Content-Aware Network for Quality Estimation of Copper Scrap Granules

Kaikai Zhao[1,2[0000-0002-8048-5145]] , Zhaoxiang Liu*[1,2[0000-0002-1267-0277]], Kai Wang[1,2[0000-0002-1171-0281]], and Shiguo Lian*[1,2[0000-0003-4308-7049]]

[1] AI Innovation Center, China Unicom, Beijing 100013, China
[2] Unicom Digital Technology, China Unicom, Beijing 100013, China
{zhaokk3,liuzx178,wangk115,liansg}@chinaunicom.cn
*Corresponding author(s)

Abstract. To determine the quality level of copper scrap granules, existing methods have to manually identify all kinds of impurities mixed in copper scrap granules relying on technicians' experience. In this paper, we pioneer a computer vision-based approach called Content-Aware Network (CANet) to estimate the quality of copper scrap granules. Specifically, CANet consists of a visual transformer-based backbone that extracts the semantic features from copper scrap granule images, a multi-layer perception-based neck that explicitly estimates the volume proportion of copper to copper scrap granules and implicitly estimates the counterparts of varieties of impurities and a well-designed head that directly outputs the quality result. Benefiting from our novel architecture and loss functions, CANet can be trained in an end-to-end manner to accurately estimate the quality of copper scrap granules only with the binary annotated images (copper area and non-copper area) without identifying these unknown impurities and their densities in advance. Experiments on real copper scrap granule datasets demonstrate the effectiveness and superiority of our proposed method.

Keywords: Copper scrap granules, Quality level, Visual transformer, Content-Aware.

1 Introduction

Copper is a kind of valuable non-renewable resource on earth, with the rise in demand for copper, copper scrap recycling which could reduce the cost and environmental pollution caused by mining and smelting new copper becomes more and more important in industry [1,2]. Copper scrap granule is an important source of copper scrap, which often contains different kinds of impurities. Fig. 1 shows some copper scrap granule images.

To determine the value and price of copper scrap granules before recycling them, it is necessary to estimate their quality. For convenience, we define the quality estimation of copper scrap granules as the task of predicting the "copper content", i.e., the mass proportion of copper to the whole copper scrap granules.

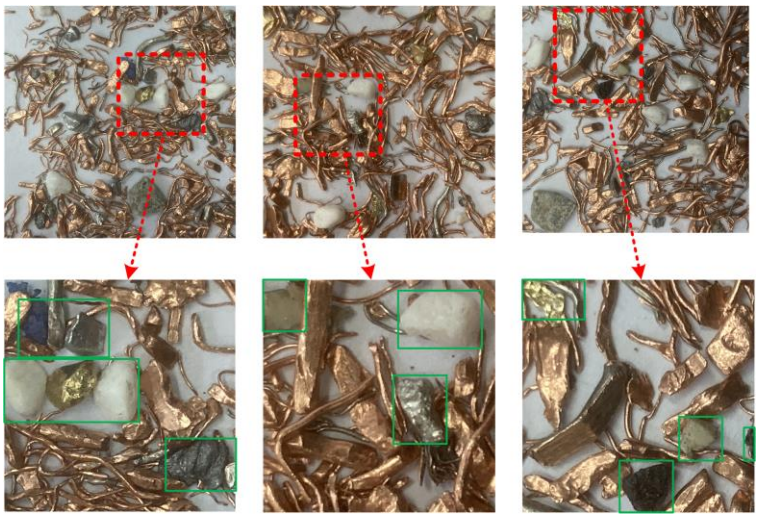

Fig. 1. Three images of copper scrap granules (first row) and their local enlarged view (second row), some impurities are marked with green rectangles. Copper scrap granules contain a variety of unknown impurities with irregular shapes in different proportions besides copper.

In the existing quality estimation method, technicians have to sample many times from the copper scrap granules, then stir over and over again and identify all kinds of impurities mixed in copper scrap granule samples relying on their experience [3]. This inefficient manual method usually results in a heavy financial loss for either the seller or purchaser due to technicians' subjectivity.

Thanks to the great progress in deep learning and computer vision [4,5,6,7], it seems possible to realize automatic quality estimation of copper scrap granules using deep learning and computer vision techniques, as long as you can enumerate all varieties of different impurities and accurately determine their densities, and then estimate the respective volume proportion of all impurities to the whole copper scrap granules using image recognition techniques. However, in the actual setting, when developing a recognition algorithm, it is quite impossible for the engineers, even the professionals to enumerate all varieties of unknown impurities in a large number of copper scrap granules, needless to say, to determine the densities of all various unknown impurities. Therefore, it is challenging but valuable to design an effective vision-based quality estimation method for copper scrap granules.

