GOSPODARKA SUROWCAMI MINERALNYMI – MINERAL RESOURCES MANAGEMENT

2023 Volume 39 Issue 1 Pages 131–148

DOI: 10.24425/gsm.2023.144630

HEHUI ZHOU[1], GAIPIN CAI[2], SHUN LIU[3]

Research on stacked ore detection based on improved Mask RCNN under complex background

Introduction

In the raw-ore-crushing process, a screen is often installed at the feeding port of the ore bin to control the feeding particle size of the crusher, so as to improve the working efficiency of the crusher, reducing energy consumption and improving the crushing quality. However, manual crushing is generally used for raw ore with a larger particle size. This manual crushing method not only has low crushing efficiency, high labor cost and high work intensity but also a very harsh working environment. Therefore, the method of manipulator crushing ore based on machine vision and image processing technology will become a development trend. However, the shape, quantity, distribution position and stacking state of large ores

✉ Corresponding Author: Gaipin Cai; e-mail: 1123615286@qq.com

[1] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, China; ORCID iD: 0000-0001-5532-0528; e-mail: zhh_0712@163.com

[2] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology; Jiangxi Province Engineering Research Center for Mechanical and Electrical of Mining and Metallurgy, China; e-mail: 1123615286@qq.com

[3] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, China; e-mail: 1711266496@qq.com

© 2023. The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike International License (CC BY-SA 4.0, http://creativecommons.org/licenses/by-sa/4.0/), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.

remaining on the barrier screen after ore unloading are random. In addition, due to the complex site environment and difficulties in identifying the ore stacking state, fuzzy image background, etc., the accurate identification of ore and the reasonable selection of the crushing point as well as the path planning of the manipulator bring great challenges. Therefore, a method that can accurately identify ores in a complex background is needed, and deep learning provides the technical means to solve the above problems.

In recent years, with the rapid development of deep learning (LeCun et al. 2015), convolutional neural networks have been widely used in the field of object detection (Zhang et al. 2019), which mainly consists of two parts. One is the one-stage target detection algorithm. The representative algorithms include YOLO (Redmon et al. 2016; Redmon and Farhadi 2017, 2018), SSD (Liu et al. 2016) and other target-detection algorithms based on a CNN (Zhang et al. 2022) network and a regression network. It adopts an end-to-end detection method and has a high detection speed, but the detection accuracy still needs to be improved (Fan et al. 2020). The second is a two-stage target-detection algorithm, the representative algorithms are RCNN (Girshick et al. 2014), Fast RCNN (Girshick 2015), Faster RCNN (Ren et al. 2015) and Mask RCNN (He et al. 2017) used in this study. These two-stage target detection algorithms based on a CNN network and a RPN network have higher detection accuracy than one-stage target detection algorithms and are suitable for occasions with high accuracy requirements (Dong et al. 2022). Huang et al. (Huang et al. 2021) achieved the detection of tension shear cracks in thermal infrared images of rocks by improving the Faster RCNN model by introducing an attention mechanism to optimize the feature pyramid network, and using a cascade structure to improve the detection frame regression accuracy. Liu et al. (Liu et al. 2020) used a simplified VGG16-based extraction network for feature extraction and the learning of rock images under the target detection framework of Faster R-CNN deep learning. This method accurately identifies rock types with similar image features and achieves type recognition of rocks such as peridotite, basalt, marble and gneiss. Mask RCNN was proposed by He Kaiming in 2017. It uses a convolutional neural network based on the mask area. It adds a branch of semantic segmentation on the basis of Faster RCNN to achieve target detection and semantic segmentation at the same time in one network. Nie et al. (Nie et al. 2020) used Mask RCNN to realize automatic detection of ship number and ship positioning under atomization and fuzzy background. Deng et al. (Deng et al. 2022) used Mask RCNN to realize the recognition and segmentation of deep-sea mineral particles. In mining engineering, accurate reserves of mineral resources need to be assessed. Liu et al. (Liu et al. 2021) proposed a method to evaluate the quality of rock mass using the Mask RCNN deep learning instance segmentation network. From the application of Mask RCNN in various fields, it can be concluded that the model based on Mask RCNN has a reasonable ability to recognize and segment in situations with complex backgrounds.

In this paper, based on the Mask RCNN model, a method that can accurately detect ore in complex background is established. Mask RCNN uses ResNet101 (He et al. 2016) to extract features from ore images, and then fuses the features through a feature fusion network. However, this simple fusion is clearly not sufficient. In order to improve the detection accuracy of the ore, an improved multi-path feature pyramid network is required and this is designed in this paper. This network can downsample shallow features to obtain feature layers of different scales, and then add them to the features of the deeper layer, so as to realize the feature fusion from the shallow layer to the deep layer, and improve the utilization efficiency of the shallow feature layer and enrich the network. Details of deep features. In order to avoid the loss of detailed information by the original network structure, the problem of misidentifying multiple stacked ores as a whole ore is caused, and the accuracy of ore identification is improved.

The first section of this paper mainly introduces the Mask RCNN model and defines its three-part loss function. In the second section, the characteristics of the backbone feature extraction network are analyzed, the shortcomings of the network for ore detection are found and an improved method is proposed. In the third section, the improved Mask RCNN is analyzed experimentally. Firstly, an ore image dataset suitable for this paper is made. The network is then trained, the evaluation index is proposed for the training results, and the results are compared and analyzed with the results of other algorithms. The results show that the proposed method is obviously better than other existing methods. The fourth section summarizes the research results of this paper.

