Article

Research on Blast Furnace Tuyere Image Anomaly Detection, Based on the Local Channel Attention Residual Mechanism

Chuanwang Song *, Ziyu Li, Yuming Li, Hao Zhang, Mingjian Jiang, Keyong Hu and Rihong Wang

School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China
* Correspondence: songchw_2006@163.com

Abstract: The channel attention mechanism is widely used in deep learning. However, the existing channel attention mechanism directly performs the global average pooling and then full connection for all channels, which causes the local information to be ignored and the feature information cannot be reasonably assigned with the proper weights. This paper proposed a local channel attention module, based on the channel attention. This module focuses on the local information of the feature image, obtains the weight of each regional channel through convolution, and then integrates the information, so that the regional information can be fully utilized. Moreover, the local channel attention module is combined with the residual module, and the local channel attention residual network LSERNet is constructed to detect the abnormal state of the blast furnace tuyere image. With sufficient experiments on the collected datasets of the blast furnace tuyere, the results show that the proposed method can efficiently extract the feature information, and the recognition accuracy of the LSERNet model reached 98.59%. Further, our model achieved the highest accuracy, compared with SE-ResNet50, ResNet50, LSE-ResNeXt, SE-ResNeXt, and ResNeXt models.

Keywords: convolutional network; blast furnace tuyere; residual network; attention mechanism; tuyere image; image recognition

1. Introduction

In the field of computer vision, deep convolutional neural networks (CNNs) have been deeply studied and widely used, such as image classification [1], object detection [2], and semantic segmentation [3], which also have achieved satisfactory results in the corresponding fields. For some difficult tasks, however, the accuracy and stability of the model are still unacceptable. Many researchers have been devoted to strengthening the feature learning ability of convolutional neural networks, to improve the expression and generalization ability of neural networks [4]. Starting from the pioneering AlexNet [5], scholars in this field have continued to study to further improve the performance of deep CNNs [6–9], such as continuously adjusting the depth and width of convolutional neural networks [10–12], adding skip connections [13], and dense connections [14] to the convolutional networks. These methods can effectively improve the stability of the network and the utilization of the feature information.

Among them, SENet (squeeze-and-excitation network) [15] plays an important role in improving the performance of deep convolutional neural networks. It learns the channel attention for each convolutional block, which greatly improves the performance of various deep CNN architectures. In SENet, the feature information is compressed into a channel descriptor by the global average pooling, which we call the squeeze operation. Then, the correlation between the channels is obtained through two fully connected layers. Finally, the sigmoid function is used to activate the correlation between channels to obtain the weight of each channel, which we call the excitation operation. Once SENet was proposed, many scholars are paying more attention to the exploration of the relationship between the channels, and are improving the SE block by extracting more complex channel dependencies, or combining channels with spatial relationships [16–19]. Park et al. [20] proposed

a hybrid module BAM (bottleneck attention module) that combines channel and spatial attention. The module is divided into two convolution parts and connects the two kinds of attention in parallel, the channel part is the SENet network structure, and the spatial part is the four-layer convolution, similar to the bottleneck structure. Finally, the two parts were integrated and the feature information weights were generated. Although this module improves the performance of the model, the complexity of the model is greatly increased, due to the four-layer convolution of the spatial attention part. In view of this defect, Woo et al. [21] improved the BAM module and proposed an improved hybrid attention module CBAM (convolutional block attention module), which changed the original parallel mode of the channel attention and spatial attention to the serial mode. Moreover, a layer of global max pooling was added to SENet, two fully connected layers were removed and two convolutional layers were added. Firstly, global pooling was used to reduce the channel dimension, and then a layer of convolution was applied to extract the spatial information, which greatly reduced the computational burden of the model. However, CBAM [21] adds two pooling operations, which improve the performance of the visual attention mechanism and increase the complexity of the model. Although these methods have achieved a higher accuracy, they usually bring a higher model complexity and bear a heavy computational burden.

Li et al. [22] proposed a grouping spatial attention module, named SGE (spatial group-wise enhance). The module divides the channel, so that each part of the division contains the feature information at different locations. Then it performs the spatial attention convolution on each part. The module bears a heavy computational burden. Zhang et al. [23] proposed a grouped hybrid attention module SA-Net (shuffle attention). The module first groups the channels of the feature map in a certain proportion, and then bisects the grouped sub-features, according to the channels, and performs channel attention and spatial attention, respectively. Finally, the output is integrated to obtain the final feature map of interest. The module increases the complexity of the model. In addition, Wang et al. [24] proposed ECA-Net (efficient channel attention networks), which integrate the channel information by using a properly sized one-dimensional convolution. Although the computational cost is reduced, the information interaction in each dimension is ignored.

To solve these problems, we propose a local channel attention residual module LSERB (local squeeze-and-excitation residual block). The module includes the local channel attention module LSEB (local squeeze-and-excitation block) and the residual module RB (residual block). To be specific, the LSEB module first divides the feature information, in order to obtain the sub-feature information, and then carries out the channel attention convolution for each partition, to assign different weights to the local information, after that it carries out the information integration to make full use of the feature information. Moreover, this paper introduces the RB module in the convolution operation, to improve the generalization ability of the model. Finally, we use the LSERB module to construct LSERNet, and the network is used in the blast furnace tuyere image anomaly detection. The experimental results show that the proposed LSERNet performs better in the collected images of the blast furnace tuyere.

The Section 2 of the article introduces the related work; Section 3 introduces the specific method and the specific structure of the model; Section 4 introduces the implementation process of the experiment; Section 5 analyzes the results; Section 6 summarizes the method we proposed and the results we obtained.

