processes

MDPI

Article

An Automatic Recognition Approach for Tapping States Based on Object Detection

Lingfeng Xue [1], Hongwei Guo [2], Helan Liang [1,*], Bingji Yan [2,*] and Hao Xu [2]

[1] School of Computer Science and Technology, Soochow University, Suzhou 215006, China; 20224227045@stu.suda.edu.cn
[2] Shagang School of Iron and Steel, Soochow University, Suzhou 215137, China; guohongwei@suda.edu.cn (H.G.); xu525167948@163.com (H.X.)
* Correspondence: hlliang@suda.edu.cn (H.L.); bjyan@suda.edu.cn (B.Y.)

Abstract: Monitoring tapping states, which reflects the smoothness of blast furnace (BF) production, is important in the blast furnace ironmaking process. Currently, these monitoring data are often recorded manually, which has limitations such as low reliability and high delays. In this study, we propose an automatic recognition approach for tapping states based on object detection, using furnace front monitoring videos combined with learning-based image processing technology. This approach addresses crucial aspects such as automatically recognizing the start and end times of iron tapping and slag discharging, accurately calculating their duration, and logging tapping sequences for multi-taphole operations. The experimental results demonstrate that this approach can meet the requirements of accurate and real-time recognition of tapping states and calculation of key monitoring data in industrial applications. The automatic recognition system developed based on this approach has been successfully applied in engineering projects, which provides real-time guidance for comprehensive monitoring, intelligent analysis, and operational optimization in blast furnace production.

Keywords: blast furnace; iron tapping; slag discharging; state recognition; object detection

check for updates

Academic Editor: Jun-Qiang Wang

Received: 19 November 2024
Revised: 21 December 2024
Accepted: 4 January 2025
Published: 7 January 2025

Citation: Xue, L.; Guo, H.; Liang, H.; Yan, B.; Xu, H. An Automatic Recognition Approach for Tapping States Based on Object Detection. *Processes* **2025**, *13*, 139. https://doi.org/10.3390/pr13010139

Copyright: © 2025 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

Blast furnace (BF) ironmaking is a crucial process in steel production. During BF operation, iron ore is reduced at high temperature to obtain hot metal and slag [1]. The process of discharging the hot metal and slag is called tapping. As shown in Figure 1, when tapping from the BF, the hot metal with larger density first flows out through the taphole. As the slag level in the furnace decreases, the slag that is composed of unreacted oxide components flows out. The hot metal and slag flow along the main trough, then they separate from each other in an iron–slag separator [2]: the slag flows in the slag ditch to water rushing slag equipment, while the hot metal flows in the hot metal ditch to the ladles.

Timely tapping provides space for the lower part of the BF, promoting stability within the furnace [3]. The duration of slag discharging and the time delay relative to iron tapping reflect the condition of the BF. Furthermore, as for different tapholes, whether the tapping time is consistent or not reflects the state of the hearth activity and affects the state of the taphole. Thus, monitoring the tapping states is critical in BF ironmaking [4]. In many facilities, these monitoring data are still manually recorded by on-site workers who observe the furnace front monitoring video or stand beside the taphole. However, manual recording has several limitations: (1) it is subjective, and incomplete, or incorrect data may occur due to workers' insufficient experience or irresponsibility; (2) such data

collection is labor-intensive and inefficient; (3) there is a long delay, which cannot meet real-time analysis requirements.

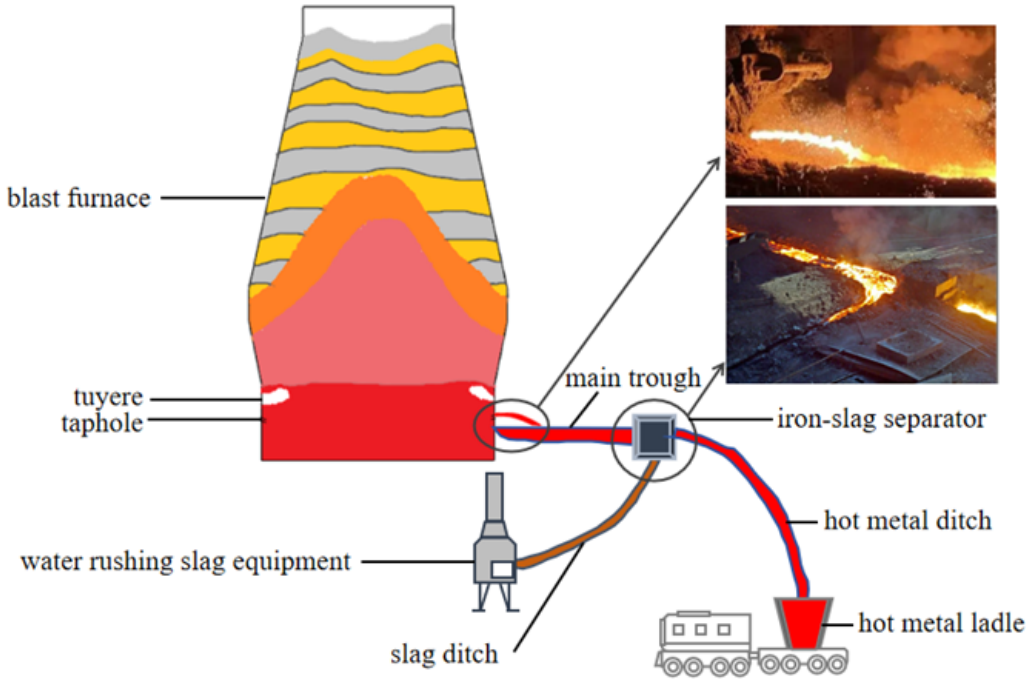


Figure 1. Schematic diagram of the tapping of the BF.

In some factories, relatively expensive equipment is used to achieve automatic identification of the iron-tapping status. This includes using a spectrometer to compare light frequencies, a high-precision infrared camera to monitor temperature changes for predicting the iron-tapping status, and a laser to measure the liquid level for status identification. While these methods can effectively enable automatic status identification, the cost of the equipment remains high.

