

Article

Visual Recognition of Coal–Biomass Blend Ratios on a Conveyor Belt Using YOLO-Series Models with Oriented Bounding Boxes

Yisheng Mao ¹, Huijin Yang ², Cuihua Zhang ³, Weihui Liao ³, Zhilong Ruan ³, Haibin Pu ³, Xu Huang ¹, Xiaolong Wu ⁴ and Zhimin Lu ^{2,5,*} 

¹ Guangdong Electric Power Development Co., Ltd., Guangzhou 510630, China; 13702779665@163.com (Y.M.); hx13802671460@163.com (X.H.)

² School of Electric Power, South China University of Technology, Guangzhou 510640, China; 202521016195@mail.scut.edu.cn

³ Guangdong Red Bay Power Generation Co., Ltd., Shanwei 516623, China; ydsw@vip.163.com (C.Z.); mxpplwh@126.com (W.L.); 18927908686@163.com (Z.R.); phb1987@126.com (H.P.)

⁴ Xian Thermal Power Research Institute Co., Ltd., Xi'an 710032, China; wuxiaolong@tpri.com.cn

⁵ State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou 310027, China

* Correspondence: zhmlu@scut.edu.cn

Abstract

Real-time perception of coal–biomass blending during conveyor-belt transport remains challenging because of local aggregation, particle overlap, and illumination variation. In this study, a laboratory-scale conveyor-belt image dataset covering different coal mass fractions, illumination conditions, and particle sizes was constructed. Whole-image classification, cropped-ROI classification, direct regression, horizontal bounding box (HBB)-based detection, oriented bounding box (OBB)-based detection, and RT-DETR-L detection baselines were compared using YOLO-series and auxiliary models. Coal mass fraction was estimated using a frequency-weighted statistical strategy that converts frame-level predictions into continuous estimates. YOLOv8-cls achieved an average RMSE of 13.98 percentage points (pp), indicating the influence of background interference in whole-image classification. Among HBB models, YOLOv8m achieved the lowest mean RMSE of 6.10 pp but required higher computational cost. Compared with YOLOv8n, YOLOv8n-OBB reduced the average RMSE from 9.02 to 6.90 pp by providing a more compact material-region representation and reducing background redundancy. These results show that OBB representation improves the stability of lightweight models. The proposed method provides a feasible vision-based soft-sensing approach for online trend monitoring of coal–biomass blending under lightweight deployment.

Keywords: coal–biomass blending; vision-based monitoring; blend-ratio estimation; YOLOv8; oriented bounding box; conveyor belt; co-firing



Academic Editor: Albert Ratner

Received: 26 May 2026

Revised: 9 June 2026

Accepted: 15 June 2026

Published: 18 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In coal-fired power plants with coal–biomass co-firing, the actual blending state of coal and biomass directly affects feeding stability, in-furnace combustion, blend-ratio regulation, and pollutant-emission control. Biomass fuels are widely available, renewable, and associated with potential reductions in fossil–carbon emissions. When blended with coal at an appropriate ratio, they provide a feasible pathway for low-carbon and flexible operation of coal-fired units. Unlike single-fuel conveying, however, coal–biomass mixtures on a conveyor belt commonly exhibit random spreading, local aggregation, and non-uniform spatial distribution. If the blend ratio and its fluctuations cannot be perceived in

time during conveying, subsequent boiler operation usually relies on delayed combustion feedback or offline sampling. Therefore, an online sensing method capable of continuously reflecting the coal–biomass blending state during fuel transport is of practical significance for co-firing process control.

Fuel composition and blending state are commonly assessed by manual visual inspection, offline sampling and analysis, or indirect inference from post-combustion behavior. These approaches suffer from delayed response, subjectivity, insufficient timeliness, and limited capability for continuous online feedback. They therefore do not fully meet the need for rapid perception and online regulation of fuel transport in practical power-plant operation. By contrast, machine-vision-based image recognition is non-contact, fast, relatively low-cost, and readily integrated with conveyor belts, silo outlets, and upstream feeding systems. In conveyor-belt imaging scenes, however, a camera does not usually capture individual coal or biomass particles, but rather a continuously distributed material region composed of mixed coal and biomass particles. This region is affected by spreading state, local stacking, occlusion, particle-size differences, and illumination variation, resulting in substantial fluctuations in color, texture, and morphological features. Moreover, the visible surface composition in a single frame does not necessarily correspond exactly to the overall mass fraction of the prepared sample. Coal–biomass blend recognition is therefore not a simple image-classification task, but an industrial vision-based soft-sensing problem that involves bulk-material region representation, robustness under complex operating conditions, and continuous ratio estimation.

For biomass-fuel image analysis, previous studies have shown that image features such as color, texture, and particle morphology can effectively support biomass-fuel classification, quality assessment, and process-variable estimation. Lukas et al. compared conventional image processing and deep learning for characterizing biomass fuels and their mixtures, noting that color and texture remain important discriminative information for fuel image recognition, whereas deep learning offers stronger feature-learning potential in complex mixed scenes [1]. Ögren et al. further applied vision recognition to fuel-feeding-rate estimation, demonstrating that image outputs can be extended from static category recognition to continuous process-variable estimation [2]. Plankenbühler et al. integrated online fuel image analysis with a biomass power-plant control system, indicating that visual analysis can provide front-end sensing information for adjusting boiler-control parameters [3]. Gudavalli et al. focused on biomass-particle quality detection and emphasized that variations in particle morphology and texture are closely associated with subsequent conveying and handling stability [4]. These studies indicate that fuel image analysis has the potential to evolve from fuel-type identification toward process-state monitoring.

In coal-industry visual recognition, deep learning has been widely applied to coal-gangue recognition, coal-content estimation, and intelligent sorting systems. Zhang et al. developed a deep-learning-based coal-gangue detection model and verified the effectiveness of object detection for recognizing coal-related bulk materials [5]. Zhang et al. further used machine vision and a genetic algorithm-backpropagation neural network (GA-BPNN) model to estimate coal content in gangue, showing that visual information can be used not only for category discrimination but also for mapping to continuous content variables [6]. Liu and Xu introduced an attention mechanism into color-image recognition of gangue to enhance the model's focus on key local and channel features [7]. With the development of real-time detection models such as YOLOv8, recent studies have increasingly focused on model optimization under complex illumination, low-light conditions, small targets, dense targets, weakly supervised detection, and lightweight deployment [8–12]. More recent work further shows that coal visual recognition is shifting from offline algorithm validation toward complex-condition adaptation, dataset construction, and lightweight

deployment. Zhang et al. proposed a one-stage coal-gangue detector for real industrial applications; Cao et al. investigated lightweight coal-gangue object detection; Lv et al. built a large-scale raw-coal image dataset for deep learning, and Cao et al. further explored compact, high-performance coal-gangue detection networks [13–16]. These studies suggest that visual recognition of coal-related bulk materials is moving from single classification or detection verification toward coordinated optimization of data, models, and deployment for engineering applications.

