is generated. The brightest 0.1% pixels are extracted according to the brightness in the dark channel image, and the value with the highest brightness point at the corresponding position is found in the original foggy image as the atmospheric light value A .

According to the atmospheric light component, the transmittance of each pixel in the original image is estimated. The formula for calculating transmittance is as follows:

$$\tilde{t}(x) = 1 - \omega \min_c \left(\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A^c} \right) \right) \quad (2)$$

where ω is 0.95, $\Omega(x)$ represents the window centered on pixel x .

Finally, the lower bound t_0 is restricted to t , and its value is 0.1, when $t < t_0$, make $t = t_0$. The final fog-free image $J(x)$ can be expressed as the following formula:

$$J(x) = \frac{I(x) - A}{\max[t(x), t_0]} + A \quad (3)$$

Based on the above formula, the billet image can be defogged according to the perspective and atmospheric light composition, and the processed picture is shown in Figure 4.

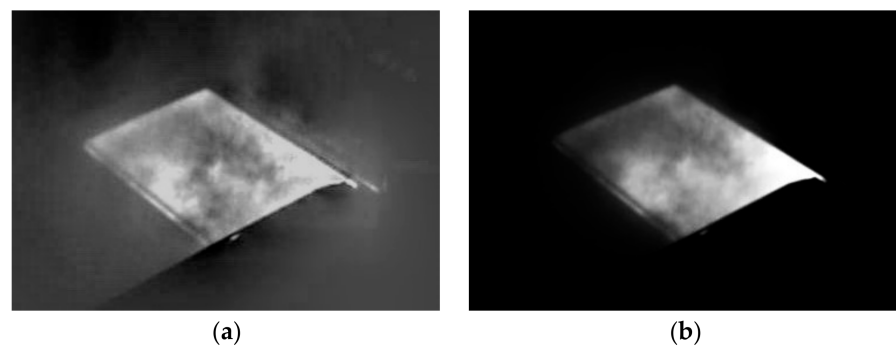


Figure 4. Schematic diagram of defogging billet surface: (a) fogged original image; (b) images after defogging.

2.3. Adaptive Enhancement of Billet Image

Due to the temperature change in the billet and the water vapor caused by the laminar cooling system, the brightness of the billet image is seriously uneven. Because of the dynamic change in the water vapor shielding process, the threshold segmentation algorithm for obtaining the billet's surface area has difficulty determining the appropriate threshold, which leads to a big difference between the obtained billet surface area and the actual one.

In this paper, an adaptive enhancement algorithm for a gray image is proposed, which can eliminate the influence of unstable factors and extract the surface area of the billet from the image background more clearly [14]. The flow of the adaptive enhancement algorithm is shown in Figure 5.

Firstly, the gray value of each pixel is changed by a linear transformation of the gray value, so as to improve the contrast between the light and dark parts of the billet image. The linear transformation formula is shown in Equation (4):

$$g' = g \times \text{Mult} + \text{Add} \quad (4)$$

In the formula, g denotes the gray value of the billet image before the gray stretching operation; g' denotes the gray value of the billet image after the gray stretching operation; Mult denotes the multiplier factor; and Add denotes the addend factor.

Then the mean filter is used to smooth the image after the linear transformation, and the average gray value of the filter mask window is calculated, which is used as the background estimation value of the local area. The billet image after contrast enhancement is compared with the billet image after smoothing one by one, and the threshold value is dynamically determined by the deviation between them. According to the dynamic threshold, the region that is several gray values higher than the smoothed image is segmented; that is, the foreground region. The background pixels are scanned in the foreground region; the holes in the foreground region are determined based on the scanning results; and the

holes are filled to make the region more continuous and complete. The isolated and small points and rough edge lines are removed by opening the operation, and the original areas of most areas are retained. The calculation formula for opening the operation is as follows:

$$A \circ B = (A \ominus B) \oplus B \quad (5)$$

In the formula, $A \circ B$ denotes the opening operation of A by using the structural element B , $A \ominus B$ denotes that A is corroded by B , and $(A \ominus B) \oplus B$ denotes that A is expanded by B .

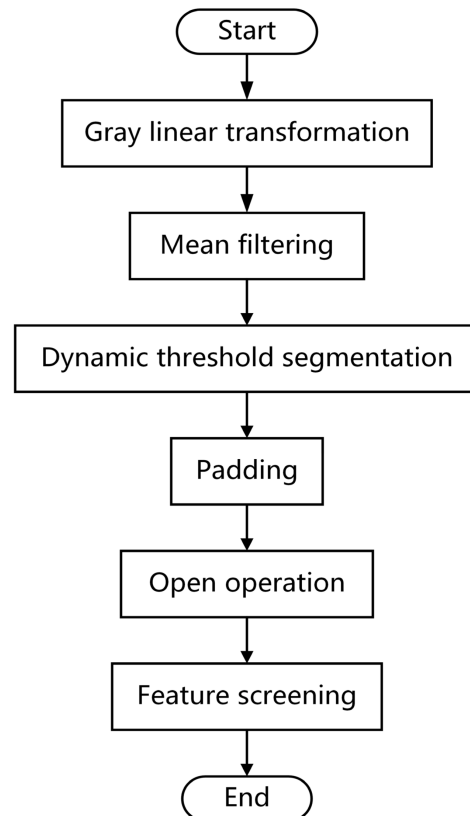


Figure 5. Flow chart of adaptive image enhancement algorithm.

Finally, the segmented regions are traversed by the 8-neighborhood adjacent relation. According to the actual production situation, the regions are screened according to the basic area range and rectangular degree of the billet; the screened connected regions are sorted according to the pixel area; and the connected region with the largest pixel area is selected as the billet feature image.

According to the steps in Figure 5, after the adaptive enhancement processing is finished for the collected original billet gray image, the billet surface with a rectangular shape in the segmented area can be obtained. Figure 6a is the original collected image; Figure 6b is the gray distribution of the original image surface; and the interference of water vapor leads to serious uneven light and shade on the surface. Figure 6c is the image processed by the adaptive enhancement algorithm, and the brightness uniformity of the billet surface is obviously improved. The image enhancement algorithm automatically adapts to billet detection under the condition of water vapor interference and provides support for the subsequent high-precision rotation angle measurement.

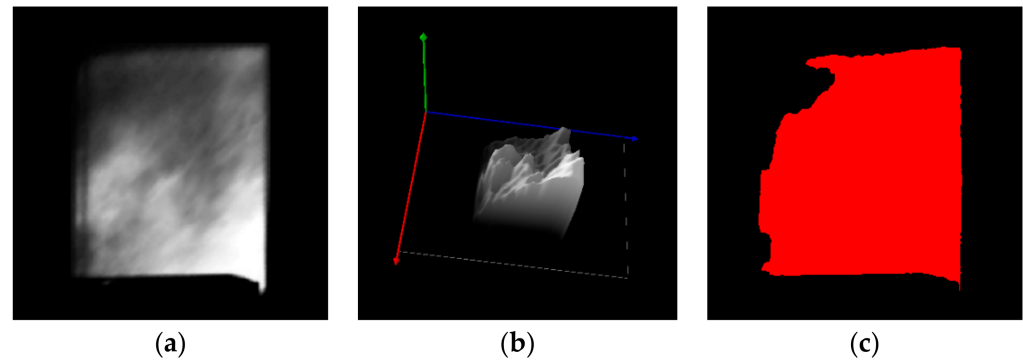


Figure 6. Processing results of the image adaptive enhancement algorithm: (a) image of the original billet; (b) gray scale distribution on the surface of the original billet; (c) surface area of the billet after adaptive reinforcement.