Although some studies [5,6] have proposed image processing-based solutions in which the state is determined according to the brightness of the images, existing methods primarily rely on traditional image processing techniques such as Gaussian blur, threshold segmentation, and morphological operations. However, these methods face several limitations in practice: (1) They struggle to precisely extract iron-tapping features under variable operating conditions, such as day–night alternation, lighting variations, and smoke emissions, which have poor anti-interference ability. (2) These methods are highly sensitive to changes in camera position. Even slight vibrations or camera movements can significantly impact the accuracy of the solutions. (3) They do not optimize the image processing workflow, thus have difficulties meeting the stringent requirements of real-time tapping-state recognition and recording.

The most recent studies [7,8] have focused on recognizing the tapping states based primarily on iron liquid-level data. In [7], a threshold-based heuristic approach was introduced, wherein a specific threshold of the liquid level determined whether the state was "iron-tapping" or not. Ref. [8] proposed a time-series model based on bidirectional gated recurrent units (Bi-GRUs), which input continuous liquid-level data and predicted the tapping states. However, these methods still have some shortcomings: (1) They require extra cost for purchasing radar-level meters, and such equipment has to work in harsh operation environment. (2) They require a large quantity of liquid-level data. As mentioned in [7,8], millions of continuous liquid-level data have to be collected. (3) They can only recognize the state of iron tapping but not the slag-discharging state.

As image processing technology has advanced, machine vision has been applied more and more widely across diverse industrial sectors. ResNet (Residual Network) [9] is a deep learning model that introduced residual connections, allowing the effective training of deeper architectures. By implementing skip connections, ResNet addresses issues of gradient vanishing and difficulties in training deep networks, significantly improving image recognition accuracy. Following this, models like DenseNet [10] and EfficientNet [11] have further innovated on this foundation, optimizing network structures and computational efficiency, while Vision Transformers [12] have begun to be applied in image recognition, broadening the scope of deep learning applications. In industry, these technologies are widely used for product defect detection, production line monitoring, and smart logistics, significantly enhancing production efficiency and quality control, thus advancing the development of Industry 4.0.

In the field of BF tapping, existing research mainly focuses on the blocking time prediction of tapholes [13,14]; tapping velocity prediction [15,16]; and tapping temperature prediction [17,18]. However, these methods are inherently task-specific, relying heavily on infrared imaging devices and high-speed cameras to capture iron flow images, and subsequently employing deep learning-based classification or regression models for problem solving. Given our focus on recognizing tapping states, exploring the potential of utilizing ordinary monitoring images and developing simplified yet efficient methodologies holds significant practical importance for engineering applications.

To address the above deficiencies, this study proposes an automatic recognition method for tapping states based on object detection. By utilizing furnace front monitoring videos and learning-based image processing technology, an object detection model for recognizing the BF iron-tapping and slag-discharging states is constructed. Furthermore, to meet real-time recognition requirements, a multithreading-based online scheduling strategy is designed.

Our approach can automatically detect the start and end times of iron tapping and slag discharging, calculate their duration, and record them in real time without manual intervention. It can adapt to different operating conditions, ensuring accurate data collection and real-time database updates. The automatic recognition system developed based on this method has been successfully running at the BF No. 7 (1750 m^3 in volume) of Tranvic Steel Co., Ltd in China for a year, which provides real-time guidance for comprehensive monitoring, intelligent analysis, and operational optimization in BF production.

Compared with the existing works, the contributions of this article are as follows.

1. We revisit the problem of tapping-state recognition and transform it into an object detection task. This approach not only enhances the recognition accuracy but also improves its interpretability. It is the first effective attempt in utilizing deep image processing technology for tapping-state recognition, which provides a new idea for future research.
2. We introduce a YOLOX-based object detection model tailored for tapping-state recognition. Additionally, we propose an online scheduling strategy that enables real-time recognition and accurate calculation of key monitoring data. Experimental results demonstrate that our model can achieve a recognition accuracy of 99.8%.
3. We have developed a comprehensive tapping-state recognition system that has been successfully applied in engineering projects. The intuitive data dashboard seamlessly integrates real-time tapping states and key monitoring data from multiple tapholes, offering a holistic view of the tapping operation and real-time insights for operational optimization in BF production.

2. Our Method

Currently, BF monitoring systems are widely applied for monitoring various BF operation statuses. Among them, the BF front monitoring system uses a camera installed near the tapholes to capture real-time operation of the BF front. Consider that the hot metal and slag flow through the main trough and slag ditch, whose image features during the tapping cycle are obvious. Thus, in this paper, we use object detection techniques to identify the states of the main trough and slag ditch, and establish a mapping between image features and the states of iron tapping and slag discharging. The framework of our approach is shown in Figure 2.

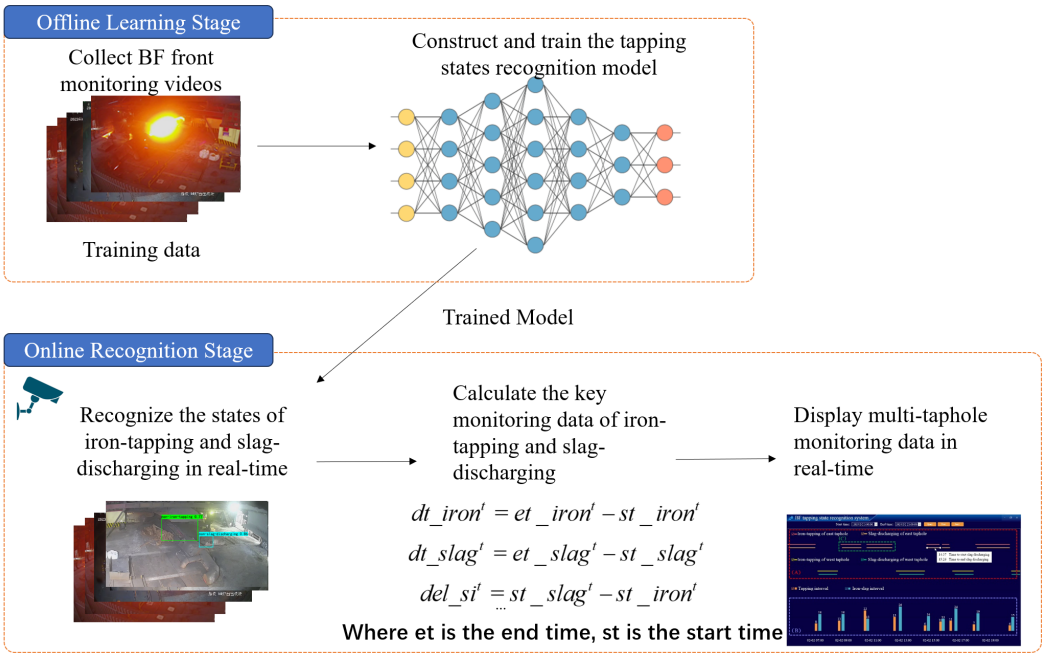


Figure 2. Framework of our approach.