2 Related work

Because of its great economic value and environmental benefits, scrap metal recycling attracts more and more attention from the computer vision community. The method in [8] proposes a combined architecture to classify the scrap metals and estimate their masses simultaneously from 3D images. Methods in [9,10,11,12,13,14,15] automatically classify scrap metals using classification [4], detection [16] and seg-

mentation [17,18,19] networks. However, the above studies focus on recognizing several limited types of scrap metal pieces which are scattered and much bigger than copper scrap granules, could be discerned relatively easily. What is more, these studies do not care about the contents of impurities, which is very important in our copper content predicting task. These methods of industrial scrap metal classification cannot solve our problem very well, and to our best knowledge, our work is the first to use computer vision and deep learning to estimate the quality of copper scrap granules so far.

Intuitively, two naive methods might be able to complete the quality estimation task. The first method employs a well-trained supervised semantic segmentation model to accurately segment the copper scrap granule images into different impurity areas and copper areas. With the assumption that the area proportion of some stuff to the whole copper scrap granules in the image is equivalent to its volume proportion [20], then the copper content can be calculated as follows:

$$p_c^m = \frac{p_c^v \times \rho_c}{\sum_{i=1}^{I}\left(\rho_i \times p_i^v\right) + p_c^v \times \rho_c} \tag{1}$$

Here p_c^m is the copper content, p_c^v is the volume proportion of copper to the whole copper scrap granules, and ρ_c is the density of copper. p_i^v represents the volume proportion of impurity to the whole copper scrap granules, ρ_i is the density of corresponding impurity, and the subscript I which ranges from 1 to I corresponds to I kinds of impurities in copper scrap granules. However, it is quite difficult for the engineers, even the professionals to enumerate all varieties of unknown impurities in a large amount of copper scrap granules in advance, needless to say, to determine the densities of all various unknown impurities. It makes it impossible to label all types of impurities and train a supervised semantic segmentation model to calculate p_i^v.

The other method estimates the copper content using Eq. (2). Simplifying the multi-class segmentation task into a binary segmentation task (copper area and non-copper area) greatly reduces the workload of labelling without classifying various unknown impurities. With the same assumption in the first method, the copper content can be calculated as follows:

$$p_c^m = \frac{\rho_c \times p_c^v \times V}{M} \tag{2}$$

V and M respectively are the volume and mass of the whole copper scrap granules. To calculate the copper content, V and M need to be known in advance. In the actual setting, M is relatively easy to acquire. However, measuring the volume of the whole copper scrap granules is quite cumbersome because it needs extra equipment and a lot of labor.

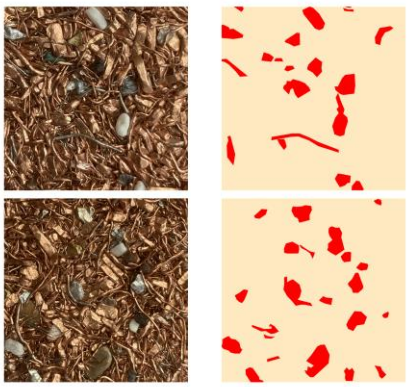

Fig. 2. Two original images (left column) and corresponding binary labeled images (right column). In binary-labeled images, the red areas represent various kinds of impurities with irregular shapes (non-copper), and the yellow areas represent copper.

To address issues of the existing manual methods and naive methods, we pioneer a novel network called Content-Aware Network (CANet) which takes a sequence of copper scrap granule images as input to directly estimate the quality of copper scrap granules. CANet is trained only with the binary annotated images (copper area and non-copper area, as shown in Fig. 2) and does not rely on any other prior information, such as densities of impurities, total mass, and total volume of granules. Our research opens the possibility of realizing accurate and automatic quality estimation of copper scrap granules.

3 Method

3.1 CANet Architecture

As shown in Fig. 3, CANet, which consists of a visual transformer-based backbone, a multi-layer perception-based neck, and a well-designed copper content calculating head, estimates the copper content using input images of copper scrap granules sample.

Firstly, CANet uses the backbone of an unsupervised semantic segmentation network (STEGO [21]) to extract pixel-level semantic features of N images belonging to each sample. This solves the pain point that supervised segmentation networks have to enumerate and label all types of impurities in advance. We discard the cluster module and the CRF module of STEGO and fine-tune the backbone to yield ViT features for the estimation of copper content.