1. Mask RCNN network

Mask RCNN is a multi-task network model. It can achieve pixel-level accurate segmentation while classifying and recognizing objects and regressing their positions. Figure 1 is a simplified diagram of the structure of Mask RCNN. The network structure of Mask RCNN is mainly composed of five parts: a backbone feature extraction network, RPN, ROI Align, a segmentation network and a loss function. Mask RCNN uses ResNet101 for feature extraction to obtain the corresponding multi-scale and multi-channel feature layers and then transfers the extracted feature layers to the FPN network for feature fusion to generate a new feature layer. The new feature layer is used for the RPN network to generate the region proposal candidate frame; furthermore, it performs the ROI Align operation together with the region proposal candidate frame. A fixed-size region proposal feature layer is obtained, and

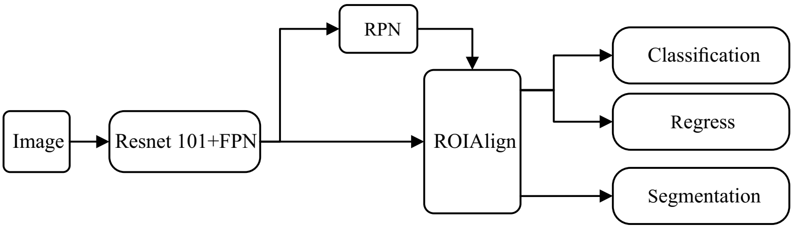


Fig. 1. Mask RCNN network structure

Rys. 1. Struktura sieci maski RCNN (konwolucyjna sieć neuronowa oparta na regionach) then it enters two task branches, one for object classification and location regression, and the other for generating segmentation masks.

Mask RCNN defines a multi-task loss function, as shown in formula (1).

$$L\left(\{p_i\},(\{t_i\}),(\{m_i\})\right)=\frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p^*_i)+ \tag{1}$$

$$+\lambda\frac{1}{N_{reg}}\sum_i p^*_i\, L_{reg}(t_i,t^*_i)+\frac{1}{N_{mask}}\sum_i L_{mask}(m_i)$$

L_{cls} represents the loss of target classification, which can be expressed as:

$$L_{cls}(p_i,p^*_i)=-\log\left[p_i p^*_i+(1-p^*_i)(1-p_i)\right] \tag{2}$$

L_{reg} represents the loss of the candidate box position regression, which can be expressed as:

$$L_{reg}(t_i,t^*_i)=smooth(t_i-t^*_i)=\begin{cases}0.5(t_i-t^*_i)^2, if\,|t_i-t^*_i|<1\\ |t_i-t^*_i|-0.5, otherwise\end{cases} \tag{3}$$

L_{mask} represents the loss of target segmentation, and the segmentation loss function is the average binary cross-entropy loss function. Each pixel in the real frame is classified into two categories to determine whether it belongs to the category of ore, which can be expressed as:

$$L_{mask}=-\sum_i^n\left[y^*_i\log(m_i)+(1-m^*_i)\log(1-m^*_i)\right] \tag{4}$$

In the appeal formula:

- N_{cls} represents the number of categories of classification,
- i is the number of region proposal candidate boxes, and p_i is the probability that the i-th region proposal candidate box is a positive sample. The ground truth label p^*_i is 1 to indicate that the candidate frame is a positive sample, and p^*_i is 0 to indicate that the candidate frame is a negative sample.
- t_i represents the center coordinates of the proposed candidate boxes in each region, and t^*_i represents the center coordinates of the target ground-truth boxes.
- m_i represents the predicted pixel value and m^*_i represents the actual pixel value.

2. Establishment of improved Mask RCNN network model

2.1. Analysis of Feature Pyramid Networks

Figure 2 is the backbone feature extraction network structure diagram of Mask RCNN. The image input to ResNet101 was firstly convolved and maximally pooled, and then entered into two residual structures Conv Block and Identity Block to get the feature layer. After different times of residual structure, four feature layers of different scales were obtained: C2, C3, C4 and C5. Four feature layers of different scales were used to construct the feature pyramid network. C2, C3, C4 and C5 were convolved and upsampled, and then the shared feature layers P2, P3, P4 and P5 were obtained by one-way top-down fusion. However, this feature pyramid network structure is relatively simple, only one-way feature fusion is performed from top to bottom, and the degree of feature fusion is low. This fusion