Although these studies provide a methodological basis for coal–biomass blend recognition, the problem addressed in this work differs from existing fuel classification, coal-gangue recognition, and ordinary object-detection tasks. First, coal–biomass blend-ratio labels have ordered and continuous properties, and the visual differences between adjacent blend ratios are small. Whole-image classification may therefore be affected by the conveyor-belt background, blank regions, local spreading variations, and illumination changes. Second, the effective target in conveyor-belt images is not an individual particle but a coal–biomass material region continuously distributed along the conveying direction; the localization method and geometric representation of the target region directly affect the stability of ratio-label discrimination. Accordingly, YOLOv8-cls was first used as a whole-image classification baseline to verify the necessity of explicit material-region localization for coal mass-fraction estimation. To further improve the rationality of task formulation and model comparison, cropped-ROI classification, direct regression, and RT-DETR-L detection baselines were also introduced. Then, within the YOLO object-detection framework, different model scales and two bounding-box representations, namely HBB and OBB, were compared in terms of detection performance, blend-ratio estimation error, and deployment cost.

YOLO-series models are characterized by single-stage detection, end-to-end inference, and good real-time performance, and they have been widely used in industrial visual detection [17–19]. However, standard YOLO detection models usually describe target locations using horizontal bounding boxes (HBBs). HBBs can satisfy general detection requirements when target shapes are regular and orientation changes are small. In coal–biomass bulk-material images, however, biomass particles are often elongated or flaky, coal particles are typically irregular blocks, particle orientations are random, and local overlap is pronounced. HBBs, therefore, tend to include irrelevant background regions, reducing localization compactness and the stability of category discrimination. Oriented bounding boxes (OBBs) introduce angular information and can more tightly fit targets with directional or elongated shapes; they have been widely applied in remote-sensing and aerial-image object detection [20,21]. The Ultralytics OBB detection task also indicates that oriented boxes are better suited for describing rotated objects and directional target regions [22]. Therefore, this study introduces OBBs into coal–biomass blend recognition to examine whether a more compact region representation can reduce background redundancy and improve the stability of ratio estimation for irregular bulk-material regions.

The main contributions of this study are as follows:

- (1) A coal–biomass mixture image dataset was constructed covering coal mass fractions from 0% to 100% at 5% intervals, yielding 21 classes. The dataset also covers morning natural light, nighttime infrared supplementary illumination, and different particle-size conditions, providing an experimental basis for mixed-fuel ratio recognition.
- (2) A comparative experimental framework was established, including whole-image classification, cropped-ROI classification, direct regression, HBB detection, OBB detection, and RT-DETR-L detection baselines, to evaluate the effects of task formulation, material-region localization, and model type on blend-ratio estimation.

- (3) A frequency-weighted statistical strategy based on predicted ratio labels was adopted to convert discrete single-frame prediction labels into continuous coal-fraction estimates under each test condition, and repeated training with different random seeds was used to evaluate the stability of core models.
- (4) The effects of model scale, bounding-box representation, illumination condition, and particle-size variation on ratio-recognition stability were analyzed. Feature-distribution differences between natural light and infrared images, representative error cases, and occlusion-based interpretability results were further examined to identify error sources and directions for subsequent improvement under complex operating conditions.

2. Materials and Methods

2.1. Visual Characteristics of Coal and Biomass Particles

Figure 1 shows the typical visual characteristics of coal, biomass, and mixed samples with different coal mass fractions in conveyor-belt images. The coal and biomass samples used in the experiments were both obtained from Guangdong Red Bay Power Generation Co., Ltd. (Shanwei, China); therefore, their color, morphology, particle size, and surface texture reflect material-appearance differences that may occur during practical conveying. Single biomass samples were generally lighter in color, and most particles appeared as short rods or strips with relatively continuous surface textures. Single coal samples were darker, had more irregular shapes, and showed more obvious local brightness variations and fractured edges. As indicated by the scale bars in Figure 1, the displayed coal particles ranged from small irregular fragments of approximately 10 mm to larger blocks of about 50 mm, with some particles approaching 80 mm, whereas biomass particles were mainly rod-like fragments with typical lengths of approximately 30–50 mm. As the coal mass fraction increased, the dominant visual features gradually shifted from light-colored strip-like biomass particles to dark irregular coal particles, with continuous changes in apparent color, texture, local morphology, and particle-size distribution.



Figure 1. Typical visual characteristics of coal, biomass, and blended samples under different blending states: (a) biomass particles with low stacking density; (b) biomass particles with high stacking density; (c) coal particles with low stacking density; (d) coal particles with high stacking density; (e) low-coal-fraction blended sample; (f) medium-coal-fraction blended sample; (g) high-coal-fraction blended sample. Scale bars are shown in each panel.

In the intermediate ratio range, coal and biomass particles are interlaced in color, texture, size, and spatial distribution, and their local visual contributions become similar. Therefore, adjacent ratio labels may exhibit only subtle visual differences, especially under local stacking, occlusion, and uneven surface exposure. These characteristics indicate that coal–biomass blend recognition is not a simple color-based or binary classification problem; instead, it is a complex bulk-material vision task affected by particle morphology, particle-size variability, accumulation density, target orientation, local occlusion, and continuous ratio variation.

2.2. Experimental Design and Image-Sample Construction

The coal–biomass blend-ratio recognition experiments were conducted using a laboratory-scale conveyor-belt platform. The experimental system consisted mainly of a conveyor belt, a feeding area, a dual-spectrum industrial camera with white-light/infrared supplemental illumination (DS-2XE6446FWD-WBRG (2.8–12 mm)/ /316L, Hangzhou Hikvision Digital Technology Co., Ltd., Hangzhou, China), and an image-acquisition terminal. The camera was fixed obliquely above the conveyor belt, and its imaging area covered the entire belt. During the experiments, coal and biomass were weighed according to preset mass ratios and then fed onto a feeding area, where the two materials formed randomly spread mixed states during feeding and transport.

To construct image samples for coal–biomass blend-ratio recognition, images were first collected under different coal mass fractions and illumination conditions. The coal mass-fraction labels ranged from 0% to 100% at 5% intervals, corresponding to 21 classes. The 5% interval was selected as a compromise between experimental controllability, engineering resolution, and visual distinguishability between adjacent blend ratios. Because the visual differences between adjacent 5% ratio labels can be subtle, especially under local stacking, occlusion, and non-uniform surface exposure, the discrete ratio labels were not used as final single-frame control outputs. Instead, multi-frame frequency-weighted statistics were used in the testing stage to obtain continuous coal-fraction estimates, thereby reducing the influence of single-frame randomness and label ambiguity. The illumination conditions included morning natural light and nighttime infrared supplementary illumination. For each coal-fraction condition, three independent feeding and video-acquisition trials were conducted, and the materials were reorganized between trials to reduce repeated spatial distributions and local spreading randomness. Supplementary particle-size tests were further performed at representative coal mass fractions of 30%, 50%, and 70%, and these samples were used only for operating-condition analysis rather than model training or validation.

Raw videos were collected by the fixed camera at 25 fps, and image samples were generated by uniform frame extraction. Considering the high similarity between adjacent video frames, one frame was extracted every 25 frames, corresponding to an interval of approximately 1 s. This strategy preserved changes in material spreading state during conveying while reducing the effect of continuous-frame redundancy on training and evaluation. The main parameters for experimental design and image-sample construction are listed in Table 1.

Images were collected using a dual-spectrum industrial camera intended for power-plant application scenarios. The main parameters of the imaging device are summarized in Table 2. The camera can continuously image under natural light and infrared supplementary illumination, providing a hardware basis for simulating different illumination conditions. The relatively high image resolution helps preserve texture, edge, and morphological differences between coal and biomass particles within the mixed-material region, thereby providing image features for subsequent ratio-label recognition. It should be noted

that infrared supplementary illumination changes the image gray-level distribution and texture visibility. Therefore, natural light and infrared supplementary illumination were treated as two independent illumination conditions in the comparative analysis.