Secondly, CANet uses a simple multi-layer perception and a softmax layer to generate the area proportion vector for each image, the softmax operation constrains the sum of the area proportion vector to be 1 and then computes the integrated area proportion vector for the input sample using a row-wise average pooling layer connected with a concat-layer. The integrated area proportion vector explicitly describes the area

proportions of copper and various impurities in the copper scrap granule sample. The length of each vector is I+1, and the first element (the yellow box in Fig. 3) in the vector represents the area proportion of copper and the others (the green boxes) indicate the area proportions of various impurities. This novel design of explicit area proportion vector facilitates the network to learn features and distributions of various impurities from a large amount of data. Here, the same assumption in naive methods is also adopted.

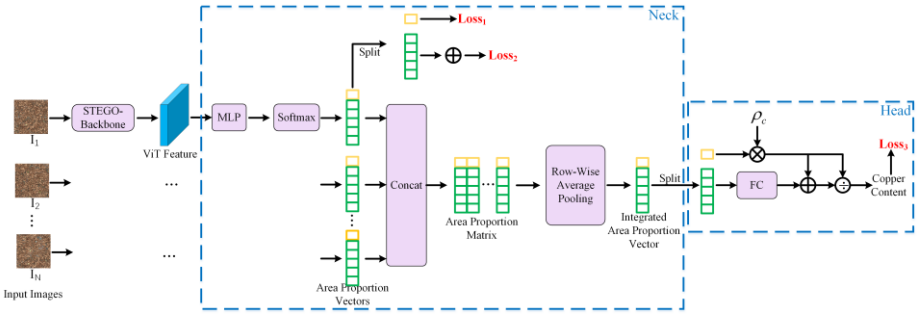

Fig. 3. Content-Aware network architecture. The area proportion vector describes the area proportions of copper and various impurities in each image and the integrated area proportion vector describes the respective average area proportions of copper and various impurities in the copper scrap granules sample.

Finally, CANet uses a well-designed head to implicitly learn the densities of various impurities and explicitly calculate the copper content. Our head consists of a simple fully connected layer with one output and some basic mathematical operations. In Fig. 3, we extract the first element of the integrated area proportion vector and multiply it by the density of copper ρ_c to get the relative mass of copper and feed the remaining elements of the vector into a single-layer fully connected network to obtain the sum of the relative mass of all impurities. Then, the relative mass of copper is divided by the total relative mass of copper scrap granules sample to get the final copper content. It should be noted that this single fully connected layer is responsible for learning the densities of various impurities from a large amount of data, which implicitly solves the problem that it is difficult to accurately determine the densities of various impurities in the naive method.

When training the network, we only need to label the copper content of each sample and the area proportion of copper which can be easily computed using annotated binary images (copper area and non-copper area), without struggling to label the distributions and contents of various impurities. During model inference, for each copper scrap granule sample, N input images are captured after stirring the sample N times. These images are then processed by the CANet, and the copper content is calculated. CANet estimates the copper content only using input images and does not rely on any other prior information, such as densities of impurities, total mass, and total volume of the sample.

3.2 Multi-task Loss

With the limited annotation work, we propose a multi-task loss to jointly learn the copper content. Separating the first element of the area proportion vector from other elements, $Loss_1$ (Eq. (3)) and $Loss_2$ (Eq. (4)) are used to optimize the backbone and the neck by learning the area proportion of copper and the sum of the area proportions of all impurities.

$$Loss_1 = \frac{1}{N}\sum_{j=1}^{N}\left(p_{c_j}^a - \hat{p}_{c_j}^a\right)^2 \tag{3}$$

$$\begin{aligned} Loss_2 &= \frac{1}{N}\sum_{j=1}^{N}\left(\left(1-p_{c_j}^a\right)-\sum_{i=1}^{I}\hat{p}_{i_j}^a\right)^2 \\ &= \frac{1}{N}\sum_{j=1}^{N}\left(\left(1-p_{c_j}^a\right)-\left(1-\hat{p}_{c_j}^a\right)\right)^2 \\ &= \frac{1}{N}\sum_{j=1}^{N}\left(p_{c_j}^a - \hat{p}_{c_j}^a\right)^2 \end{aligned} \tag{4}$$

