



OPEN

DATA DESCRIPTOR

# A large-scale open image dataset for deep learning-enabled intelligent sorting and analyzing of raw coal

Ziqi Lv<sup>1,2,5</sup>✉, Yuhan Fan<sup>2,3,5</sup>, Te Sha<sup>1</sup>, Yao Cui<sup>2,3</sup>, Yuxin Wu<sup>3</sup>, Haimei Lv<sup>4</sup>, Meijie Sun<sup>1,2</sup>, Yanan Tu<sup>1,2</sup>, Zhiqiang Xu<sup>1,2</sup> & Weidong Wang<sup>1,2</sup>✉

Under the strategic objectives of carbon peaking and carbon neutrality, energy transition driven by new quality productive forces has emerged as a central theme in China's energy development. Among these, the intelligent sorting and analysis of raw coal using deep learning constitute a pivotal technical process. However, the progress of intelligent coal preparation in China has been constrained by the absence of accurate and large-scale data. To address this gap, this study introduces DsCGF, a large-scale, open-source raw coal image dataset. Over the past five years, extensive raw coal image samples were systematically collected and meticulously annotated from three representative mining regions in China, resulting in a dataset comprising over 270,000 visible-light images. These images are annotated at multiple levels, targeting three primary categories: coal, gangue, and foreign objects, and are designed for three core computer vision tasks: image classification, object detection, and instance segmentation. Comprehensive evaluation results indicate that the DsCGF can effectively support further research into the intelligent sorting of raw coal.

## Background & Summary

China is vigorously developing new quality productive forces, requiring all industries deepen the integration of new technologies to promote comprehensive social progress<sup>1</sup>. Concurrently, under the strategic goals of carbon peaking and carbon neutrality, the energy transition characterized by intelligence, greenness, and efficiency is accelerating. However, the country's resource endowment dictates that its coal-dominated energy structure is unlikely to shift significantly in the short term. Consequently, enhancing the efficiency, cleanliness, and intelligence of the coal industry through technological innovations holds substantial practical importance and far-reaching impact. Within this context, the intelligent clean processing and utilization of coal constitute critical aspects of smart mine construction, with the intelligent sorting of gangue and foreign objects in raw coal representing a key technology for achieving intelligent transformation.

In recent years, the increasing mechanization of coal mining has resulted in significant amounts of roof rock, floor rock, and production waste being mixed into raw coal, substantially raising the gangue and foreign object content. In typical mineral processing designs, it is often necessary to include a manual sorting stage for gangue and foreign object during the pre-processing of raw coal at coal preparation plants (CPPs) to avoid equipment failures caused by them entering subsequent processes. In the past, workers were arranged along conveyor belts to manually pick out large gangue and foreign objects from the raw coal. And now, the online intelligent identification and sorting of them in raw coal has emerged as a prominent research focus in both academia and industry.

<sup>1</sup>School of Chemical & Environmental Engineering, China University of Mining & Technology (Beijing), Beijing, 100083, P. R. China. <sup>2</sup>Inner Mongolia Research Institute, China University of Mining & Technology (Beijing), Ordos, 017001, P. R. China. <sup>3</sup>School of Artificial Intelligence, China University of Mining & Technology (Beijing), Beijing, 100083, P. R. China. <sup>4</sup>State Key Laboratory of Media Convergence Production Technology and Systems, Beijing, 100803, P. R. China. <sup>5</sup>These authors contributed equally: Ziqi Lv, Yuhan Fan. ✉e-mail: lvziqi@cumt.edu.cn; wwd@cumt.edu.cn

References	Task type	Photo site	State of material	Num of Objects/images
Zhang, L. <i>et al.</i> <sup>3</sup>	Image classification	Lab	Non-production	300/1800
Cao, Z. G. <i>et al.</i> <sup>11</sup>	Image classification	Lab	Non-production	400/400
Liu, H. J. <i>et al.</i> <sup>9</sup>	Image classification	Lab	Non-production	Unspecified/3150
Pan, H. G. <i>et al.</i> <sup>15</sup>	Object detection	CPP	Non-production	Unspecified/7000
Xu, S. Y. <i>et al.</i> <sup>16</sup>	Object detection	CPP	Non-production	Unspecified/3852
Zhang, Y. C. <i>et al.</i> <sup>20</sup>	Object detection	CPP	Non-production	Unspecified/793
Wang, X. <i>et al.</i> <sup>28</sup>	Semantic segmentation	Lab	Non-production	4055/594
Cao, Z. G. <i>et al.</i> <sup>29</sup>	Semantic segmentation	Lab	Non-production	7000/1500
Gao, R. <i>et al.</i> <sup>26</sup>	Semantic segmentation	CPP	Production	Unspecified/60
Wang, X. <i>et al.</i> <sup>32</sup>	Instance segmentation	Lab	Non-production	Unspecified/600
Li, D. J. <i>et al.</i> <sup>43</sup>	Instance segmentation	Lab	Non-production	200/200

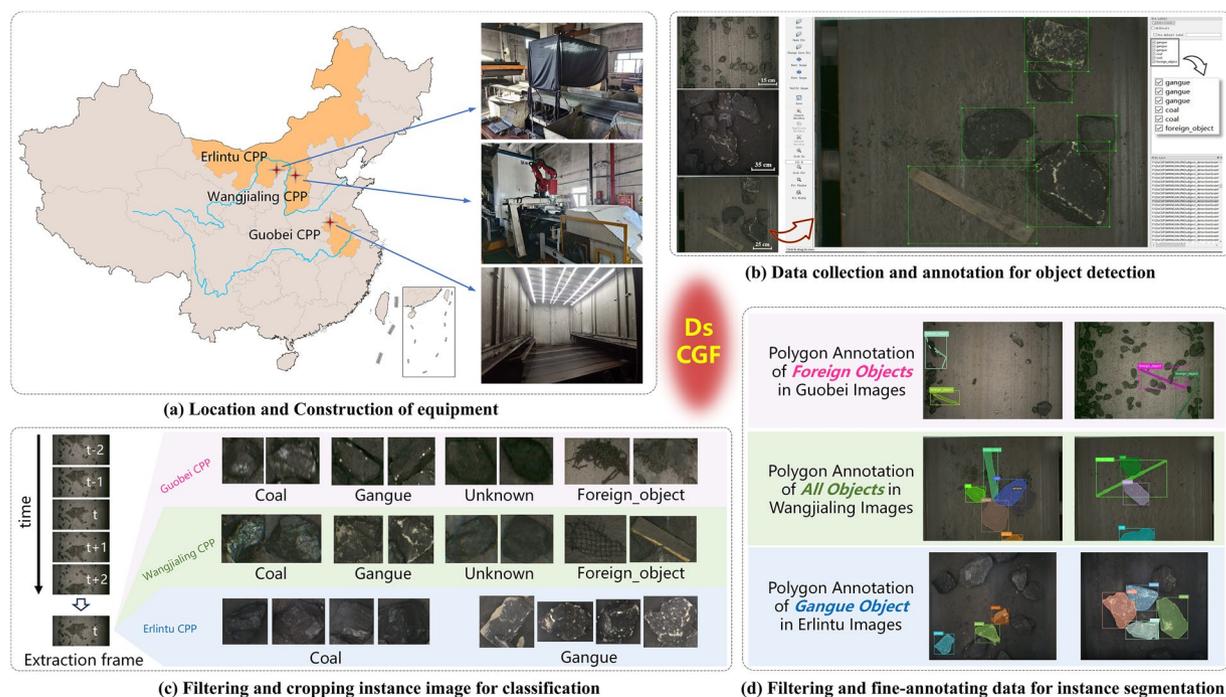
**Table 1.** A comparison of the datasets used in current studies.

