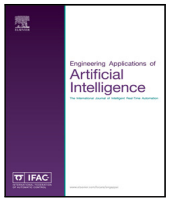
Engineering Applications of Artificial Intelligence 123 (2023) 106241

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Classification and rating of steel scrap using deep learning

Wenguang Xu [a], Pengcheng Xiao [a,d,e,*], Liguang Zhu [b,e,**], Yan Zhang [a,c], Jinbao Chang [d], Rong Zhu [e], Yunfeng Xu [c]

[a] College of Metallurgy and Energy, North China University of Science and Technology University, Tangshan 063210, Hebei, China
[b] School of Materials Science and Engineering, Hebei University of Science and Technology, Shijiazhuan 050018, Hebei, China
[c] Collage of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018 Hebei, China
[d] HBIS Group Co., Ltd (HBIS), Hebei, 050000, China
[e] School of Metallurgical and Ecological Engineering, University of Science and Technology Beijing, Beijing 100083, China

ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning
Intelligent classification and rating
Attention mechanism
Steel scrap recycling

ABSTRACT

To address the issues of high human interference and low efficiency in traditional manual methods for classifying and rating steel scrap, we propose the development of CSBFNet, a deep learning-based model for multi-category steel scrap classification and rating. Firstly, we built a 1:3 physical model of steel scrap quality inspection to simulate the unloading of a truck. We used a high-resolution vision sensor to capture the morphological characteristics of various steel scraps. Next, we trained the CSBFNet model using this data to obtain characteristic information for classifying and judging various types of scrap steel. Finally, we tested and improved the CSBFNet model at a Chinese steel mill. The results demonstrate that the model can effectively determine the automatic rating for different grades of scrap. The average accuracy rate of all types of steel scrap reaches 92.4% for the full category, with an mAP of 90.7%. Compared to traditional artificial quality detection methods, it has clear advantages in accuracy and fairness. This model solves the problem of evaluating the quality of steel scrap in the recycling process.

1. Introduction

Steel scrap (Nechifor et al., 2020) is a recyclable green resource that can replace iron ore as a raw material for steel-making and continuous casting production (Ma and Wang, 2021), reduce coking, sintering, and iron-making processes have high environmental protection value, and is an important raw material for modern steel (Sun et al., 2018). Currently, there are two main routes for steel production, namely the Blast Furnace-Basic Oxygen Furnace (BF-BOF) route and the Electric Arc Furnace (EAF) route (Sun et al., 2020). Since EAF manufacturing technology can use any kind of scrap as raw material and can remanufacture any new steel product, scrap has become the greenest material in the steel industry (Fan and Friedmann, 2021). According to the World Steel Association, global crude steel production in 2021 is 1952 million tons, an increase of 3.8% from the previous year. Of this, BF-BOF steel production is essentially flat year over year at 1.381 billion tons, and EAF steel production increases by 14.4% year over year to 563 million tons. For U.S. steel production, about two-thirds of annual production comes from scrap, including production and forming scrap, process scrap, and post-consumer scrap (Zhu et al., 2019). The share of steel production in the EU is estimated to be about 54% from scrap (Passarini et al., 2018; Fellner et al., 2018), while in Japan more than 20% of crude steel is made from scrap (Omura et al., 2016). With the further implementation of China's carbon peaking and carbon neutral strategy, the Chinese steel industry has started to increase its scrap consumption after the US, EU, and Japan (Lin et al., 2021; Yu et al., 2020). EAF has great advantages over BF-BOF in terms of energy saving and emission reduction, which has led to an increase in the world demand for scrap, and the increase in scrap consumption and exports has led to increasing pressure on recycling and quality inspection.

At present, most steel enterprises determine the scrap grade mainly by visual inspection and caliper measurement by quality management personnel, which has many problems such as high danger, low accuracy of grading, and questionable fairness. With the gradual increase of the world's scrap recovery and retention in the future, the challenges of quality evaluation in the steel enterprises' scrap acceptance process will be further aggravated. At present, the global steel industry has less research on the intelligent classification and rating of steel scrap and only involves methods related to the classification and detection of non-ferrous metals and alloying elements. In detecting and separating nonferrous metals. Mesina et al. proposed the combination of electromagnetic sensors and dual-energy X-ray transmission sensors as a new method for automatically classifying scrap nonferrous metals

* Corresponding author at: College of Metallurgy and Energy, North China University of Science and Technology University, Tangshan 063210, Hebei, China.
** Corresponding author.
E-mail addresses: xiaopc@ncst.edu.cn (P. Xiao), zhuliguang@ncst.edu.cn (L. Zhu).

https://doi.org/10.1016/j.engappai.2023.106241
Received 6 January 2023; Received in revised form 21 March 2023; Accepted 28 March 2023
Available online 12 April 2023

and classifying metals based on differences in electrical conductivity and atomic number (Mesina et al., 2007). Koyanaka et al. Proposed a new method based on differences in the apparent density and three-dimensional shape of metal parts using linear lasers and associated optical cameras to measure the width, height, volume, and projected area of irregular metal parts, which in turn separates a mixture of the cast aluminum, wrought aluminum, and magnesium fragments from steel scrap (Koyanaka and Kobayashi, 2010). Penumuru et al. proposed a general method for automatic material identification by applying machine vision and machine learning techniques in Industry 4.0 to achieve the identification of aluminum, copper, and mild steel and classification (Penumuru et al., 2020). Kashiwakura et al. applied laser-induced breakdown spectroscopy (LIBS) for copper content steel scrap classification method for fast and accurate identification by spectral wavelength (Kashiwakura and Wagatsuma, 2013). Gao et al. used a combination of optical recognition techniques and deep learning and used convolutional neural networks (CNNs) to optimize the detection of copper metal minimizing effects such as surface inhomogeneity (Gao et al., 2020). As for the task of automatic metal scrap sorting, Smirnov et al. Proposed a method for automatic cropping of scrap metal images using a convex quadrilateral in a deep learning model and applied the method to photographs of railroad vehicles containing scrap metal (Smirnov and Rybin, 2020).

It is noteworthy that several steel companies have started to develop intelligent scrap rating systems independently or cooperatively in the last two years, which have made useful explorations for scrap procurement management. However, there are few reports at home and abroad on the experimental research methods for scrap quality inspection and the scientific optimization of the algorithm model. In this paper, we construct an intelligent scrap quality inspection network model based on the attention mechanism in deep learning and conduct an experimental study on two self-built scrap data sets to realize the intelligent quality inspection of scrap category and thickness.

1.1. Deep learning in recycling

Using deep learning and computer vision to classify recycling waste may be a more efficient way to handle it. Recycling requires the sorting of solid waste, which is complex and expensive. To simplify this process using an automatic identification or sorting system based on target detection for classification recognition can help to reduce the hazards and repetitive manual sorting work in recycling management. Zhang et al. Classified household waste by computer vision and automatically classified it according to four categories of regulations. A novel two-stage waste recognition and retrieval algorithm (Waste Recognition-Retrieval algorithm, W2R) is proposed (Zhang et al., 2021). Ramsurun et al. studied proposed a deep learning approach that uses computer vision to automatically identify the type of waste and classify it into five categories: plastic, metal, paper, cardboard, and glass (Ramsurrun et al., 2021). Hu et al. Proposed a study on the application of automatic waste classification based on tensor flow and the open-source image library OpenCV for automatic waste recycling (Hu and Zhang, 2021). Bobulski et al. Created and studied automatic classification techniques to improve the overall efficiency of the recycling process (Bobulski and Kubanek, 2020). Gupta et al. proposed an artificial intelligence (AI)-based automatic solution for classifying other waste types from metal, plastic, and glass products (Gupta et al., 2019). Liu et al. Used deep learning theory and methods for the qualitative classification of waste textiles based on near-infrared spectral analysis to accomplish the automatic identification of fabric components during the classification process (Liu et al., 2020). Majchrowska et al. provided an analysis of more than 10 existing waste datasets, as well as a brief but constructive review that reveals many current problems (Majchrowska et al., 2022). Rong et al. applied two different convolutional neural network structures to walnut images to automatically segment the images and detect natural foreign objects (e.g., fleshy leaf debris, dried leaf debris, and gravel dust) and man-made foreign objects (e.g., paper scraps, packaging materials, plastic scraps, and metal parts) of different sizes (Rong et al., 2019). Deep learning and computer vision have been widely used in the field of waste recycling and have been demonstrated accordingly (Chu et al., 2018; Bircanoğlu et al., 2018; Ruiz et al., 2019; Huang et al., 2020).

1.2. Related works of classification rating

Applying deep learning and computer vision-related techniques for multi-category scrap with different characteristics is beneficial for developing scrap recycling technology and exploiting the application of artificial intelligence in metallurgy. The application of deep learning for the classification rating of related objects has been widely used in other industries, while steel scrap classification grading has been less studied. Litjens et al. summarized the application of deep learning in the medical field for image classification, and object detection, which includes nerve, retina, lung, finger pathology, breast, heart, abdomen, musculoskeletal, etc (Litjens et al., 2017). Lu et al. studied image aesthetics assessment with unified feature learning and classifier training methods (Lu et al., 2015). Existing methods build on handcrafted or generic image features and develop machine learning and statistical modeling techniques using training examples. A novel deep neural network approach is used to implement unified feature learning and classifier training for estimating image aesthetics. Hou et al. proposed Mobile Crack, an adaptive lightweight CNN model for fast object classification on asphalt pavement crack images, solving the classification tasks of cracks, sealed cracks, pavement markers, and pavement matrices (Hou et al., 2021a). Kim et al. applied deep learning techniques in the field of ground-penetrating radar (GPR) for detecting buried subsurface objects, successfully classifying cavities, pipes, manholes, and subsoil backgrounds with very low false alarm errors (Kim et al., 2021). Patel et al. applied deep convolutional neural networks (CNN) directly to regions of interest (ROI) in the radar spectrum to achieve accurate classification of different objects in the scene and to the radar classification deep learning challenges and introduces a set of novel mechanisms that can significantly improve object classification performance compared to simpler classifiers, enhancing the classification capability of automotive radar sensors (Patel et al., 2019). Deep convolutional neural networks have been successful in facial expression datasets but using general image classification networks as backbones limits their adaptability to the Facial Expression Recognition (FER) task. To address this, Li et al. propose an appropriate and lightweight Facial Expression Recognition Network, Auto-FERNet, that is automatically searched by a Neural Architecture Search (NAS) model directly on FER dataset, and a relabeling method based on Facial Expression Similarity (FES) to alleviate uncertainty in FER datasets in the wild (Li et al., 2021).