In the rest of this section, we first introduce how to construct the tapping-state recognition model in Section 2.1, and then introduce how to implement real-time tapping-state recognition in Section 2.2.

2.1. Offline Learning Stage

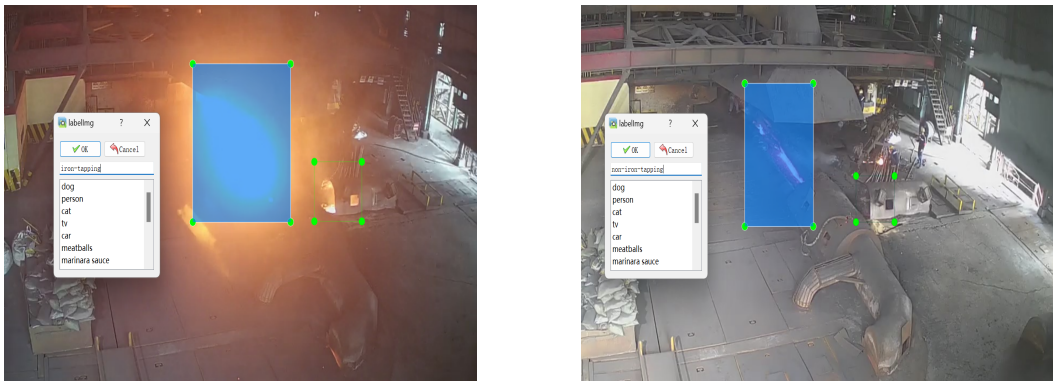
In the offline learning stage, an object detection model was constructed, which learned to automatically recognize the tapping states at the BF front. Solving this task involved annotation, model construction, and training, whose details are as follows.

2.1.1. Annotation of BF Front Monitoring Images

According to the BF ironmaking process, the states of the main trough and slag ditch can be sequentially classified as "non-iron-tapping", "non-slag-discharging", "iron-tapping", "slag-discharging", "non-iron-tapping", "non-slag-discharging", and so on. Thus, we defined four classes, namely, "non-iron-tapping", "non-slag-discharging", "iron tapping", and "slag-discharging", corresponding to the four states in a complete tapping cycle.

Take the images captured from the BF front monitoring system as an example. As shown in Figure 3a, it is evident that there is molten iron splashing in the main trough and bright light in the slag ditch. Therefore, we mark the tapping area and annotate it with the "iron-tapping" label, as indicated by the blue box in Figure 3a. Similarly, the green box on the right marks the slag-discharging target and is annotated with the "slag-discharging"

label. As a comparison, an image with "non-iron-tapping" and "non-slag-discharging" labels is shown in Figure 3b.



(**a**) (**b**)

Figure 3. Example of annotation of BF front monitoring images. (**a**) Iron-tapping and slag-discharging states; (**b**) non-iron-tapping and non-slag-discharging states.

2.1.2. Structure of the Tapping-State Recognition Model

Our purpose was to identify objects such as "iron-tapping" and "slag-discharging" in BF front monitoring images based on object detection. The key challenge was how to use a unified model for high accuracy and efficiency in both iron-tapping and slag-discharging tasks. Notably, significant differences exist between the iron tapping and slag discharging, such as the features and target sizes. To balance both accuracy and efficiency, we selected the YOLO model, which is a high-speed multi-scale detector. Furthermore, the attention mechanism has already been proven effective in many works. Thus, we used distinct attention modules to further improve recognition accuracy by self-learning critical aspects of different tasks. According to the experimental results (see Section 3 for more detail), we found that the YOLOX model [19] performed the best for the problem considered. Thus, we used the YOLOX model as the baseline and added the attention mechanism [20] to improve the detection precision, whose network structure is shown in Figure 4.

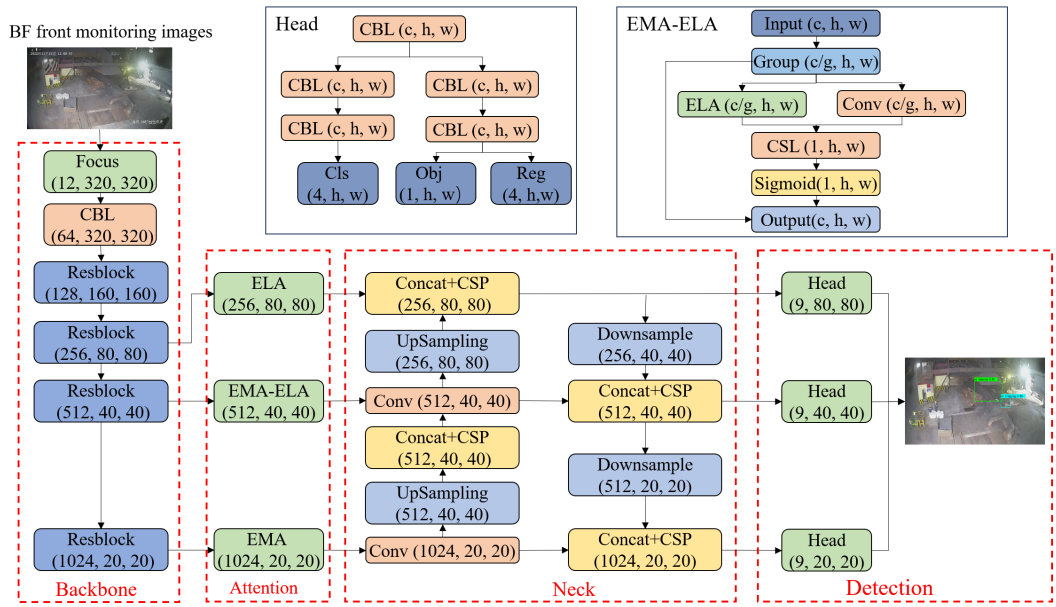


Figure 4. Network structure of the proposed model for tapping-state recognition.