Among the various input sources utilized in raw coal intelligent identification technology, visible-light images have garnered significant attention from researchers due to their low hardware deployment costs. In the initial stages of research, the primary approach was the “handcrafted feature engineering + machine learning classifier” model<sup>2,3</sup>. With the development of deep learning theory and the application of Convolutional Neural Networks (CNNs) in the field of computer vision, data-driven-based raw coal image analysis within the deep learning framework has gradually expanded, emerging as the mainstream approach. This progression mirrors the evolution of deep learning in the general field of computer vision. Numerous scholars have explored raw coal image datasets through classic algorithm frameworks, including image recognition, object detection, semantic segmentation, and instance segmentation<sup>4</sup>.

The classification of coal and gangue images using CNN is currently the most focused research area. This includes the direct introduction of mature CNN models<sup>5–7</sup>, improvements to the CNN framework based on established convolutional modules<sup>8,9</sup>, model light-weighting<sup>10,11</sup>, and edge deployment<sup>12,13</sup>. However, due to the absence of location information for targets within the images, image classification results cannot be directly used to guide automated sorting equipment for gangue removal. Nonetheless, these studies contribute to optimizing recognition results when cascaded with other task types<sup>14</sup>. Object detection is the most common task type in raw coal sorting technology field. Efficient object detection algorithms like the YOLO series<sup>15–18</sup> and RCNN series<sup>19–21</sup> are widely applied in raw coal image research, with some scholars attempting to enhance network structures for improved performance<sup>22–24</sup>. Furthermore, some researchers argue that the bounding boxes obtained from object detection are too coarse to meet the requirements of fine sorting. Therefore, discussions are increasingly focused on semantic segmentation<sup>25–28</sup> and even instance segmentation<sup>29–32</sup>. These advancements represent the latest direction in utilizing vision-guided devices to separate gangue and foreign objects from raw coal.

Currently, most related research aims to achieve high accuracy through algorithm optimization, but it is difficult to conduct substantial quantitative comparisons between different studies. Specifically, small-scale datasets fail to comprehensively represent the entire sample space, while the collection and annotation of large-scale datasets are costly. In existing research on coal-gangue image recognition, dataset sizes range from a few hundred to thousands of images (as shown in Table 1). To improve the accuracy of identifying material classes within raw coal, many researchers isolate individual materials from the coal stream for detailed examination. After categorizing each item, images are collected either at the end of the current production shift on the onsite conveyor belt or in laboratory settings after transporting the materials. These methods are collectively referred to as image acquisition under *non-production* conditions. Currently, most studies still rely on coal-gangue sample images collected in these settings. However, significant differences exist between materials in production and non-production states due to the unique environment and processes of coal preparation plants. Under *production* conditions, images of coal and gangue should be captured directly by positioning the camera over the raw coal conveyor belt. This approach enables the collection of material images without disrupting the continuous production flow or altering the materials’ surface state, ensuring consistency with the requirements of coal-gangue sorting systems. In contrast, materials examined under non-production conditions often experience surface changes, such as moisture evaporation, detachment of coal fines, and altered conveyor surfaces. These changes can lead to distortions in the materials’ surface state, reducing the relevance of such images for practical applications. Furthermore, raw coal from different mining areas exhibits different surface characteristics due to variations in coal quality, such as moisture content, sulfur content, and degree of coalification<sup>33</sup>. Although many scholars are dedicated to discussing the impact of different environmental factors on recognition performance under laboratory conditions, the uncertainty of the production environment (especially the uncertainty of material surfaces) leads to differences between the final imaging results and those simulated in the laboratory. In summary, the current studies use datasets of varying sizes and involve different levels of difficulty in material image recognition. These variations hinder the objective evaluation of findings across research teams and the effective quantitative comparison of optimized model metrics.

Therefore, our team developed the industry’s first large-scale, open-source image dataset for intelligent coal sorting, named DsCGF (Dataset for Coal, Gangue, and Foreign objects). This dataset is based on raw coal images collected over the past five years from typical mining regions in China, including Anhui, Inner Mongolia, and Shanxi. All data were collected in real production conditions and manually annotated, covering three typical



**Fig. 1** Overview of the construction process for the DsCGF.

task types: image classification, object detection, and instance segmentation, totaling nearly 270,000 visible light images (including 72,467 pieces of coal, 147,549 pieces of gangue, and 19,068 pieces of foreign objects). We hope this will accelerate the construction of intelligent mines in China.

## Methods

**The overview of workflow.** Firstly, CPPs were selected from three typical mining areas in China: Guobei CPP in Anhui Province, Erlintu CPP in Inner Mongolia Autonomous Region, and Wangjialing CPP in Shanxi Province. Image acquisition devices were installed at the manually sorting conveyor belt at these plants (Fig. 1a). Secondly, images of raw coal were continuously collected during daily production, and all images were manually annotated at the object detection level (Fig. 1b). When labeling the target object with a rectangular box, the edges of the rectangle tightly align with the target object to ensure that the rectangle is the minimum enclosing rectangle. Input the label name by connecting the two words with an underscore (i.e., `foreign_object`) instead of a space. Subsequently, on the premise of avoiding redundant sampling, and based on the object detection annotations, images were cropped. The cropped objects naturally formed a dataset for image classification tasks (Fig. 1c). Finally, considering the specific requirements for raw coal preprocessing at different CPPs, more refined instance segmentation annotations were conducted for various target objects (Fig. 1d).

**Sampling area and data sources.** The Huaibei mining area, located in the northwest of Anhui Province, China, is one of the country's main producers of coking coal and fat coal. The coal from this region is of excellent quality, characterized by low sulfur content ( $S < 1\%$ ), low phosphorus content ( $P < 0.05\%$ ), and high calorific value. The rated production capacity is 32.5 Mt/a. The Guobei CPP is one of the most representative coking coal preparation plants in this mining area, as it is a central coal preparation plant that needs to process raw coal from multiple mines within the area. We developed and installed an intelligent impurity-picking robot on the manual sorting belt at the Guobei CPP. Using its accompanying image acquisition device, we collected raw coal material with a particle size range of 25 mm to 100 mm. The image acquisition device utilizes a lighting system composed of 24 8 W strip LED lights, maintaining the illumination within the field of view at 1200 ( $\pm 100$ ) Lux. The camera used is the acA4096-40gc industrial camera from Basler, Germany, with a resolution of  $4096 \times 2168$  and a frame rate of 42 fps.

The Xinjie mining area, located in the western part of Inner Mongolia Autonomous Region, China, is a key mining area planned to enhance coal supply capacity, with a planned production capacity of 57.0 Mt/a. The main coal types produced here are long-flame coal and non-caking coal. Erlintu Coal Mine is a key mine in this area, and we chose its accompanying mine-type thermal-coal CPP as the data collection site for this mining area. After crushing and screening, raw coal larger than 300 mm enters the manual sorting belt, where an image acquisition device is installed for capturing images. Similar to the equipment used at the Guobei CPP, strip LED lights are used to maintain the illumination within the field of view at 1200 ( $\pm 100$ ) Lux. However, the camera used here is the acA4112-8gc industrial camera from Basler, Germany, with a resolution of  $4096 \times 3000$  and a frame rate of 8 fps.