1.3. Related work deployed on edge devices

Deploying neural networks on edge devices is becoming increasingly popular due to the need for real-time processing of data and the benefits of reducing latency and bandwidth usage. Lyu et al. Propose a method for designing efficient portrait parsing models suitable for edge computing or terminal devices. The method uses a multi-objective reinforcement learning-based neural architecture search to balance accuracy, parameters, FLOPs, and inference latency, and achieves state-of-the-art performance with low resource consumption and high real-time performance (Lyu et al., 2023). Bhardwaj et al. introduced a lightweight prediction-time unsupervised DNN adaptation technique to improve the model's prediction accuracy on noisy data by retuning batch normalization (BN) parameters (Bhardwaj et al., 2022). It is challenging to achieve accurate detection in real time under complex unstructured field environments. In this context Qi et al. propose a

novel lightweight convolutional neural network for medicinal chrysanthemum detection (MC-LCNN) (Qi et al., 2022). Li et al. propose a new architecture called T2IGAN, which integrates text and image features using a lightweight transformer, and achieves remarkable results on typical datasets (Li et al., 2022). Yang et al. aim to minimize the system utility of the MEC network by jointly optimizing the offloading decision and bandwidth allocation problems (Yang et al., 2022). Zhang et al. propose Edge-as-a-Service (EaaS) to enable distributed intelligence. With rapid advancements in edge AI, the time has come to realize intelligence downloading onto edge devices (e.g., smartphones and sensors) (Zhang et al., 2022). To materialize this version Huang et al. propose a novel technology called in-situ model downloading, that aims to achieve transparent and real-time replacement of on-device AI models by downloading from an AI library in the network (Huang et al., 2022). Shao et al. propose a task-oriented communication framework for edge video analytics, where multiple devices collect the visual sensory data and transmit the informative features to an edge server for processing (Shao et al., 2022). Lyu et al. proposes deploying spectral graph convolutional networks (GCNs) on memristive crossbars and an acceleration method that achieves high accuracy and reduces memristor number (Lyu et al., 2022).

1.4. Commonly used steel scrap sorting techniques/methods

Magnetic Separation: This technology uses magnets to separate ferrous metals from non-ferrous metals. Steel scrap is passed through a conveyor system that has a powerful magnet attached to it. The magnet attracts ferrous metals (like steel) and separates them from non-ferrous metals (like aluminum) (Zhang and Forssberg, 1998).

Eddy Current Separation: This method uses a high-frequency magnetic field to separate non-ferrous metals from ferrous metals. The steel scrap is passed through a conveyor system that has an eddy current separator attached to it. The separator generates an electrical current that creates a magnetic field, which interacts with the non-ferrous metals and causes them to be separated from the steel (Jujun et al., 2014).

X-ray Fluorescence: This technology uses X-rays to identify the elemental composition of the steel scrap. X-ray fluorescence detectors measure the energy of the X-rays that are emitted when the steel is exposed to a primary X-ray beam. By analyzing the energy of the emitted X-rays, the system can determine the elemental composition of the steel and sort it accordingly (Weiss, 2011).

Laser-induced Breakdown Spectroscopy (LIBS): This technology uses a laser to vaporize a small amount of the steel scrap, creating a plasma that emits light. The light is then analyzed by a spectrometer to determine the elemental composition of the steel. This method is fast and accurate and can be used to sort steel scrap based on its composition (Van den Eynde et al., 2022).

Near-Infrared (NIR) Spectroscopy: This method uses light in the near-infrared range to identify the chemical composition of the steel scrap. The steel scrap is passed through a conveyor system that has an NIR sensor attached to it. The sensor measures the reflectance of the light off the steel and uses that information to determine the chemical composition of the steel (Cesetti and Nicolosi, 2016).

Classification methods used in previous studies can be used for single-category or large target classes, while it is difficult to classify multi-category, overlapping, and small targets, while irregular, multi-category, overlapping, and multi-scale targets are typical of steel scrap. Therefore, this paper proposes a new deep learning method to classify and rate multi-category and multi-scale steel scrap, and by comparing six improved models with the classical two-stage target detection model Faster R-CNN and one-stage models YOLOv4, YOLOv5 and the latest YOLOv7, the CSBFNet model is evaluated to outperform the other models. Secondly, this paper presents the performance of the combined CSPNet + SE + BiFPN network in classification rating, laboratory scrap dataset, and field scrap dataset.

Table 1
HK_L data set number of labels per category.

Label category	Number of HK_L tags	HK_L training number of tags
<3 mm	184	162
3-6 mm	654	580
>6 mm	9596	8572
Galvanized	730	642
Greasy dirt	392	359
Paint	252	226
Inclusion	784	699

The innovative contributions of this paper are.

(1) The intelligent detection of multi-category and multi-scale scrap classification and grading has been realized. Compared with manual classification and grading, CSBFNet model has significant advantages in accuracy and efficiency.

(2) Innovative incorporation of SE attention mechanism and EfficientDet fusion network in feature extraction and feature fusion to achieve a significant improvement on the classification rating effect of the model. The model is also validated in the field, which provides the possibility of an unmanned intelligent quality inspection system at a later stage.

The paper is structured as follows: In the second part, the material collected is outlined, divided into a laboratory scrap dataset and field scrap dataset. In the third part, the scrap steel intelligent quality inspection model architecture as well as related algorithms and the data pre-processing algorithm section are presented. In the fourth part, the experimental environment and model evaluation criteria are described, and the experimental data and results of the model after training using both scrap datasets are analyzed. In the fifth and sixth parts, the envisioned industrial applications are described, as well as a summary of the paper and a discussion of future work.

2. Materials

2.1. Laboratory data sets

The artificial intelligence quality inspection technology of steel scrap is still in the development stage at home and abroad, and there is no public scrap image data set. To simulate the scrap recycling site, a total of 20 tons of a scrap of various categories with mass distribution from 1 to 90 kg, mainly including 7 categories of less than 3 mm, thickness 3–6 mm, thickness more than 6 mm, galvanized parts, painted parts, oiled parts, and impurities, were purchased from the Chinese scrap market for simulating the scrap unloading process. A 1:3 physical model of steel scrap unloading was established in the laboratory, a Hikvision vision sensor was used to collect image information, and the HK_L dataset was established. HK_L steel scrap dataset, divided into 7 labeled dataset images. The dataset is labeled with Label Image software for image labeling, the labeling method is a rectangular label box, and PASCAL VOC is used as the labeling format for the dataset. The training set and validation set are divided according to the ratio of 9:1. The number of labels in each specific category is shown in Table 1.

2.2. Steel mill dataset

To further verify the effectiveness and generalization ability of the model, a steel scrap quality inspection system is built in the steel scrap recycling area of a Chinese steel enterprise, using three Hikvision intelligent dome cameras to capture the steel scrap unloading images in real scenarios. After conducting a site survey and researching steel enterprise scrap recycling standards, the model was further optimized and adjusted. The survey revealed that galvanized and oiled parts are classified as a single category because they are secondary. Additionally, the purchase price of scrap is affected by its length, resulting in the inclusion of extra-long parts measuring 1.2–1.5 m and 1.5–2 m, as well

Table 2
HK_T data set number of labels by category.

Label category	Number of HK_T tags	HK_T training number of tags
Airtight	3330	3022
Scattered	2589	2309
Inclusion	1077	971
Ungraded	2898	2635
Overlength (1.2–1.5 m)	1482	1340
Overlength (1.5–2 m)	207	195
<3 mm	312	278
3–6 mm	36138	32437
>6 mm	86688	78271
Crane	2136	1931
Boxcar	3336	2994

Table 3
HK_L, HK_T datasets.

Type	HK_L data set	HK_T data set
Number of pictures	278	3387
Total number of tags	12592	140193
Number of training set images	250	3048
Number of training set labels	11240	126383
Number of validation set images	28	339
Number of validation set tags	1352	13810

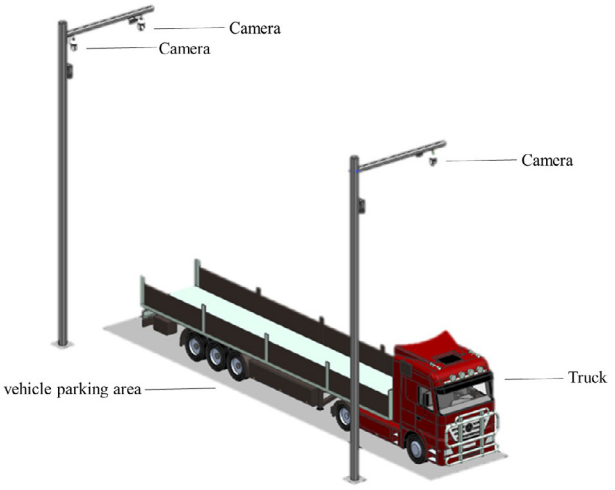


Fig. 1. On-site steel scrap collection process.

as finely crushed materials and confined parts. As a result, the HK_T dataset was established. HK_T scrap dataset, divided into 11 labels, is Airtight (confined parts), Scattered (fine crushed material), Inclusion (impurity), Ungrade (secondary), Overlength (1.2–1.5 m), Overlength (1.5–2 m), <3 mm (thickness less than 3 mm), 3–6 mm (thickness 3–6 mm), >6 mm (thickness >6 mm), Boxcar, Crane. The training set and validation set are divided according to the ratio of 9:1. The number of specific categories of labels is shown in Table 2, and the detailed data of the two datasets are shown in Table 3. The image acquisition system was built, and the steel scrap images were collected on-site by the main program developed in the laboratory, and the height of the spherical cameras (3 in total) was set according to the working height of the gripper, and the steel scrap images were collected on-site as shown in Fig. 1.

2.3. Data pre-processing

The image acquisition function is realized by the image sensor, the current image sensor mainly has charge-coupled devices, CCD and CMOS sensors, CCD sensors have the advantages of high resolution, low noise, dynamic range, etc. However, both CCD and CMOS sensors always introduce various noises and distortion when converting the actual scene into an image signal, so it is generally necessary to pre-process the image sensor image. In this paper, in addition to using conventional image cropping and rotate, a gamma image enhancement algorithm is utilized for image pre-processing.