As shown in Figure 4, our tapping-state recognition model consisted of backbone, attention, neck, and detection networks, where each block represented an operation; the

values in parentheses represent the size of the feature map after the operation. For example, the size of the image after the Focus operation was 320 × 320 × 12, where 320 × 320 is the height and width of the image and 12 is the number of channels. To address multi-scale features, Head and SENet modules whose input feature maps varied were invoked several times. Thus, no numerical value but (h, w, c) was used to represent the size, whose value was decided by the feature map input to the module.

The goal of the backbone network was to extract features of the BF front monitoring images. For the convenience of batch processing, the backbone network first used the Focus operation to rescale the images. Next, the CBL operation ran to extract the features of the images. CBL stands for Convolution, Batch Normalization, and Leaky ReLU, which are widely used in computer vision [21]. Furthermore, multiple Residual blocks (Resblock) were incorporated to enhance feature extraction.

Multi-scale feature maps are output by the backbone network. To offer more precise features for better feature fusion in the neck network, we attempted to add an attention network after the backbone. We tested several attention mechanisms and found that efficient local attention (ELA) [22] performed best in the slag-discharging task while efficient multi-scale attention (EMA) [23] performed best in the iron-tapping task, whose performance and discussions are shown Section 3.2.2. Considering that the slag-discharging region is small, attention may mainly act on shallow features, while the feature information of iron tapping is much richer, where more attention should be paid to the deep semantic features. Therefore, we applied the ELA module for the high-scale features and the EMA module for the low-scale features. For the middle-scale features, we used a model combining EMA and ELA for feature augmentation.

Specifically, we used the basic EMA and ELA modules whose structures were taken from the original paper. For the fusion module of EMA-ELA, we replaced the initial feature extraction part of EMA with ELA. After the input was grouped, the grouped data first passed through the ELA module and a convolution layer separately. The resulting outputs were then processed through the cross-spatial learning (CSL) module proposed by EMA, followed by a sigmoid activation. Finally, they were multiplied with the original grouped data, so as to output the feature maps while considering attention.

Such augmented feature maps were input into the neck network for feature fusion, enhancing its ability to understand the images. The neck network mainly utilized a fusion mechanism called Cross Stage Partial (CSP) [24], which combined CBL, Residual blocks, Convolution, Batch Normalization, and Leaky ReLU. To fuse multi-scale feature maps, upsampling and dowsampling operations were used to adjust its spatial dimensions. Finally, the fusion features entered the detection network, so as to make predictions. Considering the varying sizes of targets in the image, several prediction branches existed. Thus, it could accurately identify the categories of objects (Cls) in the images and determine their precise bounding boxes (Obj and Reg) via multi-scale predictions.

2.1.3. Model Training and Inference

Our tapping-state recognition model held the BF front monitoring images as input. After forward propagation, the detection results were output such as finding the "iron-tapping" target with precise iron-tapping position. Subsequently, predicted targets were compared with the annotated ground truth, and backpropagation was performed to update the network parameters for better detection.

After training the model, we used it to recognize the tapping states. Take the image in Figure 5 as an example. The model could successfully detect two targets: "non-iron-tapping" and "non-slag-discharging", whose confidence scores were 0.77 and 0.86, respectively. Since there was no obvious tapping features in the main trough and the slag ditch, the inference

result was consistent with the observation. Furthermore, the inference time was 0.11 s per image, which can satisfy the real-time recognition requirement.

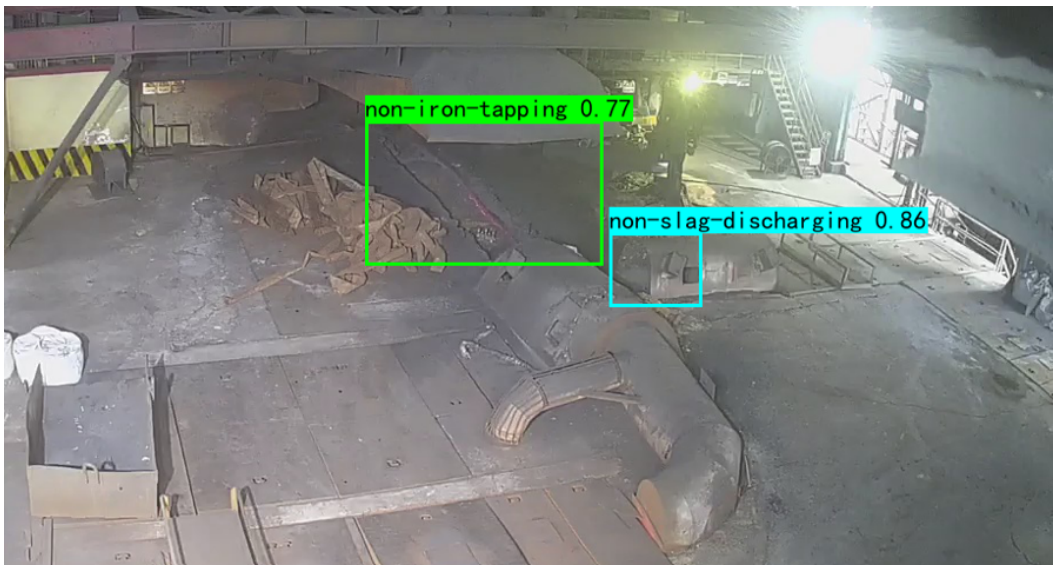


Figure 5. Example of the object detection results.

2.2. Online Recognition Stage

2.2.1. Calculation of Key Monitoring Data

Using our tapping-state recognition model, we can detect the states of iron tapping and slag discharging at any time. Note that tapping of the BF is a discrete process, with each tapping lasting 60 min to 240 min [25]. Therefore, we can determine the start time and end time of iron tapping based on changes in its state. Ideally, if the inference result is "non-iron-tapping" in the previous frame, while it becomes "iron-tapping" in the successive frame, it represents tapping is taking place. Thus, we mark that exact time as the start of iron-tapping.