Table 1. Data acquisition and sample construction strategy.

Item	Setting
Data-acquisition platform	Laboratory conveyor-belt platform
Original data format	Conveyor-belt videos
Camera frame rate	25 fps
Frame-extraction method	One frame is extracted every second
Number of videos for each experimental condition	Three independent feeding trials, with one video acquired in each trial
Illumination conditions	Morning natural light; nighttime infrared supplementary illumination
Ratio-label system	0–100% at 5% intervals; 21 classes in total
Particle-size test ratios	30%, 50%, and 70%
Small-particle range	6 mm < d ≤ 12.5 mm
Large-particle range	12.5 mm < d < 100 mm

Table 2. Camera equipment and imaging-parameter settings.

Item	Parameter
Camera type	Dual-spectrum industrial camera (white light/infrared)
Sensor	1/1.8" CMOS
Maximum resolution	2560 × 1440 px
Frame rate	25 fps
Lens focal length	2.8–12 mm
Minimum illumination	Color: 0.0005 Lux; monochrome: 0.0001 Lux
Supplemental illumination	White-light illumination/infrared illumination
Infrared wavelength	850 nm
Illumination distance	Approximately 50 m for infrared and 30 m for white light
Day/night mode	Automatic switching (ICR infrared-cut filter)
Wide dynamic range	120 dB

As shown in Table 3, the training and validation sets were not obtained by randomly splitting the same batch of images, but were divided by experimental batch. Specifically, exp1 and exp2 represent two mutually independent data-acquisition batches. In this study, the exp1 samples were used as the training set, and the exp2 samples were used as the validation set. This division reduces the risk of data leakage caused by adjacent video frames or samples from the same feeding batch appearing simultaneously in the training and validation sets. The number of morning natural-light and nighttime infrared samples in each ratio class was also approximately balanced, reducing the effect of illumination imbalance on model training and evaluation.

Table 3. Dataset split and sample statistics for coal–biomass blended images.

Dataset	Total Samples	Samples Per Class	am	pm
Training set (exp1)	7560	approximately 360	approximately 180	approximately 180
Validation set (exp2)	1722	approximately 82	approximately 41	approximately 41

Note: Two independently acquired experimental batches are denoted by exp1 and exp2. In this study, exp1 was used as the training set and exp2 as the validation set; the two batches were independent in terms of acquisition time, material-spreading process, and video samples.

Table 4 summarizes the construction method and evaluation range of the test samples. The test set was obtained from an experimental batch independent of the training and validation batches (exp1 and exp2) and was not used for model training, validation, or hyperparameter selection. Ratio testing covered coal mass fractions from 0% to 100% at 10% intervals, yielding 11 ratio conditions. Particle-size testing was performed under three representative coal mass fractions, namely 30%, 50%, and 70%, with small- and large-particle samples.

Table 4. Test-sample construction and evaluation scope.

Item	Description
Data source	Laboratory conveyor-belt platform
Used for training	No
Used for parameter tuning	No
Relationship to training/validation sets	Same platform, different experimental batch
Coal mass-fraction range	0–100% at 10% intervals; 11 ratios in total
Illumination conditions	Morning natural light; nighttime infrared supplementary illumination
Particle-size conditions	Small particles (6–12.5 mm) and large particles (12.5–100 mm), tested at 30%, 50%, and 70%

2.3. Definitions of Detection Object, Ratio Labels, and Evaluation Metrics

In this study, the visible coal–biomass mixed-material region in each conveyor-belt image was defined as the detection object, and one material-region bounding box was annotated for each frame. The bounding box was used to enclose the visible contour of the mixed material in the frame, while excluding conveyor-belt background, blank areas, and surrounding structures as much as possible. It should be emphasized that the detection object was not an individual coal or biomass particle, and the ratio label did not represent the exact local mass fraction inside the bounding box. Instead, the label corresponded to the overall coal mass fraction of the prepared sample or test condition. Therefore, the ratio labels used in this study can be regarded as sample-level weak labels with inherent uncertainty. In particular, local stacking, occlusion, and non-uniform surface exposure may cause the visible composition in a single frame to deviate from the overall sample ratio. To mitigate this frame-level fluctuation, a multi-frame frequency-weighted strategy was used to obtain a more stable continuous coal-fraction estimate. At the annotation level, HBB and OBB annotation datasets were established based on the same original images. Annotation followed a workflow of manual annotation, model pre-annotation, and manual review, with emphasis on correcting missing labels, incorrect labels, bounding-box offsets, and abnormal ratio labels.

After model training, the trained models were applied to the test videos. The test videos were sampled at fixed intervals, and the extracted images were input to the model for ratio-label prediction. For a single image frame, the detection result with the highest confidence was retained as the predicted ratio label of that frame. For the test samples under the same true coal-fraction condition, all sampled frames from the three test videos under that condition were pooled. The occurrence count of each predicted ratio label was then calculated, and a weighted average was computed according to the coal fraction corresponding to each label to obtain the predicted coal fraction under that true condition:

$$\hat{y} = \sum_{i=1}^C \frac{n_i}{N} y_i \quad (1)$$

where the predicted coal mass-fraction under a given ground-truth condition is denoted by $\hat{y}$; $C = 21$ is the number of ratio-label classes, corresponding to coal mass fraction labels from 0% to 100% at 5% intervals; y_i is the coal mass fraction corresponding to the i -th label; n_i is the occurrence count of that label among all sampled-frame results under the same ground-truth condition; N is the total number of valid predicted frames pooled from the three independent test videos under the same true coal-fraction condition.

For a given ground-truth coal mass-fraction condition, the prediction Bias was defined as follows:

$$\text{Bias} = \hat{y} - y = \sum_{i=1}^C \frac{n_i}{N} y_i - y \quad (2)$$

where y is the ground-truth coal mass fraction under the test condition. A positive Bias indicates that the model overestimates the coal mass fraction, whereas a negative Bias indicates underestimation.

Furthermore, to evaluate the overall dispersion error of all sampled-frame predictions relative to the ground-truth ratio under the same condition, RMSE was defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^C n_i (y_i - y)^2} \quad (3)$$

where y is the ground-truth coal mass fraction under the test condition, y_j is the coal mass fraction corresponding to the j -th predicted label, n_j is the number of occurrences of this label among all extracted-frame predictions under the given condition, and N is the total number of valid predicted frames under the test condition. In this study, RMSE was not calculated solely from the mean predicted coal mass fraction; rather, it was calculated from the distribution of all sampled-frame predicted labels. Therefore, RMSE reflects both the Bias between the prediction mean and the ground-truth coal mass fraction and the dispersion of frame-level predicted labels around the true ratio. A lower RMSE indicates that the model predictions are more concentrated near the true coal mass fraction and that the ratio estimation is more stable.

For the test set, RMSE was calculated separately for 11 ground-truth coal mass-fraction conditions, namely 0%, 10%, 20%, ..., 100%, and their arithmetic mean was used as the model's mean RMSE:

$$\overline{\text{RMSE}} = \frac{1}{K} \sum_{k=1}^K \text{RMSE}_k \quad (4)$$

where $\overline{\text{RMSE}}$ denotes the model's mean RMSE on the test set, $K = 11$ is the number of ground-truth coal mass-fraction conditions in the test, and RMSE_k is the RMSE under the k -th ground-truth condition. The "mean RMSE" reported in the tables was obtained by summing the RMSE values for the 11 ground-truth coal mass-fraction conditions and dividing by 11.