2. Related Work

Attention mechanisms have shown a great potential in deep learning. SENet first proposed an effective mechanism for the learning channel attention, which aims to model the correlation between the different channels and feature maps. Its structure is shown in Figure 1.
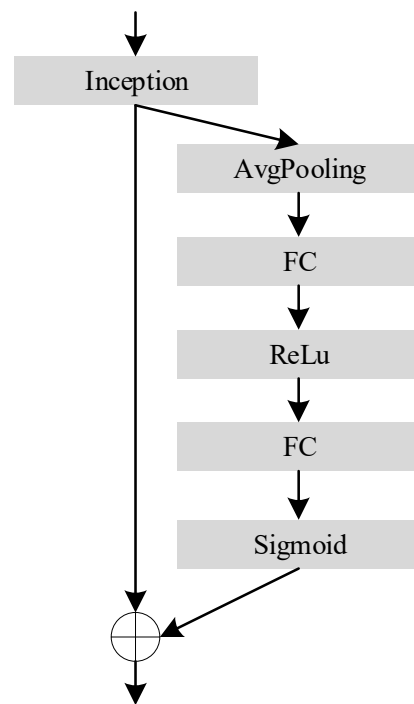
Figure 1. SE structure.

SENet captures the dependencies of all channels and integrates the global information. Firstly, global pooling was applied to the feature map, and then two convolutions were used to increase and reduce the dimension, to extract the global information of each channel. The SE module learns the weight of each channel by modeling the relationship between all channels. For the output of any convolutional layer, the SE module obtains the weighted feature map by introducing the extrusion and excitation operations.

The channel attention mechanism directly performs the global pooling operation on the information of each channel and ignores its local feature information, which makes that the local information in the feature map cannot be used reasonably. Existing methods mostly focus on developing complex attention modules. Our LSEB module aims to capture the detailed feature information of the image and achieve better feature extraction results. Compared with the original SE module, the proposed method can achieve a better performance under the same model complexity.

3. Proposed Methods

In this section, we introduce the structure and working principle of the LSEB module and the RB module, and explain how to combine the LSEB module with the RB module to build LSERNet.

3.1. The Structure and Working Principle of LSERNet

The overall structure of the proposed LSERNet is shown in Figure 2. The network consists of a convolutional layer, a pooling layer, a LSERB module, an average pooling layer, a fully connected layer, and the softmax layer. LSERB (local squeeze-and-excitation residual block) is based on the ResNet50 network structure. It is composed of the local channel attention LSEB (local squeeze-and-excitation block) combined with RB (residual block). This module is similar to the ResNet50 network structure and contains four layers, each of which is composed of RB modules. Layers from 1 to 4 consist of three, four, six, and three RB modules, respectively.
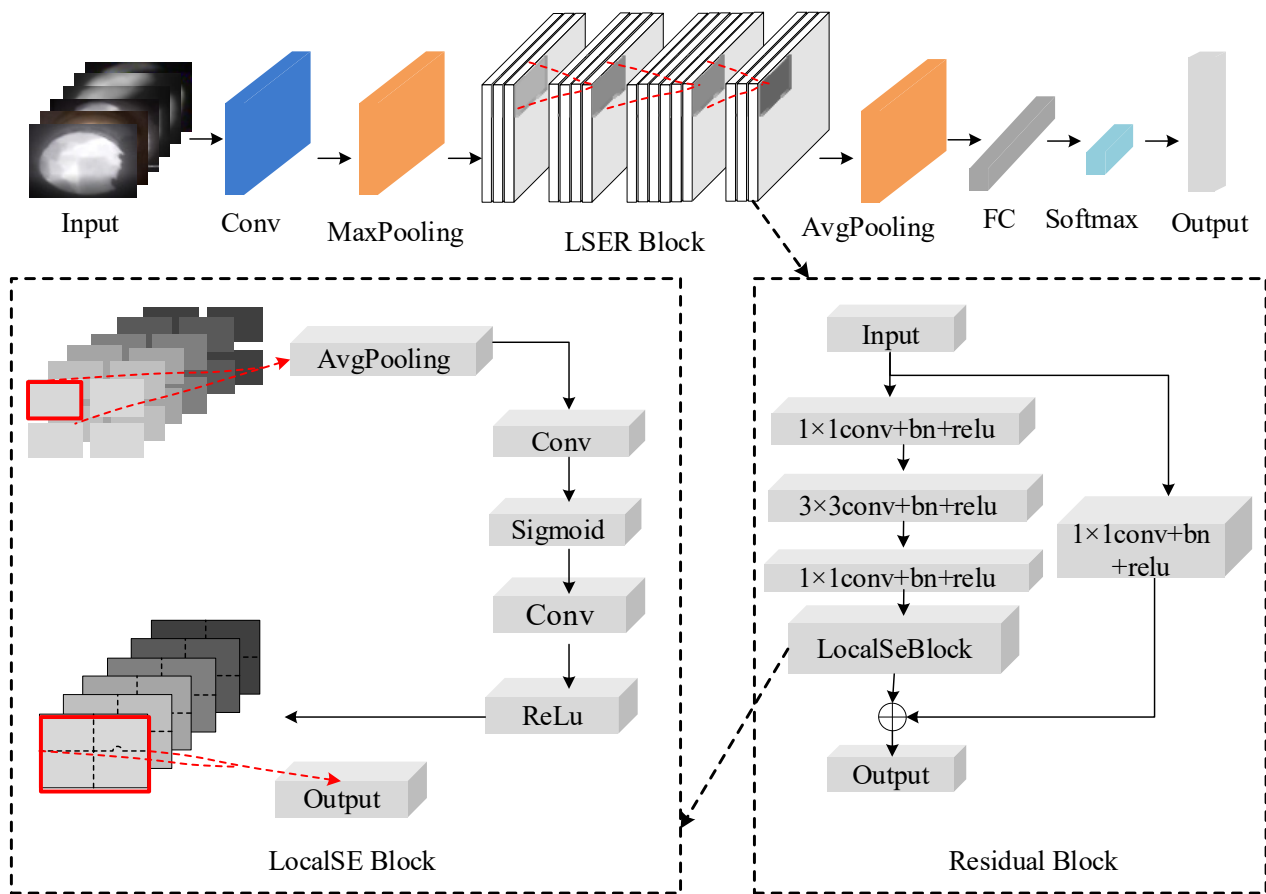
Figure 2. Structure diagram of the local channel attention residual convolutional neural network LSERNet.