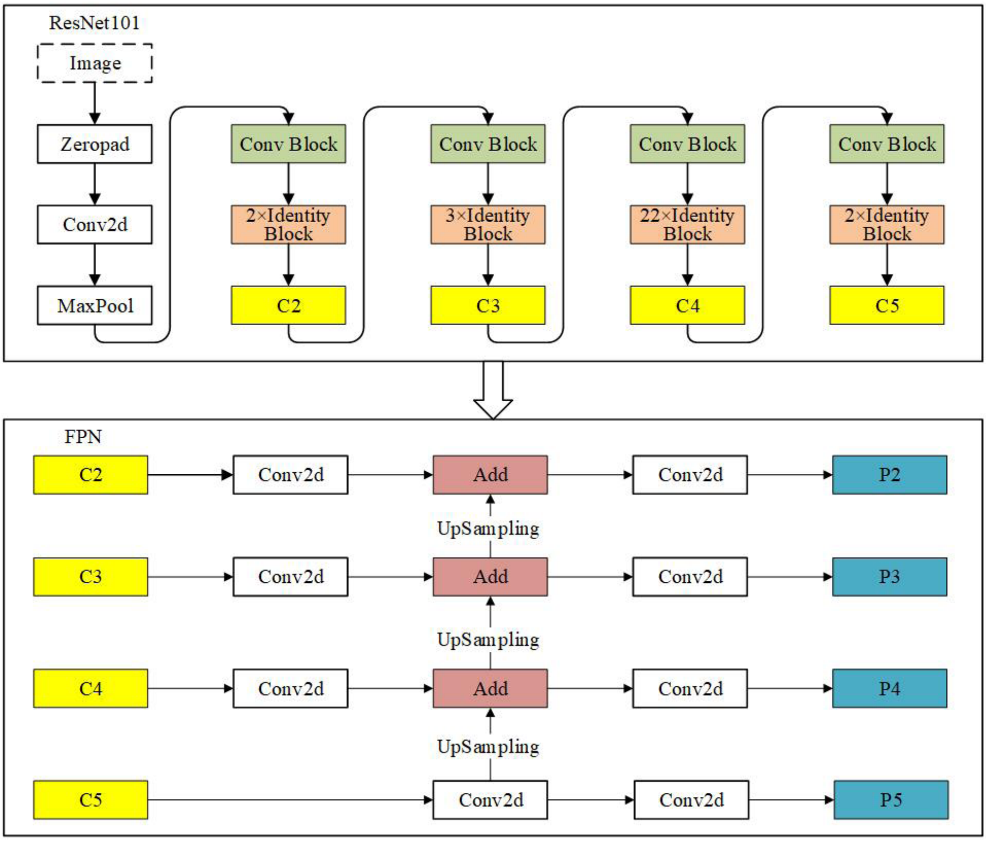


Fig. 2. Backbone network of feature extraction

Rys. 2. Sieć szkieletowa ekstrakcji cech method does not improve the deep feature layers lacking detailed information, resulting in the phenomenon that multiple stacked ores are identified as a whole, reducing the accuracy of ore identification. Therefore, the optimization scheme proposed in this paper is to add a reverse parallel feature fusion path in the original network to fully integrate the details of the shallow feature layer with the deep semantic information and enrich the details of the deep feature layer to strengthen the extraction ability of the feature extraction network for ore features.

2.2. Design of the improved multipath feature pyramid network structure

Feature layers of different scales have different semantic information and detail information. Through the analysis and comparison of various feature fusion networks, it is found that deep and shallow features are complementary in the information contained. In the target detection network, feature fusion of different levels can greatly improve the detection performance of the network. Mask RCNN uses FPN to perform unidirectional feature fusion from deep features to shallow features, which enriches the semantic information of shallow features to a certain extent but does not make any fusion for deep features. Obviously, such a feature fusion method is not sufficient.

To further solve the above problems, an improved multipath feature pyramid network (MFPN) is proposed – Figure 3 is the improved feature pyramid network proposed in this paper. A single bottom-up feature fusion path is added on the basis of FPN. The feature fusion path directly acts on the feature layer extracted by the feature-extraction network, and forms two parallel fusion paths together with the top-down feature fusion path in FPN. The dotted box in the figure is the newly added bottom-up feature fusion path, which downsamples the shallow features and adds the deeper features. This can realize the feature fusion from shallow to deep, enrich the detailed information of deep features, and then add the

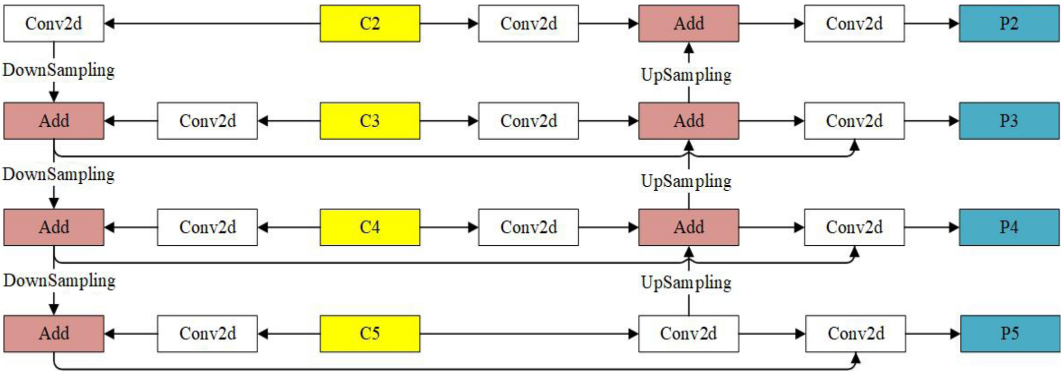


Fig. 3. Improved feature pyramid network structure

Rys. 3. Ulepszona struktura sieci piramidy cech features of each level fused from the bottom to the top and the features of each level fused from the top to the bottom through the horizontal connection. A feature layer rich in semantic information and detail information at each level is obtained. Since the two paths are operated in parallel, it can not only enhance the feature fusion of ore images and improve the accuracy of ore detection but also ensure the running speed of the network without increasing the network computing time.

2.3. The process of improving the Mask RCNN algorithm

According to the obtained improved feature pyramid network structure, Figure 4 shows the flow of the improved Mask RCNN algorithm. First, the input ore image feature layer is extracted by ResNet101, and four feature layers of different scales C2, C3, C4, and C5 are obtained, and then the four feature layers are up-sampled and down-sampled respectively through two parallel paths in MFPN. The feature layers of the same scale fused by the two sampling paths are added to obtain a new feature layer rich in semantic information and detail information, namely P2, P3, P4, and P5. The new shared feature layer enters the RPN to generate a candidate frame and performs the ROI Align operation together with the candidate frame to obtain a fixed-size candidate region feature layer, and finally enters the two branches of Mask RCNN. One of the branches enters the fully connected layer FC to classify and regress the candidate frame so that the candidate frame is infinitely close to the real frame of the ore. The other branch is to perform a fully convolutional FCN on the feature layer in the region with the help of the candidate frame and obtain the mask segmentation result of the ore image.