2.4. Candidate YOLO Models and Bounding-Box Representations

Figure 2 shows the basic structure of the YOLOv8 object-detection network. YOLOv8 uses a one-stage detection framework consisting of a Backbone, Neck, and Head and can perform target-region localization, class prediction, and confidence estimation in a single forward pass. The Backbone extracts color, texture, edge, and local morphological features from coal-biomass mixture images. The Neck enhances the representation of different particle sizes, accumulation densities, and local material distributions through multi-scale feature fusion. The Head outputs the target-region position, ratio class, and confidence. Considering the high requirements of conveyor-belt online monitoring for inference efficiency and continuous-frame processing, YOLO models were selected as

the main candidate detection framework, and different model scales and bounding-box representations were introduced for comparative analysis.

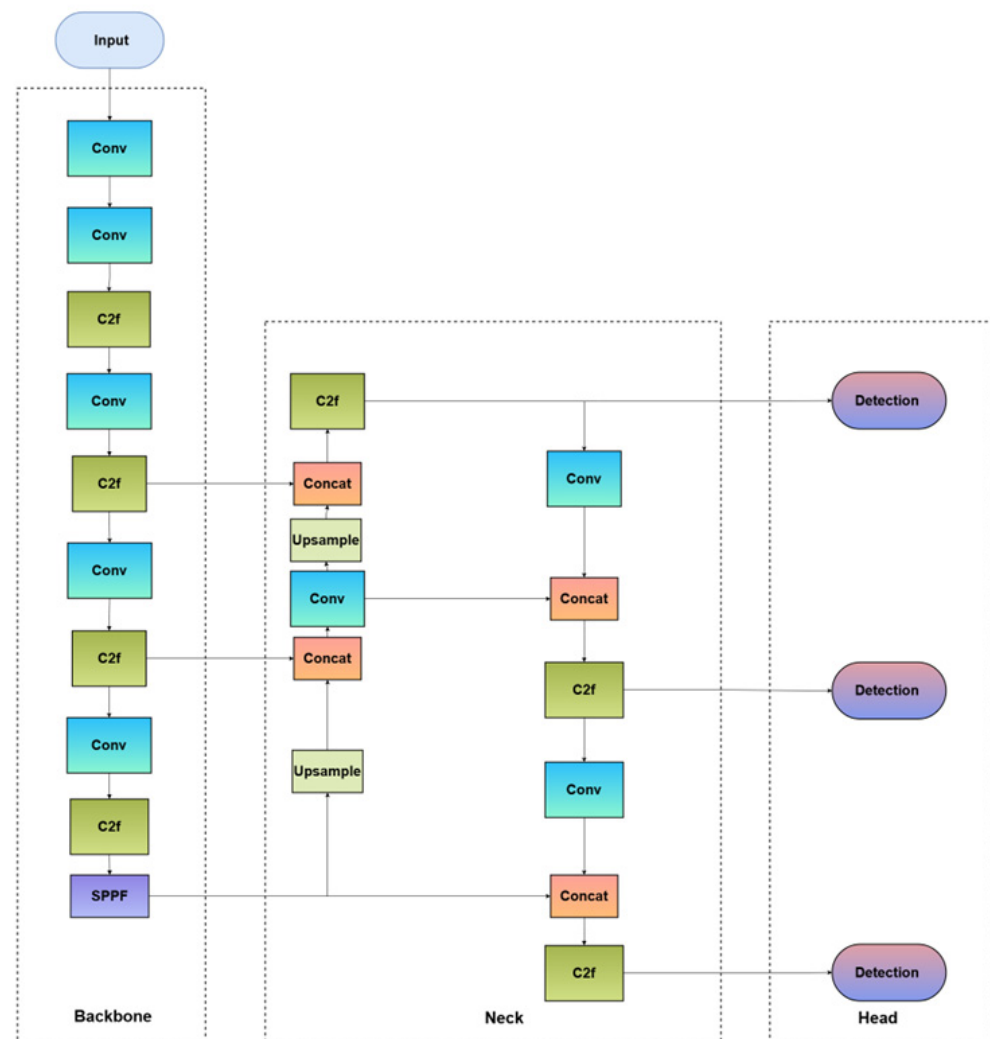


Figure 2. Schematic architecture of the YOLOv8 object-detection network.

This study did not modify the YOLO network architecture itself. Instead, it constructed a comparative experimental framework for coal–biomass blend-ratio recognition consisting of a whole-image classification baseline, HBB detection, and OBB detection. YOLOv8n-cls was first used as a whole-image classification baseline. This model does not perform explicit material-region localization; it directly takes the entire conveyor-belt image as input and predicts one of the 21 coal mass-fraction classes, thereby evaluating whether whole-image classification alone can satisfy the requirements of blend-ratio estimation. Subsequently, YOLOv5n, YOLOv8n, YOLOv8s, and YOLOv8m were selected as HBB detection models to analyze the influence of model generation and increased model capacity on detection performance and ratio-estimation error. On this basis, YOLOv8n-OBB and YOLOv8s-OBB were further selected as oriented-box detection models to evaluate the effect of bounding-box representation on localization compactness and ratio-recognition stability for visible mixed-material regions. The characteristics and roles of the candidate models are summarized in Table 5.

Table 5. Comparison of candidate YOLO-series models and the image-level classification baseline.

Model	Model Scale	Main Characteristics	Role in This Study
YOLOv8-cls	Classification baseline	Whole-image classification without explicit material-region localization	Evaluates the necessity of object-detection/material-region localization
YOLOv5	Baseline model	Mature architecture, widely used, with good general detection capability	Classical YOLO-model reference
YOLOv8n	Nano	Small parameter count, fast inference, and low deployment cost	Speed-prioritized lightweight HBB baseline
YOLOv8s	Small	Stronger feature-extraction capability than YOLOv8n, but with higher computational cost	Analysis of performance changes after increasing the scale of HBB models
YOLOv8m	Medium	Further increased model capacity and stronger feature representation	Analysis of the effect of larger model capacity on detection performance
YOLOv8n-OBB	Nano	Introduces oriented bounding boxes into a lightweight model, balancing accuracy and deployment cost	Representative lightweight oriented-box model
YOLOv8s-OBB	Small	Increases model capacity based on oriented boxes and has the potential for higher detection accuracy	Analysis of performance changes after increasing the scale of OBB models

Figure 3 and Table 6 compare HBB and OBB representations. An HBB can be represented as (x_c, y_c, w, h) , and its boundaries are always parallel to the image coordinate axes. An OBB introduces a rotation angle θ and can be represented as (x_c, y_c, w, h, θ) , allowing the box to rotate with the dominant direction of the target region. For the coal–biomass mixture images in this study, the annotation object was the visible mixed-material region in the conveyor-belt image; therefore, the difference between HBB and OBB was mainly reflected in their coverage compactness for inclined material regions. Compared with HBBs, OBBs can reduce the inclusion of conveyor-belt background and surrounding structures. Thus, the expected advantage of OBBs in this task does not mainly arise from using the orientation angle as an additional physical descriptor, but from enabling more compact target-region coverage and reducing background redundancy. However, OBB annotation requires an additional orientation parameter and involves rotated-box regression and post-processing during training and inference; therefore, the accuracy improvement and deployment cost need to be evaluated jointly.

Table 6. Comparison between horizontal and oriented bounding boxes.