Prior to the channel attention convolution, we locally divide the image feature information, input it into the channel attention module, respectively, and use the residual module for the convolution operation, to assign weights to the different regions. Then, the obtained information is integrated, and the final output is obtained by the softmax operation.

3.2. Construction of the Local Channel Attention Module LSEB (Local Squeeze-and-Excitation Block)

The channel attention mechanism focuses on modelling the relationship between the channels, and uses the dependencies to strengthen the feature screening and extraction ability of the network. The network can filter out the useful information by capturing the global information, strengthening the extraction of this part of the features, and suppressing the useless features. The channel attention mechanism includes two operations, squeeze and excitation. Firstly, the global spatial features of each channel are compressed into a descriptor to represent the information of each channel. Then, the dependency of the channels is learned, the weights of the different channels are obtained, and the feature map is adjusted, according to the weight information. In this paper, based on the channel attention mechanism, the LSEB attention convolution block is constructed, and the structure of the LSEB module is shown in Figure 3. The module focuses on the local feature information of the images. Prior to the channel attention convolution, the LSEB module first divides the local image feature information into blocks of the same size. For the channel attention convolution operation, it connects the sigmoid activation function to study the nonlinear interaction between the channels. Finally, the output of the LSEB module is obtained by the information integration, which highlights the effective information and makes full use of the feature information.
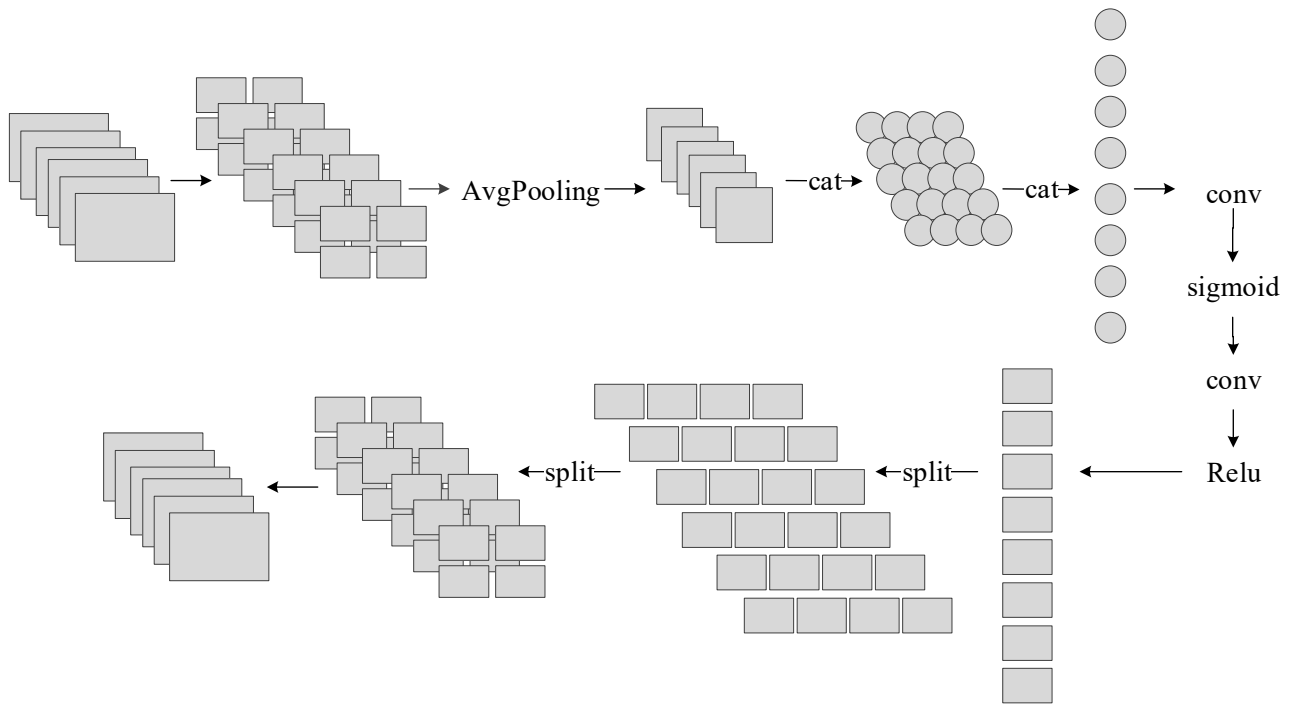
Figure 3. LSEB module structure.