The Hebaopian mining area, located in the northwest of Shanxi Province, China, was re-planned and adjusted in 2022, increasing its total production capacity to 104.9 Mt/a. The main coal types produced here are long-flame coal and gas coal. Similar to the Erlintu CPP, the Wangjialing CPP is a mine-type thermal-coal CPP supporting its coal mine. After crushing and screening, raw coal larger than 200 mm enters the manual sorting belt. We collected material images during production using an intelligent gangue-picking robot developed and installed by our team on this sorting belt. The robot's vision system is similar to the image acquisition devices used in the previous scenarios, maintaining the illumination within the area at 1000 ( $\pm 100$ ) Lux using LED spotlights. Images are captured by the acA1440-220uc industrial camera from Basler, Germany, with a resolution of 1440  $\times$  1080 and a frame rate of 227 fps.

**Labelling.** High-quality manual annotation is crucial for constructing a reliable dataset. To achieve this, we assembled a team of approximately 30 annotators, consisting of graduate students trained internally. Over the past five years, these annotators have undergone hands-on training at coal preparation plants. The training involves observing the surface characteristics of coal and gangue in real production settings alongside manual sorting belts, while simultaneously reviewing real-time images captured by industrial cameras to establish consistent annotation standards. After the training, annotators will undergo an identification assessment. Each annotator will be given 150 coal gangue images (75 of coal and 75 of gangue) for identification. The assessment will be considered passed if the number of errors in each category does not exceed 2 (with a total of no more than 4 errors). Annotators who successfully pass the coal-gangue identification assessment are then authorized to begin official annotation work. Notably, training alongside the conveyor belt allows annotators to inspect uncertain materials more closely, whereas the assessment is based solely on image identification. Although visual criteria for distinguishing coal from gangue are difficult to quantify, this comprehensive training process ensures that the annotation quality of the dataset is maintained to the highest possible standard.

The object detection task is currently the most commonly used approach for intelligent raw coal sorting. It annotates each instance in the image with bounding boxes and categories, providing both the location and category information of the targets. Therefore, we started with this task type. The DsCGF dataset includes four categories: coal, gangue, unknown, and foreign\_object. Due to the minimal visual differences between coal and gangue in industrial environments, materials that are difficult to classify in the images were uniformly labeled as unknown. This approach aligns with the established knowledge in the mineral processing field, where mixed coal with gangue naturally occurs and cannot be strictly classified as either coal or gangue. Therefore, we added the unknown category to facilitate flexible algorithm adjustments based on the coal preparation process. For example, in the pre-sorting stage, where the goal is to avoid coal being mixed with gangue, unknown can be treated as coal; in reverse sorting stages (such as recovering coal from gangue belts), unknown can be treated as gangue or coal. Additionally, foreign\_object refers to all items on the belt other than coal and gangue. This includes production waste such as anchor bolts, support nets, and wood bricks produced during coal seam mining, as well as living waste such as water bottles, discarded gloves, and protective equipment left by workers. The Labeling software was used for all raw coal image annotations at the object detection level, and the generated annotation information was saved in XML files in VOC format.

Single-label image classification is one of the most fundamental tasks in computer vision. Datasets for this task typically require each image to contain only one class of the primary target. This task is not only used for conventional classification but also for studying the visual feature differences between two types of substances, making the creation of image classification datasets very important. Our approach involves directly generating classification datasets using object detection annotation information. Specifically, we crop the target from the original image based on the bounding box information and then classify these cropped objects according to their corresponding category labels. We sample raw coal images at a rate of "one frame every 5 frames" from the continuously captured images. The 5-frame interval was determined based on the actual data collection situation. Specifically, in this dataset, regardless of the scene, selecting every 5th frame ensures that there is no overlapping area between the selected frames. There are two key considerations: First, within the same data volume, we aim to include a greater variety of distinct samples, rather than multiple views of the same sample from different angles (although such diversity is also valuable). Second, without a "sampling" approach to ensure the independence of instance samples, several issues may arise when studying image classification with these samples. For example, if a training batch consists solely of images of the same target object from different angles, using that batch for supervised learning could hinder the network's ability to learn generalized features. Therefore, we use only the sampled original images and their corresponding object detection annotations to create the image classification dataset.

Different processing workflows often require the intelligent sorting technology for raw coal to target different objects. Specifically, the Guobei CPP aims to remove foreign objects from raw coal on the manual sorting belt, the Erlintu CPP aims to remove large pieces of gangue on the manual sorting belt, and the Wangjialing CPP requires the simultaneous separation of both large pieces of gangue and miscellaneous objects. Accordingly, the DsCGF dataset provides refined instance segmentation annotations tailored to these different scenarios. Labelme software was used to create polygon annotations for miscellaneous objects at Guobei CPP, gangue at Erlintu CPP, and all objects on the belt at Wangjialing CPP. The generated annotation information is saved in COCO format within JSON files.

### Data Records

The generated DsCGF is available at the Zenodo<sup>34-36</sup>, and comprises raw coal images from three typical mining areas. Due to the capacity limit for data uploads on the Zenodo website, the entire DsCGF needs to be downloaded through three separate links:

Area	State	Task	Images	Category	Instances		
Anhui-Guobei	Production	Img-cls	108027	C	4877		
				G	96331		
				U	3669		
				F	3150		
	Production	Obj-det	12828	C	19625		
				G	385688		
				U	14735		
	Non-production	Ins-seg	12423	F	14771		
				Img-cls	100924	C	52764
						G	48160
Non-production	Obj-det	1322	C	52764			
			G	48160			
Inner Mongolia-Erlintu	Production	Img-cls	14841	C	13481		
				G	1360		
		Obj-det	8921	C	67134		
				G	6694		
		Ins-seg	3063	G	6694		
				C	1354		
Shanxi-Wangjialing	Production	Img-cls/Obj-det/Ins-seg	1200	G	1698		
				U	309		
				F	1147		
				C	1354		

**Table 2.** Details of sample in the DsCGF.

Part1: <https://doi.org/10.5281/zenodo.13268358>, including the DsCGF-1.zip (the image classification and object detection data of Anhui-Guobei).

Part2: <https://doi.org/10.5281/zenodo.13255355>, including the DsCGF-2.zip (a part of the instance segmentation data of Anhui-Guobei), the DsCGF-4.zip (the Inner Mongolian-Erlintu data) and the DsCGF-5.zip (the Shanxi-Wangjialing data).

Part3: <https://doi.org/10.5281/zenodo.13270674>, including the DsCGF-3.zip (the rest of the instance segmentation data of Anhui-Guobei).

The DsCGF includes manual annotations for all images, catering to the needs of image classification, object detection, and instance segmentation tasks. The specific data quantities are detailed in Table 2. Figure 2 provides a structured description of the dataset's file structure and annotation file format.