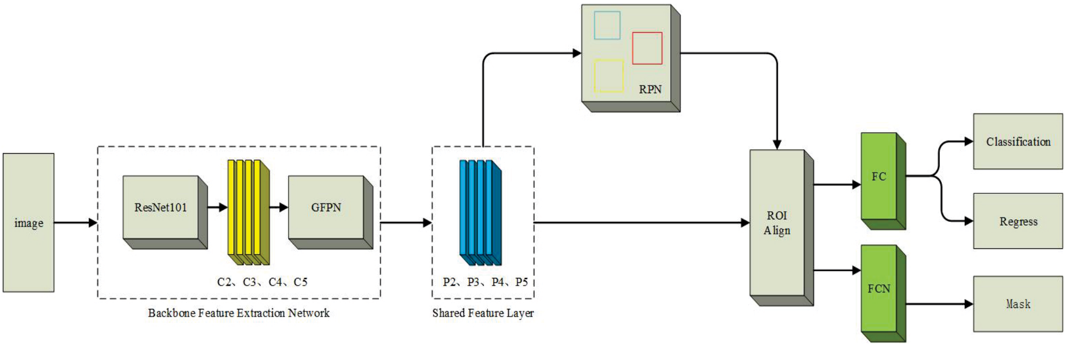


Fig. 4. Flowchart of the improved Mask RCNN algorithm

Rys. 4. Schemat działania ulepszonego algorytmu maski RCNN

3. Experiment analysis

3.1. Construction of the dataset

The original ore images are from the ore images published by Lampinen, S. et al. (Lampinen et al. 2021). The distribution of ore on the fence is divided into three states: single ore, scattered distribution of multiple ores, and stacked distribution of ore. The original image of ore is equalized so that the number of ore images distributed in the three states is consistent.

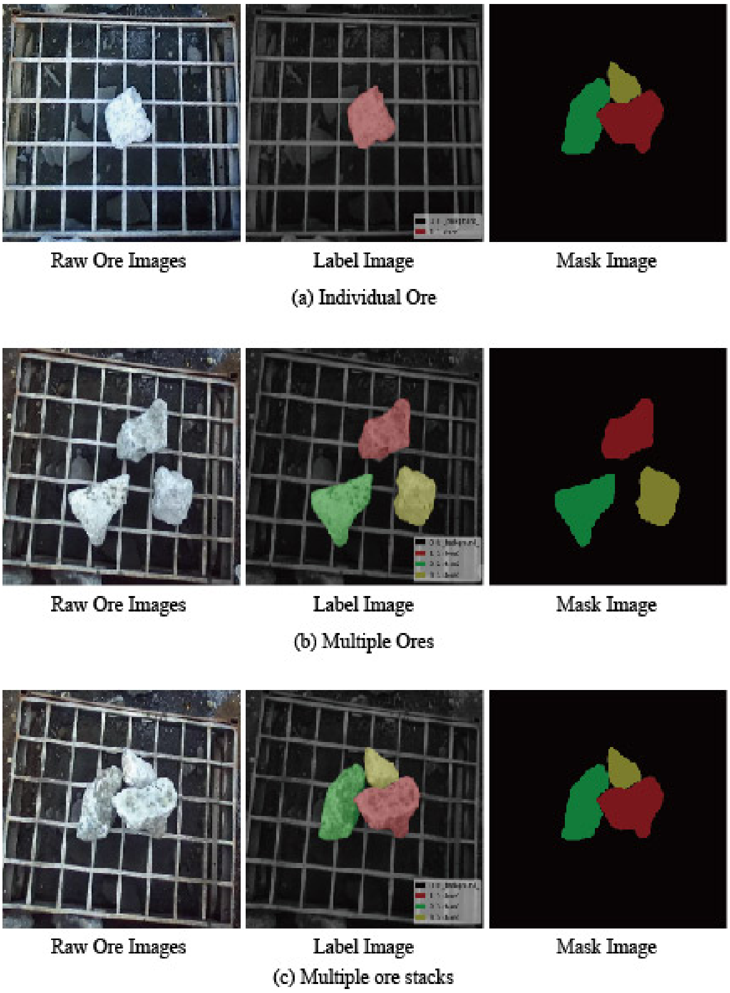


Fig. 5. Example of ore dataset
a – individual ore, b – multiple ores, c – multiple ore stackls

Rys. 5. Przykładowy zbiór danych dotyczących rudy
a – pojedyncza ruda, b – wiele rud, c – wiele stosów rudy

Label the original images with Labelme software, and make a total of 1,000 ore image datasets suitable for the Mask RCNN network. The ore image data set is divided into a training set and a validation set according to the ratio of 9 : 1. Figure 5 shows part of the ore image dataset made in this paper.

3.2. Training platform construction and model training

Table 1 presents the hardware system parameters used for training the network and testing in this paper.

On the basis of the computer shown in Table 1, build a suitable Tensorflow-GPU deep learning framework and configure the function library required in the network training process. The training and testing of the improved Mask RCNN network is implemented in the VScode compiler using the Python language.

In order to realize the training of the Mask RCNN network, it is first necessary to adjust the image of the training set to 512 × 512 pixels and set the network parameters of the pre-training model. The RPN anchor is set to (16, 32, 64, 128, 256), the number of iterations (max epoch) is set to 200, the initial learning rate is set to 0.01, and the momentum factor is set to 0.9. Table 2 shows some other parameters at the beginning of the training.

Table 1. Hardware System Parameters

Tabela 1. Parametry systemu sprzętowego

Hardware Name	Parameter configuration
CPU	AMD Ryzen 7 4800 H with Radeon Graphics 2.90 GHz
GPU	GeForce GTX 1060 Ti
VRAM	6 G

Table 2. Network Parameters

Tabela 2. Parametry sieciowe

Parameter Name	Parameter configuration
Initial Learning Rate	0.01
Momentum Factor	0.9
Training Times	200
Optimizer	Adam
Activation Function	ReLU
Output Image Size	512 × 512

Figure 6 shows the change of loss value during the training process of the improved Mask RCNN in this paper. The change of the total loss function value of the improved Mask RCNN network on the training dataset is shown in Figure 6a. According to the multi-task loss function defined by Mask RCNN, the change of the loss function value of the three parts of the total loss of classification, regression and segmentation are shown in Figures 6b, c and d, respectively.