2.4. External Rectangle Fitting Algorithm Based on Tukey Weight

The enhanced billet contour is extracted by sub-pixel contour to obtain contour point sets at all edge positions [15]. Firstly, the traditional least square fitting method [16,17] is used to initially fit the pretreated contour point set. Its basic principle is to minimize the error by calculating the sum of squares of the distances from all points to the fitting straight line. Assuming that the straight line is expressed as $y = kx + b$, calculate the residual from each contour point to the initial fitting rectangle as follows:

$$R = \sum_{i=1}^N [y_i - (kx_i + b)]^2 \quad (6)$$

In some cases with large edge noise, the data contain a large number of noise points, and the traditional least square method cannot obtain the optimal result, so this paper adopts the least squares method based on Tukey's weight for fitting. The line fitting with increasing weight function takes into account the distance from different points to the line, which can improve the fitting accuracy. The Tukey-based weight function is defined as follows:

$$t(\gamma) = \begin{cases} \left[1 - \left(\frac{|\gamma|}{\eta}\right)^2\right]^2, & |\gamma| \leq \eta \\ 0, & |\gamma| > \eta \end{cases} \quad (7)$$

In the formula, γ represents the distance from a point to a straight line, and η is the clipping coefficient, which represents the distance, and its function is to judge outliers.

When $|\gamma| > \eta$, the value of the weighting function is 0, it means that a larger residual gives a smaller weight. When $|\gamma| \leq \eta$, the value of the weighting function varies with the distance in the (0, 1) interval. The smaller the distance from the point to the line, the larger the value of the weighting function, which means that a smaller residual gives a larger weight.

Under the conditions of the weighted least square method, the expression for finding the minimum residual is as follows:

$$R' = \sum_{i=1}^N [(y_i - (kx_i + b))t(\gamma)]^2 \quad (8)$$

According to the weighted residual, the rectangle is re-fitted. When the residual error is the smallest, the parameters of the linear equation are obtained.

Fitting based on the Tukey weight function can eliminate the influence of outliers and achieve more accurate fitting results. The results processed by the external minimum rectangle fitting algorithm are shown in Figure 7, and the included angle between the rectangular long axis and the x axis (-89.45°) is the real-time rotation angle of the billet.

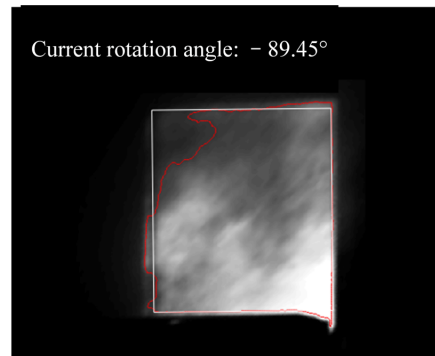


Figure 7. Schematic diagram of the circumscribed minimum rectangle fitting.

2.5. Smoothing Treatment Method of the Billet Angle

In the process of billet angle detection, due to the influence of partial occlusion, environmental interference, and other factors, the external rectangle fitting algorithm may fail to measure the angle and jump the angle in the implementation process, and the instability of the feedback angle seriously affects the real-time control process of plate turning.

In order to eliminate the unstable phenomenon of angle detection, this paper adopts the queue management method of “first-in, first-out” [18] to track the billet rotation angle. Through the logic relation of the billet rotation angle at different times, the angle is smoothed so as to avoid the influence of angle detection failure on plate turning control. When the billet reaches the turning position, the trigger event is given, and the billet angle is set near 0° as the initial angle. After starting the automatic turning control, the billet rotation angle is tracked according to the queue management method of “first in, first out” in each detection period. Before the billet detection angle is added to the queue, the angle is comprehensively judged by the stored preamble information in the queue. If the continuous condition is met, it is allowed to join the queue, thus realizing the adaptive processing of billet angle smoothing and detection failure. A threshold value is set to judge whether the current detection angle is normal, and if it exceeds the given threshold value, it indicates that the angle detection fails.

If angle detection fails, according to the preceding n groups of measurement information in the queue, a prediction model is established, and the Newton interpolation method [19] is adopted in the prediction model, as shown in Equation (9). The output value of the current angle α is calculated as the actual angle output after the smoothing processing and is added to the angle tracking queue.

$$f(x) = f(x_{t-n}) + (x - x_{t-n})f[x_{t-n}, x_{t-(n-1)}] + \dots + (x - x_{t-n})(x - x_{t-(n-1)}) \dots (x - x_{t-1})f[x_{t-n}, x_{t-(n-1)}, \dots, x_t] \quad (9)$$

where x_t is the last time to join the queue, x_{t-n} is the time to join the queue at time $t - n$, and the corresponding $f(x_{t-n})$ is the angle to join the queue at time $t - n$.

The abnormal handling process of the measuring angle based on the tracking queue is shown in Figure 8. For the angle jump caused by long and short edge detection errors caused by billet occlusion and the detection failure within protection time, an effective and credible feedback angle can be obtained by using the angle smoothing processing algorithm.

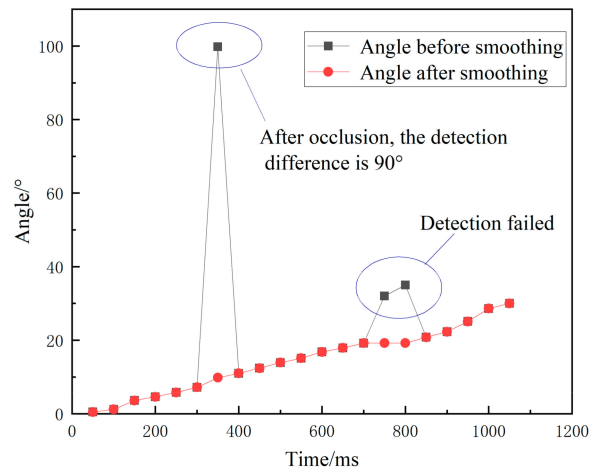


Figure 8. Smoothing treatment of billet measurement angle.

3. Automatic Plate Turning Control Model

Improving the efficiency of plate turning is the research and development goal of the automatic plate turning control system. In the actual production process, the plate turning efficiency is affected by many interwoven parameters, such as billet size, roller speed, equipment capacity, production environment, etc. There is no unified operation specification and control logic, which makes it difficult to further improve the plate turning efficiency using traditional model control methods. In recent years, with the continuous development and improvement to big data and artificial intelligence, data-driven intelligent optimal control methods have been widely used, and more and more engineering tasks have begun to try to use reinforcement learning algorithms to choose optimal control strategies. For sequential decision-making tasks, reinforcement learning can interact with the environment in real time and consider long-term returns. Compared with traditional control methods, reinforcement learning has special advantages in self-adaptation and self-learning, and it has achieved great success in solving some complex problems. It has been widely used in industries and academia [20–22].

Based on the reinforcement learning algorithm, this paper takes the time of plate turning as the constraint condition and obtains the optimal setting strategy for the plate turning speed by analyzing and summarizing the data of manual plate turning. The relationship between the setting speed and feedback speed of a conical roller table is simulated by establishing the roller table speed simulation model. The billet angle change formula is derived by combining the billet size, the real-time rotation angle, the roller table speed, and the conical roller size. Based on the above conditions, the simulation environment of plate turning is established to simulate the actual plate turning process. Then, based on the reinforcement learning theory, the reinforcement learning elements are defined for the motion model of plate turning; the forms of state space and action space are defined; the state is updated according to the state transition equation; and the reward function is designed to construct the reinforcement learning model of the intelligent control of plate turning [23]. The reinforcement learning intelligent decision algorithm is used to train the decision model for plate turning. When state information such as billet length and width and real-time angle is inputted, the most suitable speed turning angle of the billet is determined, and the speed setting of the conical roller table is controlled to realize rapid plate turning and improve the control accuracy.