(1) Gamma image enhancement algorithm

In the case of practical applications, the intensity of the output to the voltage signal is nonlinear between the research shows that the intensity of the display's output increases exponentially with the input voltage. Usually, we will refer to this power exponent as gamma (γ). Where, I refer to the light intensity of the display output; P refers to the beam voltage loaded on the display, the general beam voltage P is determined by the pixel value of the corresponding position of the image; γ is the gamma value (Eq. (1)).

$$I = P^{\gamma} \tag{1}$$

In order to match the image displayed on the monitor, it is necessary to perform gamma correction after the camera acquires the image, so that this non-linear correction is linear, as follows (Eq. (2)):

$$P_{\text{new}} = \left(P_{old}\right)^{\frac{1}{\gamma}} \tag{2}$$

P_{new} is the corrected image's pixel value; P_{old} Is the pixel value before the correction; the gamma value impacts the degree of correction, $\gamma = 1$, no correction, the bigger, the greater the amount of pixel value correction. In this study, the gamma transform is applied to the steel scrap photos in the gamma value range of 0.25–1.75, with 0.25 as the span to do the steel scrap image enhancement, and finally, a gamma value of 1.5 is chosen to improve the dataset. One of the scrap images taken by one of the camera positions is selected for gamma image enhancement, as shown in Fig. 2.

3. Methods

3.1. CSBFNet model

The proposed CSBFNet model based on the attention mechanism (Vaswani et al., 2017) is used for the quality-checking rating of the steel scrap unloading process, i.e., the size, thickness, and presence or absence of plating are rated according to the relevant criteria, and the network framework is shown in Fig. 3. For the scrap quality inspection scenario, the Cross Stage Partial Networks (CSP) (Wang et al., 2020) structure is first used for feature extraction of multiple categories of steel scrap images. Squeeze-Excitation (SE) (Hu et al., 2018) attention mechanism module is used to solve the problem of the different channels in the convolutional (Kim, 2014) pooling process by the attention mechanism. EfficientDet (Bidirectional Feature Pyramid Network, BiFPN) (Tan et al., 2020) is used to balance feature information from different scales for more efficient multi-scale feature fusion. The functions and principles of each module in the CSBFNet model are described below in turn.

3.2. Related technologies

(1) Cross Stage Partial Networks

In the picture feature extraction step, Cross Stage Partial Networks (CSP) enable the model to learn additional feature information. CSP incorporates gradient variations into the feature map, resulting in gradient information with a substantial change in significance and hence more feature information. It also provides substantial advantages in terms of minimizing the processing effort while enhancing inference speed and accuracy. The CSP structure separates the input into two branches, one of which travels through the CBL (Wang et al., 2021) structure before being convolution again after several residual structures (He et al., 2016) (Resunit). A concat (Dettmers et al., 2017) operation merges the CBL of the other branch with it before performing a CBL operation. Conv (convolution) + BN (Batch Normalization) +

Fig. 2. Gamma data enhancement effect. (a) gamma = 0.25; (b) gamma = 0.5; (c) gamma = 0.75; (d) gamma = 1.25; (e) gamma = 1.5; (f) gamma = 1.75.

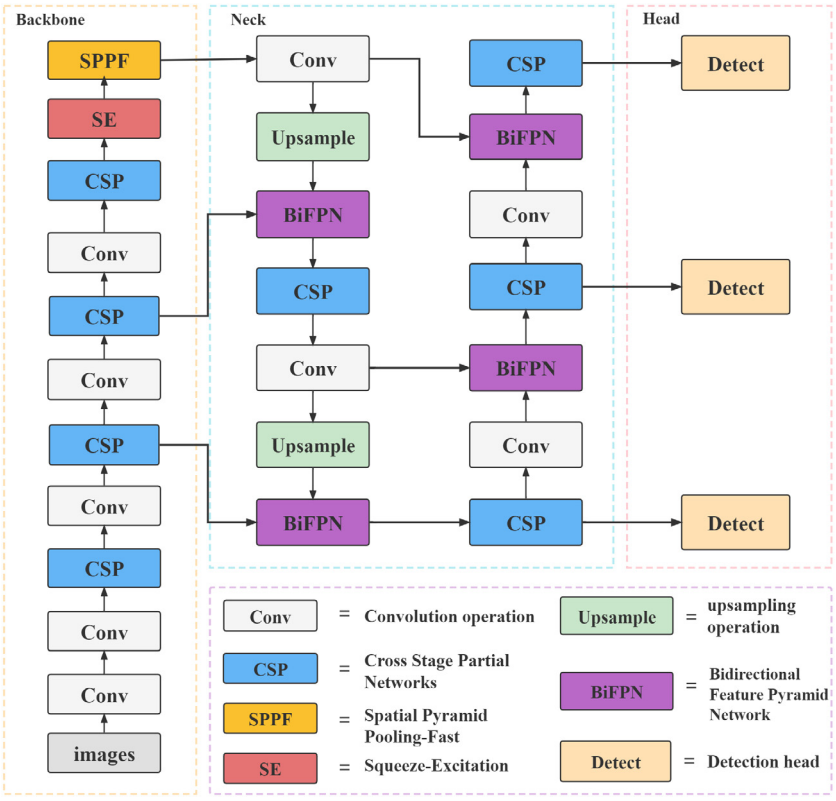


Fig. 3. CSBFNet model network diagram.

Leaky Relu (activation function) = CBL. Batch Normalization is a technique used in deep neural networks to improve the training speed and stability of the model. It involves normalizing the input data to each layer of the network, typically by subtracting the mean and dividing by the standard deviation of the data within a mini-batch of training examples; Leaky ReLU is an activation function commonly used in deep neural networks. It is similar to the standard Rectified Linear Unit (ReLU) activation function, but instead of setting negative values to zero, it allows a small negative slope. Fig. 4 depicts the structure of a CSP.

(2) Squeeze-Excitation

The Squeeze-Excitation (SE) attention method addresses the issue of feature loss caused by the convolutional pooling process's varying percentage of distinct channels. The global features are then subjected to the Excitation operation to learn the relationship between the channels, and the weights are applied to the feature map to create the final features, i.e., each channel is multiplied by its own weight to obtain the final features. The weights show each channel's contribution to feature extraction. The SE module is essentially a channel attention operation that allows the network to choose to enhance the helpful

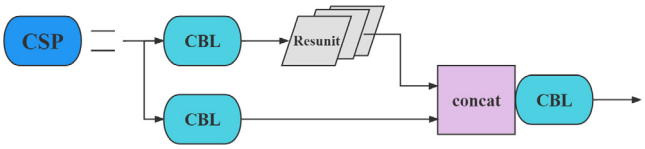


Fig. 4. CSP structure diagram.

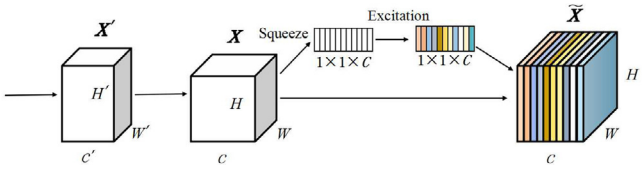


Fig. 5. SE module structure diagram.

feature channels and suppress the useless feature channels using global information, allowing adaptive calibration of the feature channels. $\boldsymbol{X'}$ is the original input feature, H' is the height of the original input space, W' is the width of the original input space. $\boldsymbol{X}$ is the input data feature map after the convolution operation, H' is the space height, W' is the space width, c' is the number of channels. The feature map is compressed by the Squeeze operation, and the feature map is excited by the Excitation operation. $\widetilde{\boldsymbol{X}}$ is the feature recalibration result. The units of H and W in all formulas in this article are pixels. Fig. 5 depicts the SE module's construction. Fig. 6. depicts the SE-ResNet module. Among them, FC stands for full connection, ReLU and Sigmoid are activation functions, scale is the weighted result of feature map, and $\oplus$ is concat operation.

The Squeeze operation employs global average pooling to broaden the perceptual field, encode the full spatial features in one dimension as a global feature, and calculates the weights of each channel, as illustrated in (Eq. (3)).

$$\boldsymbol{z} = f_{sq}(\mathbf{x}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{x}_{i,j} \tag{3}$$

In the formula, $\boldsymbol{z}$ represents the global feature, f_{sq} represents the squeeze operation, $\mathbf{x}$ represents the input feature map, H is the height of the feature map, W is the width of the feature map, and $\mathbf{x}_{i,j}$ is the pixel in the ith row and jth column eigenvectors of. The squeeze (Sequeeze) operation obtains the global description feature, and the excitation (Excitation) operation is used to obtain the correlation between channels, and retain the channel with the largest feature information, and suppress the channel with a small amount of information. Its formalization is shown in (Eq. (4)).

$$\boldsymbol{s} = f_{ex}(\mathbf{z}, \mathbf{W}) = \sigma\left(\boldsymbol{W}_2 \delta\left(\boldsymbol{W}_1 \boldsymbol{z}\right)\right) \tag{4}$$

In the formula, $\boldsymbol{s}$ represents the incentive score vector, f_{ex} represents the incentive operation, $\boldsymbol{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ represents the weight matrix of $\frac{C}{r}$ rows and C columns, $\boldsymbol{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ represents C the weight matrix of row $\frac{C}{r}$ column, σ is the sigmoid function, and δ is the ReLU activation function.