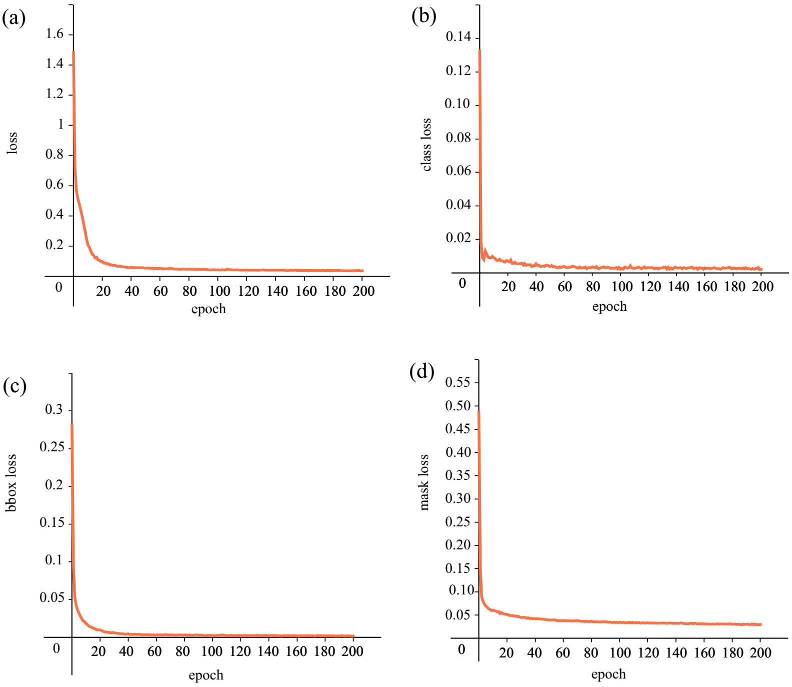


Fig. 6. Improved network training loss curve

Rys. 6. Poprawiona krzywa straty uczenia sieci

The abscissa in Figure 6 represents the number of iterations during network training and the ordinate represents the loss value during network training. The time of each epoch is affected by various parameters such as GPU, anchor size of RPN and batch. With the parameters set in this paper, the time of each epoch is about 390 sec. As the number of iterations increases, the loss value of the network model on each part of the training dataset decreases. In Figure 6a, the total loss value decreases rapidly in the first forty iterations, indicating that there is a large difference between the initial parameters and the optimal parameters set by the network training model. The loss value between 40 and 140 iterations declines relatively gently, indicating that the parameters of the network training model are close to the optimal value at this time. In order to avoid the phenomenon of network training overfitting, the network model parameters are fine-tuned by reducing the learning rate. When the number of iterations tends to 140, the loss value of the training set stabilizes around 0.04, and the change trend of the loss value curve of the three parts (b), (c) and (d) is roughly the same as the change trend of the total loss value, indicating that the network model continuously optimizes the parameters through training. Finally, it approximates the optimal value so as to achieve the effect of convergence and realize the training process of the network model.

3.3. Determination of evaluation indicators

For deep learning network models, precision and recall are usually used to evaluate the detection model accuracy. The precision rate represents the correct rate of the model's prediction of positive samples, and the recall rate represents the ratio of the number of positive samples predicted by the model to all positive samples. The calculation formulas of precision rate and recall rate are shown in Formula (5) and Formula (6), respectively.

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

In the above formula:

- *TP* is the number of ore that is actually ore and is correctly identified as ore by the model;
- *FP* is the number of ores recognized by the model as the actual background;
- *FN* is the number that is actually a single ore model but is recognized as a background.

The ore detection model based on Mask RCNN will output a confidence level for each predicted ore, which represents the probability that the target is a positive sample, that is, the probability that the predicted result is the original ore. Usually, confidence is proportional to precision and inversely proportional to recall, in other words, the higher the confidence, the higher the precision of the prediction result and the lower the recall rate of the prediction result. Therefore, when the confidence level is different, the detection results are very different. Thus, it is reasonable to combine the precision rate and the recall rate as the evaluation index of the model's detection effect on raw ore. In order to better describe the prediction performance of the network model under all confidence thresholds, a *PR* curve is drawn with the recall rate on the abscissa and the precision rate on the ordinate, which represents the maximum precision rate under a certain recall rate. The average precision can describe the performance of the network more intuitively, it represents the area enclosed by the *PR* curve and the abscissa, and the calculation formula is shown in Formula (7).

$$AP = \int_{0}^{1} P(R)dR \tag{7}$$

In this experiment, the IoU thresholds are set to 0.5 and 0.75 to calculate the AP value of the network, which are denoted as AP_{50} and AP_{75}, respectively. The IoU threshold is interpolated ten times from 0.5 to 0.95, and the AP value is calculated every 0.05, denoted as AP', and the AP' value calculated for 10 times is averaged and denoted as AP_m. The larger the value of AP_m, the more accurate the prediction result of the network.

3.4. Experimental results and analysis

In order to verify the accuracy of the improved Mask RCNN network model, the trained network model was tested on the test set. Table 3 shows the precision (P), recall (R) and the metric precision and recall function values ($F1$) obtained after training this model.

Table 3. Mask RCNN precision (P), recall (R) and $F1$ on the test set

Tabela 3. Precyzja maski RCNN (P), czułość (R) i $F1$ na zestawie testowym

Evaluation Indicators	Precision (P)	Recall (R)	$F1$
Test Set	96.5%	97.4%	96.95%

The data in Table 3 shows that the Mask RCNN network model has a high ability to identify positive samples in the data set, that is, the ability of the model to correctly identify the actual ore as an ore is more prominent. The model is suitable for different particle sizes of ores, and the accuracy of segmentation and identification mainly depends on the number and density of distribution of ores in the image and the accuracy of segmentation is lower for the overlapping and stacked ores. However, the improved model in this paper improves the accuracy of this situation, so the model in this paper has a better performance in the background of complex working conditions and has a certain industrial application potential.