3.1. Reasoning Model for Setting Optimal Plate Turning Speed

The size and real-time rotation angle of the billet are collected by industrial cameras installed near the entry and exit rolling tables of the rolling mill. Combined with the control instructions of the manual plate turning process and the change in the roller table speed, the optimal plate turning data of operators are collected based on the data preprocessing

algorithm. The optimal plate turning data are the optimal empirical strategy with the shortest plate turning times and the highest plate turning efficiency.

The related variables and operation-specified change curves of operators in the process of primary plate turning are shown in Figure 9. By analyzing and summarizing the operator's experience of plate turning, the optimal setting rules of the roller table speed in the process of plate turning are as follows: setting odd and even conical roller table speeds at the beginning of plate turning; setting the roller table speed to 0 at a suitable billet rotation angle; making the billet decelerate by inertia; and finally, making the billet rotate to a position near 90° when the roller table speed is 0°. It is necessary to determine the optimal turning point of roller table acceleration and roller table deceleration, that is, to determine the rotation angle of the billet at the turning point of speed.

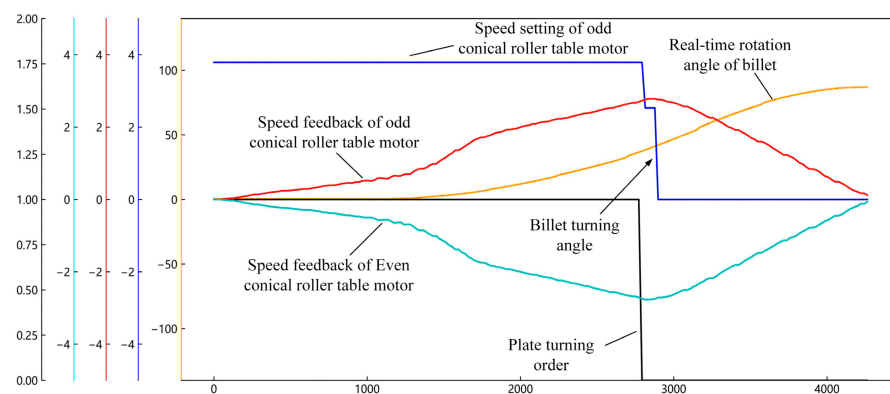


Figure 9. Curve chart of related variables in plate turning operation.

The plate turning system is a highly nonlinear and strongly coupled complex system. The speed turning point is affected by the motor capacity of the conical roller table, the thickness, width, and length of the rolled piece, and other factors, and the corresponding billet rotation angle is extremely uncertain. This paper relies on production data and data mining based on the reinforcement learning algorithm; establishes a digital reasoning model based on big data, automatically reasoning the action setting in the process of plate turning; and determines the optimal speed system under the condition of high-speed plate turning by comparing with the actual billet measurement angle.

3.2. Model of Setting Speed and Feedback Speed of Conical Roller Table Motor

By analyzing the change in related variable curves in the plate turning process, it can be known that the deceleration stage depends on inertia deceleration, and the slope of the deceleration curve keeps near a certain constant value, so the first-order linear control system can be used to describe the relationship between the set speed of the roller table motor and the actual feedback speed in the plate turning process.

Firstly, according to the characteristics of the roller table deceleration process and the system requirements, the transfer function model of the first-order control system is defined, and the simulation experiment of the transfer function model is carried out using multiple optimal plate turning data. Among them, the first-order control system is composed of a first-order transfer function, which can be used to dynamically show the relationship between input and output as a first-order linear differential equation, as shown by the following Equation (10) [24]:

$$T \frac{dy(t)}{dt} + k_1 y(t) = k_2 x(t) \quad (10)$$

where T is the time constant; k_1 and k_2 are proportional gains; k_2 is the input signal; $y(t)$ is the output signal; and $\frac{dy(t)}{dt}$ is the derivative (rate of change) of the output variable $y(t)$ with respect to time t .

The simulation experiment selects the same step signal as the actual plate turning process and makes the simulation results close to the actual results by adjusting the two key parameters: proportional gain and time constant, as shown in Figure 10, where $x(t)$ is the given value and $y(t)$ is the feedback result. Finally, the specific parameters of the first-order simulation system can be determined as follows: $k_1= 1.5$; $k_2= 1.2$; and $T= 0.5$. The simulation results show that the system index can meet the control requirements of the actual roller table deceleration process.

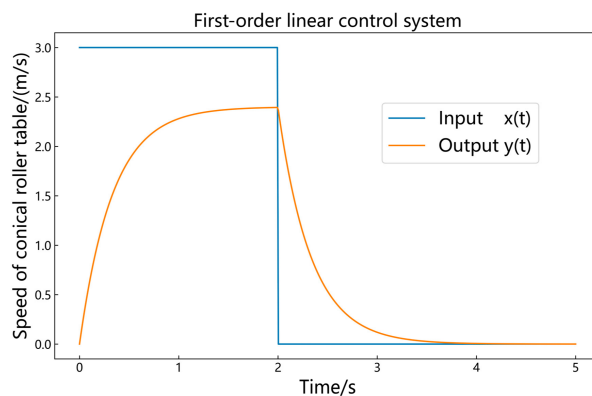


Figure 10. Schematic diagram of simulation of the first-order linear control system.

The transfer function model is adjusted by several model parameters determined based on the simulation results of simulation experiments, and the set speed and feedback speed model of the tapered roller table motor are obtained, which can accurately simulate the relationship between the set speed and feedback speed of the motor in the deceleration stage of the roller table in the plate turning process.

3.3. Theoretical Model of Billet Rotation Speed

In the process of plate turning, the motor drives the tapered roller to rotate. Because of the linear friction between the tapered roller and the billet, and the fact that the linear friction is distributed differently in the longitudinal direction, it will exert horizontal rotating torque on the billet. A rotational moment brings horizontal rotational acceleration, which causes the billet to rotate. Based on the physical changes in the actual process of plate turning, a mathematical equation is constructed to determine the relationship between the angular velocity of the billet and other factors and then accurately predict the change in the billet rotation angle. The dimensions of the billet and the conical roller table are shown in Figures 11 and 12.

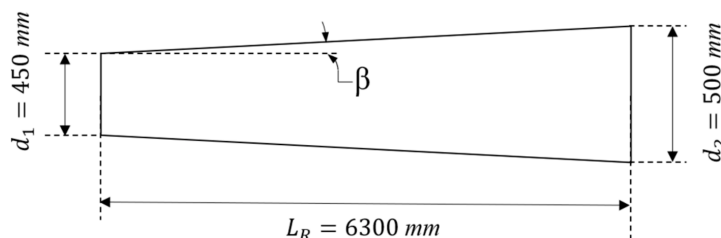


Figure 11. Dimension drawing of the conical roller table.

When the billet rotation angle is α , the projection of the billet diagonal in the axial direction of the roller table is as follows:

$$L_1 = L \sin \alpha + W \cos \alpha \tag{11}$$

where L_1 is the projection of the billet diagonal in the axial direction of the roller table, L is the billet length, W is the billet width, and α is the real-time rotation angle.