(3) Convolutional Block Attention Module

CBAM (Convolutional Block Attention Module) (Woo et al., 2018) is a feature refinement method that performs an adaptive feature mapping on both channel and spatial dimensions of the feature map. The channel attention module first performs global maximum and mean pooling on the feature map $\boldsymbol{F}$, resulting in a 1-dimensional channel attention $\boldsymbol{M}_{\mathrm{C}} \in \mathbb{R}^{C \times 1 \times 1}$ through a fully connected layer. The element-wise multiplication of $\boldsymbol{F}$ and $\boldsymbol{M}_{\mathrm{C}}$ generates the refined feature map $\boldsymbol{F'}$, where $\boldsymbol{M}_{\mathrm{C}}$ represents the attention weight for each channel, and C is the number of channels. The spatial attention module applies the same pooling operation on $\boldsymbol{F'}$ to generate a 2-dimensional spatial

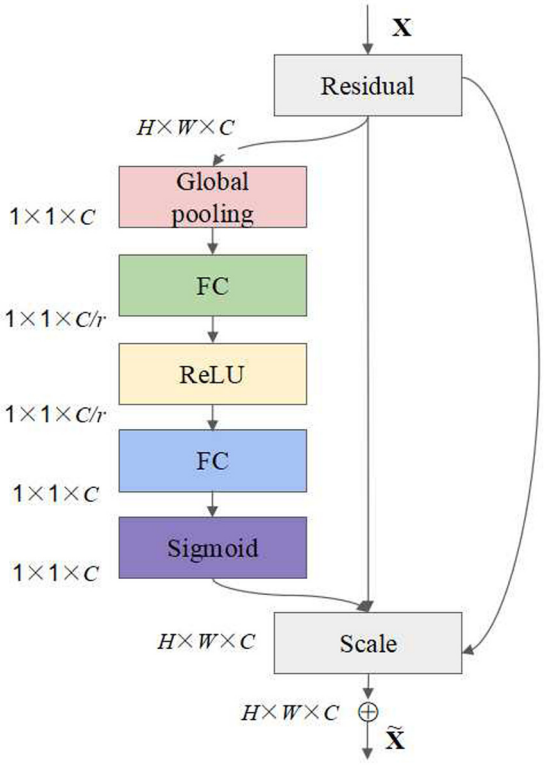


Fig. 6. SE-ResNet module.

attention $\boldsymbol{M}_{\mathrm{S}} \in \mathbb{R}^{1 \times H \times W}$, which is then element-wise multiplied with $\boldsymbol{F'}$ to obtain $\boldsymbol{F''}$, where $\boldsymbol{M}_{\mathrm{S}}$ represents the attention weight for each spatial position, and H and W denote the height and width of the input feature map, respectively. The tensor $\boldsymbol{M}_{\mathrm{S}}$ can be interpreted as the attention weight for each spatial position, with each element representing the importance of the corresponding spatial position. These weights can be used to adjust the contribution of different positions and improve the model performance. $\otimes$ denotes the corresponding element-by-element multiplication. The attention process of CBAM can be described by Eq. (5).

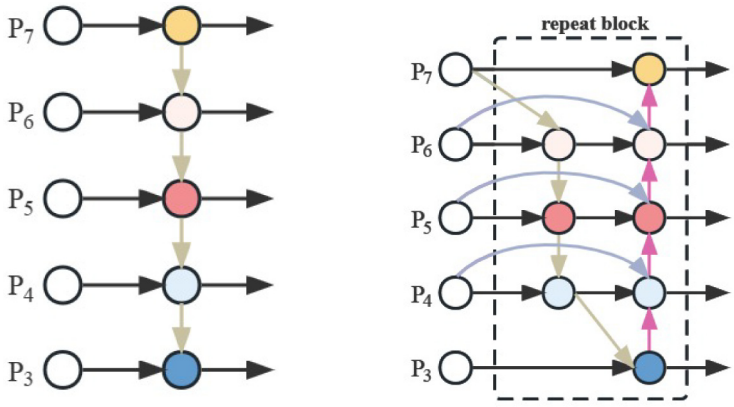
$$\begin{aligned} \boldsymbol{F'} &= \boldsymbol{M}_{\mathbf{C}}(\boldsymbol{F}) \otimes \boldsymbol{F} \\ \boldsymbol{F''} &= \mathrm{M}_{\mathrm{s}}\left(\boldsymbol{F'}\right) \otimes \boldsymbol{F'} \end{aligned} \tag{5}$$

(4) Coordinate Attention

Coordinate Attention (CA) (Hou et al., 2021b) incorporates location data into channel attention. Enabling mobile networks to reach out to bigger areas while avoiding considerable computational overhead. In order to obtain attention to the image width and height and encode the precise location information, the input feature maps are first divided into two directions, width, and height, for global averaging pooling to obtain feature maps in both width and height directions, respectively, specifically. Given the input $\boldsymbol{X}$, two spatial ranges (H, 1) or (1, W) of the pooling kernel is used to encode each channel along the horizontal and vertical coordinates, respectively. Thus, the output at channel c at height h can be expressed as (Eq. (6)).

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i) \tag{6}$$

In this formula, $x_c(h, i)$ represents the feature value of input data at channel c, position h, and spatial dimension i, and W is the spatial dimension of input data. The formula computes the average pooling of feature values along the spatial dimension, resulting in a feature vector z_c^h that represents the channel c and position h. This feature vector can be viewed as a compressed and abstract representation of input data on the spatial dimension.



(a) FPN (b) BiFPN

Fig. 7. FPN and BiFPN structure diagram.

The output of the Cth channel width w can be written as (Eq. (7)):

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w) \tag{7}$$

In this formula, $x_c(j, w)$ represents the feature value of input data channel c, position w, and spatial dimension j, and H is the spatial dimension of input data. The formula computes the average pooling of feature values along the channel dimension, resulting in a feature vector $z_c^w(w)$ that represents the spatial position w and channel c. This feature vector can be viewed as a compressed and abstract representation of input data on the channel dimension.

In CA attention, $z_c^h(h)$ and $z_c^w(w)$ are used to aggregate information from different dimensions and generate corresponding attention weights. These attention weights can be used to compute the importance of input data at different positions and channels, enabling the effect of local perception and global aggregation.

(5) EfficientDet

EfficientDet utilizes the Bidirectional Feature Pyramid Network (BiFPN), which is a simple and efficient weighted feature fusion network. The figure shows the structure diagrams of FPN and BiFPN, where P3 to P7 are feature maps of different scales. P3 has the high resolution but less abstract information, while P7 has low resolution but more abstract information. In feature pyramid network (FPN), the low-resolution and high-information features are up-sampled to the previous layer from top to bottom, enabling the fusion of features. This fusion allows high-resolution and low-information features from the bottom level to be integrated into the upper layer to provide abstract information.

The BiFPN structure builds on the FPN by increasing the connection from the original node to the bottom-up node and integrating the top-down and bottom-up processes into a reusable block. The optimization includes removing the single-input edge nodes that cannot provide more information to simplify the network, increasing the connection from the input node to the output node of the same layer to include the information of the upper and lower layers, and adding a weight to each scale feature of the fusion to adjust the contribution of each scale and improve detection speed. The network structure of FPN and BiFPN is illustrated in the Fig. 7.

(6) Spatial Pyramid Pooling Fast

Spatial Pyramid Pooling Fast (SPPF) is a pooling technique that allows for variable input sizes to be fed into a fully connected neural network. This is accomplished by dividing the input image into multiple sub-regions and performing max pooling within each sub-region. Unlike traditional max pooling, which requires a fixed input size, SPPF generates a fixed-length output regardless of the input size.

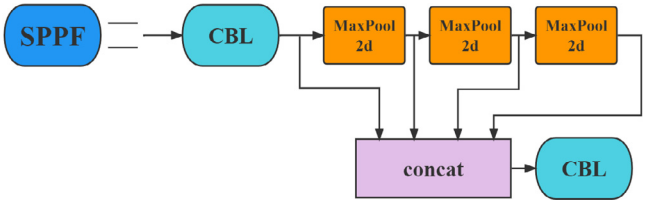


Fig. 8. SPPF module structure diagram.

SPPF works by dividing the input feature map into a set of fixed-size rectangular bins, and then pooling within each bin to generate a fixed-length output. The size of the bins is determined by the output size of the final fully connected layer in the network. Each bin is then pooled using max pooling to produce a fixed-length vector. The final output is the concatenation of all of the pooled vectors from each bin.

One advantage of SPPF is that it allows for flexible input sizes, which is important in applications where input sizes can vary widely, such as in object detection and recognition. By pooling features from multiple spatial scales, SPPF can capture both fine-grained and coarse-grained information, making it well-suited for object detection tasks. The SPPF structural diagram is shown in Fig. 8. In the SPPF structure, CBS = conv + BN + SiLU, where conv is convolution, BN is Batch Normalization, SiLU is the activation function, and MaxPool is the maximum pooling.

4. Experiment and analysis

4.1. Experimental environment

The experiments in this paper build a model training and verification environment for the PyTorch framework on an Ubuntu 18.04 system, using an NVIDIA GeForce RTX 3090 (24G) *8 graphics card accelerated by an Intel (R) Xeon (R) Silver 4310 CPU @ 2.10 GHz, and the Python programming language.

4.2. Model evaluation criteria

The evaluation metrics of the model include Precision, Recall, Accuracy, Average Precision (AP), and Mean Average Precision (mAP). The formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

Table 4
Positive and negative examples.

Type	P (Positive)	N (Negative)
T (True)	True positive (TP)	True negative (TN)
F (False)	False positive (FP)	False negative (FN)

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

The correspondence of TP, FP, FN, and TN is shown in Table 4.

AP is the average precision, which is calculated by plotting the P-R curve and selecting the maximum precision greater than or equal to each R value, and then averaging the AP values. mAP is the average of AP of each scrap category, and the formula is shown in (Eq. (11)). In the formula, i is the ith AP, and N is the total number of categories.

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{11}$$

4.3. Results and analysis

4.3.1. Comparison experiment

To demonstrate the superiority of the CSBFNet model compared with other target detection algorithms in identifying and rating steel scrap categories, the laboratory image dataset HK_L was selected and compared with the classical two-stage model Faster R-CNN and one-stage models YOLOv4 (Bochkovskiy et al., 2020), YOLOv5s and the latest YOLOv7 are in target detection algorithms, respectively, and the detection precision of the four models were compared in Table 5. GFLOPS is the evaluation index of model complexity. The results show that the improved CSBFNet model has a significant advantage in detection precision.

As shown in Table 5, the CSBFNet model has a lower number of network layers and parameters than the YOLOv4 model, but the detection precision is much higher than YOLOv4, with a 19.5% difference between them. Although CSBFNet uses a CSP network in the backbone network, which leads to an increase in model computation and a 3 ms slower inference time than YOLOv5s, the detection precision is 23.5% better than YOLOv5s, and CSBFNet has a higher inference time and detection precision than Faster R-CNN with the similar number of parameters. Although CSBFNet has a higher number of network layers and parameters than YOLOv7, it has lower inference time and higher mAP. From a computational complexity perspective, the CSBFNet model and the latest YOLOv7 model have similar computational complexities, but the CSBFNet model has higher mAP and faster inference time. Although the computational complexity of the CSBFNet model is higher than that of the other three comparison models, its inference time is faster than that of Faster R-CNN and YOLOv4, while also achieving higher mAP. These results indicate that the CSBFNet model is better optimized and more efficient. The detection time of CSBFNet for a single image is only 11 ms, which can meet the real-time requirements of scrap steel identification and classification. To sum up, CSBFNet has the best comprehensive effect, which meets the accuracy and real-time requirements of scrap classification and rating.