In order to verify the effectiveness of the improved feature pyramid fusion module (MFPN) with the addition of the reverse fusion path in improving the network performance, AP_{50}, AP_{75} and AP_m are used as the evaluation indicators for the improved network structure. Table 4 shows the comparison results of the evaluation indexes before and after the improvement of the model.

The data in Table 4 shows that the detection performance of the network can be effectively improved by using the improved feature pyramid module with the addition of reverse fusion paths. Compared with the original algorithm, the accuracy indexes AP_{50}, AP_{75} and

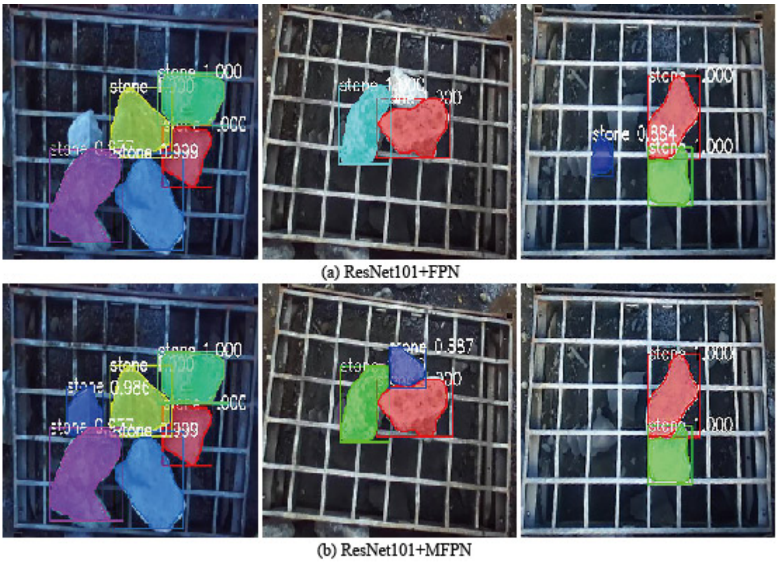
AP_m of ore detection by the improved network are improved, among which AP_{75} is improved by 6.64%, indicating that the use of the improved feature fusion network MFPN can improve the accuracy of Mask RCNN for ore detection.

Table 4. Test results of improved feature fusion network

Tabela 4. Wyniki testów ulepszonej sieci połączonych funkcji

Network Structure	AP_{50}	AP_{75}	AP_m
ResNet101+FPN	0.9662	0.6648	0.6376
ResNet101+MFPN	0.9731	0.7312	0.6827

Figure 7 shows the test results using two feature fusion networks. From Figure 7(a), it can be seen that the detection of ore by the original network model has the situation of missed identification and misidentification. Figure 7(b) is the test result using the improved network model, which can accurately identify and detect small ores that are partially occluded and reduce misidentification. Therefore, compared with the original algorithm, the improved pyramid network with the reverse parallel feature fusion path can improve the detection performance of the network and solve the problems of the network's misidentification of ore, missed recognition and the prediction frame not being close to the edge of the ore.



(a) ResNet101+FPN

(b) ResNet101+MFPN

Fig. 7. Test results of two feature fusion networks
a – ResNet101 + FPN, b – ResNet101 +MFPN

Rys. 7. Wyniki badań dwóch sieci fuzji cech
a – ResNet101 + FPN, b – ResNet101 +MFPN

In order to verify the superiority of the proposed improved Mask RCNN algorithm, the improved Mask RCNN network is compared with the original algorithm and two common target detection algorithms under the same experimental environment and the same data set. AP_{50}, AP_{75} and AP_m are used as the evaluation indicators for this comparative experiment. Table 5 shows the comparison of the experimental results of the four networks.

Table 5. Comparison of four network test results

Tabela 5. Porównanie wyników czterech testów sieci

Network Model	AP_{50}	AP_{75}	AP_m
yolo v3	0.8743	0.6359	0.6273
SSD	0.8132	0.6028	0.5833
Mask RCNN	0.9562	0.6748	0.6476
The improved model in this paper	0.9731	0.7312	0.6827