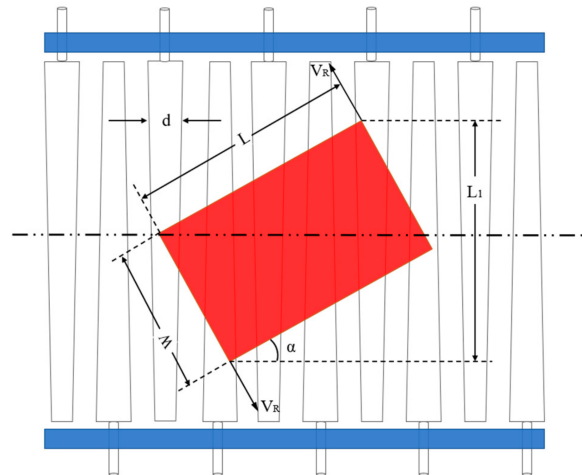


Figure 12. Schematic diagram of the billet and the roller table dimensions.

During the rotating process of the rotary roller table, the contact position between the billet and conical roller table changes constantly, and the diameter of the contact roller table changes accordingly. The calculation formula for the roller table diameter at the contact position between the billet and the conical roller table is as follows:

$$\tan \beta = \frac{d_2 - d_1}{2L_R} \quad (12)$$

$$d = d_1 + \tan \beta (L_R + L_1) \quad (13)$$

where d is the roller table diameter at the contact position between the billet and the conical roller table, d_1 is the small end roller diameter, d_2 is the large end roller diameter, L_R is the roller table length, and β is the taper of the conical roller.

The horizontal linear velocity V_P at the contact position between the billet and the conical roller table is calculated by Equation (14):

$$V_P = \frac{d}{d_2} V_R \quad (14)$$

where V_P is the horizontal linear velocity at the contact position between the billet and the conical roller table, and V_R is the linear velocity at the lap position between the conical roller table and the billet.

Thus, the angular velocity of the billet at the rotation angle α is as follows:

$$\omega = \frac{V_P}{r} = \left[2 + \frac{(d_2 - d_1)(L_R - L_1)}{L_R d_1} \right] \frac{V_R}{L_1} \quad (15)$$

where ω is the angular velocity of the billet at the rotation angle α , $r = \frac{L_1}{2}$.

Based on the above calculation process, the angular velocity of the billet can be calculated according to the billet size, the roller table linear velocity, and the rotation angle information, and then the angle change in the whole plate turning process can be predicted.

3.4. Building Simulation Environment of Automatic Plate Turning Based on Gym

In order to simulate the rotating process of the billet on the conical roller table, a set of simulation environments for automatic plate turning were built using the Gym framework of OpenAI, which can meet the experimental scene training of the speed and billet angle changes in the conical roller table in a real environment and can display the graphical interface. Using a virtual simulation experiment can save manpower and material resources for scene building in real scenes, and it also avoids actual losses caused by trial and error in a real environment, thus saving resources to a great extent.

(1) State space:

The state space describes the environment to the agent, and the state space needs to provide enough information to make the agent take appropriate actions. In the simulation environment of this paper, the length, width, and real-time rotation angle of the billet are obtained as state information, as shown in Table 1. According to the actual plate turning situation and the stability of the control system, the maximum and minimum values are set for the state information [25].

Table 1. State design.

State	Minimum Value	Maximum Value
Current angle	0°	180°
Billet length	1000 mm	3000 mm
Billet width	1000 mm	3000 mm

In order to obtain the optimal operation output under any billet size, the initial angle is set to 0 after each environment reset, and a random billet size is generated according to the length-width ratio of 0.5~2.0, which can meet the rapid plate turning operation of billets with different specifications.

(2) Action space:

In this paper, the action space of the conical roller table is decomposed into two stages: acceleration and deceleration, to control the conical roller table continuously. In order to maintain consistency with the action behavior of the conical roller table in the real world as much as possible, 1 is used to denote the acceleration process of the roller table, and 0 is used to denote the inertia deceleration process of the roller table, as shown in Table 2.

Table 2. Action design.

Action	Meaning
1	Acceleration stage curve with step length of 20 ms
0	Deceleration stage curve with step length of 20 ms

(3) Transition of environmental state:

After the environment receives the action setting of the conical roller table transmitted by the agent, it outputs the feedback speed of the roller table according to the set speed and feedback speed model of the conical roller table motor; inputs the feedback speed of the roller table and the billet state into the billet speed model; and outputs the real-time angular speed, the product of the angular speed, and the preset time interval plus the initial angle of the billet as the current angle of the billet, where the default time interval is $dt = 20$ ms. In the training process, the plate turning simulation environment continuously detects the feedback speed of the roller table and ends the iterative training when the feedback speed of the roller table is detected to be zero.

(4) Design of the reward function:

The deep reinforcement learning neural network model needs to guide training according to the reward value, and the design of the reward function needs to be consistent with the task goal. The stronger the proportional relationship between the reward function and goal state, the faster the agent learns to deal with the task.

In the simulation environment, the closer the real-time angle of the billet is after one-time plate turning to the target value, that is, an angle of 90° rotation, the better the performance of the action sequence, and the less the plate turning steps are expected to ensure the shortest control time, so negative rewards and guiding rewards are set to make the agent finish the exploration as soon as possible. Therefore, the reward function in this article is designed according to the following steps:

Take the current angle as A , the target angle as T , $T = 90^\circ$, and $action \in \{0, 1\}$. The reward function can be defined as follows:

$$R_{action} = \begin{cases} 1, & action = 1 \text{ and } 0 < A < T/3 \\ -1, & action = 0 \text{ and } 0 < A < T/3 \\ 1, & action = 0 \text{ and } \frac{2}{3}T < A < T \\ -1, & action = 1 \text{ and } \frac{2}{3}T < A < T \end{cases} \quad (16)$$

$$R_{done} = \begin{cases} (reward_max - reward_min) * (1 - abs(T - A)/T) + reward_min \\ -10, & A > T \end{cases} \quad (17)$$

$$R_{total} = R_{action} + R_{done} \quad (18)$$

R_{done} is a sparse bonus. When a sparse bonus is triggered, the billet will end this round of training [26]. $reward_min$ and $reward_max$ are the minimum and maximum values of the reward value, respectively. The score, according to the difference between the target angle and the current angle, is combined with the reward range. If the billet exceeds the target angle during rotation, it will receive a negative reward of -10 .

R_{action} is a guiding reward, and its purpose is to speed up the convergence of the algorithm. According to the actual experience of setting the speed of plate turning, it is in the stage of roller table acceleration within $0^\circ \sim 30^\circ$, and in the stage of deceleration, it is within $60^\circ \sim 90^\circ$. If you select the corresponding action, you will receive a positive reward of 1, and if you make a mistake, you will receive a negative reward of -1 .

All awards in the current round add up to the total award R_{total} .

3.5. Reinforcement Learning Algorithm Training

The DeepMind team put forward the DQN (Deep Q-Network) [27] algorithm in 2013. Based on the Q-Learning algorithm, it introduced an experience replay mechanism and a target network structure, and it achieved amazing performance in an Atari game environment, which kicked off the large-scale research of deep reinforcement learning.