Where, $\hat{p}_{c_j}^a$ is the prediction of copper area proportion in j_{th} image, and $p_{c_j}^a$ is the labelled copper area proportion for j_{th} image. $\hat{p}_{i_j}^a$ is the prediction of area proportion of i_{th} impurity in j_{th} image. It is clear that $Loss_1$ and $Loss_2$ are the same, this is because of the softmax used before outputting the area proportion vectors. Then, for the copper content, $Loss_3$ is used to optimize the entire network.

$$Loss_3 = \left(p_c^m - \hat{p}_c^m\right)^2 \tag{5}$$

Where $\hat{p}_c^m$ is the prediction of copper content. Finally, the sum of $Loss_1$, $Loss_2$ and $Loss_3$ jointly optimize the parameters of the CANet, as shown in Eq. (6).

$$L = \frac{1}{N}\lambda_a\sum_{j=1}^{N}\left(p_{c_j}^a - \hat{p}_{c_j}^a\right)^2 + \lambda_m\left(p_c^m - \hat{p}_c^m\right)^2 \tag{6}$$

Where λ_a and λ_m control the balance of the learning signals, and in practice, we find that $\lambda_a = 0.2$ and $\lambda_m = 0.8$ work well. The copper content guides the network to learn the area proportion vectors, which in turn helps the calculation of copper content. After training, our CANet can directly calculate the copper content only using images of the input copper scrap granules sample.

4 Experiments

In this section, we evaluate the proposed CANet on real copper scrap granule data. We build a copper scrap granule dataset using real copper scrap granules and conduct extensive experiments on the dataset to verify the effectiveness of the CANet.

4.1 Datasets

Based on recycled copper scrap granules of different quality, we get 1100 samples using simple sampling without replacement. For each copper scrap granule sample, we stir it 64 times and collect corresponding images, which means $N <= 64$ in Fig. 3. We do not limit the specific implementation method of stirring as long as most of the copper granules and impurities have the same opportunity to appear on the top surface and can be captured by the camera.

When generating ground truth for each sample, we only label a copper content scalar and 64 binary semantic segmentation images. The copper content of each sample is labelled and checked by 10 professional technicians. The copper area proportion of each image is automatically calculated by the computer based on its corresponding binary segmentation image. Finally, We split the entire dataset into 800 samples as the training set and 300 samples as the test set. The copper scrap granules dataset contains 51200 training images and 19200 test images.

4.2 Experimental Results

To verify the effectiveness of our proposed method, we compare the copper area proportion and copper content estimation results with ground truth. Furthermore, we discuss the hyperparameter search process and analyze the impacts of the backbone and decoder through ablation studies.

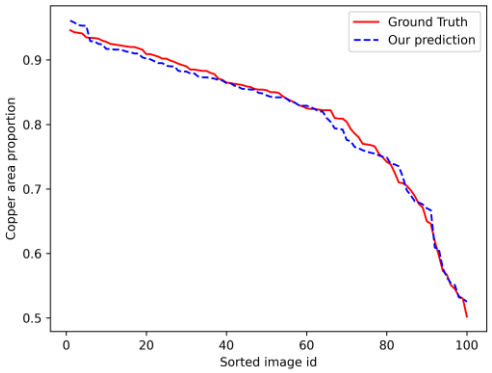

Fig. 4. Copper area proportion results of 100 images. Without directly predicting segmentation results, CANet can accurately calculate the area proportion of copper. For convenience, we sort images in descending order according to the real area proportion.

Copper Area Proportion Estimation. To verify the effectiveness of the backbone and neck in CANet, we first analyze the accuracy of copper area proportion in the area proportion vector. On random 100 images belonging to different test samples, we calculate the area proportion of copper using the output of the Content-Aware neck and compare it with ground truth. The comparison results are shown in Fig. 4. It demonstrates that the neck of CANet is reliable, and the ViT-based backbone can extract effective features to distinguish copper and non-copper.

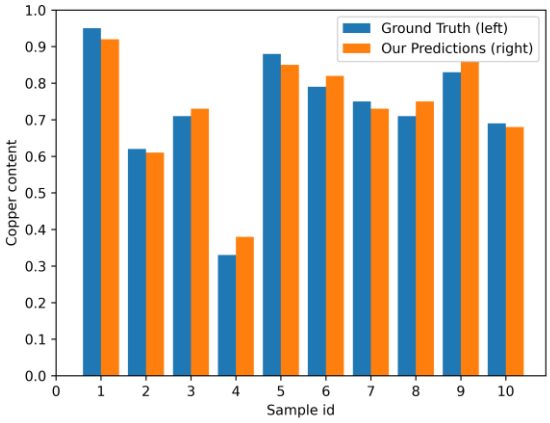

Fig. 5. Copper content results predicted by our best CANet and corresponding ground truth of random 10 test samples.