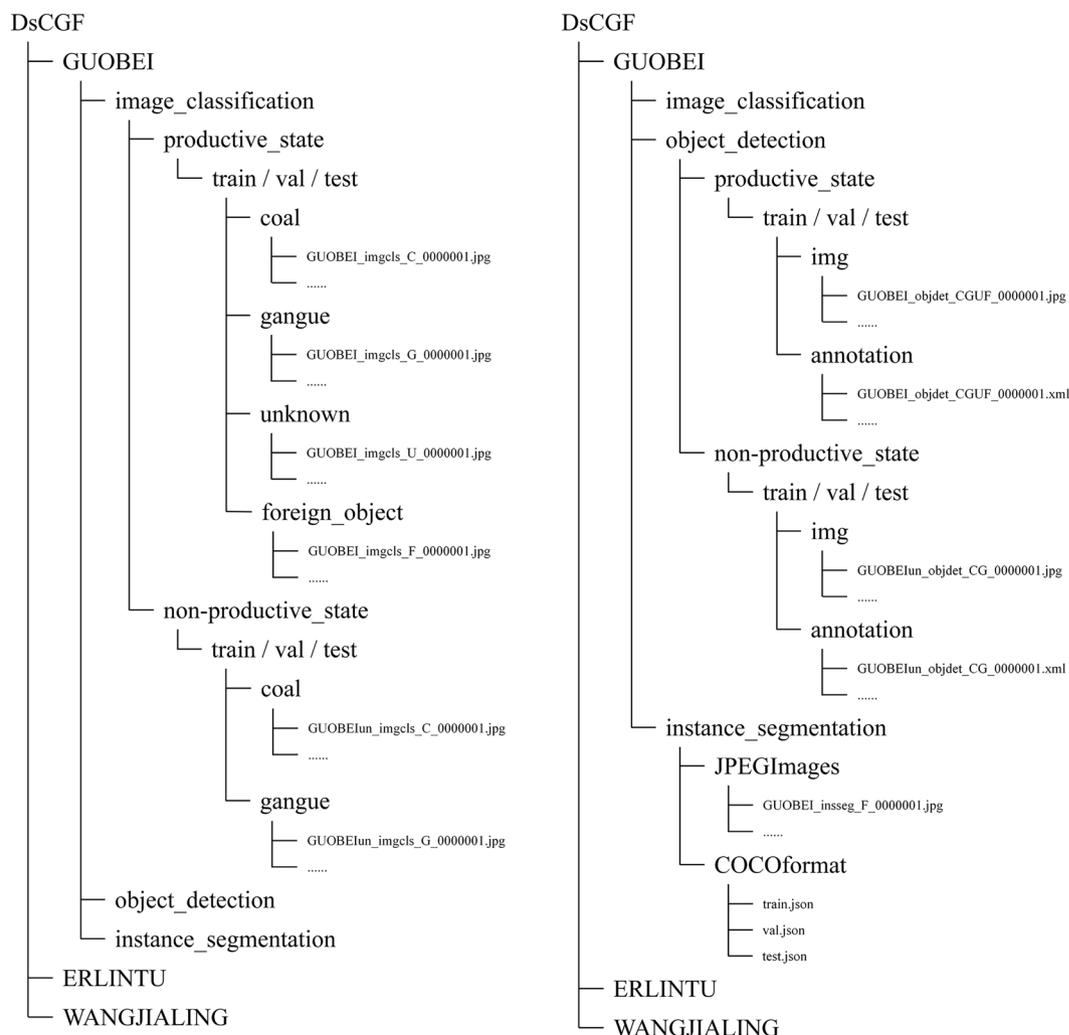
In Table 2, the “Task” column uses the following abbreviations: Img-cls stands for image classification, Obj-det stands for object detection, and Ins-seg stands for instance segmentation. In the “Category” column, C represents coal, G represents gangue, U represents unknown, and F represents foreign\_object. In the DsCGF dataset, we divided the data into training, validation, and test sets in a 6:2:2 ratio. The dataset division is based on the chronological order of image acquisition. All images were encoded and arranged in ascending order according to their collection sequence, and then split into training, validation, and test sets at a ratio of 6:2:2. Data leakage, referring to external factors that compromise the independence of training and test sets, is a critical consideration when constructing datasets. By strictly separating training and test sets based on the acquisition time, this issue can be effectively mitigated. Moreover, this approach ensures that the DsCGF dataset objectively reflects real-world conditions in the CPP, including potential class imbalances and temporal shifts in material surface characteristics. It should be noted that across different areas, the ratio of categories within the training, validation, and test sets remains relatively consistent. This prevents scenarios where a specific class is severely underrepresented in one subset, or where the class distribution between the training and test subsets differs significantly. An exception to this is the test subset for the Guobei-production-state dataset, which was collected independently. Detailed explanations for this exception are provided in the Usage Notes.

### Technical Validation

**Employing various advanced model to evaluating for different tasks.** In this paper, we carefully selected state-of-the-art (SOTA) approaches in each task domain and conducted comprehensive training and evaluation using the DsCGF dataset. The goal is to provide the industry with objective and comparable benchmarks.

Image classification, a fundamental task, often serves as the backbone for deep learning networks used in other tasks. In this paper, we chose SwinViT<sup>37</sup> for image classification tasks due to its robust feature representation capabilities demonstrated across various visual tasks in general domains. Additionally, its shifted window attention mechanism effectively captures both local details and global relationships, which is crucial for distinguishing subtle differences between fine-grained categories like “coal and gangue”.

The YOLO series, renowned for its single-stage object detection networks, is widely praised in both industry and academia for its strong performance and excellent scalability. YOLOv8<sup>38</sup> was selected for object detection



**Fig. 2** The schematic diagram of the DsCGF dataset structure (using Guobei CPP as an example). In the dataset, we archive data using the names of the coal preparation plants rather than the mining areas. This approach is intended to facilitate future expansions of the dataset.

tasks in this study due to its powerful multi-scale feature extraction capabilities, optimized head supporting finer classification and regression, and Focal Loss which alleviates the issue of class imbalance. These features make YOLOv8 particularly suitable for the DsCGF dataset.

For instance segmentation, we did not select a particularly novel network as the benchmark method. Instead, we chose a more established, robust network that has been validated across various domains and is easier to improve upon, making it more suitable as a benchmark. In this paper, the classic Mask-RCNN<sup>39</sup> was employed to handle the instance segmentation dataset in the DsCGF.

**Implementation details and metrics.** All experiments were carried out on NVIDIA GeForce RTX 4090 GPU by using the PyTorch framework. Details on hyper-parameter settings are provided in Table 3. Given the fine-grained nature of the coal and gangue targets, a conservative data augmentation strategy was adopted during model training. This strategy included horizontal flipping, vertical flipping, and random rotation—transformations that do not affect the imaging quality. Since color is a crucial intrinsic feature of raw coal materials, our preliminary research indicated that introducing color perturbations can improve bounding box regression in object detection tasks but negatively impact the classification of coal and gangue. Consequently, a conservative adjustment strategy was adopted in this study. Specifically, images were converted from the RGB color space to the HSV color space, where random perturbations within  $\pm 10\%$  of the original values were applied to the hue, saturation, and value channels. The results were then converted back to the RGB color space.

For the image classification task, we use the confusion matrix and F1 score as metrics. The DsCGF dataset involves up to four categories, so for a four-class scenario, the confusion matrix would appear as shown in Table 4.

To provide a comprehensive evaluation, the Micro-Averaged F1 Score<sup>40</sup> is calculated as follows: First, sum the True Positives (TP), False Positives (FP), and False Negatives (FN) across all categories. Then, compute the

Task	Image classification	Object detection	Instance segmentation
Model versions	SwinViT-base	YOLOv8-m	Mask-RCNN
Using pretrained weights	√ (ImageNet-1K)	√ (COCO-Train2017)	√ (COCO-Train2017)
Image size	224	416	800
Batch size	32	32	8
Epochs	50	100	100
Workers	4	4	1
Optimizer	Adam	SGD	SGD
Learning rate decrement method	None	Cosine annealing	Step (gamma = 0.1)
Learning rate	1E-4	5E-2 ~ 5E-4	1E-2
Weight decay	1E-4	1E-2	1E-4
Momentum	None	0.937	0.9
Data augmentation strategy	Horizontal flipping	Horizontal flipping	Horizontal flipping
	Vertical flipping	Vertical flipping	Vertical flipping
	Random rotating	Adjusting in HSV color space	Random rotating