DQN randomly samples N tuples $\{(s_i, a_i, r_i, s_{i+1})\}_{i=1, \dots, N}$ from the replay buffer as a small batch of training data, takes the loss function of Formula (19) as the optimization objective of the current network, and uses the gradient descent method to solve its weight.

$$L = \frac{1}{N} \sum_{i=1}^N [(r + \gamma \max_{a'} Q(s_{i+1}, a_{i+1}, \theta) - Q(s, a, \theta))^2] \quad (19)$$

However, both the current Q value and the target Q value in Formula (19) use the same parameters and are updated synchronously, which makes the model training unstable and difficult to converge. In order to solve this situation, the DQN algorithm uses the old network parameters to evaluate the state Q value of the next time step and updates the parameters every certain time step, as shown in Figure 13. The current network $Q(s, a, \theta)$ is used to evaluate the value function of the current state and action, and the target network $Q(s_{i+1}, a_{i+1}, \theta')$ is used to calculate the temporal difference. In this way, it provides a relatively clear reference target for the current Q network to be fitted and adjusts the optimization target to the Formula (20).

$$L = \frac{1}{N} \sum_{i=1}^N [(r + \gamma \max_{a_{i+1}} Q(s_{i+1}, a_{i+1}, \theta') - Q(s, a, \theta))^2] \quad (20)$$

In the construction of the algorithm network, reinforcement learning needs to use the trained model immediately after training, so it needs faster training and inference speed. In order to achieve this goal, we chose a shallow network structure; that is, a three-layer network. Generally, a slightly redundant network capacity is selected to keep good fitting, and the data of 2 to the N power is selected as much as possible, which can be completely put into the GPU for calculations, so 128 neurons were selected in this paper. ReLU has linear characteristics, a fast calculation speed, and is helpful for solving the problem of

gradient disappearance, so the ReLU function was used as the activation function in this paper. Because PPO and SAC algorithms also need to meet similar requirements, these three hyperparameters were set in the same way as DQN. The hyperparameters of the DQN algorithm are shown in Table 3.

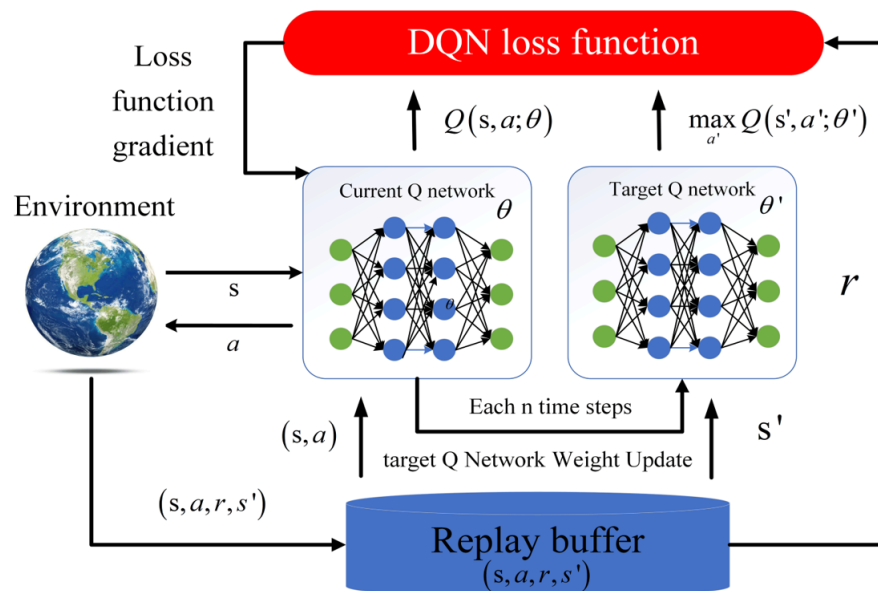


Figure 13. DQN algorithm framework.

Table 3. DQN algorithm hyperparameter settings.

Hyperparameter	Value
Q network learning rate	2×10^{-3}
Discount factor	0.98
Replay buffer size	10,000
Batch size	64
Target network update parameters	10
Number of neural network layers	3
Number of hidden neurons per layer	128
Activation function	ReLU
Exploration factor	0.01

The PPO (proximal policy optimization) [28] near-end policy optimization algorithm is a reinforcement learning benchmark algorithm based on the Actor-Critic framework proposed by OpenAI in 2017. It not only has good performance, but is also easier to implement than the previous TRPO (trust region policy optimization) [29].

The PPO algorithm uses iterative learning of the original data to ensure that the loss of function value in the iterative process is minimized to prevent the gradient value from updating too fast. It uses the proportion of the old and new strategies to control the update range of the new strategy to obtain a more optimized strategy and increase the robustness of the algorithm. The ratio r_t is recorded as follows:

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \tag{21}$$

In the learning process, the replay buffer stores the training value data of the state, action, and reward obtained when executing the current strategy of the Actor network.

These training data are used to calculate the loss function needed to update the network weights. The PPO algorithm loss function is shown in Formula (22).

$$L^{CLIP}(\theta) = E_t[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)A_t)] \tag{22}$$

where θ represents the policy parameter; $r_t(\theta)$ represents the ratio of the occurrence probability of the old and new strategies, which is used to constrain the old and new strategies so that they will not differ too much; A_t is the advantage function at time t , which is used to express the advantage of the updated strategy compared with the old strategy; ϵ is the super parameter controlling the cutting interval, which represents the maximum difference between the old and new strategies, and usually takes the values of 0.1 and 0.2; E_t denotes empirical expectations; $\text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)$ is a clipping function that restricts the strategy ratio $r_t(\theta)$ in the range of $[1 - \epsilon, 1 + \epsilon]$, and outputs $1 - \epsilon$ if $r_t(\theta) < 1 - \epsilon$; if $r_t(\theta) > 1 + \epsilon$, $1 + \epsilon$ is output.

Among many reinforcement learning algorithms, PPO has the advantages of strong adaptability and stable training. The objective function proposed by PPO can be iteratively updated with a small number of samples in multi-round training, which solves the problems of the step size being difficult to determine and the update difference being too large in the strategy gradient algorithm.

The specific framework of the PPO algorithm is shown in Figure 14.

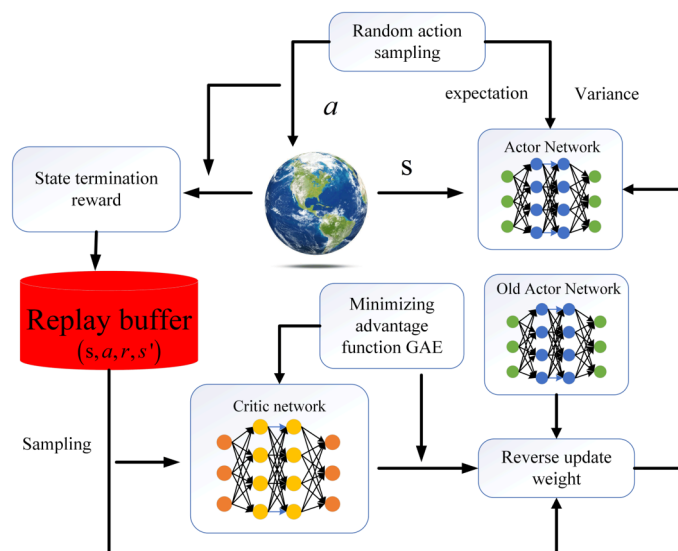


Figure 14. PPO algorithm framework.

On the construction of algorithm network, the hyperparameters of the PPO algorithm are shown in Table 4.

Table 4. PPO algorithm hyperparameter settings.

Hyperparameter	Value
Actor network learning rate	1×10^{-4}
Critic network learning rate	5×10^{-3}
Discount factor	0.95
Number of neural network layers	3
Number of hidden neurons per layer	128
Activation function	ReLU
GAE parameters	0.95
Estimate the clipping coefficient of advantage function	0.2

The SAC (Soft Actor Critic) [30] algorithm introduces entropy into the objective function to reward exploration so as to achieve a balance between exploration and utilization.

At the same time, a parallel architecture can be established based on off-policy, which accelerates the learning speed. The increase in entropy improves the exploration ability of agents in the algorithm. The randomness of the strategy can prevent the strategy from converging prematurely to the local optimal value, and it has considerable performance. The specific framework is shown in Figure 15.