Copper Content Estimation. Our best CANet achieves a 2.87% mean absolute error (MAE) of copper content on test samples, and we randomly select 10 test samples to demonstrate the copper content results predicted by our best CANet and corresponding ground truth in Fig. 5. The predicted copper content on test samples is fairly close to ground truth, demonstrating the utility of our CANet on the copper granules dataset.

The experimental results show that even if we do not know the prior information such as types of impurities, contents of impurities, total mass, and total volume of samples, CANet can still accurately predict the copper content scrap granules by learning and perceiving in real dataset.

Hyperparameters Analysis. To analyze the impacts of λ_a and λ_m in the loss function (Eq. (6)), we first set both N and I+1 at their maximum values: N=64, I+1=14. Then, by adjusting λ_a ($\lambda_m = 1 - \lambda_a$), we train CANet and calculate the MAE of copper content on the test set. The corresponding results are presented in Table 1. It is evident that $Loss_1$ plays an important role in model convergence, and the intermedi-

ate supervision signal helps model learning. The smallest MAE is achieved when $\lambda_a = 0.2$ and $\lambda_m = 0.8$.

Table 1. Copper content errors of different λ_a and λ_m when N=64, I+1=14.

λ_a	0	0.1	0.18	0.2	0.22	0.25	0.3	0.5	1
MAE (%)	13.1	3.55	2.93	**2.87**	2.92	3.3	4.01	4.36	48.2

Secondly, fixing $\lambda_a = 0.2$ and setting N to its maximum value (N = 64), we train different variants of CANet as we adjust the length of the area proportion vector to search the optimal value for I+1. The test results are presented in Table 2. When I+1=4, the copper content error reaches 14.3%, and it gradually decreases as the length increases. At I+1=10, CANet achieves the minimum error rate of 2.87%. However, there is no improvement in the estimation accuracy of copper content if $I+1>10$, suggesting that our dataset contains about 9 different impurities. When the length of the area proportion vector is less than the number of impurity types, CANet can not represent the distribution of all types of impurities very well, resulting in less accurate copper content prediction.

Table 2. Copper content errors of different I+1 when $\lambda_a = 0.2$, N = 64.

I+1	4	8	9	10	11	14
MAE(%)	14.3	6.26	4.33	**2.87**	2.87	2.87

Finally, with $\lambda_a = 0.2$ and I+1 = 10, we train different variants of CANet as we adjust the number of input images for each sample. The test results are shown in Table 3. If $N < 32$, the model fails to extract sufficient information and features from input samples, resulting in a sharp decline in accuracy. The proposed minimum number of input images is 32.

Table 3. Copper content errors of different N when $\lambda_a = 0.2$, I+1 = 10.

N	64	40	32	28	16	8
MAE(%)	2.87	2.87	**2.87**	3.1	10.21	36

Ablation Study on Backbone. To better understand the influence of CANet's backbone, we use several SOTA backbones to extract the features of input images (replac-

ing only the STEGO-Backbone module in Fig. 3). The final MAE results for copper content on the test set are presented in Table 4.

Table 4. Copper content errors of different backbone.

Back-bone	STEGO-ViT-Base	STEGO-ViT-Small	ResNet50	ResNet101
MAE(%)	**2.87**	3.41	13.32	11.57

Both ViT-based backbones and CNN-based backbones listed in Table 4 are pre-trained on ImageNet and fine-tuned on our training data. STEGO-ViT-B/8 and STEGO-ViT-S/8 respectively refer to the STEGO backbone with ViT-Base and ViT-Small (8×8 patches) in [21]. The results indicate that the ViT-based STEGO backbone significantly outperforms ResNet in our copper content estimation task.

Ablation Study on Decoder. At the beginning of our research, we designed a simple and intuitive network as a baseline. Following the same backbone in Fig. 3, this baseline network uses a fully convolution-based decoder to directly predict the copper content from concatenated features of the input images. The decoder integrates N ViT features using five ResNet50 blocks (each block consists of 1×1 conv, 3×3 conv, and 1×1 conv) and outputs the final copper content using a single convolution kernel. The size of the input features is 256×256×1024, the output size of the five ResNet50 blocks is respectively 128×128×2048, 64×64×2048, 32×32×2048, 16×16×1024, and 8×8×512 and the size of the single convolution kernel is 8×8×512. This baseline network yields an MAE of 20.63% on our dataset, indicating that directly using the CNN network to predict copper content is unreliable.