However, such a strategy is not practical since the accuracy of the model cannot reach 100%. On the one hand, smoke and dust may be generated at the ironmaking site due to the high temperatures and pollutants, which will interfere with recognition. On the other hand, the end of tapping is not an instantaneous state, but is likely to last for a while, during which the flame irregularly decreases until it disappears. Since the feature is vague, the results of the model may hover between tapping and non-tapping, which obviously does not match the actual production situation.

To avoid misjudgment and guarantee that only one start time and one end time are recorded for each tapping cycle, a dynamic window strategy was designed. Suppose that the current state is "non-iron-tapping" and the threshold is eight of ten. Once eight out of ten consecutive results are "iron-tapping" which are opposite to the current state, we consider the state has changed, and record the time of the first changed result out of ten as the start time of iron tapping. Accordingly, the current state is changed to "iron-tapping". As time goes by, once eight out of ten consecutive detection results are "non-iron-tapping", the state changes again, and the time is recorded as the end time of iron tapping. We can obtain the start time and end time of slag discharging in a similar way.

After determining the start time and the end time of iron-tapping, the duration of iron tapping can be calculated by Equation (1).

$$dt^t_{iron} = et^t_{iron} - st^t_{iron} \tag{1}$$

where dt^t_{iron} is the iron-tapping duration of the tth tapping cycle. st^t_{iron}, et^t_{iron} are the start time and the end time of the iron tapping of the tth tapping cycle, respectively.

Similarly, we can calculate the duration of slag discharging according to the start time and the end time of slag discharging, as shown in Equation (2).

$$dt^t_{slag} = et^t_{slag} - st^t_{slag} \tag{2}$$

where dt^t_{slag} is the slag-discharging duration of the tth tapping cycle. st^t_{slag}, et^t_{slag} are the start time and the end time of the slag discharging of the tth tapping cycle, respectively.

Additionally, we can calculate the interval between iron tapping and slag discharging by Equation (3).

$$del^t_{iron-slag} = st^t_{slag} - st^t_{iron} \tag{3}$$

where $del^t_{iron-slag}$ is the iron-to-slag interval of the tth tapping cycle.

As for continuous tapping cycles, we can calculate the intervals between consecutive iron tapping and slag discharging by Equations (4) and (5).

$$del^t_{iron} = st^t_{iron} - et^{t-1}_{iron} \tag{4}$$

$$del^t_{slag} = st^t_{slag} - et^{t-1}_{slag} \tag{5}$$

where del^t_{iron} is the iron-tapping interval between the tth and $(t-1)$th tapping cycles. del^t_{slag} is the slag-discharging interval.

Note that an extended period of iron tapping and slag discharge signifies an effective removal of hot metal and slag from the furnace. However, this prolonged process can also lead to a drop in furnace temperature and accelerate taphole erosion. Conversely, a shorter duration of these operations indicates the accumulation of hot metal and slag within the furnace, potentially elevating the furnace temperature. By grasping these crucial monitoring data, we can enhance the precision of assessing the furnace's thermal state, a vital aspect for optimizing the BF operation.

2.2.2. Real-Time Scheduling with Multiple Threads

When applying tapping-state recognition, two tasks needed to be performed simultaneously: retrieving the monitoring videos from the camera and analyzing the images by the tapping-state recognition model. However, since the image retrieval time was nearly 0.02 s while the inference efficiency was nearly 0.1 s, the processing rate of the model lagged behind the camera frame rate. As the number of images to be processed increased, issues such as detection delay and program freezing could occur.

To address this issue, we proposed an online scheduling strategy based on multiple threads. One thread was responsible for retrieving images and constructing the candidate image sequence to be detected. Another thread was used for image processing based on object detection. After the model processed a frame, it extracted the latest image from the candidate sequence, rather than processing the next frame of the image it just processed. Additionally, at regular intervals, the candidate sequence was cleared to avoid unlimited growth. This scheduling strategy could not only improve system processing efficiency but also prevent issues such as data inconsistency, detection delay, and program freezing caused by asynchrony between image retrieval and detection tasks.

Finally, once the state change was detected, the corresponding monitoring data, such as the end time of iron tapping and its duration, were automatically stored in a database without manual intervention. Thus, operators could grasp real-time states about iron tapping and slag discharging. It also helped other programs to access them for further analysis and BF control.

3. Method Validation

3.1. Data Preparation and Evaluation Metrics

In our experiment, we selected several iron-casting processes with various working conditions and captured 600 frames at average intervals so that these images contained features of various tapping phases and working conditions. The annotated VOC (Visual Object Classes) image dataset was then split into training and validation sets in a ratio of 9:1. We used the PyTorch 1.13.0 framework to build the tapping-state recognition model, and the typical Mosaic and MixUp packages were used for data enhancement [26]. The training process consisted of 300 epochs with a batch size of eight. The confidence threshold was 0.5, the learning rate was 10^{-4}, and the optimizer was SGD. We trained the model on a computer with an i7-12700 processor, 64 GB memory, and an RTX 3060 GPU.

We collected another ten iron tapping videos for validation, which were not from the same day as the training one. This helped avoid inflated test results due to similar environmental conditions on the same day. For these ten videos, we employed three experiments. First, we randomly selected nearly two thousand images for a comprehensive accuracy test. Second, we clipped ten three-minute videos before and after the state change, which contains nearly a thousand images, assessing the model's recognition accuracy at error-prone detection points. Finally, we compared with the actual changing times of these ten iron tapping processes, so as to estimate the time errors of different methods.

Two criteria including Accuracy (Acc) and F1-Score ($F1$) were used to evaluate the performance of the proposed tapping-state recognition model, as shown in Equations (6) and (7). Furthermore, the root-mean-square error ($RMSE$) in Equation (8) was used for evaluating the deviation between our method and experienced on-site workers in determining the time of state change (e.g., the start time of iron-tapping).

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{8}$$

where TP represents true positives, FN represents false negatives, FP represents false positives, and TN represents true negatives. P represents precision, $P = \frac{TP}{TP+FP}$. R represents recall, $R = \frac{TP}{TP+FN}$. n is the number of samples. y_i is the ground-truth time of state change, and $\hat{y}_i$ is the time output by our method.