Comparison Aspect	HBB (Horizontal Bounding Box)	OBB (Oriented Bounding Box)	Relevance to Coal–Biomass Blend Recognition
Bounding-box representation	(x, y, w, h)	(x, y, w, h, θ)	OBB introduces orientation information to support compact region representation.
Region coverage	Parallel to the image coordinate axes	Rotates with the dominant direction of the material region	OBB covers inclined mixed-material regions more compactly.
Background redundancy	More likely to include conveyor-belt background regions	Usually reduces redundant background	Helps reduce interference unrelated to ratio discrimination.
Annotation and application cost	Simple annotation and lower post-processing complexity	Requires orientation-angle annotation and higher post-processing complexity	Accuracy improvement and real-time requirements must both be considered.

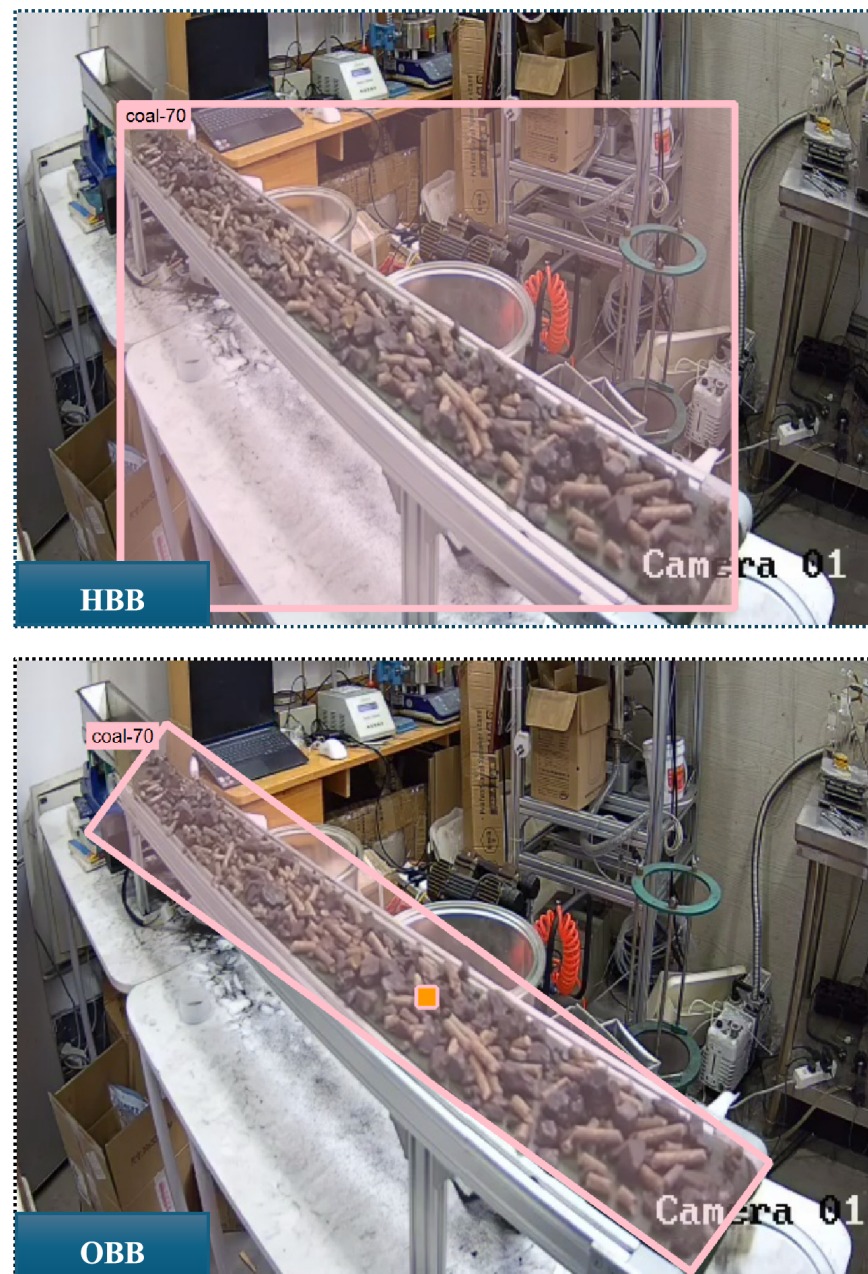


Figure 3. Comparison of horizontal bounding box (HBB) and oriented bounding box (OBB) representations for the visible mixed-material region: the upper panel shows HBB, and the lower panel shows OBB.

2.5. Model-Training Settings and Reproducibility

To ensure comparability among the models, the training-data source, experimental-batch split, input size, number of epochs, data-augmentation strategy, and test workflow were controlled as consistently as possible. Except for the whole-image classification model, which did not use bounding-box annotations, all detection models were trained and evaluated using conveyor-belt images from the same source and the same ratio-label system. The whole-image YOLOv8n-cls model, HBB detection models, and OBB detection models corresponded to different task formulations; therefore, they were not regarded as direct architecture-level equivalents. Instead, they were used to examine whether whole-image classification was sufficient for blend-ratio estimation and whether model scale and bounding-box representation affected ratio-estimation accuracy when explicit material-region localization was available. Accordingly, YOLOv8n-cls was evaluated only in terms of

ratio-estimation error, whereas HBB and OBB detection models were evaluated in terms of detection performance, ratio-estimation error, and deployment cost. For detection models, precision, recall, mAP50, and mAP50-95 were used to evaluate detection performance. Because the ratio labels are ordered and continuous, discrete detection metrics such as mAP cannot fully represent the practical error between adjacent ratio predictions; therefore, Bias and RMSE were used as the main metrics for ratio-estimation performance.

In addition to the YOLO-series models, several auxiliary baselines were included to evaluate the rationality of the task formulation and the fairness of model comparison. First, a cropped-ROI YOLOv8n-cls model was trained as a classification baseline with reduced background interference. The ROI images were manually cropped from the visible material regions in the same original conveyor-belt images, and the corresponding sample-level coal-fraction labels were kept unchanged. The cropped-ROI dataset followed the same experimental-batch split as the detection datasets, with exp1 used for training, exp2 used for validation, and the independent test batch used only for final evaluation. Second, a ResNet18 regression model was trained on the same ROI dataset to examine whether direct mapping from material-region images to continuous coal mass fraction was suitable for this sample-level estimation task. The ResNet18 model was initialized with ImageNet pretrained weights, and its final fully connected layer was replaced with a single-output regression layer. The model was trained using mean squared error loss. In addition, RT-DETR-L was included as a transformer-based detector baseline to provide a non-YOLO reference. These auxiliary baselines were used mainly for task-formulation analysis, whereas the main comparison focused on YOLO-series models and the effects of model scale and bounding-box representation. All auxiliary baselines followed the same experimental-batch split and independent test protocol as the YOLO models, while model-specific training settings were adjusted according to the corresponding task formulation.

During testing, YOLOv8n-cls and cropped-ROI YOLOv8n-cls output the Top-1 ratio class for each frame, whereas the ResNet18 regression model directly outputs a continuous coal-fraction value. For HBB, OBB, and RT-DETR-L detection models, the detection result with the highest confidence in each frame was retained, and its class was used as the predicted ratio label of that frame. For each true coal-fraction condition, the prediction results from the three independent test videos were pooled, and Bias, RMSE, and mean RMSE were calculated according to Section 2.3. This statistical strategy reduced the influence of single-frame local spreading differences, particle occlusion, and random stacking on the evaluation results and was more consistent with continuous monitoring requirements in conveyor-belt scenarios.