5 Conclusions

In the field of copper scrap granules recycling, we pioneer a ViT-based network that can accurately estimate the copper content only with the binary labelled images without identifying these unknown impurities and their densities. Experiments on real copper scrap granule datasets verify the effectiveness and superiority of the proposed network. Our method makes it possible to realize automatic quality estimation of copper scrap granules and exhibits a big potential to apply to the quality estimation of other scrap metals.

Acknowledgments. This work is supported by the Funding of Beijing Association of Science and Technology Outstanding Engineer Growth Plan[1].

[1] The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Glöser S, Soulier M, Tercero Espinoza L A. Dynamic analysis of global copper flows. Global stocks, postconsumer material flows, recycling indicators, and uncertainty evaluation[J]. Environmental science & technology, 2013, 47(12): 6564-6572.
2. Gomez F, Guzmán J I, Tilton J E. Copper recycling and scrap availability[J]. Resources Policy, 2007, 32(4): 183-190.
3. Chenglong Z, Yujia C, Jingwei W, et al. Recovery of copper from bio-leaching solutions of waste printed circuit boards waste by ion exchange[C]//2010 International Conference on Digital Manufacturing & Automation. IEEE, 2010, 2: 138-140.
4. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
5. Chen Y, Li Y, Kong T, et al. Scale-aware automatic augmentation for object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 9563-9572.
6. Mehta D, Skliar A, Ben Yahia H, et al. Simple and efficient architectures for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2628-2636.
7. Chavan A, Shen Z, Liu Z, et al. Vision transformer slimming: Multi-dimension searching in continuous optimization space[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4931-4941.
8. Díaz-Romero D J, Van den Eynde S, Sterkens W, et al. Simultaneous mass estimation and class classification of scrap metals using deep learning[J]. Resources, Conservation and Recycling, 2022, 181: 106272.
9. Diaz-Romero D, Sterkens W, Van den Eynde S, et al. Deep learning computer vision for the separation of Cast-and Wrought-Aluminum scrap[J]. Resources, Conservation and Recycling, 2021, 172: 105685.
10. Chen S, Hu Z, Wang C, et al. Research on the process of small sample non-ferrous metal recognition and separation based on deep learning[J]. Waste Management, 2021, 126: 266-273.
11. Smirnov N V, Rybin E I. Machine learning methods for solving scrap metal classification task[C]//2020 International Russian Automation Conference (RusAutoCon). IEEE, 2020: 1020-1024.
12. Gao Z, Sridhar S, Spiller D E, et al. Applying improved optical recognition with machine learning on sorting Cu impurities in steel scrap[J]. Journal of Sustainable Metallurgy, 2020, 6: 785-795.
13. Huang B, Liu J, Zhang Q, et al. Identification and classification of aluminum scrap grades based on the Resnet18 model[J]. Applied Sciences, 2022, 12(21): 11133.
14. Han S D, Huang B, Ding S, et al. Toward fully automated metal recycling using computer vision and non-prehensile manipulation[C]//2021 IEEE 17th International Conference on Automation Science and Engineering (CASE). IEEE, 2021: 891-898.
15. Zhang Y, Ye J, Chen X, et al. Study on scrap steel classification using deep learning[C]//Eighth Symposium on Novel Photoelectronic Detection Technology and Applications. SPIE, 2022, 12169: 3172-3178.
16. Li C, Xu C, Cui Z, et al. Feature-attentioned object detection in remote sensing imagery[C]//2019 IEEE international conference on image processing (ICIP). IEEE, 2019: 3886-3890.

17. Hu X, Yang K, Fei L, et al. Acnet: Attention based network to exploit complementary features for rgbd semantic segmentation[C]//2019 IEEE International conference on image processing (ICIP). IEEE, 2019: 1440-1444.
18. Wang Y, Zhou Q, Liu J, et al. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation[C]//2019 IEEE international conference on image processing (ICIP). IEEE, 2019: 1860-1864.
19. Yu C, Xiao B, Gao C, et al. Lite-hrnet: A lightweight high-resolution network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10440-10450.
20. Zhang C, Li T, Zhang W. The detection of impurity content in machine-picked seed cotton based on image processing and improved YOLO V4[J]. Agronomy, 2021, 12(1): 66.
21. Hamilton M, Zhang Z, Hariharan B, et al. Unsupervised semantic segmentation by distilling feature correspondences[J]. arXiv preprint arXiv:2203.08414, 2022.