4.3.2. Ablation experiments

During the training process, Adam (Kingma and Ba, 2014) optimizer is used as the parameter, and CIoU Loss (Zheng et al., 2019) is used as the loss function. "image_size" indicates the input image size and is set to 640 × 640 pixel, "epochs" indicates the number of training sessions, and "batch_size" indicates the number of images trained in a single batch and is set to 16. The parameter settings are shown in Table 6.

The precision, recall, and mAP were obtained on the training set of the HK_L dataset using the CSP model, 6 improved models, and CSBFNet model respectively, and the same experiments were done on the later steel mill dataset to further verify the superiority of CSBFNet model, the training results of each model under both datasets are shown in Table 7, and the details of each improved model are shown in Table 8. From Table 7, it can be seen that the mAP value of CSBFNet, the network model improved by this paper, reaches 90.7%. The improved model 1 aimed to enhance the original backbone feature extraction network by adding a CA attention mechanism. CA is known to focus on the location information (width and length) of the features. However, since the scrap locations are primarily concentrated in the carriage, and the various scrap items are stacked more densely, this approach did not yield a significant improvement in the model's performance. In fact, the mAP of this improved model was even lower than that of the original model.

The improved model 2 is to add the CBAM attention mechanism to the original backbone feature extraction network, CBAM has both spatial and channel attention operation modules, but the improvement effect of the model is not obvious, the reason is that too many attention modules instead increases the computational volume and complexity of the model, and its GFLOPS increases by 0.3 compared with the improved model 3 which adds SE module, thus causing the model not to be The improved model 3 adds the SE attention mechanism to the original backbone feature extraction network, and the model complexity GFLOPS is slightly smaller than that of the original model and each improved model, but the mAP improves 0.8% and 6.9% under the two datasets compared with the original model, respectively, due to the small number of images in the HK_L dataset, which affects the model training effect, while the HK_T dataset with a larger number of samples The maximum enhancement effect of the model is exploited. The intrinsic principle is that the SE attention mechanism performs Squeeze operation on the steel scrap feature map obtained after the convolution operation to obtain the global steel scrap category features and the weights of each channel at the channel level, and then performs an Excitation operation on the global features to learn the relationship between each channel, retaining the important features of each steel scrap category as well as suppressing the irrelevant feature information (non-steel scrap, background, etc.).

The improved model 4 is in replacing the original feature fusion network with the new EfficientDet feature fusion network BiFPN, comparing with the feature fusion network of CSP model, BiFPN enhances the representation of features by residual operation, and removes the nodes with single input edges with less information to reduce the computation, adds different weights for each scale of fused features, and adjusts the contribution of each The CSP model without the EfficientDet feature fusion network BiFPN is compared, and its mAP is increased by 0.7% and 6.6% under two data sets, but the computational effort does not increase, and the model complexity GFLOPS is the same as the original model.

CSBFNet is a model that combines the attention mechanism and EfficientDet feature fusion module BiFPN into the network. Compared to the original model, CSBFNet does not increase GFLOPS, but the mAP increases by 1.5% and 7.9%, and precision and recall also substantially increase. The reason for this improvement is that the SE attention mechanism provides more accurate image features, while the EfficientDet feature fusion module BiFPN increases the weights of multi-scale features. This makes important features more prominent and reduces the interference of irrelevant features, resulting in richer and more comprehensive features of steel scrap categories. In contrast, although both improved model 5 and improved model 6 add BiFPN, the enhancement effect of CA and CBAM on the steel scrap dataset is not obvious, and it increases the computation and model complexity, thus affecting the model training results.

There are two main reasons why adding SE attention mechanism is better than adding CBAM attention: (1) Importance of channel features: Objects in the recycled iron and steel raw material dataset vary in shape and size, so the importance of channel features is higher. The SE attention mechanism evaluates and weights the importance of channel features, which helps the model to better attend to important features in the recycled steel raw material dataset and improve detection

Table 5
Comparison of detection precision of different network models.

Models	mAP/%	Number of network layers	Parameters/M	Inference time/ms	GFLOPS
Faster R-CNN	64.1	244	44	44	23.9
YOLOv4	68.1	516	52	26	59.7
YOLOv5s	64.1	270	7	8	16
YOLOv7	85.8	415	38	12	105.2
CSBFNet	87.6	474	46	11	108.5

Table 6
Training parameters.

Optimizer	Loss function	Image-size	Epoch	batch_size
Adam	CIoU	(640 × 640 pixel)	200	16

Table 7
Evaluation indices of different models under two data sets.

Model	Datasets	mAP/%	GFLOPs
CSP	HK_L	86.1	108.5
	HK_T	82.8	
Improved Model 1	HK_L	85.9	108.6
	HK_T	82.9	
Improved Model 2	HK_L	86.7	108.7
	HK_T	83.1	
Improved Model 3	HK_L	87.0	108.4
	HK_T	89.8	
Improved Model 4	HK_L	86.8	108.5
	HK_T	89.4	
Improved Model 5	HK_L	85.9	109.4
	HK_T	85.7	
Improved Model 6	HK_L	87.2	109.4
	HK_T	87.1	
CSBFNet	HK_L	**87.6**	108.5
	HK_T	**90.7**	

Table 8
Table of each improvement model.

Model	CA	CBAM	SE	BiFPN
Improved Model 1	✓			
Improved Model 2		✓		
Improved Model 3			✓	
Improved Model 4				✓
Improved Model 5	✓			✓
Improved Model 6		✓		✓
CSBFNet			✓	✓

performance. Additionally, SE attention is computationally efficient compared to CBAM on this dataset. The SE attention only needs one global average pooling layer and two fully connected layers, while CBAM attention requires more computing resources, including two average pooling layers, two fully connected layers and two convolutional layers. Therefore, using SE attention can improve the computational efficiency of the model without affecting the detection accuracy.

The Fig. 9 shows the mAP curves of each model before and after improvement and the CSBFNet model in training HK_L and HK_T datasets. The horizontal axis represents the training epochs, and the vertical axis represents the mAP value. Panels (a) and (b) represent the mAP curves of HK_L and HK_T datasets, respectively. It can be observed that the improved model 1 with a CA attention mechanism has the most drastic fluctuation and a sharp decrease at 75–90 epochs on the HK_L dataset, indicating that it is difficult to achieve model improvement for scrap classification and rating with CA attention mechanism. On the other hand, the improved model 6 with the addition of BiFPN in CBAM attention has a better improvement effect in the first 40 epochs, but the training that has a better effect. Finally, the CSBFNet model after adding SE attention mechanism and BiFPN of EfficientDet feature fusion network has a lower category imbalance rate due to

Table 9
Performance of the HK_L validation set for each category.

Class	Images	Labels	P/%	R/%	AP/%
<3 mm	28	22	82.0	62.3	71.6
3–6 mm	28	74	92.3	95.9	94.5
>6 mm	28	1024	87.5	82.2	87.5
Galvanized	28	88	88.8	89.7	91.6
Paint	28	26	94.6	87.1	91.3
Greasy dirt	28	33	85.7	78.8	88.6
Inclusion	28	85	98.8	100	99.5

Table 10
Performance of the HK_T validation set for each category.

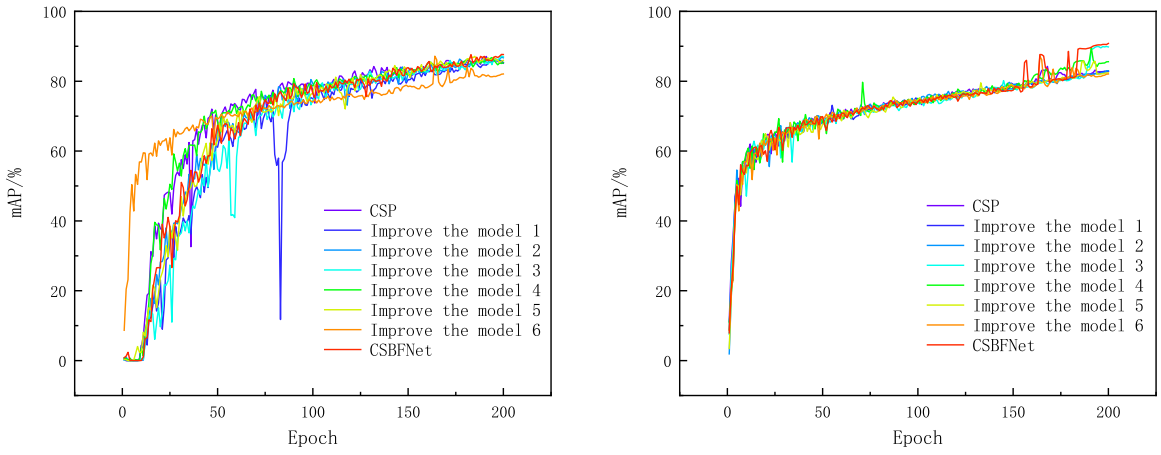
Class	Images	Labels	P/%	R/%	AP/%
Airtight	339	308	98.8	92.6	95.9
Scattered	339	280	97.9	90.1	92.9
Inclusion	339	106	93.9	83.8	89.1
Ungraded	339	263	100	95.8	97.6
Overlength (1.2–1.5 m)	339	142	98.2	75.2	85.3
Overlength (1.5–2 m)	339	12	87.4	96.2	96.7
<3 mm	339	34	85.4	92.4	94.2
3–6 mm	339	3701	96.3	89.6	94.3
>6 mm	339	8417	96.0	93.7	96.3

the attention and retention of scrap category information by the SE attention mechanism, and the model computation is relatively small compared to the other two attention improvement models. Therefore, it has a higher mAP and better training effect. Figs. 10 and 11 shows the precision and recall curves of the model training, and the curve of CSBFNet is higher and more stable than other models. The two curves of CSBFNet are higher and more stable than other models, which also proves the excellent effect of the model from another aspect.