3.3. Construction of the RB (Residual Block) Module

In order to avoid the degradation of learning and training process, as the layer number of the network model increases, our convolutional network introduces the RB (residual block) module, based on the bottleneck of ResNet50. ResNet was proposed by He et al. [13], in 2015. Its main contribution is to solve the problem of the classification accuracy decline with the depth of the CNN, accelerate the CNN training process through the proposed residual learning idea, and solve the problem of the gradient disappearance and gradient explosion. Using the idea of residual learning, He et al. proposed a shortcut connection structure of identity mapping, and the shortcut connection structure is shown in Figure 4. Here x is the input, F(x) is the residual map, H(x) is the ideal map, H(x) = F(x) + x. By transforming the fitting residual map F(x) into the fitting reasonable mapping H(x), we generate the output as a superposition of the input and residual maps, making the network more sensitive to changes between the input x and the output H(x).
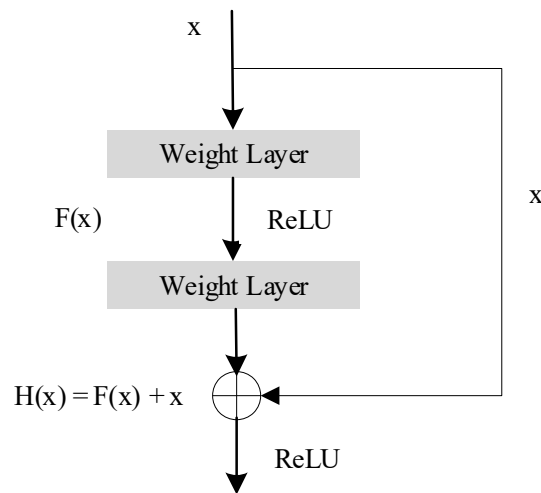

Figure 4. Shortcut Connections structure.

In order to build a deeper network structure, He et al. also proposed the bottleneck structure, as shown in Figure 5. The bottleneck is constructed to reduce the number of parameters and adapt to the deeper network structure. For ResNet50 or deeper networks, the number of calculations and parameters can be reduced, and the training time can be reduced. A 1×1 convolution is added to the bottleneck structure, in order to reduce the dimension of the input. The bottleneck structure is used in all ResNet50/101/152 networks. In order to avoid the degradation of the learning and training ability in the process of the model training, we introduce the LSEB module before the output of the bottleneck, which realizes the combination of the local channel attention and the residual mechanism. It ensures that the network performance is improved while the number of layers of the network model is continuously deepened.
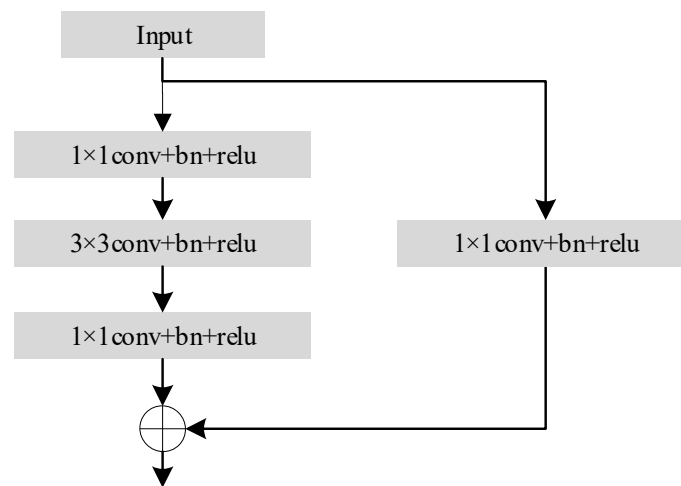

Figure 5. Bottleneck structure.

4. Experiments

This part describes the dataset, the preprocessing of the dataset and the process of the experiment.

4.1. Data Set and Preprocessing

The experiment is carried out on the blast furnace tuyere dataset collected by us. We collected 167 video data, the overall length is about 26.5 h. The video frame captures the blast furnace tuyere images in the artificial state of the blast furnace tuyere, while working. The six different categories consist of large, block, normal, and coal breaking, hang slag, and damping down. The description of the different working status categories of the blast furnace tuyere is shown in Figure 6.

The uneven amount of data obtained by each category may affect the training effect of the model. In order to obtain a better training effect, the following operations are performed on the data before the model input data:

1. The data should be augmented to make the model have a strong generalization ability and to avoid overfitting. The images were rotated randomly, and slight changes were added to enrich the dataset;
2. In order to reduce the influence of the image noise, the region of interest of the image is intercepted;
3. The size of the input image is 256×256, which reduces the amount of calculation and speeds up the operation of the model.

Therefore, the image dataset of the blast furnace tuyere after the screening and expansion, contains 21,120 images, including 3520 images of large chunks falling, material block, normal, coal breaking, slag hanging, and wind off [16].
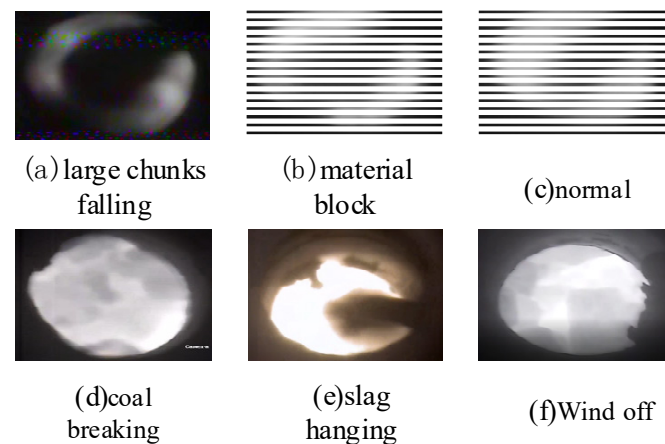
(a) large chunks falling   (b) material block   (c) normal

(d) coal breaking   (e) slag hanging   (f) Wind off

Figure 6. Blast furnace tuyere image categories. (**a**) large chunks falling: large darkened areas appear at the air outlet, and the large chunks are melted and disappear in the blast furnace, and light returns in the air outlet; (**b**) material block: gray coke in the blast furnace at the wind outlet for the winding movement; (**c**) normal: the edge of the tuyere is smooth; there are no impurities in the tuyere and no dark areas, and the dark areas of the coal gun and coal injection are clearly visible; (**d**) coal breaking: the tuyere boundary is smooth, the tuyere is clear and bright, and there are no dark areas produced by the coal injection; (**e**) slag hanging: there is uneven residue at the edge of the tuyere, and there is a small dark area at the slag hanging, and the gray level is large; (**f**) wind off: the coke rotation speed is slow, the carbon block gradually accumulates, and the wind is not bright and gradually darkens.