**Table 3.** Hyper parameter setting during training.

	Predicted coal	Predicted gangue	Predicted unknown	Predicted foreign_object
Actual coal	$TP_{cc}$	$FP_{gc}$	$FP_{uc}$	$FP_{fc}$
Actual gangue	$FP_{cg}$	$TP_{gg}$	$FP_{ug}$	$FP_{fg}$
Actual unknown	$FP_{cu}$	$FP_{gu}$	$TP_{uu}$	$FP_{fu}$
Actual foreign_object	$FP_{cf}$	$FP_{gf}$	$FP_{uf}$	$TP_{ff}$

**Table 4.** The calculation method of the four-class confusion matrix involved in the DsCGF.

overall precision and recall, and finally calculate the F1 score. For example, considering the “coal” class as shown in Table 4:  $TP_c = TP_{cc}$ ,  $FP_c = FP_{gc} + FP_{cu} + FP_{cf}$ ,  $FN_c = TP_{gc} + FP_{uc} + FP_{fc}$ ,

$$\text{Micro } TP = \sum_{i=1}^C TP_i, \quad (1)$$

$$\text{Micro } FP = \sum_{i=1}^C FP_i, \quad (2)$$

$$\text{Micro } FN = \sum_{i=1}^C FN_i, \quad (3)$$

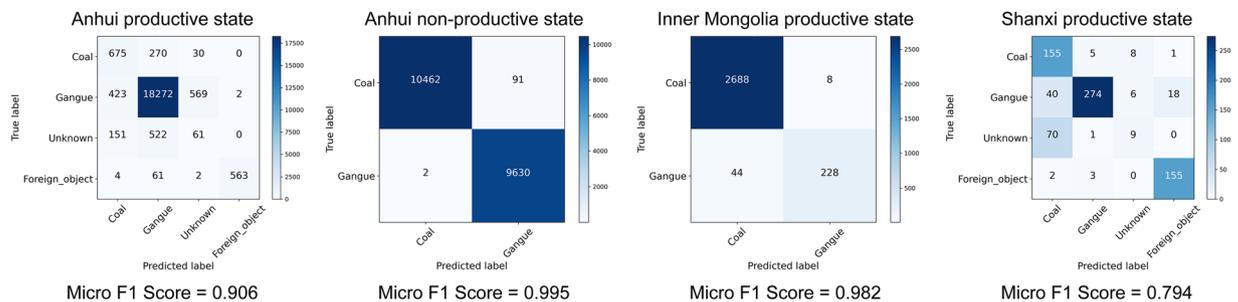
$$\text{Micro Precision} = \frac{\text{Micro } TP}{\text{Micro } TP + \text{Micro } FP}, \quad (4)$$

$$\text{Micro Recall} = \frac{\text{Micro } TP}{\text{Micro } TP + \text{Micro } FN}, \quad (5)$$

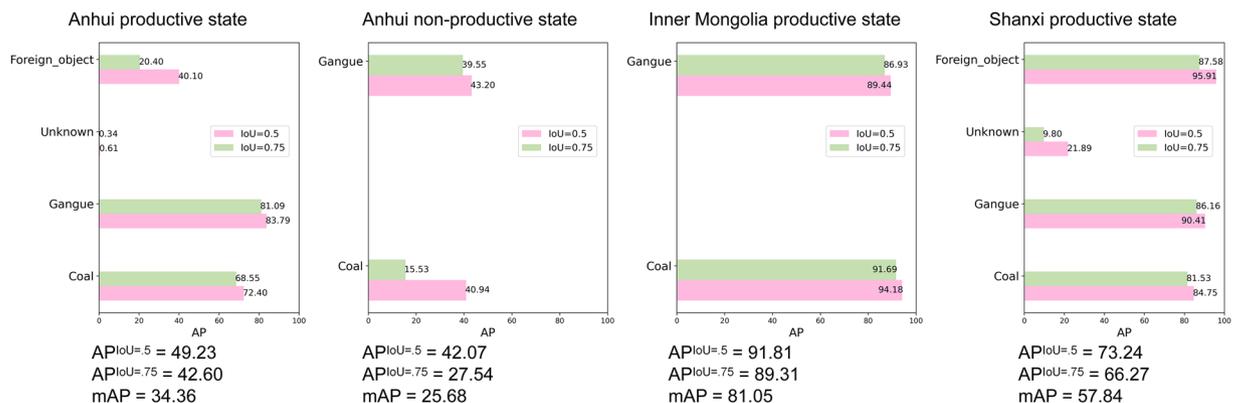
$$\text{Micro F1 Score} = 2 \cdot \frac{\text{Micro Precision} \cdot \text{Micro Recall}}{\text{Micro Precision} + \text{Micro Recall}}. \quad (6)$$

For the object detection task, the mean Average Precision (mAP) is a commonly used metric to measure the average precision of a detector across different categories. In this paper, we report the mAP calculated following the method described in the VOC dataset<sup>41</sup>. For the instance segmentation task, we adopt a series of evaluation metrics introduced in the COCO dataset<sup>42</sup>, including indicators related to average precision and average recall.

**Test results for the DsCGF.** Using the methods described in the past two section to validate the DsCGF, we obtained results on the test set. In four scenarios across the three mining regions, the SwinViT base model achieved Micro F1 Scores of 0.906, 0.995, 0.982, and 0.794 for image classification tasks (Fig. 3). The YOLOv8 medium model achieved mAP scores of 34.36%, 25.68%, 81.05%, and 57.84% for object detection tasks (Fig. 4). For the instance segmentation tasks specific to the three mining regions, Mask-RCNN demonstrates the ability to effectively segment numerous targets; however, its performance metrics show significant potential for



**Fig. 3** Confusion matrix results of classification tasks under four scenarios in the DsCGF.



**Fig. 4** The test results of object detection tasks under four scenarios in the DsCGF.

improvement. It is important to note in Fig. 5 that, due to the lack of strictly small targets (area  $< 32 \times 32$  pixels) in the instance segmentation dataset, only medium ( $32 \times 32$  pixels  $<$  area  $< 96 \times 96$  pixels) and large (area  $> 96 \times 96$  pixels) targets were considered when calculating metrics based on target area. As the metrics proposed by COCO can be calculated either based on bounding boxes or segmentation masks, two separate sets of outcomes are generated. Therefore, we present two rows of subplots in Fig. 5 to illustrate both sets of results. In the three scenarios, instances where more than ten targets appear simultaneously in a single image are rare. Therefore, in the subplots of the maxDets column in Fig. 5, the data at the 10 and 100 marks on the x-axis show almost no change. Several demos of test results were shown in Figs. 6, 7.

### Usage Notes

Additionally, we must highlight some specific considerations in the creation of the DsCGF to help future researchers understand and utilize the dataset effectively:

- In the coking coal samples collected from Anhui-Guobei, the quantity of gangue significantly exceeds that of coal. In contrast, the thermal coal samples collected from Inner Mongolia-Erlintu have a higher proportion of coal relative to gangue, resulting in a notable class imbalance. We consider this imbalance to be a common occurrence in practical production; therefore, we did not employ specialized methods (e.g., undersampling) to mitigate it.
- The coking coal samples from Anhui-Guobei are visually difficult to distinguish, and accurately annotating all gangue samples during continuous production presents significant challenges. To ensure the reliability of the test set, we collected material samples toward the end of each production day and re-collected images of these samples on the belt after they had been classified by skilled operators. The rationale for collecting samples towards the end of production is that if materials are removed from the belt and left for too long, their visual characteristics may change due to factors such as moisture, coal slurry, and dust. A total of 900 images were selected using these procedures to form the test set.
- Some coal and gangue samples from Anhui-Guobei were collected under non-production conditions. Initially, a batch of materials was obtained from the manual sorting belt during production, and each item was classified individually to determine its category. After the production shift ended, these materials were laid flat on the belt, and images were captured using a camera. While this method ensures complete accuracy in material classification, the extended period of on-site identification caused changes in the surface characteristics of the materials, such as the transition from being coated with coal slurry to becoming covered in dust. As a result, discrepancies were observed compared to the materials' appearance during actual production.