3.2. Experimental Results of Tapping States Recognition

3.2.1. Comparing to Existing Methods

We first validated the accuracy of our proposed model in recognizing tapping states. Given that SSD [27] and Faster R-CNN [28] are widely renowned in object detection, we used them as competing models. They underwent training and validation using the same dataset as our model. Additionally, we report on the recognition outcomes achieved using the traditional image processing method (TIP) [6]. Notably, previous studies [7,8] have proposed state recognition methods for iron tapping based on liquid-level data. We include their reported results in the first and second rows of Table 1, citing directly from the original papers.

As shown in Table 1, image-based methods demonstrated exceptional performance, achieving an effectiveness of over 99%, surpassing the results obtained by methods reliant on liquid-level data. Furthermore, liquid-level-based methods had a limitation in that they

could only recognize the iron-tapping state but not the slag-discharging state. Conversely, image-based methods excelled in simultaneously recognizing both the iron-tapping and slag-discharging states, making them a more logical and comprehensive solution for the problem considered.

Table 1. Performance of different methods.

Model	Iron Tapping		Slag Discharging	
	Accuracy	F1-Score	Accuracy	F1-Score
DWAA * [7]	0.966	0.972	-	-
BI-GRU * [8]	0.981	0.981	-	-
TIP	0.996	0.995	0.991	0.988
SSD	0.998	0.997	0.995	0.993
Faster R-CNN	0.998	0.997	0.997	0.995
Ours	0.998	0.998	0.998	0.997

*: Results are drawn from the original papers. Since no slag discharging was considered, it is represented by "-".

3.2.2. Model Validation During the State Changing

Due to the obvious image features during the stable tapping process, it is easily recognizable. We found that the biggest difficulty lay in the period at the beginning and end of tapping during which the iron-tapping and slag-discharging states were changed. Thus, we cropped the above ten consecutive tapping cycles, each of which kept only 3 min recording the state changing of iron tapping and slag discharging for the test, so as to distinguish the performance of various image-based methods. The experimental results are shown in Table 2.

Table 2. Model validation during the state changing.

Model	Iron Tapping		Slag Discharging		Inference Time (s)
	Accuracy	F1-Score	Accuracy	F1-Score	
TIP	0.829	0.854	0.749	0.668	0.017
SSD	0.916	0.923	0.853	0.864	0.078
Faster R-CNN	0.913	0.920	0.891	0.890	0.11
Ours	0.950	0.950	0.964	0.964	0.077

The first row of Table 2 shows that the accuracy of the TIP method was much worse than the learning-based counterparts. Since the brightness changes between different states are vague and easily affected by environment, it was difficult for the TIP method to apply to various working conditions. To illustrate this more clearly, take Figure 6 as an example. There happened to be a cloud of smoke in the middle, which caused the calculated pixel value to be significantly lower, resulting in misclassification by the TIP method. In contrast, thanks to adaptive localization and deep feature extraction, all the learning-based methods could correctly recognize the tapping states.

Table 2 shows that the accuracy of iron tapping was higher than that of slag discharging in most methods, due to the more obvious image features of iron tapping. Moreover, we found that the SSD and Faster R-CNN models had a higher recognition error to "non-iron-tapping" samples, resulting in a more likely longer delay in the end time of iron tapping. Thus, their performance was worse than that of our method.

Furthermore, to verify the effectiveness of our proposed attention network, we compared the results of combining different YOLOX models with different attention mechanisms. Specifically, we used new attention mechanisms in recent years, including squeeze-

and-excitation network (SENet) [29], convolutional block at attention module (CBAM) [30], efficient channel attention (ECA) [31], ELA [22], and EMA [23].



Figure 6. Iron tapping with smoke interference.

Table 3 shows that incorporating an attention network could significantly improve the model's accuracy. From a single-strategy perspective, ELA performed best for slag discharging, while EMA was more effective for iron tapping. The reason may be that the ELA employs band pooling in the spatial dimension to extract feature vectors in both horizontal and vertical directions, maintaining an elongated kernel shape to capture long-range dependencies while avoiding interference from irrelevant regions in label prediction. Such mechanism allows the model to retain fine-grained positional information, which is advantageous for the precise localization required in the slag-discharging task. In contrast, in the iron-tapping task, the region of interest is larger and contains richer feature information, so the focus is on how to precisely augment iron-tapping features across different channels. Thanks to its parallel subnetworks and cross-spatial learning structure, ELA strengthens the feature fusion of different channels, resulting in better performance on the iron-tapping task.

Table 3. Results of using different attention modules.

Model	Iron-Tapping		Slag-Discharging		Inference Time (s)
	Accuracy	F1-Score	Accuracy	F1-Score	
YOLOX	0.926	0.926	0.930	0.947	0.075
YOLOX-SE	0.942	0.942	0.936	0.937	0.077
YOLOX-CBAM	0.944	0.944	0.943	0.946	0.10
YOLOX-ECA	0.931	0.932	0.945	0.946	0.077
YOLOX-EMA	0.949	0.950	0.938	0.940	0.077
YOLOX-ELA	0.944	0.945	0.948	0.951	0.077
Ours	0.950	0.950	0.964	0.964	0.077

Considering that obtaining more detailed image information should focus more on processing shallow features, while cross-channel fusion is more suitable for handling complex and abstract deep feature representations to achieve comprehensive feature enhancement, our method effectively utilized different attention mechanisms at various feature extraction layers. This fully leveraged the advantages of the aforementioned strategies in enhancing feature attention in both spatial and channel dimensions. The experimental results indicated that this design was effective.

3.2.3. Model Validation Under Changing Camera Position

In the real-world application, the camera position can experience small shifts over time, and after maintenance, the camera may not be reinstalled in the exact same location. To simulate these minor camera position changes, we performed 10% translations, 10% rotations, and 10% reduction scaling operations on the test images which kept only 3 min of recording the state changing of iron tapping and slag discharging. The results are summarized in Table 4.

Table 4. Results of various image transformations.