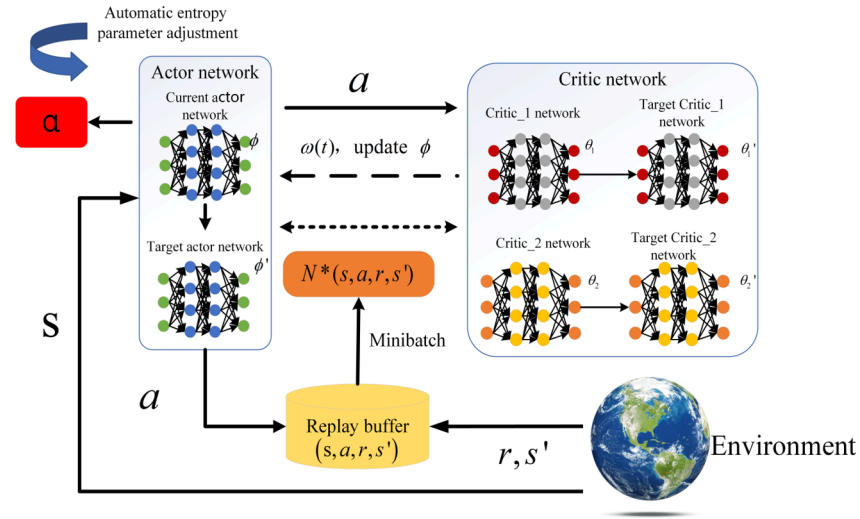


Figure 15. SAC algorithm framework.

The SAC samples n tuples from the empirical buffer, and for each tuple $\{(s_i, a_i, r_i, s_{i+1})\}_{i=1, \dots, N}$ it calculates y_i with the target network:

$$y_i = r_i + \gamma \min_{j=1,2} Q_{\theta_j'}(s_{i+1}, a_{i+1}) - \alpha \log \pi_{\theta}(a_{i+1}|s_{i+1}) \quad (23)$$

where $a_{i+1} \sim \pi_{\theta}(\cdot|s_{i+1})$.

Update both Critic networks as follows: For $j = 1, 2$, minimize the loss function:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - Q_{\theta_j}(s_i, a_i))^2 \quad (24)$$

Sample action \tilde{a}_i using the re-parameterization technique, and then update the current Actor network using gradient ascending:

$$L_{\pi}(\theta) = \frac{1}{N} \sum_{i=1}^N \left(\alpha \log \pi_{\theta}(\tilde{a}_i|s_i) - \min_{j=1,2} Q_{\theta_j}(s_i, \tilde{a}_i) \right) \quad (25)$$

And automatically update the entropy regularity coefficient α :

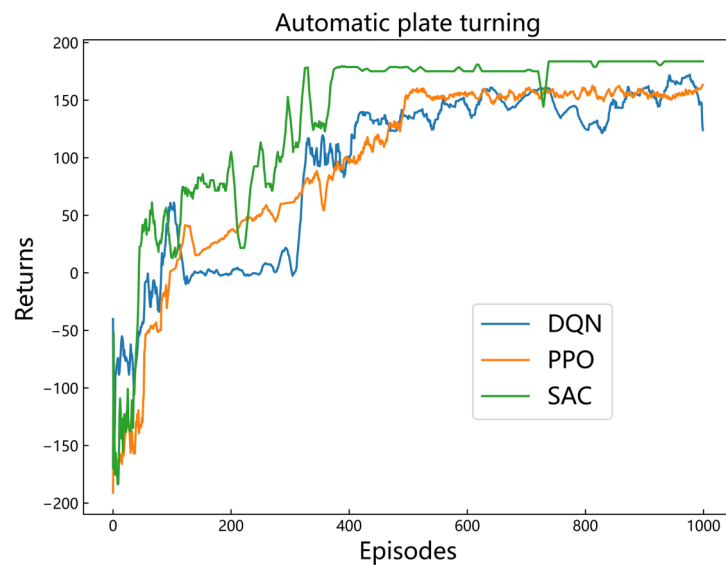
$$L(\alpha) = E_{s_t \sim R, a_t \sim \pi(\cdot|s_t)} [-\alpha \log \pi(a_t|s_t) - \alpha H_0] \quad (26)$$

In the construction of the algorithm network, the hyperparameters of the SAC algorithm are shown in Table 5.

In each plate turning task, the conical roller table will drive the billet to rotate in a random action selection mode according to the action space. After 1000 rounds of training, the average value of the reward function is output every 10 rounds in the training process. After the model converges, the network parameters of the current best training results are saved, and the reward value change graph in the training process is obtained, as shown in Figure 16. The abscissa is the number of rounds, and the ordinate is the reward value of the algorithm.

Table 5. SAC algorithm hyperparameter settings.

Hyperparameter	Value
Actor network learning rate	1×10^{-3}
Critic network learning rate	1×10^{-2}
entropy regularity coefficient learning rate	1×10^{-2}
Discount factor	0.98
Replay buffer size	100,000
Number of neural network layers	3
Number of hidden neurons per layer	128
Activation function	ReLU
Entropy regularization coefficient	0.01

**Figure 16.** Training process of reinforcement learning algorithm.

It can be seen from the figure that the training results of the three algorithms in the first 100 rounds do not increase significantly, and the reward value will fluctuate up and down during the training process. This is because the samples used by the neural network in the training process are all from the experience data obtained when the conical roller table randomly executes the rotating task during the exploration process, while the number of successful experience samples is small at the initial stage of training. When the successful experience samples accumulate to a certain extent, the action variance gradually decreases, and there is an obvious leap in the 100th to 300th training rounds, and the cumulative return rises steadily.

The SAC algorithm reached its first peak in the 390th training round, and the reward value reached 175 in the 400th training round. The model basically learned the optimal control strategy for steel transfer, the model converged in the following rounds, and the reward value remained stable. Then, after a short decline, the reward value stabilized at 183 after the 680th training round.

The PPO algorithm reached its first peak in the 510th training round. Because of the importance of sampling in the PPO reinforcement learning model, the reinforcement learning model converges well, the reward function curve grows steadily without excessive fluctuation, and it stabilizes at about 158 after the first peak.

The DQN algorithm reached its first peak in the 370th training round, and the reward value tended to be 128. After the first peak, although the reward value increased obviously, it fluctuated greatly; even some rounds were lower than the reward value of the first 400 rounds, and the convergence was poor.

Therefore, as can be seen from Figure 16, the SAC algorithm in this paper is obviously superior to the DQN and PPO algorithms in the task of automatic steel transfer because of its faster convergence speed, smaller fluctuation of rewards, and larger reward value after convergence.

3.6. Online Strategy Optimization

In this paper, the action strategy output by the SAC algorithm is taken as the initial strategy, and the strategy will continue to be automatically optimized in the subsequent plate turning action. After starting the plate turning, the state of the billet is input into the Actor network, and its output action is applied to the speed control of the roller table. When each plate turning process is finished, each plate turning sequence is added into the experience buffer D , and the Critic network is trained first using the plate turning gap time, and then the Actor network is trained using Critic network to obtain the extraction of the latest dataset of plate turning strategy. Through online datasets and continuous training, the plate turning strategy output by the Actor network can adapt to the production environment continuously, obtain higher control efficiency and better control accuracy, and be widely used in the plate turning control process of plates.

4. Application of Automatic Plate Turning System

As shown in Figure 17, the automatic plate turning system has been successfully applied in a 4100 mm plate production line that uses a double-stand, four-high reversible rolling mill. For the control range, where the rotation angle is $85^{\circ}\sim 95^{\circ}$, the original operator needs frequent manual intervention, and the success rate of one-time plate turning is low. After the automatic plate turning is put into operation, the one-time rotation rate is $\geq 98.5\%$.