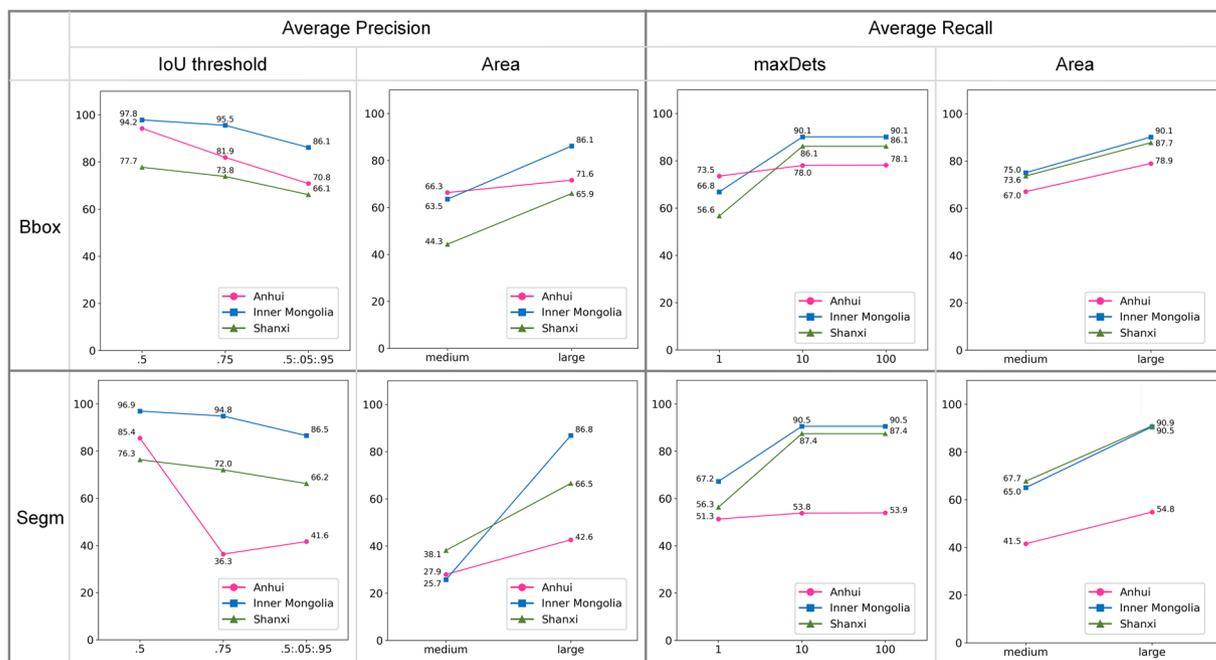


Fig. 5 The test results of instance segmentation tasks under three scenarios in the DsCGF.

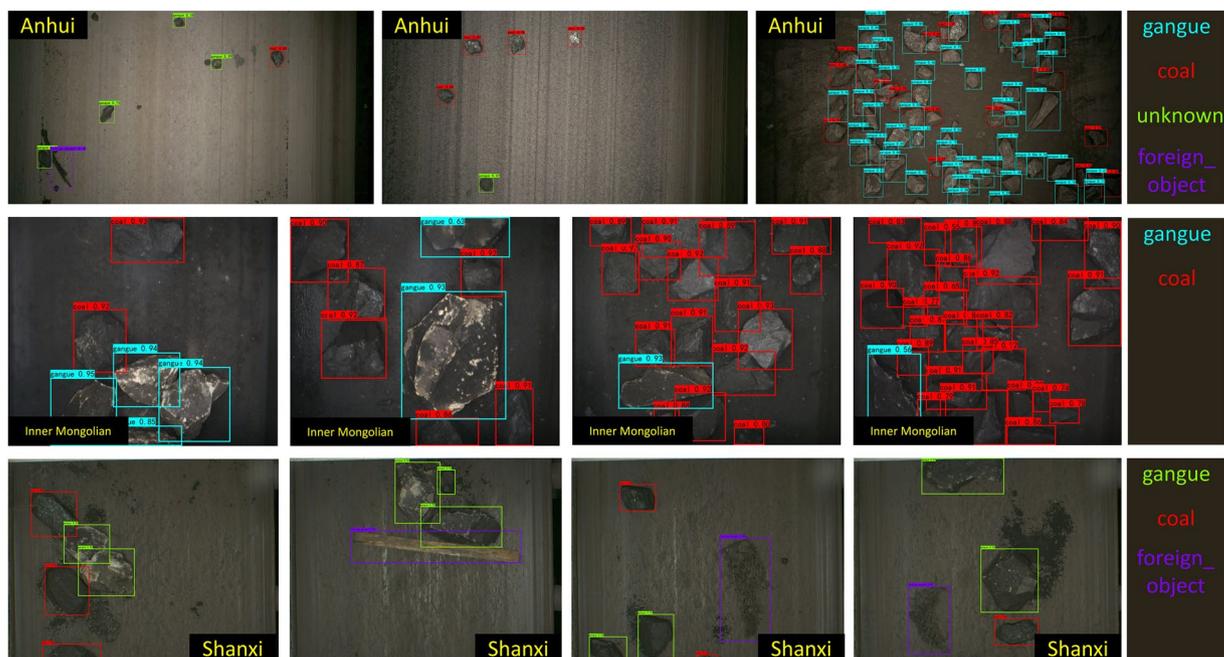


Fig. 6 The samples of test results of object detection.

- d) Foreign objects are infrequently encountered in the Guobei CPP's production. Therefore, the raw coal images used for instance segmentation in this scenario were not collected during the same period as those used for object detection. The collection of images containing foreign objects spanned a longer period.
- e) When creating the instance segmentation dataset for thermal coal from Inner Mongolia-Erlintu, images without gangue were excluded. Consequently, the number of images in the instance segmentation dataset is fewer than that in the object detection dataset.
- f) The Inner Mongolia-Erlintu dataset does not include examples of foreign objects or the "unknown" category, which is due to the specific conditions at the site. At the Erlintu CPP, foreign objects are almost entirely absent in the raw coal being processed, so no separate category for such materials was defined during dataset construction. Additionally, the coal and gangue in the raw coal processed at Erlintu CPP



**Fig. 7** The samples of test results of instance segmentation.

are particularly distinct due to the geological conditions of the mining area. This results in straightforward identification, with no materials that are difficult for the human eye to distinguish, thereby eliminating the need for an “unknown” category.

- g) In creating the image classification dataset for thermal coal from Shanxi-Wangjialing, the “one frame every 5 frames” sampling method described in the Labelling-section was not used due to insufficient quantity after sampling. This is purely because the data collection time in Shanxi was relatively short, resulting in a smaller dataset. However, this does not affect the quality of the data. Therefore, the dataset for Shanxi-Wangjialing is consistent across the three tasks, and we have combined them into a single column in Table 2. However, it should be noted that this leads to data duplication. We recommend that users train the model according to the default division method in our link (which avoids the issue of “test set leakage” caused by duplicated data). Additionally, when shuffling the data, please do not mix and shuffle the data crossing over the train/val/test folders. The impact of data duplication on model performance can be a scientific issue for users to explore in the future.
- h) When dividing the dataset according to the 6:2:2 ratio, the data for image classification tasks were proportionally allocated by category, while the data for object detection and instance segmentation tasks were proportionally allocated by image count. Due to the aforementioned reason (b), the object detection tasks for Anhui-Guobei data were not divided according to this ratio. Instead, images collected during continuous production were split into training and validation sets in a 7:3 ratio, with the 900 specifically annotated images mentioned in (b) designated as the test set.