Method	Model	Iron-Tapping		Slag-Discharging	
		Accuracy	F1-Score	Accuracy	F1-Score
Ours	Original	0.950	0.950	0.964	0.964
	Translation only	0.938	0.938	0.937	0.938
	Translation + rotation	0.926	0.925	0.935	0.937
	Translation + rotation + scaling	0.914	0.919	0.922	0.918
TIP	Original	0.829	0.854	0.753	0.725
	Translation only	0.729	0.787	0.607	0.376
	Translation + rotation	0.642	0.737	0.547	0.484
	Translation + rotation + scaling	0.646	0.738	0.490	0.001

Table 4 shows that in the TIP method, recognition accuracy significantly declined as the images underwent changes. The reason was that as the camera position changed, the target could move out of the preset fixed area. When the pixel values that satisfied the prominent tapping features were less than the threshold, a judgment error occurred. Such problem was more severe for the slag-discharging task, as the slag-discharge area was small, and the features were not obvious. In contrast, our method could adaptively adjust the position of the bounding box according to the extracted image features, so that the tapping object could be accurately located even when its position changed. Thus, it demonstrated robust performance, maintaining relatively high recognition accuracy.

Note that our method has been running continuously for one year in engineering projects without diagnostic errors, which demonstrates our method can deal with various complex working conditions such as the camera position changing.

3.3. Experimental Results of Deviation of Monitoring Data

To further verify the correctness of the monitoring data, we compared the iron-tapping and slag-discharging monitoring data output by our approach and that manually recorded by on-site workers for the above 10 consecutive tapping cycles. We used the *RMSE* as the evaluation criterion, which can determine whether a method can make correct judgments. The experimental results are shown in Table 5.

Table 5. *RMSE* of different learning-based methods (s).

Model	Start Time		End Time		Average
	Iron Tapping	Slag Discharging	Iron Tapping	Slag Discharging	
SSD	10.5	180.2	19.1	38.9	62.2
Faster R-CNN	37.8	19.7	28.2	48.5	33.6
YOLOX	6.1	17.5	23.3	35.1	20.5
Ours	6.1	12.4	26.2	26.4	17.8

Table 5 shows that the overall performance of our method was the best, as it had the least average error of the monitoring data. The RMSE of the start time of iron tapping was the smallest for each model, while that of the end time of slag discharging was the largest.

The start time error was smaller than its end time, which means it was easier to recognize the start than the end of tapping. Moreover, the error of slag discharging was larger than that of iron tapping, which means the recognition of slag discharging was much harder. This is in line with the actual situation on site. Once tapping takes place, it changes from no brightness to strong brightness, whose features are easy to grasp. In contrast, finishing tapping is a process in which the brightness of the hot metal and slag gradually decreases. Therefore, it is difficult to determine the specific time when it stopped, resulting in a greater time error. Furthermore, the brightness and intensity of iron tapping are much greater than those of slag discharging, so its state recognition is relatively easy, and the corresponding error is smaller.

4. Engineering Application

An automatic recognition system for BF tapping states based on the proposed approach has been applied to the BF No. 7 (1750 m^3 in volume) of Tranvic Steel Co., Ltd. in China. Our system runs on a computer with an i7-12700 processor, 64 GB of memory, and an RTX 3060 GPU. Figure 7 shows one of its user interfaces, which has been translated.

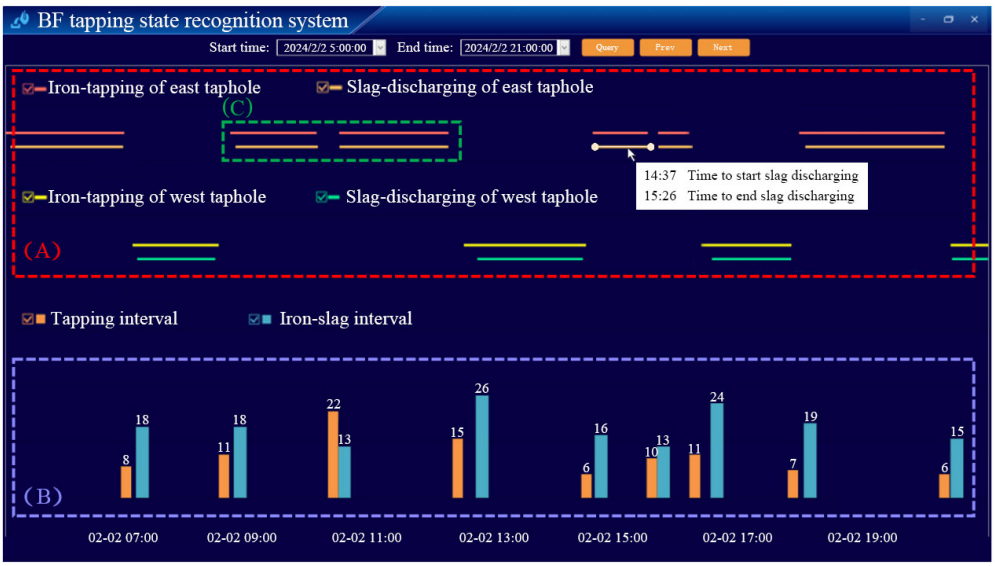


Figure 7. Screenshot of the system interface.

The BF is equipped with two distinct tapholes: the eastern and western ones. The upper section of Figure 7, labeled as zone "A", depicts the status of each tapping cycle. Distinct horizontal lines, each in a unique color, represent the iron-tapping and slag-discharging activities in the respective tapholes. This allows operators to swiftly identify the start and end times, duration, and potential alternatives for multi-taphole operations. Specifically, when the mouse cursor is hovered over a line, details of the specific tapping cycle emerge. For instance, Figure 7 shows that the slag discharge commences at 14:37 and terminates at 15:26.

Meanwhile, the lower section of Figure 7, designated as zone "B", presents a bar chart that quantifies the intervals between consecutive tapping cycles. For instance, the first two bars indicate an 8 min interval between two tapping cycles, and the interval of iron and slag is 18 min. By comparing the bar heights with the accompanying numerical values, operators gain a precise understanding of the transitions between the iron-tapping and slag-discharging states, along with the duration of consecutive tapping cycles.

Furthermore, the system provides functions such as intelligent statistics and alarm, facilitating the optimization of BF tapping management. For example, an effective tapping

process involves alternating between multiple tapholes to maintain production continuity and heat balance, whereas, as shown in zone 'C', continuous tapping from a single taphole took place, which invoked an alarm to operators for potential anomalies. Furthermore, a high slag ratio and delayed slag discharging may indicate issues like increased pipeline pressure. These statistical data contribute to a comprehensive assessment of BF operating conditions.