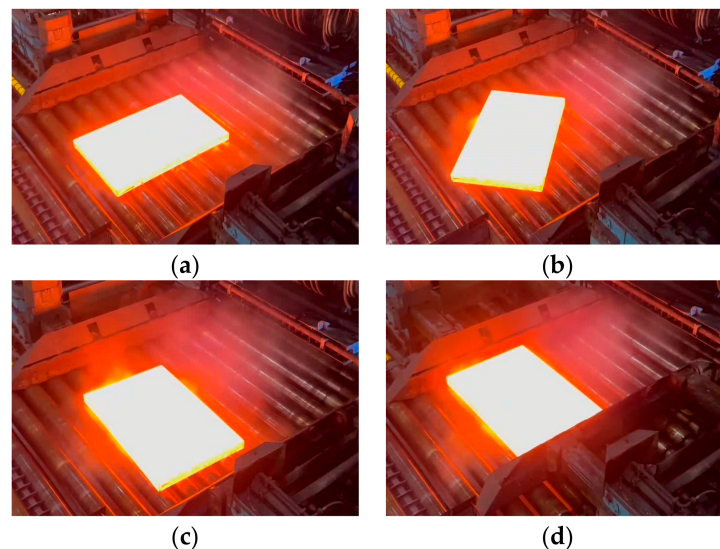


Figure 17. Processing results of the image adaptive enhancement algorithm: (a) the billet is in place and stopped; (b) start plate turning; (c) the plate turning in place; and (d) side guide clamping.

During the field test of the automatic plate turning control system, the reasoning time for the plate turning strategy is less than 20 ms, and the closed-loop control time, including image processing, data communication, and setting plate turning commands, is less than 50 ms, which meets the real-time control requirements for automatic plate turning in the production process. The intelligent control method based on the whole process of plate turning reduces the time consumption of plate turning passes to the greatest extent and makes the plate turning process more stable and smooth.

Figure 18 displays the comparative data of manual plate turning and automatic plate turning tests during a period of the production process. In the process of plate turning, the turning time is closely related to the length/width ratio of the billet. In order to better

analyze and evaluate the efficiency of plate turning, we classified and counted the billets with different length/width ratios by taking the manual and automatic plate turning times as indexes. Figure 18a takes the start of plate turning to the completion of plate turning as the time statistical range, and Figure 18b takes the billet throwing out signal before plate turning to the billet biting in signal after plate turning as the time statistical range. Compared with manual plate turning, the average plate turning time of each billet during automatic plate turning is obviously reduced, and the fastest time can be shortened by more than 1 s. The average rolling rhythm of a billet is about 150 s, and the average time saved by automatic plate turning is 0.7 s, so the rolling efficiency can be improved by 0.46%. The overall efficiency can be improved by 0.23% when the contribution of overall efficiency improvement is calculated at 50%. Taking the annual output of a steel mill of 1 million tons as an example, the efficiency is improved and the production is increased by about 2300 tons, which greatly improves the economic benefits.

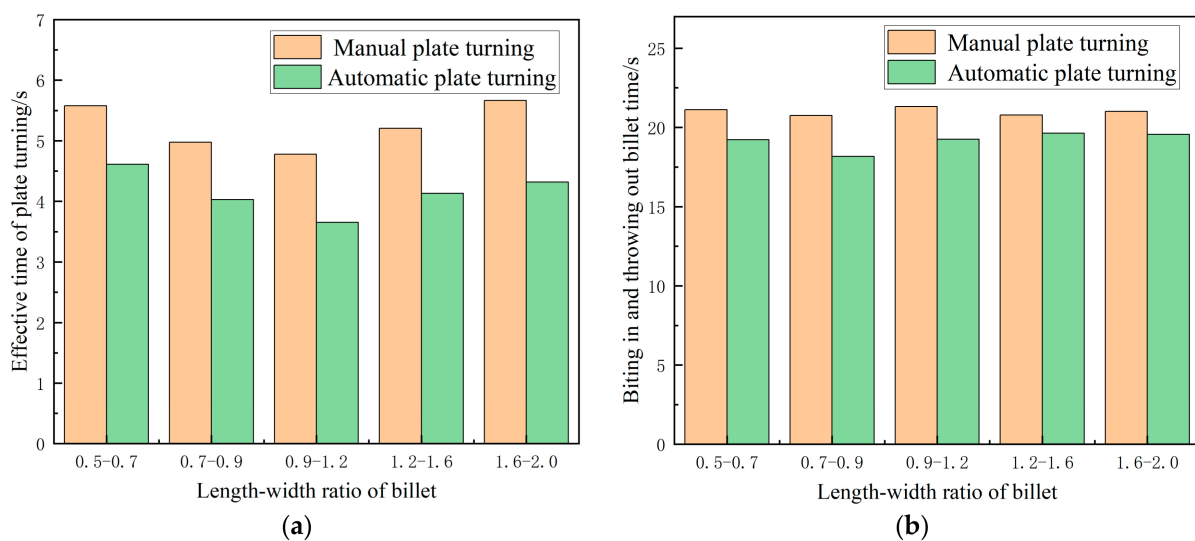


Figure 18. Comparison of plate turning time: (a) effective time of plate turning; (b) biting in and throwing out billet time.

5. Conclusions and Prospect

5.1. Conclusions

To solve the problem of plate turning in the plate rolling production line still needing manual control by operators, an automatic plate turning control system was designed and developed in this paper. The improved image processing algorithm can realize the billet angle measurement function in a complex production environment. The digital model and optimization algorithm for automatic plate turning based on reinforcement learning were established, and the automatic optimization of plate turning speed and precision was completed. The main conclusions of this paper are as follows:

- (1) The image defogging algorithm, improved image adaptive enhancement algorithm, external rectangle fitting algorithm based on Tukey weight, and angle smoothing algorithm were proposed in order to resolve the problem of billet angle detection in complex production environments. The fusion processing of the above algorithms can adapt to changes in influencing factors such as water vapor interference and partial occlusion in the rolling process, automatically eliminate water vapor interference and abnormal angle situations, and provide real-time and stable angle detection values for the automatic plate turning control system.
- (2) Using the big data of manual plate turning operations, the optimal roller table speed setting rules were obtained. The set speed and feedback speed models of the conical roller table motor and the theoretical model of billet speed were constructed to simulate the plate turning process. Based on reinforcement learning theory, the elements of

reinforcement learning were defined for the motion model of plate turning. The forms of state space and action space were defined; the state was updated according to the state transition equation; and the reward function was designed. The reinforcement learning model for intelligent control of plate turning was constructed.

- (3) Three reinforcement learning algorithms, DQN, PPO, and SAC, were used to train the model in the simulation environment of automatic plate turning. The simulation results showed that the SAC algorithm has a faster convergence speed, fewer fluctuations of rewards, and reward value after convergence, and it has better online optimization performance than the DQN and PPO algorithms in automatic plate turning tasks.
- (4) In the testing process of the automatic plate turning system, the closed-loop control time, including image processing, data communication, and plate turning command setting, was less than 50 ms, and the plate turning angle detection error was less than $\pm 2^\circ$. For the control range of $85^\circ \sim 95^\circ$, the rate of one-time plate turning was more than 98.5%, and the average plate turning time of each billet was greatly shortened compared with manual plate turning, and the fastest time could be shortened by more than 1 s.