Despite the significant focus on intelligent sorting and analysis technologies for raw coal in China’s energy sector, many fundamental and technical challenges remain unresolved. The purpose of publicly releasing the DsCGF dataset is to encourage more researchers to address these issues. By combining previous work with the experiments presented in this paper, we outline a series of research directions based on DsCGF that have the potential for breakthrough progress:

- i) In the Anhui-Guobei portion of the DsCGF dataset, there are two completely independent parts: Production and Non-Production. It is advised that users not mix data from these two parts indiscriminately, but instead systematically analyze the surface differences in terms of image-level features under these two states. Users are encouraged to conduct broader and deeper research based on data from these two scenarios, which is expected to promote advances in coal gangue image recognition algorithms, particularly in the area of directional feature extraction.
- j) Certain types of coal and gangue are inherently difficult to distinguish under actual production conditions. However, for the same dataset, image classification networks generally achieve more accurate coal-gangue recognition results compared to object detection and instance segmentation networks. Therefore, utilizing the Anhui-Guobei portion of the DsCGF dataset can facilitate an in-depth exploration of how to maintain model precision for fine-grained targets in complex tasks. Additionally, the “unknown” category of materials in the DsCGF can be leveraged to explore the potential of visual detection methods for identifying “inter-band coal” mixed within raw coal.

- k) Accurate instance segmentation of large gangue and foreign objects in raw coal is essential for sorting using mechanical arms or other actuators. Currently, research in this area remains insufficient. The Inner Mongolian-Erlintu instance segmentation dataset within the DsCGF can be used to explore ways to enhance pixel-level segmentation accuracy. The Anhui-Guobei instance segmentation dataset can be leveraged to study the precise segmentation of irregular foreign objects under challenging conditions, such as the presence of complex coal slime and water stains on the conveyor belt. Additionally, it could serve as a basis for exploring the detection and segmentation of “rare foreign objects” through incremental learning approaches. The Shanxi-Wangjialing instance segmentation dataset can be applied to investigate specific object grasping strategies under various material stacking and distribution conditions.
- l) The release of the DsCGF dataset, which includes visible light images of raw coal from multiple regions, provides a valuable opportunity to study the properties of raw coal across diverse geographical areas from a visible light perspective. This has the potential to stimulate broader application research for intelligent sorting algorithms. Researchers can leverage the large-scale data in DsCGF to train pre-trained weights for custom models and conduct transfer learning using smaller datasets from new scenarios. This may involve exploring weakly supervised, semi-supervised, or even unsupervised deep learning techniques for intelligent coal sorting tasks. Furthermore, in-depth analysis of the characteristics of raw coal from different regions and types, using a “large model” approach, can contribute to advancements in cross-regional intelligent sorting technologies. It is important to note, however, that the imaging conditions across the three scenarios in the dataset differ. While this variability provides an opportunity to evaluate the cross-scenario generalization ability of trained models, caution is necessary when using DsCGF’s coal and gangue images to study inter-regional or inter-coal-type image characteristics. Specifically, spurious correlations may arise due to differences in lighting and color temperature, which can affect color features more than texture features. To minimize these influences, appropriate image preprocessing should be applied to mitigate the impact of imaging conditions on the analysis.

### Code availability

The codes of deep learning approaches in this study are shared at the Github: SwinVIT and YOLOv8: <https://github.com/CUMTB-IMP-Lab/the-verification-of-the-DsCGF>. Mask-RCNN: [https://github.com/WZMIAOMIAO/deep-learning-for-image-processing/tree/master/pytorch\\_object\\_detection/mask\\_rcnn](https://github.com/WZMIAOMIAO/deep-learning-for-image-processing/tree/master/pytorch_object_detection/mask_rcnn). Users only need to adjust the hyperparameters provided in Table 3 within the open-source code referenced in the three links, download the dataset and trained weights, and proceed with model training or testing to reproduce the results presented in this paper. No modifications to the model architecture are required.

Received: 1 October 2024; Accepted: 27 February 2025;

Published online: 08 March 2025

### References

- Liu, Y. & He, Z. C. Synergistic industrial agglomeration, new quality productive forces and high-quality development of the manufacturing industry. *International Review of Economics & Finance* **94**, <https://doi.org/10.1016/j.iref.2024.103373> (2024).
- Wang, W. D., Lv, Z. Q. & Lu, H. R. Research on methods to differentiate coal and gangue using image processing and a support vector machine. *International Journal of Coal Preparation and Utilization* **41**, 603–616, <https://doi.org/10.1080/19392699.2018.1496912> (2021).
- Zhang, L., Sui, Y. P., Wang, H. S., Hao, S. K. & Zhang, N. B. Image feature extraction and recognition model construction of coal and gangue based on image processing technology. *Scientific Reports* **12**, <https://doi.org/10.1038/s41598-022-25496-5> (2022).
- Yin, J. Q. *et al.* Intelligent photoelectric identification of coal and gangue - A review. *Measurement* **233**, <https://doi.org/10.1016/j.measurement.2024.114723> (2024).
- Pu, Y. Y., Apel, D. B., Szmigiel, A. & Chen, J. Image Recognition of Coal and Coal Gangue Using a Convolutional Neural Network and Transfer Learning. *Energies* **12**, <https://doi.org/10.3390/en12091735> (2019).
- Su, L., Cao, X., Ma, H. & Li, Y. Research on coal gangue identification by using convolutional neural network. in *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. 810–814 (IEEE).
- Liu, Q., Li, J. G., Li, Y. S. & Gao, M. W. Recognition Methods for Coal and Coal Gangue Based on Deep Learning. *Ieee Access* **9**, 77599–77610, <https://doi.org/10.1109/access.2021.3081442> (2021).
- Liu, Y., Zhang, Z. L., Liu, X., Wang, L. & Xia, X. H. Deep Learning Based Mineral Image Classification Combined With Visual Attention Mechanism. *Ieee Access* **9**, 98091–98109, <https://doi.org/10.1109/access.2021.3095368> (2021).
- Liu, H. J. & Xu, K. Recognition of gangues from color images using convolutional neural networks with attention mechanism. *Measurement* **206**, <https://doi.org/10.1016/j.measurement.2022.112273> (2023).
- Xu, Z. Q., Lv, Z. Q., WANG, W. D., Zhang, K. H. & Lv, H. M. Machine vision recognition method and optimization for intelligent separation of coal and gangue. *Journal of China Coal Society* **45** (2020).
- Cao, Z. G. *et al.* Research on image classification of coal and gangue based on a lightweight convolution neural network. *Energy Science & Engineering* **11**, 3042–3054, <https://doi.org/10.1002/ese3.1501> (2023).
- Wu, J. *et al.* A Lightweight Small Object Detection Method Based On Multi-Layer Coordination Federated Intelligence for Coal Mine IoVT. *IEEE Internet of Things Journal* (2024).
- Yan, P. C. *et al.* Lightweight detection method of coal gangue based on multispectral and improved YOLOv5s. *International Journal of Coal Preparation and Utilization* **44**, 399–414, <https://doi.org/10.1080/19392699.2023.2193739> (2024).
- Lv, Z. Q., Wang, W. D., Xu, Z. Q., Zhang, K. H. & Lv, H. M. Cascade network for detection of coal and gangue in the production context. *Powder Technology* **377**, 361–371, <https://doi.org/10.1016/j.powtec.2020.08.088> (2021).
- Pan, H. G., Shi, Y. H., Lei, X. Y., Wang, Z. & Xin, F. F. Fast identification model for coal and gangue based on the improved tiny YOLO v3. *Journal of Real-Time Image Processing* **19**, 687–701, <https://doi.org/10.1007/s11554-022-01215-1> (2022).
- Xu, S. Y., Zhou, Y. J., Huang, Y. R. & Han, T. YOLOv4-Tiny-Based Coal Gangue Image Recognition and FPGA Implementation. *Micromachines* **13**, <https://doi.org/10.3390/mi13111983> (2022).
- Chen, K. Y., Du, B., Wang, Y. W., Wang, G. X. & He, J. X. The real-time detection method for coal gangue based on YOLOv8s-GSC. *Journal of Real-Time Image Processing* **21**, <https://doi.org/10.1007/s11554-024-01425-9> (2024).