4.3.3. Analysis of test results

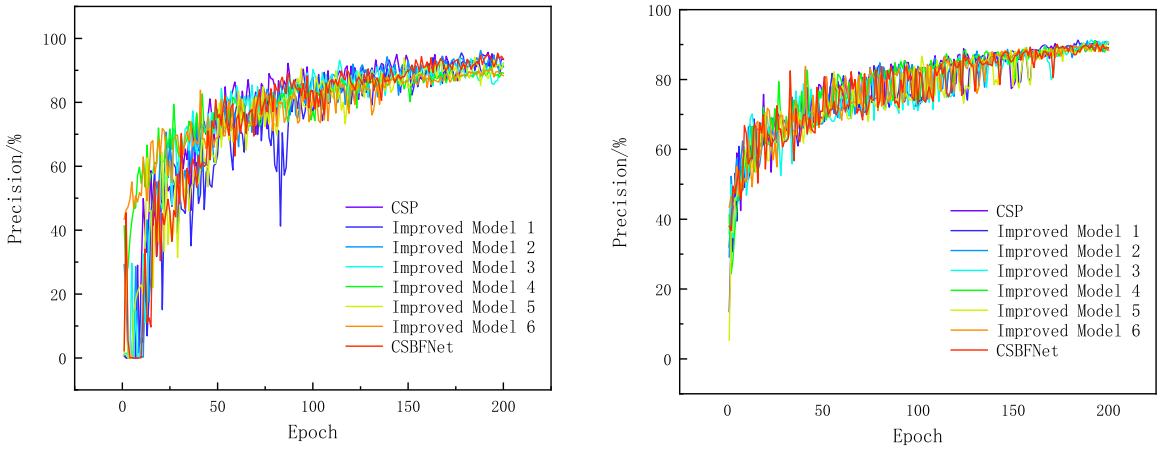
In this paper, the CSBFNet model was validated on the validation sets of two datasets, and the evaluation metrics of each scrap category were analyzed, and the validation results are shown in Tables 9 and 10. The images of HK_L and HK_T validation datasets are 28 and 339, and the labels are 1352 and 13263, respectively. Among them, in the verification results of the HK_L verification set, the highest single-category precision rate of steel scrap is 100%, the highest recall rate is 96.2%, the highest AP is 97.6%, and the mAP is 93.5%. Fig. 12 shows the detection effect plots, (a) and (b) for laboratory scrap and (b) and (c) for the CSBFNet model at the scrap recycling site. The information in the recognition rating box in the figure contains the name of the scrap category and the confidence level of the scrap category recognition rating detection. From the figure, it can be seen that the model trained on the HK_L dataset has a high confidence level for each category of scrap detection, accurate recognition, and a good recognition rating effect for the categories with fewer training labels; (d) the CSBFNet used in the figure also has a good recognition effect. From the graphical analysis, it can be obtained that the model has been trained with a large amount of data and has high precision for the identification and rating of multiple categories of scrap, and meets the actual production requirements in terms of real-time and detection efficiency.

Under the conditions of the HK_T data set, the number of training times is 200, and the batch-size is 16, the confusion matrix of different types of recycled iron and steel raw materials is obtained, as shown



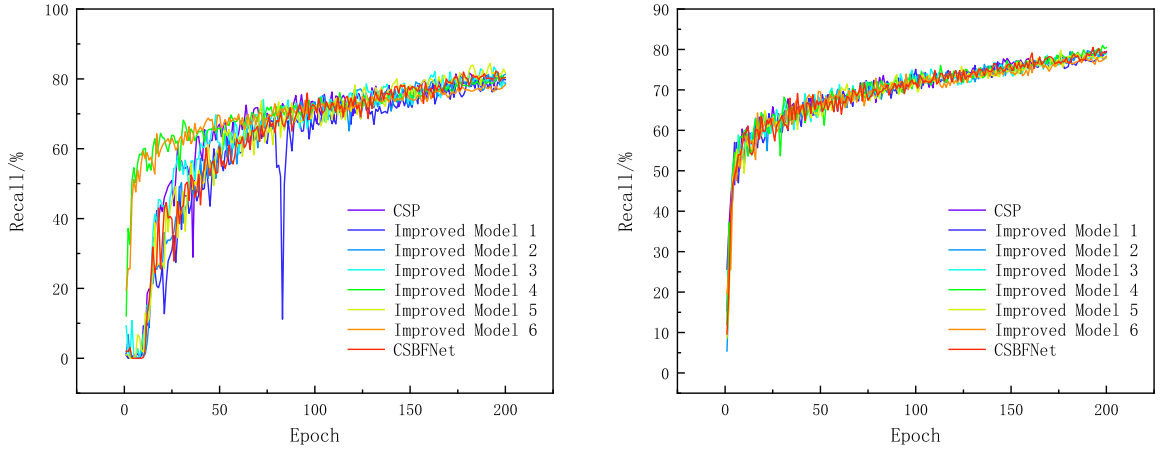
(a) Datasets=HK_L (b) Datasets=HK_T

Fig. 9. mAP curve for each model.



(a) Datasets=HK_L (b) Datasets=HK_T

Fig. 10. Precision curves for each model.



(a) Datasets=HK_L (b) Datasets=HK_T

Fig. 11. Recall curves for each model.

in Fig. 13. The horizontal axis is the true category label, and the vertical axis is the predicted category label. background FN represents the number of objects missed by the model, that is, the number of objects that are predicted to have no objects in the background category, but actually exist; background FP represents the number of false detections of the model, that is, the background or non-existing

Fig. 12. Model testing effect. (a) and (b) are laboratory scrap, (a) is the original figure (b) is the result after detection; (c) (d) is the detection effect of the CSBFNet model on the scrap recycling site, (c) is the original figure (d) is the result after detection.

objects are mistakenly identified as existing objects, as can be seen from Fig. 13, regarding the false detection rate, compared to all categories, there are more false detections for scrap with thickness greater than 6 mm, due to the fact that this category has a higher proportion, thus contributing more to the overall false detection rate. However, for scrap with thickness greater than 6 mm, the proportion of false detections is relatively low. The numbers in each graphic box in the figure represent the proportion of prediction results of different categories. In addition It can be seen from the figure that the CSBFNet model has good results in predicting various types of recycled iron and steel raw materials, and the average accuracy rate of all types of steel scrap reaches 92.4%. Due to the limited number of sample data, individual categories of scrap, such as Overlength (1.2–1.5 m), may have lower classification and rating effectiveness compared to other categories. Subsequently increasing the number of samples in these categories can improve the classification and rating accuracy of these categories.

4.3.4. Result of discussion

Overall, these experiments show that combining the SE attention mechanism with EfficientDet is a new and promising scheme for smart scrap classification. The proposed method can be applied to different classes of steel scrap and used as a quality inspection system in the first step of scrap recycling, classifying and grading scrap steel. In addition, the system can be connected to the feeding system of the electric furnace, and the steel tapping rate can be predicted through the scrap steel that has been classified and graded.

The CSBFNet model further validates the effect of classification and rating on the field dataset (HK_T). The results show that the CSBFNet model is suitable for the scene of scrap steel quality inspection and shows superior results. CSBFNet model can effectively extract the categorical features of steel scrap on the backbone network with the addition of a SE attention mechanism, and perform an attention operation on the channels to increase the weight of important features and retain important feature information; The EfficientDet structural fusion network (BiFPN) on the SE processed feature map according to the resolution and the degree of feature contribution, the feature fusion operation is performed so that the high-resolution low-information features at the bottom layer can be fused to the abstract information at the top layer to balance the feature information at different scales. From the experimental data, the performance of BiFPN with the addition of SE attention mechanism and EfficientDet feature fusion network is better than multiple comparison models, CSBFNet model in each category of steel scrap, has reached the optimal effect. It can effectively solve the safety, accuracy and fairness problems of the traditional manual steel quality inspection mode, and help to get rid of the technical difficulties of intelligent quality inspection.

5. Envisioned industrial applications

The first envisaged industrial application is the classification and grading of a wide range of scrap using the developed method. Since the length and thickness of the scrap directly affect the purchase price and the ability to meet the specification criteria for inclusion in the furnace, scrap that meets the specification can reduce the operating costs of the steel mill and simplify the scrap recovery process. On the other hand, the accuracy of quality inspection in scrap recycling will also directly reflect the cost of steel-making. Therefore, it is of great importance to conduct an intelligent quality inspection for scrap recycling accurately and efficiently. The deep learning model of intelligent scrap inspection introduced in this study helps to simplify the quality inspection process, improve the safety, fairness, and accuracy of scrap quality inspection, and achieve the effect of replacing manual quality inspection to realize unmanned intelligent scrap quality inspection.

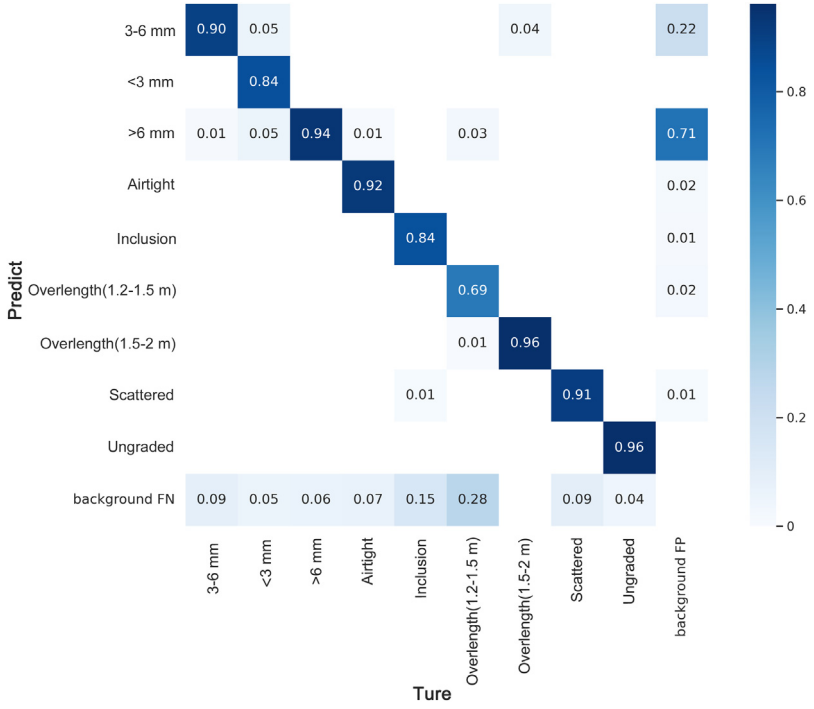


Fig. 13. Scrap category confusion matrix.