4.2. Experimental Parameter Setting

The experiments are conducted on PyCharm software (https://www.jetbrains.com/pycharm/) and the model is implemented using PyTorch (https://pytorch.org/). We set the number of iterations to 150, the optimization algorithm is set to be the stochastic gradient descent (SGD), and the loss function to be CrossEntropyLoss. In order to obtain the best effect of the model, we determine the values of the hyperparameter batch size and the learning rate through the experiments, the batch size is set as 16, 32, and 64, and the learning rate is set as 0.1, 0.01, and 0.001. There were nine combinations. The experimental results of the nine combinations were compared to determine the batch size and the learning rate.

Since the dataset is randomly divided in the model training, the training results of the model will be extremely unstable. In order to avoid such a situation, we use the k-fold cross validation to train the model. The k-fold cross-validation randomly divides the dataset into k parts, it takes k−1 parts as the training set and the remaining one part as the validation set, and trains the model k times. Each time, the dataset will be randomly divided into k parts again. Having repeated the results of the k experiments and taking the average, we obtain a good approximation of the generalization result. Therefore, we choose the result of the k-fold cross-validation as the final result of the model to improve the accuracy of the model evaluation.

4.3. Model Evaluation Indicators

We evaluate the complexity of the model by the number of model parameters. In the case of the same complexity, we compared the accuracy of the different models. In this paper, the average accuracy is used to compare the accuracy of each model, which is defined as follows:

$$\overline{A} = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{n_{ii}}{n_i}, \tag{1}$$

where, n_c represents the total number of sample categories, and we set it as 6 in our work; i is the category label, which is selected from 1 to 6. n_i represents the total number of samples of category i; n_{ii} represents the total number of accurate predictions in class i.

5. Analysis of the Results

This part introduces the k-fold cross-validation results of the model and the comparison of the experimental results with the different models.

5.1. k-Fold Cross-Validation Results

Since our model is based on ResNet50, the SE module is improved to construct the LSER-Net model, so we compare the proposed LSERNet model with SE-ResNet50 and ResNet50. The SE-ResNet50 model is the combination of the original SE module and ResNet50 model. Except for the difference of SE module, the rest is consistent with the LSERNet model. In order to verify the transferability of our proposed module, we combine the LSEB module with the ResNeXt model to construct the LSE-ResNeXt model, and conduct the comparative experiments with the SE-ResNeXt model and the ResNeXt model, respectively.

In the k-fold cross-validation experiment, we set k as 5, we randomly divide the dataset into 5 evenly, and take 4 as the training set and 1 as the validation set. The whole experiment should be run five times, and the average of the five validation results is taken as the validation error of this model. The k-fold cross-validation results of the different models are shown in Figure 7. There were nine combinations of batch sizes and learning rates for each model, and the number of experiments was carried out 45 times. Finally, the optimal parameter combination and the optimal accuracy of each model were obtained.
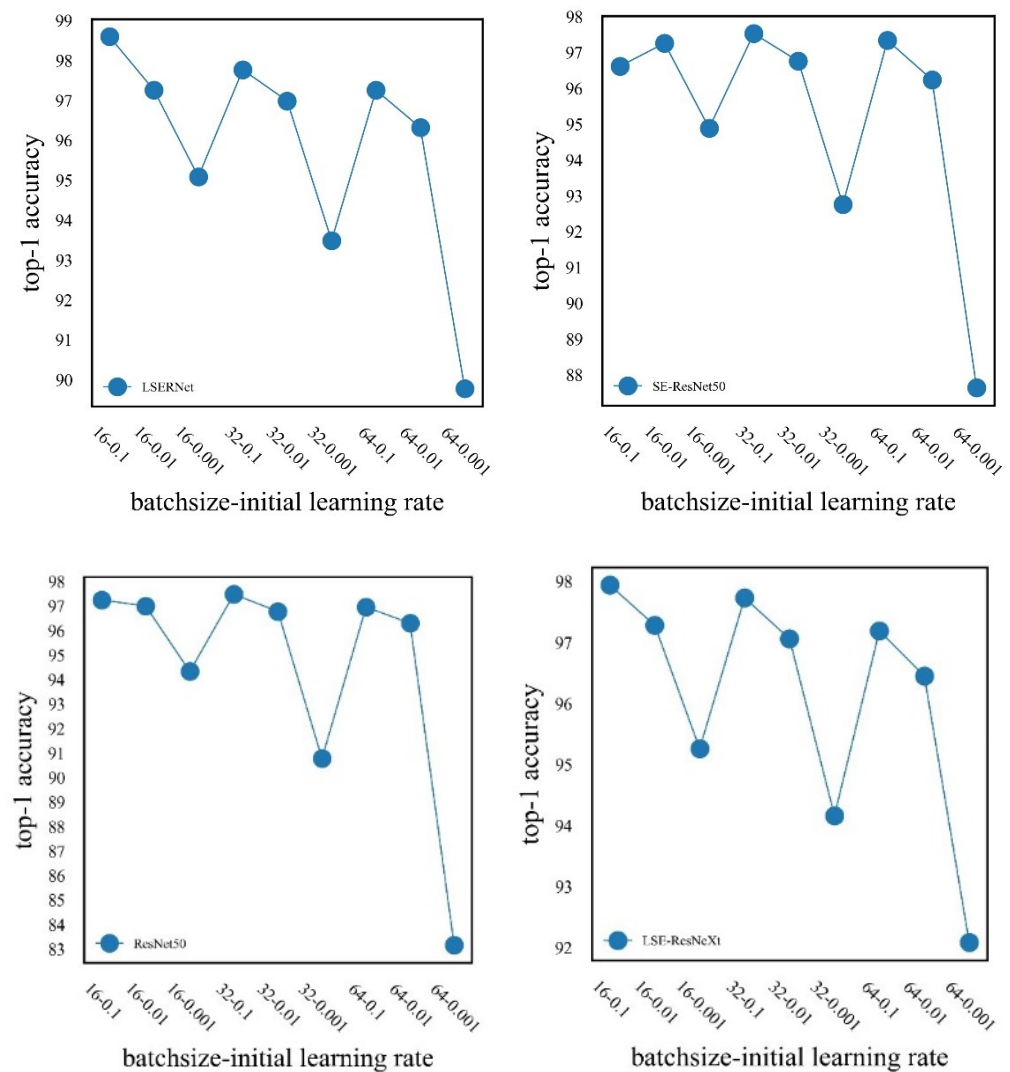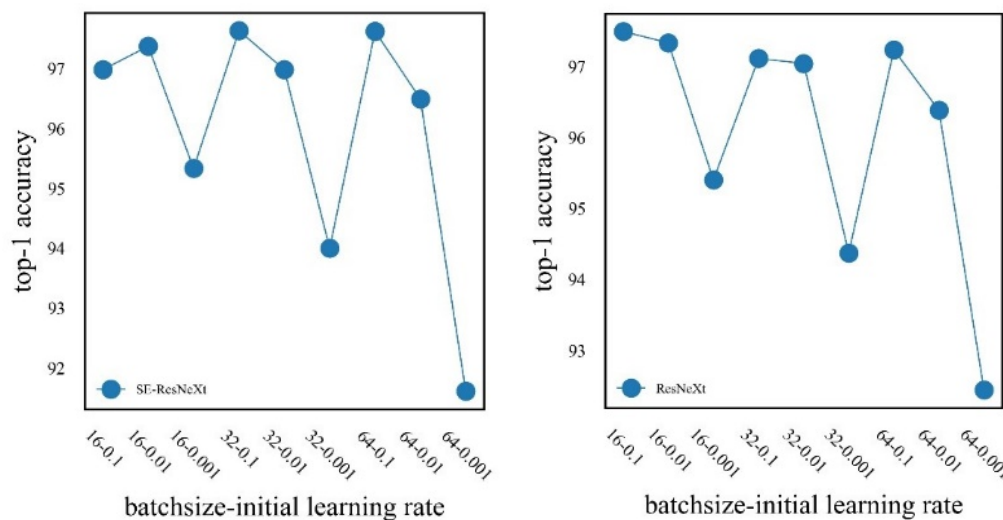