Model deployment cost was analyzed in terms of parameter size, GFLOPs, theoretical inference frame rate, actual inference frame rate, and training time. Table 7 summarizes the key training and evaluation configurations, and Table 8 lists the experimental hardware and software environment.

Table 7. Model training and evaluation configuration.

Item	Setting/Description
Training/validation split	Split by experimental batch; exp1 for training and exp2 for validation
Test samples	Not used for training or parameter selection; a different experimental batch on the same experimental platform
Ratio labels	0–100% at 5% intervals; 21 classes in total
HBB models	YOLOv5n, YOLOv8n, YOLOv8s, and YOLOv8m
OBB models	YOLOv8n-OBB and YOLOv8s-OBB

Table 7. Cont.

Item	Setting/Description
Classification baselines	Whole-image YOLOv8n-cls and manually cropped ROI-YOLOv8n-cls
Regression baseline	ResNet18 regression trained on manually cropped ROI images
Transformer-based detector baseline	RT-DETR-L
Repeated training	Selected baselines and core models were trained with three random seeds where applicable, and the mean $\pm$ SD was reported
Input image size	1024 $\times$ 1024
Batch size	16
Epochs	100
Optimizer/initial learning rate/weight decay	Optimizer: Auto; initial learning rate $lr_0 = 0.01$; $weight_decay = 0.0005$
Data augmentation	HSV augmentation, translation, scaling, horizontal flipping, Mosaic augmentation, RandAugment, and random erasing
Confidence threshold/IoU-NMS threshold	IoU-NMS threshold: 0.7; confidence threshold not specifically set (default value used)
ResNet18 regression settings	ImageNet-pretrained ResNet18; input size 224 $\times$ 224; batch size 32; 100 epochs; Adam optimizer; learning rate (1×10^{-4}); MSE loss

Table 8. Experimental hardware and software environment.

Configuration Name	Configuration Parameter
Operating system	Windows 11
CPU	Intel (R) Xeon (R) Platinum 8470Q
GPU	NVIDIA GeForce RTX 4090
Python	3.12
PyTorch	2.7.0
CUDA	12.8

3. Results

3.1. Comparison of Whole-Image Classification, Detection-Box Representation, and Engineering Cost

As shown in Table 9, the whole-image YOLOv8n-cls baseline yielded a mean RMSE of 13.98 pp, which was substantially higher than that of the material-region detection-based estimation models. After cropping the material ROI, the mean RMSE of YOLOv8n-cls decreased to 12.81 ± 0.64 pp, indicating that removing background and blank regions reduced the error of classification-based ratio recognition. However, this result was still inferior to those of the detection-based models, suggesting that ROI-level classification alone was insufficient for stable blend-ratio estimation. The direct regression baseline, ResNet18 regression, produced a mean RMSE of 23.35 ± 1.79 pp, further indicating that directly mapping single-frame ROI features to continuous ratio values was less effective than the detection-based estimation strategy using ordered ratio classes.

Among the detection-based formulations, YOLOv8n HBB, YOLOv8n-OBB, and RT-DETR-L achieved mean RMSE values of 9.02 ± 0.46 pp, 6.90 ± 0.39 pp, and 8.72 pp, respectively, all outperforming the whole-image classification and direct-regression baselines. These results demonstrate that explicit material-region localization is important for reducing ratio-estimation error. RT-DETR-L was included as an additional transformer-based

detector baseline to verify the effectiveness of the detection-based formulation. However, considering the objective of this study, namely evaluating the influence of YOLO model scale and bounding-box representation on coal–biomass blend-ratio estimation, the subsequent comparison focuses on YOLO-series models. Overall, the results in Table 9 suggest that coal–biomass blend recognition is more appropriately formulated as a material-region detection and ratio-estimation problem rather than as direct whole-image classification or single-frame regression.

Table 9. Comparison of different task formulations for coal–biomass blend-ratio estimation.

Method	Input Representation	Role	RMSE/pp
YOLOv8n-cls	Whole image	Diagnostic baseline for background interference	13.98
Cropped YOLOv8n-cls	Cropped material ROI	Fair ROI-classification baseline	12.81 ± 0.64
ResNet18 regression	Cropped material ROI	Direct regression baseline	23.35 ± 1.79
YOLOv8n	HBB material region	Detection + weighted estimation	9.02 ± 0.46
YOLOv8n-OBB	OBB material region	Detection + weighted estimation	6.90 ± 0.39
RT-DETR-L	HBB material region	Transformer-based detector baseline	8.72

Based on the above task-formulation comparison, the following analysis focuses on YOLO-series models to further evaluate the effects of model scale, bounding-box representation, ratio-estimation error, and engineering cost. Table 10 summarizes the detection performance, ratio-estimation error, and computational cost of the selected YOLO models. mAP50 and mAP50-95 were used to evaluate the ability of the models to localize visible mixed-material regions and distinguish ratio labels, whereas average RMSE was used to evaluate test-set ratio-estimation performance. The theoretical inference frame rate mainly reflects the speed of the neural network inference stage, while the actual inference frame rate was measured under the experimental hardware and test workflow of this study, including image reading, preprocessing, model inference, and post-processing. Therefore, the actual inference frame rate is more informative than the theoretical inference frame rate for engineering deployment analysis.

Table 10. Key performance indicators of YOLO models in the coal–biomass blend-recognition task.

Model	mAP50	mAP50-95	Mean RMSE	Parameters/M	GFLOPs	FPS_T	FPS_A	Training Time/s
YOLOv5n	0.461	0.419	9.50	2.507	7.1	909.09	53.46	12,474
YOLOv8n	0.503	0.451	9.02	3.01	8.1	769.23	58.63	12,200.4
YOLOv8s	0.557	0.508	6.70	11.13	28.5	769.23	20.93	15,073.2
YOLOv8m	0.590	0.540	6.10	25.85	78.8	454.55	16.02	15,300
YOLOv8n-OBB	0.568	0.512	6.90	3.08	8.4	384.61	24.53	23,744.6
YOLOv8s-OBB	0.569	0.528	6.45	11.42	29.5	250	14.26	26,395.2

Note: FPS_T denotes the theoretical inference frame rate, and FPS_A denotes the actual inference frame rate measured under the experimental hardware environment and complete testing workflow in this study. The mean RMSE is the arithmetic mean of the RMSE values for 11 ground-truth coal mass-fraction conditions from 0% to 100% at 10% intervals.

In terms of detection performance, the YOLOv8 series outperformed YOLOv5n overall. The mAP50 of YOLOv8n was 0.503, representing an increase of approximately 9.1% compared with the 0.461 of YOLOv5n. As the model scale increased, the detection performance of YOLOv8s and YOLOv8m further improved, with YOLOv8m achieving the highest mAP50 and mAP50-95 values among all candidate models, namely 0.590 and 0.540, respectively. It should be noted that the absolute mAP values of the models remained relatively limited, which is closely related to the ratio-label setting used in this study. The labels covered 21 classes from 0% to 100% at 5% intervals; adjacent ratios had small visual differences, and coal-fraction changes were continuous. For example, for a sample with a true coal fraction of 50%, predictions of 45% or 55% are close to the true value from the perspective of continuous ratio estimation, but they are not counted as correct detections for the true class in discrete mAP evaluation. Thus, relying only on mAP cannot fully reflect the model response to true coal-fraction trends; Bias and RMSE were therefore further used to evaluate ratio-estimation performance.