Another envisioned application is to use the proposed method to enhance the efficiency of robotic scrap sorting. Today, the field of artificial intelligence is growing rapidly and is used in various industries. The problems of high danger, low efficiency, and poor accuracy in the field of the scrap recycling, combining intelligent inspection and intelligent sorting to achieve a truly unmanned and intelligent scrap recycling system, will help to improve the efficiency and quality of scrap recycling, accelerate the development of the scrap recycling field, provide high-quality raw materials for the steel industry and guarantee the green and low carbon development of the steel industry.

6. Conclusion and future work

This paper addresses the problems of many types of steel scrap, complex actual inspection scenarios, and the difficulty of manual system interface, and transforms the current method of determining scrap grades in most steel enterprises, which is mainly determined by visual inspection by quality managers and caliper measurement together, into intelligent grading. How to achieve openness, fairness, and equity in scrap grade inspection and solve the problems of human factors interference and inefficiency, we propose a deep learning model of CSBFNet scrap quality inspection based on attention mechanism, in which, the SE attention mechanism is added to the feature extraction network and the feature fusion network is replaced with EfficientDet module (BiPFN), which makes the feature extraction in scrap category and multi-scale feature fusion with significant advantages, which improves the overall performance of the network.

The experiments were conducted on the laboratory dataset HK_L and the field-collected dataset HK_T for model training and optimization. The results showed that the mAP of the CSBFNet model reached 90.7%, and the average accuracy rate for all types of steel scrap reached 92.4%. The evaluation index was higher than that of the model without the SE attention mechanism and BiPFN module, as well as mainstream target detection models such as Faster R-CNN. The detection time for a single image is only 11 ms, which can meet the real-time requirements of scrap identification and grading in actual production scenarios. This indicates that the CSBFNet model has high performance and excellent detection and grading capabilities. Compared with the traditional manual scrap inspection, the CSBFNet steel scrap inspection model has obvious advantages in accuracy, safety, and fairness.

The future research direction mainly includes three aspects. First, to reduce the influence of the complex background of the scrap steel image on the classification and rating effect, image segmentation technology could be employed to establish a compartment segmentation model and accurately determine the scrap steel classification and rating area. Second, a quality prediction model could be applied to estimate the quality of steel scrap after classification and grading. Finally, the steel scrap after quality prediction could be fed into the electric furnace to predict the tapping rate based on the predicted steel scrap quality data.

In future experiments, the dataset will be expanded and balanced to reduce the problem of inaccurate ratings for individual categories of scrap due to insufficient samples, while the model will be further optimized and improved for more complex steel scrap recycling quality inspection scenarios in the future. In addition, the dataset will provide a finer breakdown of the steel scrap classification and add a quality assessment model to the subsequent work to deduct weight from impurities in scrap recycling to make the inspection more complete. Finally, the system will be integrated into the real-time system of scrap inspection, which will help reduce the problem of inaccurate scrap rating and get more value in scrap recycling.

Furthermore, intelligent scrap rating not only affects cost settlement, but also dovetails with electric furnace charging, which has a catalytic effect on improving the steelmaking process regime.

CRediT authorship contribution statement

Wenguang Xu: Ideas, Creation of models, Writing – original draft. **Pengcheng Xiao:** Formulation or evolution of overarching research goals and aims, Provision of study materials, Reviewing and editing. **Liguang Zhu:** Supervision. **Yan Zhang:** Management and coordination responsibility for the planning and execution of research activities. **Jinbao Chang:** Provision of laboratory samples. **Rong Zhu:** Conducting a research and investigation process. **Yunfeng Xu:** Provision of computing resources.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Pengcheng Xiao reports financial support was provided by National Natural Science Foundation of China. Pengcheng Xiao reports financial support was provided by the Natural Science Foundation of Hebei Province Project. Pengcheng Xiao reports financial support was provided by the Science and Technology Research Project of Hebei Higher Education Institution. Pengcheng Xiao reports financial support was provided by the Key Project of Tangshan City Talent Funding.

Data availability

The data that has been used is confidential.

Acknowledgments

The research was completed under the National Natural Science Foundation of China 51904107; Natural Science Foundation of Hebei Province, China E2020209005, E2021209094; Science and Technology Research Project of Hebei Higher Education Institution, China BJ2019041; "San San San Talent Project" of Hebei Province, China A202102002; Key Project of Tangshan City Talent Funding, China A202010004.

References

Bhardwaj, K., Diffenderfer, J., Kailkhura, B., Gokhale, M., 2022. Benchmarking test-time unsupervised deep neural network adaptation on edge devices. In: 2022 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). pp. 236–238. http://dx.doi.org/10.1109/ISPASS55109.2022.00033.

Bircanoğlu, Cenk., Meltem, Atay, Fuat, Beşer, Özgün, Genç, Kızrak, Merve Ayyüce, 2018. RecycleNet: Intelligent waste sorting using deep neural networks. In: 2018 Innovations in Intelligent Systems and Applications. INISTA, IEEE, pp. 1–7. http://dx.doi.org/10.1109/INISTA.2018.8466276.

Bobulski, Janusz, Kubanek, Mariusz, 2020. Project of sorting system for plastic garbage in sorting plant based on artificial intelligence. In: CS & IT Conference Proceedings (Vol. 10, No. 9). CS & IT Conference Proceedings. http://dx.doi.org/10.5121/csit.2020.100903.

Bochkovskiy, A., Wang, C.Y., Liao, Hym, 2020. YOLOv4: Optimal speed and accuracy of object detection. http://dx.doi.org/10.48550/arXiv.2004.10934.

Cesetti, M., Nicolosi, P., 2016. Waste processing: new near infrared technologies for material identification and selection. J. Instrum. 11 (09), C09002. http://dx.doi.org/10.1088/1748-0221/11/09/C09002.

Chu, Yinghao, Huang, Chen, Xie, Xiaodan, Tan, Bohai, Kamal, Shyam, Xiong, Xiaogang, 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Comput. Intell. Neurosci. 2018, http://dx.doi.org/10.1155/2018/5060857.

Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S., 2017. Convolutional 2D knowledge graph embeddings. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32, http://dx.doi.org/10.1609/aaai.v32i1.11573, no. 1 2018.

Fan, Zhiyuan, Friedmann, S. Julio, 2021. Low-carbon production of iron and steel: Technology options economic assessment, and policy. Joule 5, 829–862. http://dx.doi.org/10.1016/j.joule.2021.02.018.

Fellner, Johann, Laner, David, Warrings, Rainer, Schustereder, Kerstin, Lederer, Jakob, 2018. Potential Impacts of the EU circular economy package on the utilization of secondary resources. Detritus 2, 16. http://dx.doi.org/10.31025/2611-4135/2018.13666.

Gao, Zhijiang, Sridhar, S., Erik, D., Patrick, R. Taylor, 2020. Applying improved optical recognition with machine learning on sorting Cu impurities in steel scrap. J. Sustain. Metall. 6, 785–795. http://dx.doi.org/10.1007/s40831-020-00300-8.

Gupta, Praveen Kumar, Shree, Vidhya, Hiremath, Lingayya, Rajendran, Sindhu, 2019. The use of modern technology in smart waste management and recycling: artificial intelligence and machine learning. In: Recent Advances in Computational Intelligence. Springer, Cham, pp. 173–188. http://dx.doi.org/10.1007/978-3-030-12500-4_11.

He, K., Zhang, X., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 770–778. http://dx.doi.org/10.1109/CVPR.2016.90.

Hou, Yue, Li, Qiuhan, Han, Qiang, Peng, Bo, Wang, Linbing, Gu, Xingyu, Wang, Dawei, 2021b. MobileCrack: Object classification in asphalt pavements using an adaptive lightweight deep learning. J. Transp. Eng. B Pavements 147, 04020092. http://dx.doi.org/10.1061/JPEODX.0000245.

Hou, Q., Zhou, D., Feng, J., 2021a. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp. 13713–13722. http://dx.doi.org/10.48550/arXiv.2103.02907.

Hu, Jie, Shen, Li, Sun, Gang, 2018. Squeeze-and-excitation networks squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 7132–7141. http://dx.doi.org/10.48550/arXiv.1709.01507.

Hu, Jing, Zhang, Bo, 2021. Application research of automatic garbage sorting based on TensorFlow and OpenCV. J. Phys. Conf. Ser. 1883 (1), 012169. http://dx.doi.org/10.1088/1742-6596/1883/1/012169.

Huang, Guang-Li, He, Jing, Xu, Zenglin, Huang, Guangyan, 2020. A combination model based on transfer learning for waste classification. Concurr. Comput. Pract. Exper. 32, e5751. http://dx.doi.org/10.1002/cpe.5751.

Huang, K., Wu, H., Liu, Z., Qi, X., 2022. In-situ model downloading to realize versatile edge AI in 6G mobile networks. http://dx.doi.org/10.48550/arXiv.2210.03555, arXiv preprint arXiv:2210.03555.

Jujun, Ruan, Yiming, Qian, Zhenming, Xu, 2014. Environment-friendly technology for recovering nonferrous metals from e-waste: Eddy current separation. Resour. Conserv. Recy. 87, 109–116. http://dx.doi.org/10.1016/j.resconrec.2014.03.017.

Kashiwakura, Shunsuke, Wagatsuma, Kazuaki, 2013. Characteristics of the calibration curves of copper for the rapid sorting of steel scrap by means of laser-induced breakdown spectroscopy under ambient air atmospheres. Anal. Sci. 29, 1159–1164. http://dx.doi.org/10.2116/analsci.29.1159.

Kim, Y., 2014. Convolutional Neural Networks for Sentence Classification (Master's thesis). University of Waterloo, http://hdl.handle.net/10012/9592.

Kim, Namgyu, Kim, Sehoon, An, Yun-Kyu, Lee, Jong-Jae, 2021. A novel 3D GPR image arrangement for deep learning-based underground object classification. Int. J. Pavement Eng. 22, 740–751. http://dx.doi.org/10.1080/10298436.2019.1645846.