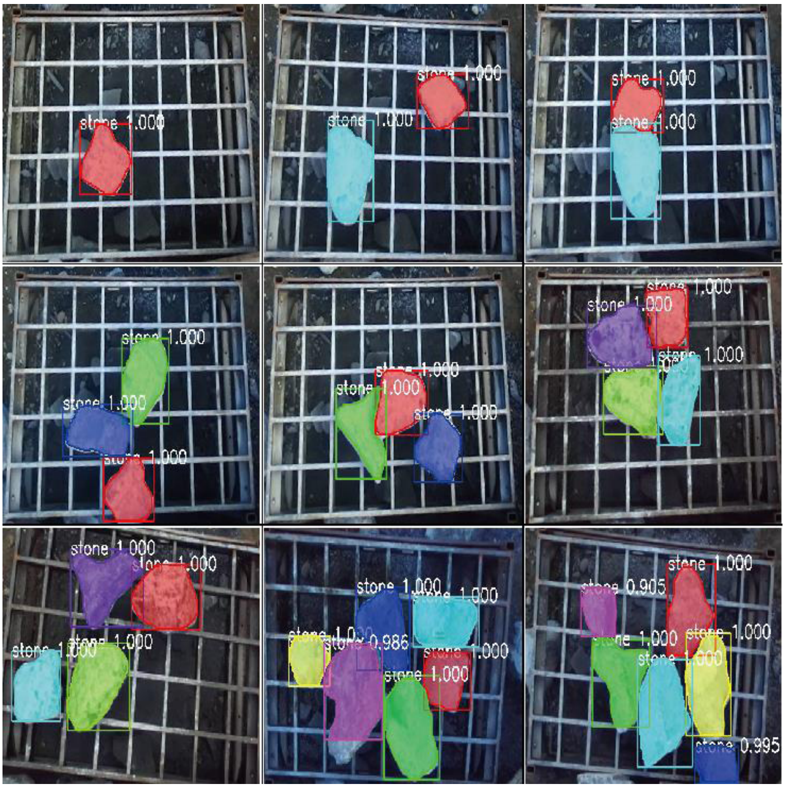


Fig. 8. The improved Mask RCNN ore identification and segmentation results

Rys. 8. Ulepszone wyniki identyfikacji i segmentacji rudy maski RCNN

According to the analysis in Table 5, it can be found that the improved Mask RCNN algorithm has outstanding performance in AP_{50}, AP_{75} and AP_m indicators. Among them, the AP_{75} indicator has the best improvement effect, which is about 6.64% higher than the original algorithm. In order to more intuitively observe the outstanding performance of the improved model in this paper with regard to ore detection, Figure 8 shows the detection results of ore in some actual scenes. It can be seen from the figure that the size of the mask segmented by the network is close to the actual size of the ore. It can be seen that the improved network model has achieved good performance in the detection of ore with different illumination, pose and background.

In the site working environment, the amount of ore on the fence and the stacking situation are more random, and the site lighting environment is also more complex. However, from the results of ore segmentation in the actual scene, the method in this paper can solve the above problems. Therefore, by using the model in this paper, the crushing robot arm can automatically identify and crushed ore from the above complex background, providing a certain automatic basis for the work on site and improving the efficiency of the work.

Conclusions

In order to improve the degree of automation in the ore crushing process, a manipulator equipped with machine vision is used to automatically identify and crush the ore. For the task of ore identification and segmentation under complex working conditions, an improved Mask RCNN target detection instance segmentation algorithm is used. This algorithm is used to train and test the collected ore image. The experimental results show that the trained model can identify ore with an accuracy rate of 96.5%. For ore with different illumination and pose, this method has high accuracy and good robustness under complex working conditions. Therefore, the method proposed in this paper has a certain industrial application potential.

Aiming at the problem that the feature pyramid network FPN in Mask RCNN loses the detailed information of deep features, which leads to the problem of insufficient recognition accuracy of stacked ore, a parallel reverse fusion path is added to the feature fusion network FPN of Mask RCNN. This enriches the detailed information of the deep feature layer, thereby improving the network's ability to extract ore features and the utilization efficiency of feature layers of different scales, effectively improving the problem of the missed detection of stacked ores, and improving the detection accuracy of the network. The experimental results show that the AP_{75} of ore recognition by Mask RCNN with an improved feature fusion network is about 73.12%, which is about 6.64% higher than the original algorithm. In particular, it has improved recognition and segmentation of stacked ore. Therefore, the method proposed in this paper has good application potential for the identification of stacked ores under complex working conditions.

National Natural Science Foundation of China(51464017), "Research on energy consumption of vibration crushing based on multi-scale cohesive particle model" Jiangxi Province Key Research and Development Project (20181ACE50034), "Research and demonstration on key technology of intelligent equipment for medium-low speed selective high energy laminated cone crushing".

REFERENCES

Deng et al. 2022 – Deng, J.L., Dong, L.H., Song, W., Zhao, X.B., Liu, T.M. and Pang, Y.T. 2022. Processing of Seabed Polymetallic Nodule Images Based on Sea-thru and Mask R-CNN. *Mining and Metallurgy Engineering*, pp. 9–13 (*in Chinese*).

Dong et al. 2022 – Dong, W.X., Ling, H.T., Liu, G.D., Hu, Q. and Yu, X. 2022. Review of Deep Convolution Applied to Object Detection Algorithms. *Computer Science and Exploration*, pp. 1025–1042 (*in Chinese*).

Fan et al. 2020 – Fan, L.L., Zhao, H.W., Zhao, H.Y., Hu, H.S. and Wang, Z.A. 2020. Review of Object Detection Based on Deep Convolutional Neural Networks. O*ptical Precision Engineering*, pp. 1152–1164 (*in Chinese*).

Girshick et al. 2014 – Girshick, R., Donahue, J., Darrell, T. and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE conference on computer vision and pattern recognition*, pp. 580–587.

Girshick, R. 2015. Fast r-cnn. *IEEE international conference on computer vision*, pp. 1440–1448.

He et al. 2016 – He, K., Zhang, X., Ren, S. and Sun, J. 2016. Deep residual learning for image recognition. *IEEE conference on computer vision and pattern recognition*, pp. 770–778.

He et al. 2017 – He, K., Gkioxari, G., Dollár, P. and Girshick, R. 2017. Mask r-cnn. *IEEE international conference on computer vision*, pp. 2961–2969.

Huang et al. 2021 – Huang, X.H., Lu, Y., Zhang, R.D. and Dong, S.Q. 2021. Shear crack detection in thermal infrared images of rock based on improved Faster RCNN. *Metal Mine*, pp. 1–10 (*in Chinese*).

Lampinen et al. 2021 – Lampinen, S., Niu, L., Hulttinen, L., Niemi, J. and Mattila, J. 2021. Autonomous robotic rock breaking using a real-time 3D visual perception system. *Journal of Field Robotics* 38(7), pp. 980–1006.

LeCun et al. 2015 – LeCun, Y., Bengio, Y. and Hinton G. 2015. Deep learning. *Nature* 521(7553), pp. 436–444.