In terms of ratio-estimation error, YOLOv8m achieved the lowest average RMSE of 6.10 pp among all candidate models, indicating that larger model capacity improves the feature representation of complex mixed-material regions. Among lightweight models, YOLOv8n-OBB achieved an average RMSE of 6.90 pp, markedly lower than the 9.02 pp of YOLOv8n, corresponding to a relative reduction of approximately 23.5%. YOLOv8s-OBB achieved an average RMSE of 6.45 pp, slightly lower than YOLOv8n-OBB, but with substantially increased parameter size, GFLOPs, training time, and actual inference cost. These results indicate that both increasing model capacity and improving bounding-box representation can reduce ratio-estimation error, while OBBs improve the representation ability and ratio-estimation stability of lightweight models for irregular bulk-material regions under a smaller parameter scale.

From the perspective of deployment cost, YOLOv8n achieved the highest actual inference frame rate, with FPS_A reaching 58.63 FPS, making it suitable for rapid monitoring scenarios with high real-time requirements and some tolerance for estimation error. YOLOv8n-OBB had parameter size and GFLOPs close to those of YOLOv8n, but its FPS_A decreased to 24.53 FPS because of rotated-box prediction and post-processing. Although YOLOv8m achieved the highest detection accuracy and lowest average RMSE, its parameter size reached 25.85 M, its GFLOPs reached 78.8, and its FPS_A was only 16.02 FPS, resulting in a substantially higher deployment cost than that of lightweight models.

To further explain why OBB detection achieved a lower RMSE in lightweight models, the geometric compactness of HBB and OBB annotations was quantitatively compared. As shown in Table 11, the average annotated box area of HBBs was 78,161.98 px², whereas that of OBBs was 16,697.98 px². The OBB/HBB area ratio was 0.21, corresponding to a box-area reduction of 78.63%. This indicates that OBB annotations can fit the visible coal–biomass mixed-material region more tightly and reduce potential redundant regions and conveyor-belt background interference inside the box. This provides a plausible explanation for the lower RMSE achieved by YOLOv8n-OBB among lightweight models.

Table 11. Quantitative analysis of HBB and OBB annotation compactness.

Metric	HBB	OBB
Average box area/pixels	78,161.98	16,697.98
OBB/HBB area ratio	1	0.21
Background reduction rate	—	78.63%

Considering detection accuracy, ratio-estimation error, and actual inference efficiency, YOLOv8n-OBB achieved a favorable accuracy–efficiency balance among lightweight models. Therefore, YOLOv8n and YOLOv8n-OBB were selected as representative HBB and OBB models for subsequent analysis of the influence of bounding-box representation on coal–biomass blend-ratio recognition.

Figure 4 compares the precision–recall curves of YOLOv8n and YOLOv8n-OBB. The curve of YOLOv8n-OBB is generally above that of YOLOv8n, indicating that it maintained higher precision at different recall levels. This result suggests that the performance improvement of the OBB model was not caused only by local results under a particular confidence threshold but was observed across a wider threshold range. This further indicates that oriented bounding boxes improve the geometric representation of irregularly stacked bulk-material regions and thereby enhance model recognition of visible mixed-material regions.

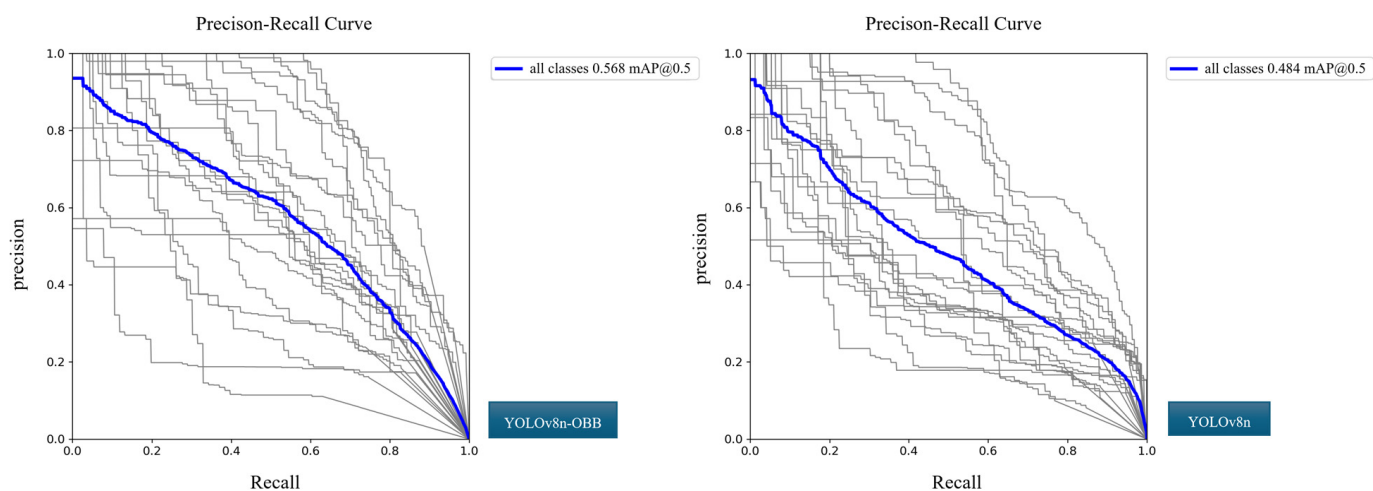


Figure 4. Comparison of precision–recall curves for YOLOv8n and YOLOv8n-OBB.

Figure 5 shows that the errors of both models were mainly concentrated between adjacent ratio labels, especially in the intermediate ratio range of 40–60%. This indicates that coal–biomass blend recognition has an obvious ordered-ratio discrimination characteristic: under extreme ratio conditions such as 0% or 100%, the visual features of a single material dominate, allowing the model to form a more stable judgment; under intermediate ratios, the visual contributions of coal and biomass particles are similar, the local mixing state is more complex, and the boundaries between adjacent ratio labels are more ambiguous.

Compared with YOLOv8n, the confusion matrix of YOLOv8n-OBB showed a more concentrated diagonal response and weaker off-diagonal interference, indicating that oriented bounding boxes helped improve the discrimination stability of the model for coal–biomass mixed-material regions. The misclassifications of YOLOv8n-OBB were mainly concentrated in adjacent or near-adjacent ratio classes rather than classes far from the true ratio, indicating that although discrete-class confusion remained, the model still exhibited a degree of ordered response to changes in coal fraction. Some intermediate ratio classes still had relatively low diagonal values, mainly because coal and biomass particles were highly interlaced in the intermediate blending range and visual differences between adjacent 5% ratio labels were limited. This result is consistent with the limitations of the mAP metric discussed above and further confirms that discrete detection metrics alone cannot fully evaluate ratio-estimation performance. Therefore, Bias and RMSE were used in subsequent analyses to comprehensively evaluate coal-fraction estimation performance.