Kingma, D., Ba, J., 2014. Adam: a method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980, arXiv preprint arXiv:1412.6980.

Koyanaka, S., Kobayashi, K., 2010. Automatic sorting of lightweight metal scrap by sensing apparent density and three-dimensional shape. Resour. Conserv. Recycl. 54, 571–578. http://dx.doi.org/10.1016/j.resconrec.2009.10.014.

Li, S., Li, W., Wen, S., Shi, K., Yang, Y., Zhou, P., Huang, T., 2021. Auto-FERNet: A facial expression recognition network with architecture search. IEEE Trans. Netw. Sci. Eng. 8 (3), 2213–2222. http://dx.doi.org/10.1109/TNSE.2021.3083739.

Li, W., Wen, S., Shi, K., Yang, Y., Huang, T., 2022. Neural architecture search with a lightweight transformer for text-to-image synthesis. IEEE Trans. Netw. Sci. Eng. 9 (3), 1567–1576. http://dx.doi.org/10.1109/TNSE.2022.3147787.

Lin, Yuancheng, Yang, Honghua, Ma, Linwei, Li, Zheng, Ni, Weidou, 2021. Low-carbon development for the iron and steel industry in China and the world: Status Quo future vision, and key actions. Sustainability 13 (12548), http://dx.doi.org/10.3390/su132212548.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. http://dx.doi.org/10.1016/j.media.2017.07.005.

Liu, Zhengdong, Li, Wenxia, Wei, Zihan, 2020. Qualitative classification of waste textiles based on near infrared spectroscopy and the convolutional network. Textile Res. J. 90, 1057–1066. http://dx.doi.org/10.1177/0040517519886032.

Lu, Xin, Lin, Zhe, Jin, Hailin, Yang, Jianchao, Wang, James Z., 2015. Rating image aesthetics using deep learning. IEEE Trans. Multimed. 17, 2021–2034. http://dx.doi.org/10.1109/TMM.2015.2477040.

Lyu, B., Wen, S., Shi, K., Huang, T., 2023. Multiobjective reinforcement learning-based neural architecture search for efficient portrait parsing. IEEE Trans. Cybern. 53 (2), 1158–1169. http://dx.doi.org/10.1109/TCYB.2021.3104866.

Lyu, B., et al., 2022. Efficient spectral graph convolutional network deployment on memristive crossbars. IEEE Trans. Emerg. Top. Comput. Intell. http://dx.doi.org/10.1109/TETCI.2022.3210998.

Ma, Yiqun, Wang, Junhao, 2021. Time-varying spillovers and dependencies between iron ore scrap steel, carbon emission, seaborne transportation, and chinas steel stock prices. Resour. Policy 74, 102254. http://dx.doi.org/10.1016/j.resourpol.2021.102254.

Majchrowska, Sylwia, Mikołajczyk, Agnieszka, Ferlin, Maria, Klawikowska, Zuzanna, Plantykow, Marta A., Kwasigroch, Arkadiusz, Majek, Karol, 2022. Deep learning-based waste detection in natural and urban environments. Waste Manag. 138, 274–284. http://dx.doi.org/10.1016/j.wasman.2021.12.001.

Mesina, M.B., Jong, Tprd, Dalmijn, W.L., 2007. Automatic sorting of scrap metals with a combined electromagnetic and dual energy X-ray transmission sensor. Int. J. Miner. Process. 82, 222–232. http://dx.doi.org/10.1016/j.minpro.2006.10.006.

Nechifor, Victor, Calzadilla, Alvarom, Bleischwitz, Raimund, Winning, Matthew, Tian, Xu, Usubiaga, Arkaitz, 2020. Steel in a circular economy: Global implications of a green shift in China. World Dev. 127, 104775. http://dx.doi.org/10.1016/j.worlddev.2019.104775.

Omura, Akihiro, Todorova, Neda, Li, Bin, Chung, Richard, 2016. Steel scrap and equity market in Japan. Resour. Policy 47, 115–124. http://dx.doi.org/10.1016/j.resourpol.2016.01.001.

Passarini, F., Ciacci, L., Nuss, P., Manfredi, S., 2018. Material flow analysis of aluminium, copper, and iron in the EU-28. Publications Office, Luxembourg, https://rmis.jrc.ec.europa.eu/uploads/MFA_Report_2018_FINAL.pdf.

Patel, Kanil, Rambach, Kilian, Visentin, Tristan, Rusev, Daniel, Pfeiffer, Michael, Yang, Bin, 2019. Deep learning-based object classification on automotive radar spectra. In: 2019 IEEE Radar Conference (RadarConf). IEEE, pp. 1–6. http://dx.doi.org/10.1109/RADAR.2019.8835775.

Penumuru, Durga Prasad, Muthuswamy, Sreekumar, Karumbu, Premkumar, 2020. Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. J. Intell. Manuf. 31, 1229–1241. http://dx.doi.org/10.1007/s10845-019-01508-6.

Qi, C., Chang, J., Zhang, J., Zuo, Y., Ben, Z., Chen, K., 2022. Medicinal Chrysanthemum Detection under Complex Environments Using the MC-LCNN Model. Plants 11 (7), 838–841. http://dx.doi.org/10.3390/plants11070838.

Ramsurrun, Nadish, Suddul, Geerish, Armoogum, Sandhya, Foogooa, Ravi, 2021. Recyclable waste classification using computer vision and deep learning. In: 2021 Zooming Innovation in Consumer Technologies Conference. ZINC, IEEE, pp. 11–15. http://dx.doi.org/10.1109/ZINC52049.2021.9499291.

Rong, Dian, Xie, Lijuan, Ying, Yibin, 2019. Computer vision detection of foreign objects in walnuts using deep learning. Comput. Electron. Agric. 162, 1001–1010. http://dx.doi.org/10.1016/j.compag.2019.05.019.

Ruiz, Victoria, Sánchez, Ángel, Vélez, José F., Raducanu, Bogdan, 2019. Automatic image-based waste classification. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, Cham, pp. 422–431. http://dx.doi.org/10.1007/978-3-030-19651-6_41.

Shao, J., Zhang, X., Zhang, J., 2022. Task-oriented communication for edge video analytics. http://dx.doi.org/10.48550/arXiv.2211.14049, arXiv preprint arXiv:2211.14049.

Smirnov, N.V., Rybin, E.I., 2020. Machine learning methods for solving scrap metal classification task. In: 2020 International Russian Automation Conference (RusAutoCon). http://dx.doi.org/10.1016/j.compag.2019.05.019.

Sun, Liangliang, Jin, Hang, Li, Ye, 2018. Research on scheduling of iron and steel scrap steelmaking and continuous casting process aiming at power saving and carbon emissions reducing. IEEE Robot. Autom. Lett. 3, 3105–3112. http://dx.doi.org/10.1109/LRA.2018.2849500.

Sun, Wenqiang, Wang, Qiang, Zheng, Zhong, Cai, Jiuju, 2020. Material–energy–emission nexus in the integrated iron and steel industry. Energy Convers. Manage. 213, 112828. http://dx.doi.org/10.1016/j.enconman.2020.112828.

Tan, M., Pang, R., Le, Q.V., 2020. Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp. 10781–10790. http://dx.doi.org/10.48550/arXiv.1911.09070.

Van den Eynde, Simon, et al., 2022. Assessing the efficiency of laser-induced breakdown spectroscopy (LIBS) based sorting of post-consumer aluminium scrap. Procedia CIRP 105, 278–283. http://dx.doi.org/10.1016/j.procir.2022.02.046.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. http://dx.doi.org/10.48550/arXiv.1706.03762, In arXiv.

Wang, J., Chen, Y., Gao, M., Dong, Z., 2021. Improved yolov5 network for real-time multi-scale traffic sign detection. http://dx.doi.org/10.48550/arXiv.2112.08782, arXiv preprint arXiv:2112.08782.

Wang, C.Y., Liao, Hym., Wu, Y.H., Chen, P.Y., Yeh, I.H., 2020. CSPNet: A new backbone that can enhance learning capability of CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. IEEE, pp. 390–391. http://dx.doi.org/10.1109/CVPRW50498.2020.00203.

Weiss, Martin, 2011. Resource recycling in waste management with x-ray fluorescence. Diss. Univ. Leoben https://pure.unileoben.ac.at/portal/files/2459623/AC08703426n01vt.pdf.

Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. CBAM: Convolutional block attention module. In: Proceedings of the European conference on computer vision. ECCV, pp. 3–19. http://dx.doi.org/10.48550/arXiv.1807.06521.

Yang, S., Lee, G., Huang, L., 2022. Deep learning-based dynamic computation task offloading for mobile edge computing networks. Sensors 22 (11), 4088. http://dx.doi.org/10.3390/s22114088.

Yu, Biying, An, Runying, Zhao, Guangpu, 2020. Spatial and temporal disparity of the in-use steel stock for China. Resour. Conserv. Recy. 155, 104667. http://dx.doi.org/10.1016/j.resconrec.2019.104667.

Zhang, M., Cao, J., Sahni, Y., Chen, Q., Jiang, S., Wu, T., 2022. Eaas: A service-oriented edge computing framework towards distributed intelligence. In: 2022 IEEE International Conference on Service-Oriented System Engineering. SOSE, pp. 165–175. http://dx.doi.org/10.1109/SOSE55356.2022.00026.

Zhang, Song, Chen, Yumiao, Yang, Zhongliang, Gong, Hugh, 2021. Computer vision based two-stage waste recognition-retrieval algorithm for waste classification. Resour. Conserv. Recy. 169, 105543. http://dx.doi.org/10.1016/j.resconrec.2021.105543.

Zhang, Shunli, Forssberg, Eric, 1998. Mechanical recycling of electronics scrap-the current status and prospects. Waste Manag. Res. 16 (2), 119–128. http://dx.doi.org/10.1177/0734242X9801600204.

Zheng, Z., Wang, P., Liu, W., Li, J., Ren, D., 2019. Distance-IoU loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, vo. 34 no. 07. pp. 12993–13000. http://dx.doi.org/10.1609/aaai.v34i07.6999.

Zhu, Yongxian, Syndergaard, Kyle, Cooper, Daniel R., 2019. Mapping the annual flow of steel in the United States. Environ. Sci. Technol. 53, 11260–11268. http://dx.doi.org/10.1021/acs.est.9b01016.