Figure 7. *Cont.*

Figure 7. k-fold cross-validation results for the different models.

For the LSERNet model, when the batch size is 16 and the learning rate is 0.1, the accuracy of the model is highest with 98.59%; when the learning rate is 0.01 and 0.001, the accuracy decreases. When the batch size is 32 and 64, the accuracy is relatively higher when the learning rate is 0.1, and the accuracy decreases when the learning rate is 0.01 and 0.001, respectively.

For the SE-ResNet50 model, when the batch size is 32 and the learning rate is 0.1, the accuracy of the model is 97.52%, which is the highest; when the learning rate is 0.01 and 0.001, the accuracy decreases. When the batch size is 16, the accuracy is relatively higher when the learning rate is 0.01. The accuracy decreases for both learning rates of 0.1 and 0.001. When the batch size is 64, the accuracy is relatively highest when the learning rate is 0.1; when the learning rate is 0.01 and 0.001, the accuracy decreases.

For the ResNet50 model, when the batch size is 32 and the learning rate is 0.001, the accuracy of the model is 97.42%, which is the highest accuracy; when the learning rate is 0.1 and 0.01, the accuracy decreases. When the batch size is 16 and 64, the accuracy is relatively higher when the learning rate is 0.1; when the learning rate is 0.01 and 0.001, the accuracy decreases.

For the LSE-ResNeXt model, when the batch size is 16 and the learning rate is 0.1, the accuracy of the model is 97.94%, which is the highest accuracy; when the learning rate is 0.01and 0.001, the accuracy decreases. When the batch size is 32 and 64, the accuracy is relatively higher when the learning rate is 0.1; when the learning rate is 0.01 and 0.001, the accuracy decreases.

For the SE-ResNeXt model, when the batch size is 32 and the learning rate is 0.1, the accuracy of the model is 97.63%, which is the highest accuracy; when the learning rate is 0.01 and 0.001, the accuracy decreases. When the batch size is 16, the accuracy is relatively higher when the learning rate is 0.01. The accuracy decreases for both learning rates of 0.1 and 0.001. When the batch size is 64, the accuracy is relatively higher when the learning rate is 0.1, and the accuracy decreases when the learning rate is 0.01 and 0.001.

For the ResNeXt model, when the batch size is 16 and the learning rate is 0.1, the accuracy of the model is 97.49%, which is the highest accuracy; when the learning rate is 0.01 and 0.001, the accuracy decreases. When the batch size is 16 and 64, the accuracy is relatively higher when the learning rate is 0.1; when the learning rate is 0.01 and 0.001, the accuracy decreases. The optimal parameter combinations for the different models are shown in Table 1.

Table 1. The optimal parameter combinations of the different models.