18. Yang, Y. H. *et al.* Research on coal gangue recognition method based on XBS-YOLOv5s. *Measurement Science and Technology* **35**, <https://doi.org/10.1088/1361-6501/acfab1> (2024).
19. Li, D. J. *et al.* An Image-Based Hierarchical Deep Learning Framework for Coal and Gangue Detection. *Ieee Access* **7**, 184686–184699, <https://doi.org/10.1109/access.2019.2961075> (2019).
20. Zhang, Y. C., Wang, J. S., Yu, Z. W., Zhao, S. & Bei, G. X. Research on intelligent detection of coal gangue based on deep learning. *Measurement* **198**, <https://doi.org/10.1016/j.measurement.2022.111415> (2022).
21. Lai, W. H., Hu, F., Kong, X. X., Yan, P. C. & Dai, X. X. The study of coal gangue segmentation for location and shape predicts based on multispectral and improved Mask R-CNN. *Powder Technology* **407**, <https://doi.org/10.1016/j.powtec.2022.117655> (2022).
22. Li, D. Y., Wang, G. F., Zhang, Y. & Wang, S. Coal gangue detection and recognition algorithm based on deformable convolution YOLOv3. *Iet Image Processing* **16**, 134–144, <https://doi.org/10.1049/ipr2.12339> (2022).
23. Li, M., He, X. L., Yuan, Y. X. & Yang, M. L. Multiple factors influence coal and gangue image recognition method and experimental research based on deep learning. *International Journal of Coal Preparation and Utilization* **43**, 1411–1427, <https://doi.org/10.1080/19392699.2022.2118260> (2023).
24. Zeng, Q. L. *et al.* Detection of Coal and Gangue Based on Improved YOLOv8. *Sensors* **24**, <https://doi.org/10.3390/s24041246> (2024).
25. Liu, X. *et al.* A Diverse Environment Coal Gangue Image Segmentation Model Combining Improved U-Net and Semi-supervised Automatic Annotation. in *International Conference on Cognitive Systems and Signal Processing*. 167–179 (Springer).
26. Gao, R. *et al.* Automatic coal and gangue segmentation using u-net based fully convolutional networks. *Energies* **13**, 829 (2020).
27. Liu, Y., Zhang, Z. L., Liu, X., Wang, L. & Xia, X. H. Efficient image segmentation based on deep learning for mineral image classification. *Advanced Powder Technology* **32**, 3885–3903, <https://doi.org/10.1016/j.apt.2021.08.038> (2021).
28. Wang, X. *et al.* Rapid detection of incomplete coal and gangue based on improved PSPNet. *Measurement* **201**, <https://doi.org/10.1016/j.measurement.2022.111646> (2022).
29. Cao, Z. G. *et al.* Research on coal and gangue segmentation based on MFCCM-Mask R-CNN. *Energy Science & Engineering* **12**, 2958–2973, <https://doi.org/10.1002/ese3.1797> (2024).
30. Lv, Z. Q. *et al.* A synchronous detection-segmentation method for oversized gangue on a coal preparation plant based on multi-task learning. *Minerals Engineering* **187**, <https://doi.org/10.1016/j.mineng.2022.107806> (2022).
31. Lv, Z. Q. *et al.* A high-confidence instance boundary regression approach and its application in coal-gangue separation. *Engineering Applications of Artificial Intelligence* **132**, <https://doi.org/10.1016/j.engappai.2024.107894> (2024).
32. Wang, X. *et al.* Multi-scale coal and gangue detection in dense state based on improved Mask RCNN. *Measurement* **221**, <https://doi.org/10.1016/j.measurement.2023.113467> (2023).
33. Lv, Z. Q. *et al.* Investigating comparisons on the coal and gangue in various scenarios using multidimensional image features. *Minerals Engineering* **204**, <https://doi.org/10.1016/j.mineng.2023.108450> (2023).
34. Lv, Z. Q., Wang, W. D. & Xu, Z. Q. DsCGF: A large-scale open image dataset for deep learning enabled intelligent sorting and analyzing of raw coal (part 1). *Zenodo* <https://doi.org/10.5281/zenodo.13268358> (2024).
35. Lv, Z. Q., Wang, W. D. & Xu, Z. Q. DsCGF: A large-scale open image dataset for deep learning enabled intelligent sorting and analyzing of raw coal (part 2). *Zenodo* <https://doi.org/10.5281/zenodo.13255355> (2024).
36. Lv, Z. Q., Wang, W. D. & Xu, Z. Q. DsCGF: A large-scale open image dataset for deep learning enabled intelligent sorting and analyzing of raw coal (part 3). *Zenodo* <https://doi.org/10.5281/zenodo.13270674> (2024).
37. Liu, Z. *et al.* Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. in *18th IEEE/CVF International Conference on Computer Vision (ICCV)*. 9992–10002 (2021).
38. Varghese, R. & Sambath, M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. 1–6 (IEEE).
39. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. in *Proceedings of the IEEE international conference on computer vision*. 2961–2969.
40. Takahashi, K., Yamamoto, K., Kuchiba, A. & Koyama, T. Confidence interval for micro-averaged F1 and macro-averaged F1 scores. *Applied Intelligence* **52**, 4961–4972, <https://doi.org/10.1007/s10489-021-02635-5> (2022).
41. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* **88**, 303–338, <https://doi.org/10.1007/s11263-009-0275-4> (2010).
42. Lin, T.-Y. *et al.* Microsoft coco: Common objects in context. in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V* **13**. 740–755 (Springer).
43. Li, D., Meng, G., Sun, Z., Xu, L. & Cui, W. Weight Estimation Method of Coal and Gangue on Conveyor Belt Based on Instance Segmentation. in *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*. 432–435.

## Acknowledgements

This work was supported by “the Fundamental Research Funds for the Central Universities” (2024XJHH02). The author would like to thank the DADI ENGINEERING DEVELOPMENT CO.,LTD and the Tianjin Branch&Coal Preparation Design Institute for help in this work. We are especially grateful to all those who participated in the dataset construction and label annotation process.

## Author contributions

Ziqi Lv mainly be responsible for the overall planning for organize dataset, writing the manuscript and programming the codes, Yuhan Fan responsible for revising the manuscript and maintenance dataset, Te Sha responsible for abnormal data filtering and partial coding work, Yao Cui, Yuxin Wu and Haimei Lv in charge of collect dataset pictures and the dataset label annotation, manuscripts of the dataset were reviewed by Meijie Sun, Yanan Tu, Zhiqiang Xu and Weidong Wang.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04719-0>.

**Correspondence** and requests for materials should be addressed to Z.L. or W.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025