5. Conclusions

An automatic recognition approach for tapping states was proposed in this paper. This approach utilized an improved YOLOX model with attention to determine the start and end times of iron tapping and slag discharging. An image processing workflow using multiple threads was designed, ensuring accurate and real-time data processing. Experimental results demonstrated that our model could achieve a recognition accuracy of 99.8%. The automatic recognition system developed based on this method has been successfully running at the BF No. 7 (1750 m^3 in volume) of Tranvic Steel Co., Ltd. in China for a year, which provides real-time guidance for comprehensive monitoring, intelligent analysis, and operational optimization in BF production.

In our future work, we aim to encompass a broader range of iron-tapping environments across multiple steel plants. To achieve this goal, we intend to intensify our research efforts in transfer learning. This approach has the potential to significantly reduce the need for new training data while enhancing the model's generalization capabilities. Additionally, we will delve deeper into the analysis of molten iron flow images to guide the BF production optimization. At the same time, we aspire to integrate it with automatic recognition programs for tasks such as mud volume detection and taphole depth detection, enabling a comprehensive evaluation of the ironmaking status from multiple perspectives.

Author Contributions: L.X., conceptualization, software, validation, writing—original draft preparation; H.G., methodology, formal analysis, project administration; H.L., conceptualization, writing—review and editing; B.Y., visualization; H.X., visualization, data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China under Grant No. 52474364, Grant No. 52074185, and Grant No. 61902269; Science and Technology Major Project of WuHan (2023020302020572); Suzhou Science and Technology Plan Project (No. SYG202127); the Priority Academic Program Development of Jiangsu Higher Education Institutions, China.

Data Availability Statement: Because our data are related to the factory's production privacy, we cannot disclose the data.

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

References

1. Smith, M.P. Blast furnace ironmaking—A view on future developments. *Procedia Eng.* **2017**, *174*, 19–28. [CrossRef]
2. Lu, Z.; Gu, H.; Chen, L.; Liu, D.; Yang, Y.; McLean, A. A review of blast furnace iron-making at Baosteel facilities. *Ironmak. Steelmak.* **2019**, *46*, 618–624. [CrossRef]
3. Frischer, R.; Grycz, O.; Hlavica, R. Concept industry 4.0 in metallurgical engineering. In Proceedings of the 26th International Conference on Metallurgy and Materials (METAL), Brno, Czech Republic, 24–26 May 2017.
4. Zhou, P.; Song, H.; Wang, H.; Chai, T. Data-driven nonlinear subspace modeling for prediction and control of molten iron quality indices in blast furnace ironmaking. *IEEE Trans. Control Syst. Technol.* **2016**, *25*, 1761–1774. [CrossRef]
5. Li, H. Online Detection Method for Iron Status in Blast Furnace Slag Discharge. Patent CN110184401A, 25 December 2020.
6. Li, X. A Monitoring Method and System for the Status of Blast Furnace Tapping Hole. Patent CN 113122669 B, 16 July 2021.
7. Zong, Y.; Wang, Z.; Liu, X.; Nian, Y.; Pan, J.; Zhang, C.; Wang, Y.; Chu, J.; Zhang, L. Judgment of blast furnace iron-tapping status based on data differential processing and dynamic window analysis algorithm. *Prog. Nat. Sci. Mater. Int.* **2023**, *33*, 450–457. [CrossRef]

8. Zong, Y.; Hu, S.; Qin, D.; Wang, Z.; Zhang, C.; Chu, J.; Zhang, L. Iron-Tapping State Recognition of Blast Furnace Based on Bi-GRU Composite Model and Post-processing Classifier. *IEEE Sens. J.* **2023**, *23*, 22006–22018. [CrossRef]
9. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
10. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
11. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
12. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
13. Jiang, Z.; Dong, J.; Pan, D.; Wang, T.; Gui, W. A new monitoring method for the blocking time of the taphole of blast furnace using molten iron flow images. *Measurement* **2022**, *204*, 112155. [CrossRef]
14. Jiang, Z.; Dong, J.; Pan, D.; Wang, T.; Gui, W. A novel intelligent monitoring method for the closing time of the taphole of blast furnace based on two-stage classification. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105849. [CrossRef]
15. He, L.; Jiang, Z.; Xie, Y.; Chen, Z.; Gui, W. Velocity measurement of blast furnace molten iron based on mixed morphological features of boundary pixel sets. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
16. He, L.; Jiang, Z.; Xie, Y.; Gui, W.; Chen, Z. Mass flow measurement of molten iron from blast furnace, based on trusted region stacking using single high-speed camera. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]
17. Pan, D.; Jiang, Z.; Chen, Z.; Jiang, K.; Gui, W. Compensation method for molten iron temperature measurement based on heterogeneous features of infrared thermal images. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7056–7066. [CrossRef]
18. Pan, D.; Jiang, Z.; Xu, C.; Gui, W. Polymorphic temperature measurement method of molten iron after skimmer in ironmaking process. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [CrossRef]
19. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
20. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **2022**, *8*, 331–368. [CrossRef]
21. Ding, Y.; Zhu, Y.; Wu, Y.; Jun, F.; Cheng, Z. Spatio-temporal attention LSTM model for flood forecasting. In Proceedings of the 2019 International Conference on Internet of Things (IThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 458–465.
22. Xu, W.; Wan, Y. ELA: Efficient Local Attention for Deep Convolutional Neural Networks. *arXiv* **2024**, arXiv:2403.01123.
23. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
24. Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-attention with relative position representations. *arXiv* **2018**, arXiv:1803.02155.
25. Geerdes, M.; Chaigneau, R.; Lingiardi, O. *Modern Blast Furnace Ironmaking: An Introduction*; IOS Press: Amsterdam, The Netherlands, 2020.
26. Flusser, J.; Farokhi, S.; Höschl, C.; Suk, T.; Zitova, B.; Pedone, M. Recognition of images degraded by Gaussian blur. *IEEE Trans. Image Process.* **2015**, *25*, 790–806. [CrossRef] [PubMed]
27. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef] [PubMed]
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
30. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
31. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.

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.