Model	Batch Size	Learning Rate	Average Accuracy
LSERNet	16	0.1	98.59%
SE-ResNet50	32	0.1	97.52%
ResNet50	32	0.001	97.42%
LSE-ResNeXt	16	0.1	97.94%
SE-ResNeXt	32	0.1	97.63%
ResNeXt	16	0.1	97.49%

5.2. Comparison and Analysis of the Accuracy of the Different Models

The different models are trained with their own optimal parameter settings, and the complexity of each model is consistent. Figure 8 shows the average classification accuracy of the different models under the same complexity. The horizontal axis represents the number of parameters of the model, and each model has the same number of parameters and the same complexity. The results show that in the blast furnace image dataset collected in this paper, based on the ResNet50 model, the accuracy of the proposed LSERNet is 98.59%, the accuracy of the SE-ResNet50 model is 97.52%, and the accuracy of the ResNet50 model is 97.42%. In comparison, the LSERNet model constructed by ResNet50 combined with our proposed LSEB module has the highest accuracy. At the same time, based on the ResNeXt model, the accuracy of LSE-ResNeXt model is 97.94%, the accuracy of the SE-ResNeXt model is 97.63%, and the accuracy of the ResNeXt model is 97.49%. For comparison, the LSE-ResNeXt model constructed by combining ResNeXt with our proposed LSEB module, also achieved the highest accuracy. In general, the combination of our proposed LSEB module with ResNet50 and ResNeXt achieves the highest accuracy. This is because the LSEB module focuses on the local information of the image, so that the regional information can be fully utilized and the accuracy of the model is improved.
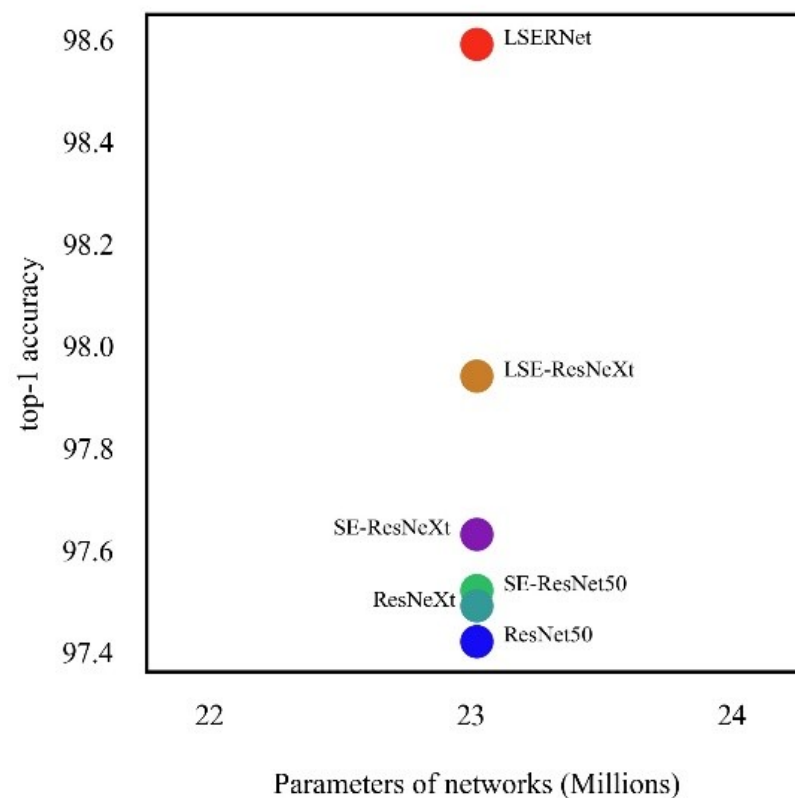
Figure 8. Average classification accuracy of the different models.

5.3. Comparative Analysis of the Feature Extraction

We randomly selected four channels a, b, c, and d, to compare the feature extraction capability of the proposed LSEB module with that of the SE module, and the feature extraction heat map is shown in Figure 9. It can be seen from the figure that some important features of channels b, c, and d are ignored after the convolution of the SE module, so that the model cannot efficiently recognize the different states of the blast furnace tuyere image. Following the LSEB module convolution, the model will strengthen the extraction of the local feature information, the feature information is reasonably assigned the corresponding weights, and the important information can be fully utilized. Since the original channel attention is to pool the entire channel and then do the full connection, and the weight of each part of the feature map is not the same, it should not be treated the same, so that some important features will be ignored. The LSEB module makes the regional information fully utilized and strengthens the feature extraction ability.
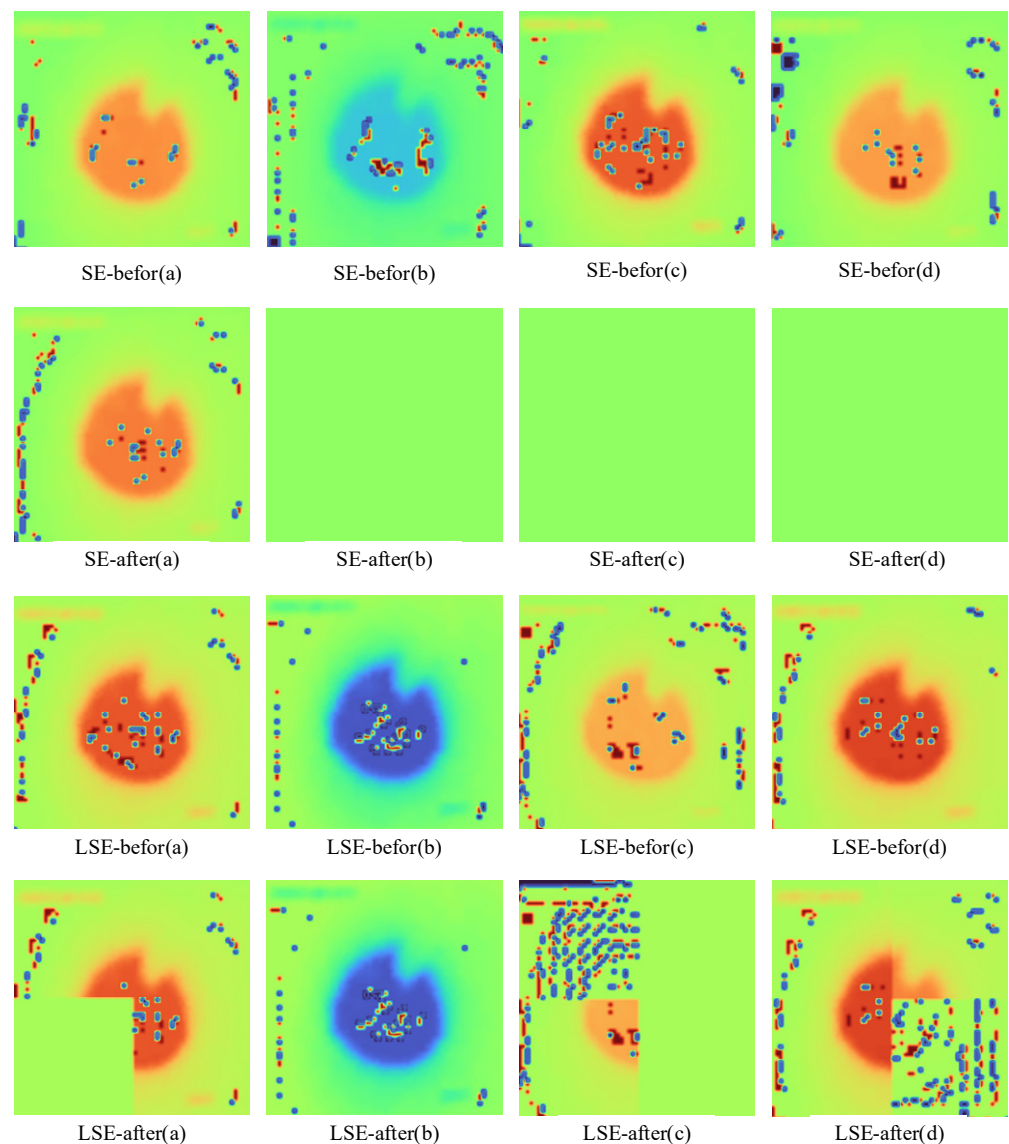