Liu et al. 2016 – Liu, W., Anguelov, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C. 2016. SSD: single shot multibox detector. *European conference on computer vision*, pp. 21–37.

Liu et al. 2020 – Liu, X., Wang, H., Jing, H., Shao, A. and Wang, L. 2020. Research on intelligent identification of rock types based on faster R-CNN method. *Ieee Access* 8, pp. 21804–21812.

Liu et al. 2021 – Liu, F.Y., Liu, Y.H., Yang, T.H., Xin, J.C., Zhang, P.H., Dong, X. and Zhang, H.T. 2021. Meticulous evaluation of rock mass quality in mine engineering based on machine learning of core photos. *Chinese Journal of Geotechnical Engineering,* pp. 968–974.

Nie et al. 2020 – Nie, Z.G., Ren, J. and Lu, J.H. 2020. Mask RCNN for detection of ship flow under the background of atomization. *Journal of Beijing Institute of Technology*, pp. 1223–1229 (*in Chinese*).

Redmon et al. 2016 – Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. 2016. You only look once: Unified, real-time object detection. [In:] Las Vegas et al. eds. *IEEE Conference on Computer Vision and Pattern Recognition*, June 27–30, 2016, USA, New York: IEEE, pp. 779–788.

Redmon, J. and Farhadi, A. 2017. YOLO9000: better, faster, stronger. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271.

Redmon, J. and Farhadi, A. 2018. YOLOv3: an incremental improvement. *arXiv preprint arXiv:1804.02767.*

Ren et al. 2015 – Ren, S., He, K., Girshick, R. and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28.

Zhang et al. 2019 – Zhang, S., Gong, Y.H. and Wang, J.J. 2019. Development of Deep Convolutional Neural Networks and Its Application in Computer Vision. *Journal of Computer Science* 42(03), pp. 453–482 (*in Chinese*).

Zhang et al. 2022 – Zhang, J., Nong, C.R. and Yang, Z.Y. 2022. Review of Target Detection Algorithms Based on Convolutional Neural Networks. *Chinese Academy of Weapons and Equipment Engineering*, pp. 37–47 (*in Chinese*).

RESEARCH ON STACKED ORE DETECTION BASED ON IMPROVED MASK RCNN UNDER COMPLEX BACKGROUND

Keywords

feature fusion, Mask RCNN, feature pyramid, ore detection

Abstract

In order to achieve accurate identification and segmentation of ore under complex working conditions, machine vision and neural network technology are used to carry out intelligent detection research on ore, an improved Mask RCNN instance segmentation algorithm is proposed. Aiming at the problem of misidentification of stacked ores caused by the loss of deep feature details during the feature extraction process of ore images, an improved Multipath Feature Pyramid Network (MFPN) was proposed. The network firstly adds a single bottom-up feature fusion path, and then adds with the top-down feature fusion path of the original algorithm, which can enrich the deep feature details and strengthen the fusion of the network to the feature layer, and improve the accuracy of the network to the ore recognition. The experimental results show that the algorithm proposed in this paper has a recognition accuracy of 96.5% for ore under complex working conditions, and the recall rate and recall rate function values reach 97.4% and 97.0% respectively, and the AP75 value is 6.84% higher than the original algorithm. The detection results of the ore in the actual scene show that the mask size segmented by the network is close to the actual size of the ore, indicating that the improved network model proposed in this paper has achieved a good performance in the detection of ore under different illumination, pose and background. Therefore, the method proposed in this paper has a good application prospect for stacked ore identification under complex working conditions.

BADANIA NAD WYKRYWANIEM USYPANEJ (STOS) RUDY W OPARCIU O ULEPSZONĄ MASKĘ RCNN W ZŁOŻONYM TLE

Słowa kluczowe

połączenie funkcji, maska RCNN, piramida funkcji, wykrywanie rudy

Streszczenie

Aby uzyskać dokładną identyfikację i segmentację rudy w złożonych warunkach pracy, do prowadzenia inteligentnych badań wykrywania rudy wykorzystywane są technologie wizji maszynowej i sieci neuronowych, zaproponowano udoskonalony algorytm segmentacji obrazu Mask RCNN (Region Convolutional Neural Networks). Mając na celu rozwiązanie problemu błędnej identyfikacji ułożonych rud, spowodowanego utratą głębokich szczegółów cech podczas procesu ekstrakcji cech z obrazów rudy, zaproponowano ulepszoną sieć wielościeżkową piramidy cech MFPN (*Multipath Feature Pyramid Network*). Sieć najpierw dodaje pojedynczą ścieżkę łączenia funkcji od dołu do góry, a następnie dodaje ścieżkę łączenia funkcji od góry do dołu oryginalnego algorytmu, co może wzbogacić głębokie szczegóły funkcji i wzmocnić połączenie sieci z warstwą funkcji (obiektową) i poprawić dokładność sieci do rozpoznawania rudy. Wyniki eksperymentalne pokazują, że algorytm zaproponowany w niniejszej pracy ma dokładność rozpoznawania na poziomie 96,5% dla rudy w złożonych warunkach pracy, a wartości współczynnika czułości i współczynnika czułości funkcji osiągają odpowiednio 97,4 i 97,0%, a wartość AP75 jest wyższa o 6,84% niż oryginalny algorytm. Wyniki wykrywania rudy w rzeczywistej scenie pokazują, że rozmiar maski podzielonej na segmenty przez sieć jest zbliżony do rzeczywistego rozmiaru rudy, co wskazuje, że ulepszony model sieci zaproponowany w tym artykule osiągnął dobrą efektywność w wykrywaniu rudy przy różnym oświetleniu, ułożeniu i tle. Dlatego zaproponowana w pracy metoda ma dobre perspektywy aplikacyjne do identyfikacji usypanych rud w złożonych warunkach pracy.