5.2. Prospect

In the process of plate turning, due to the warping of the billet, the buckling head, and the contact part of the roller table, the slip between the billet and the roller table occasionally occurs, and the plate turning area becomes limited, which may cause the billet to deviate from the center of the roller table and exceed the limit, which affects the rate of plate turning in a place at one time. In the future, it is necessary to continue to collect more manual plate turning operation instructions under abnormal conditions. By optimizing and improving the offline reinforcement learning algorithm, the ability to identify abnormal characteristics of plate turning quickly and accurately is enhanced, and the self-organization and self-optimization abilities of the data-driven model are used to find the optimal plate turning strategy, so as to further improve the efficiency of the automatic plate turning system.

Author Contributions: Conceptualization, C.H. and Z.J.; methodology, C.H. and S.X.; software, S.X.; validation, S.X.; formal analysis, Z.W. and Z.Z.; investigation, Z.W. and Z.Z.; data curation, S.X.; writing—original draft preparation, C.H. and S.X.; writing—review and editing, C.H. and S.X.; visualization, Z.W.; supervision, C.H. and Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities (N160704003, N170708020, N2107007).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Wang, G.D. Status and prospects of research and development of key common technologies for high-quality heavy and medium plate production. *Steel Roll.* **2019**, *36*, 1. [[CrossRef](#)]
2. Jiao, Z.J.; He, C.Y.; Wang, L.X.; Cai, Y.L.; Wang, X.; Sun, X.D. Torque Model in Plate Rolling Process with Biting Impact Considered. *ISIJ Int.* **2021**, *61*, 239–247. [[CrossRef](#)]
3. Jiao, Z.J.; Luo, J.Y.; Wang, Z.Q.; Xu, Z.P.; He, C.Y.; Zhao, Z. Research and application of the angular rolling technology for plate mill. *Adv. Manuf.* **2023**, *11*, 462–476. [[CrossRef](#)]
4. Dong, Z.S.; Li, X.; Luan, F.; Zhang, D.H. Prediction and analysis of key parameters of head deformation of hot-rolled plates based on artificial neural networks. *J. Manuf. Process.* **2022**, *77*, 282–300. [[CrossRef](#)]
5. Schausberger, F.; Steinboeck, A.; Kugi, A.; Jochum, M.; Wild, D.; Kiefer, T. Vision-Based Material Tracking in Heavy-Plate Rolling. *IFAC-Pap.* **2016**, *49*, 108–113. [[CrossRef](#)]
6. Jiao, Z.J.; He, C.Y.; Zhao, Z.; Wu, Z.Q.; Wang, J. Research progress and application of high precision intelligent control system for plate rolling. *Steel Roll.* **2022**, *99*, 52. [[CrossRef](#)]

7. Men, Q.L. An image identification based automatic plate turning scheme for wide heavy plate mills. *Metall. Ind. Autom.* **2010**, *34*, 55–60.
8. He, C.Y.; Jiao, Z.J.; Wu, X.G.; Wang, J. Research on Control Method of Turning Plate Based on Image Processing Technology. In Proceedings of the 2015 6th International Conference on Manufacturing Science and Engineering, Guangzhou, China, 28–29 November 2015; Atlantis Press: Qingdao, China, 2015. [[CrossRef](#)]
9. Wang, L. Fully automatic steel conversion and widening technology based on image processing system during the rolling process of wide and thick plates. *Shanxi Metall.* **2018**, *41*, 88. [[CrossRef](#)]
10. Mukhopadhyay, P.; Chaudhuri, B.B. A survey of Hough Transform. *Pattern Recognit.* **2015**, *48*, 993–1010. [[CrossRef](#)]
11. Liu, H. Survey for application of conventional artificial intelligence technologies in the steel & iron field. *Metall. Ind. Autom.* **2019**, *43*, 24–28.
12. Jiao, Z.J.; Gao, S.W.; Liu, C.J.; Luo, J.Y.; Wang, Z.Q.; Lang, G.Y.; Zhao, Z.; Wu, Z.Q.; He, C.Y. Digital Model of Plan View Pattern Control for Plate Mills Based on Machine Vision and the DBO-RBF Algorithm. *Metals* **2024**, *14*, 94. [[CrossRef](#)]
13. He, K.M.; Sun, J.; Tang, X.O. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [[CrossRef](#)] [[PubMed](#)]
14. Choi, D.; Yun, J.P.; Jeon, Y.; Kim, S.W. Pinhole detection in steel slab images using machine vision. *IFAC Proc. Vol.* **2009**, *42*, 397–401. [[CrossRef](#)]
15. Yang, H.; Zhang, Y.; Gao, Y.J.; Zhou, P.; Wang, C.Z.; Huo, X.G. Development and Application of Steel Plate Contour Online Detection System Based on Machine Vision. *Shandong Metall.* **2022**, *44*, 49. [[CrossRef](#)]
16. Qin, D.D.; Lu, J.; Song, Y.Q. The Research on Appearance Recognition and Visual Inspection Technology of Workpiece. *Modul. Mach. Tool&Autom. Manuf. Tech.* **2018**, *9*, 84–87. [[CrossRef](#)]
17. Wang, R.Z.; Gao, Q.Y.; Wang, D.J. Research of Least Square Curve Fitting and Simplified Algorithm. *Sens. World* **2021**, *27*, 8–10. [[CrossRef](#)]
18. Wu, Z.; Cai, P. An improved data preprocessing method in dynamic measurement. *Chin. J. Sens. Actuators* **2010**, *23*, 558–561.
19. Shang, X.L.; Tan, L.; Yu, K.P.; Zhang, J.; Kaur, K.; Hassan, M.M. Newton-interpolation-based zk-SNARK for Artificial Internet of Things. *Ad Hoc Netw.* **2021**, *123*, 102656. [[CrossRef](#)]
20. Yu, Y. Reinforcement Learning from Offline Data: Approaches and Advances. *China Basic Sci.* **2022**, *24*, 35.
21. Sun, H.N.; Liu, N.; Tan, L.; Du, P.; Zhang, B.C. Digital twin-based online resilience scheduling for microgrids: An approach combining imitative learning and deep reinforcement learning. *IET Renew. Power Gener.* **2024**, *18*, 1–13. [[CrossRef](#)]
22. Udekwe, D.; Ajayi, O.; Ubadike, O.; Ter, K.; Okafor, E. Comparing actor-critic deep reinforcement learning controllers for enhanced performance on a ball-and-plate system. *Expert Syst. Appl.* **2024**, *245*, 123055. [[CrossRef](#)]
23. Song, S.Y.; Hu, J.B.; Wang, Y.Y.; Han, X.L. Actor-Critic Learning Algorithm for Parameter Tuning of Sliding Mode Controller. *Electron. Opt. Control* **2020**, *27*, 24.
24. Sindhuja, S.; Lovenia, J.D.L.; Lavanya, A.P.; Jawahar, G. Qualitative Research of First Order Linear Difference Equations. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 1021–1025. [[CrossRef](#)]
25. Cui, G.M.; Zhu, J.T. Hot continuous rolling intelligent PID gauge control based on reinforcement learning. *Mod. Electron. Technol.* **2022**, *45*, 78–82. [[CrossRef](#)]
26. Zhang, J.; Jiang, X.; Shi, X.Y.; Cheng, J.; Zheng, Y.B. Offline reinforcement learning for eco-driving control at signalized intersections. *J. Southeast Univ.* **2022**, *52*, 762.
27. Mnih, V.; Kavukcuoglu, K.; Silver, D. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [[CrossRef](#)]
28. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
29. Schulman, J.; Levine, S.; Moritz, P.; Jordan, M.; Abbeel, P. Trust region policy optimization. In Proceedings of the 2015 International Conference on Machine Learning, Lille, France, 7–9 July 2015; ACM Press: New York, USA, USA, 2015; pp. 1889–1897.
30. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the International Conference on Machine Learning 2018, Stockholm, Sweden, 10–15 July 2018; p. 1856.

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.