Figure 9. Heat map of the feature extraction. SE-befor is the feature map before the convolution of the SE module, SE-after is the feature map after the convolution of the SE module. LSE-befor is the feature map before the LSE module convolution, LSE-after is the feature map after the LSE module convolution. (**a**–**d**) are the four randomly selected channels.

6. Conclusions

In this paper, a local channel attention module LSEB is proposed, and combined with the residual mechanism, a local channel attention residual network LSERNet is constructed. Extensive experiments are carried out with other models on the blast furnace tuyere image dataset collected by us. The results show that:

1. Compared with the original SE module, the LSEB module strengthens the ability of the feature extraction, does not ignore some important feature information, and makes full use of the regional information;
2. Compared with other models, the LSERNet model constructed by us has achieved the highest accuracy in the case of the same model complexity as other models, which can effectively identify the abnormal working state of the blast furnace tuyere.

The current limitation of the proposed method is that we only applied our own collected dataset with a high accuracy. When migrating to a new dataset, the model may need to be retrained, due to the different resolution of the sensors and different camera positions. All we need to do is retrain the model with a new dataset. Although retraining is required, the training time is acceptable and does not necessarily affect the applicability of the method.

The model has been built, and the effect is very good, but it has not been actually installed in the factory. The next step is that, before the model is applied to the factory, a high-intensity test will be carried out to minimize the error rate, and a complete early warning system will be developed.

References

1. Liu, Y.; Lei, Y.B.; Fan, J.L.; Wang, F.; Gong, Y.C.; Tian, Q. Survey on image classification technology based on small sample learning. *Acta Autom. Sin* **2021**, *47*, 297–315.
2. Wu, S.; Xu, Y.; Zhao, D. Survey of object detection based on deep convolutional network. *Pattern Recognit. Artif. Intell.* **2018**, *31*, 45–46.
3. Jing, Z.W.; Guan, H.Y.; Peng, D.F.; Yu, Y.T. Survey of research in image semantic segmentation based on deep neural network. *Comput. Eng.* **2020**, *46*, 1–17.
4. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
6. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
7. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

8. Li, P.; Xie, J.; Wang, Q.; Zuo, W. Is second-order information helpful for large-scale visual recognition? In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2070–2078.
9. Li, Y.; Wang, N.; Liu, J.; Hou, X. Factorized bilinear models for image recognition. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2079–2087.
10. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F.F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
11. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7794–7803.
12. Gao, Z.; Xie, J.; Wang, Q.; Li, P. Global second-order pooling convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
13. 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.
14. Huang, G.; Liu, Z.; 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. 2261–2269.
15. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. *arXiv* **2017**, arXiv:1709.01507.
16. Wang, R.; Li, Z.; Yang, L.; Li, Y.; Zhang, H.; Song, C.; Jiang, M.; Ye, X.; Hu, K. Application of Efficient Channel Attention Residual Mechanism in Blast Furnace Tuyere Image Anomaly Detection. *Appl. Sci.* **2022**, *12*, 7823. [CrossRef]
17. Chen, Y.; Kalantidis, Y.; Li, J.; Yan, S.; Feng, J. A2-Nets: Double attention networks. In Proceedings of the NIPS, Montréal, QC, Canada, 3–8 December 2018.
18. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3141–3149. [CrossRef]
19. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Vedaldi, A. Gather-excite: Exploiting feature context in convolutional neural networks. In Proceedings of the NeurIPS, Montréal, QC, Canada, 3–8 December 2018.
20. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. Bam: Bottleneck attention module. *arXiv* **2018**, arXiv:1807.06514.
21. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
22. Li, X.; Hu, X.; Yang, J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. *arXiv* **2019**, arXiv:1905.09646.
23. Zhang, Q.L.; Yang, Y.B. Sa-net: Shuffle attention for deep convolutional neural networks. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 2235–2239.
24. 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 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.