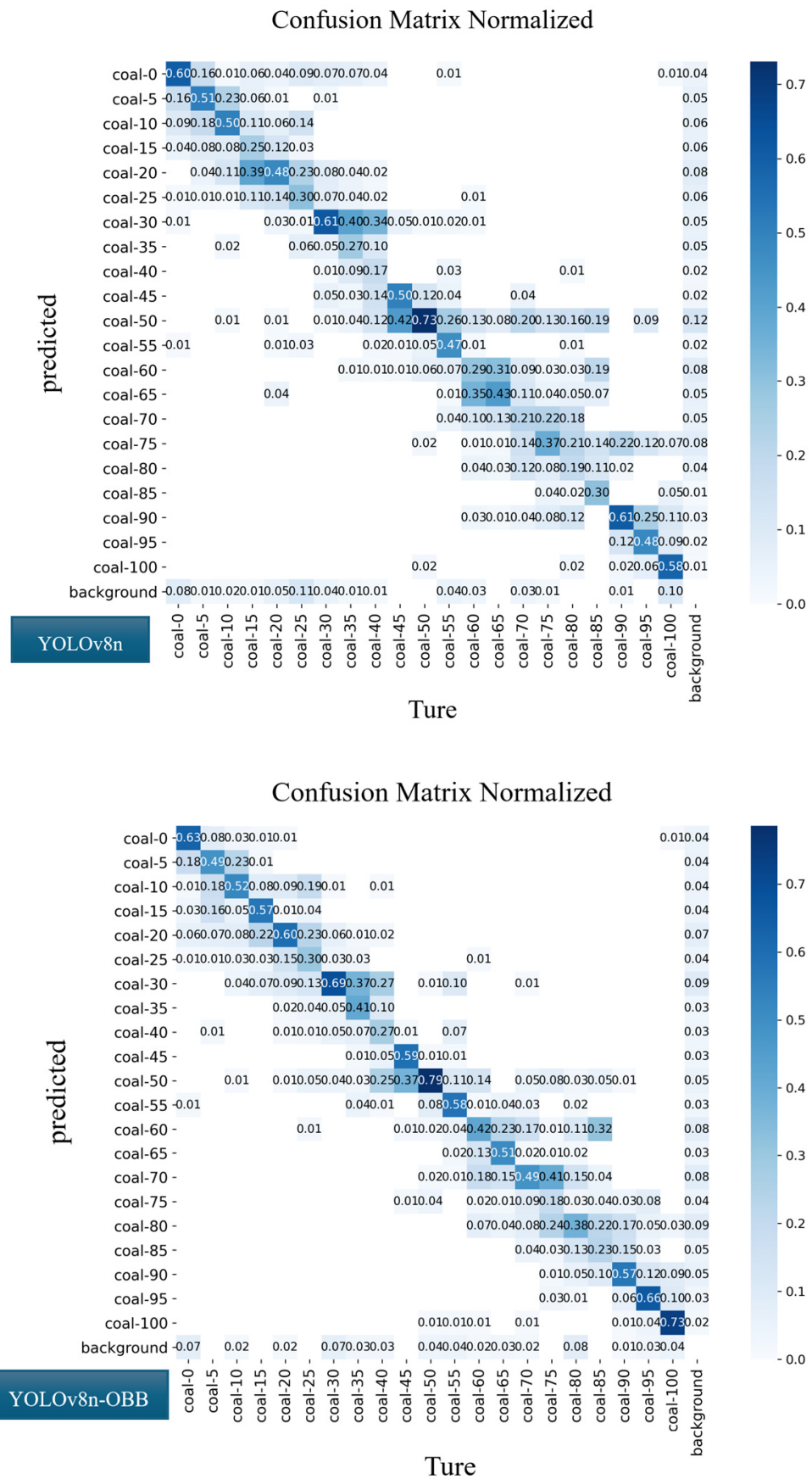


Figure 5. Comparison of normalized confusion matrices for YOLOv8n and YOLOv8n-OBb.

3.2. Blend-Ratio Recognition Performance

After comparing the basic detection performance, the ability of YOLOv8n and YOLOv8n-OBB to estimate the coal–biomass blend ratio was further evaluated under 11 true coal-fraction test conditions from 0% to 100% at 10% intervals. According to Equation (1), the ratio labels output by each model under each test condition were statistically weighted by frequency to obtain the predicted coal fraction, which was then compared with the true coal fraction.

Figure 6 shows that the predicted coal fractions of both models were generally distributed near the ideal reference line $y = x$, indicating that both models could basically reflect changes in coal fraction. Further comparison shows that the predictions of YOLOv8n-OBB were generally closer to the ideal reference line, whereas YOLOv8n showed more obvious deviations under some medium-to-high coal-fraction conditions. This indicates that both models had a certain ability to estimate coal fraction, but the use of oriented bounding boxes improved the ability of the model to follow the true blending ratio and enhanced prediction stability.

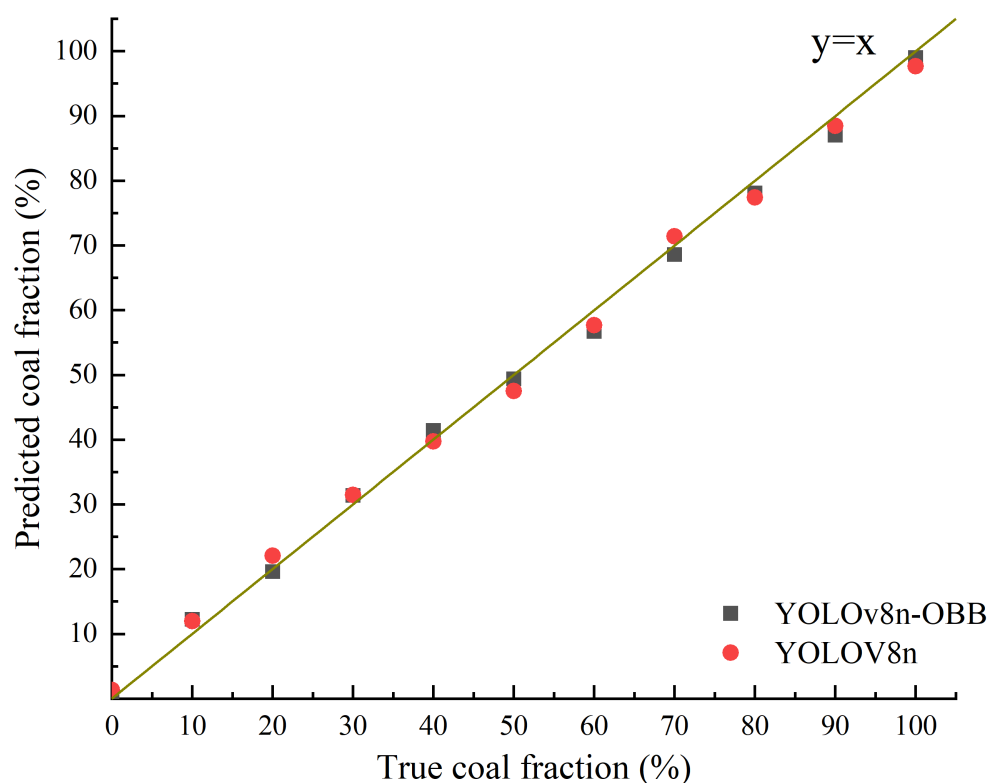


Figure 6. Relationship between true and predicted coal mass fractions for the two models.

Figure 7 further confirms the observations from the confusion matrices from the perspective of ratio-estimation error. Compared with YOLOv8n, YOLOv8n-OBB achieved a lower RMSE under all test-ratio conditions, reducing the average RMSE from 9.02 pp to 6.90 pp, a relative reduction of approximately 23.5%. The RMSE curves of both models showed higher errors in the intermediate ratio range and lower errors at the two ends, consistent with the adjacent-ratio label confusion described above. This result indicates that a more compact region representation helps improve ratio-estimation stability, although OBBs cannot completely eliminate the uncertainty caused by local mixing state, particle stacking, and continuous ratio transitions in the intermediate ratio range.

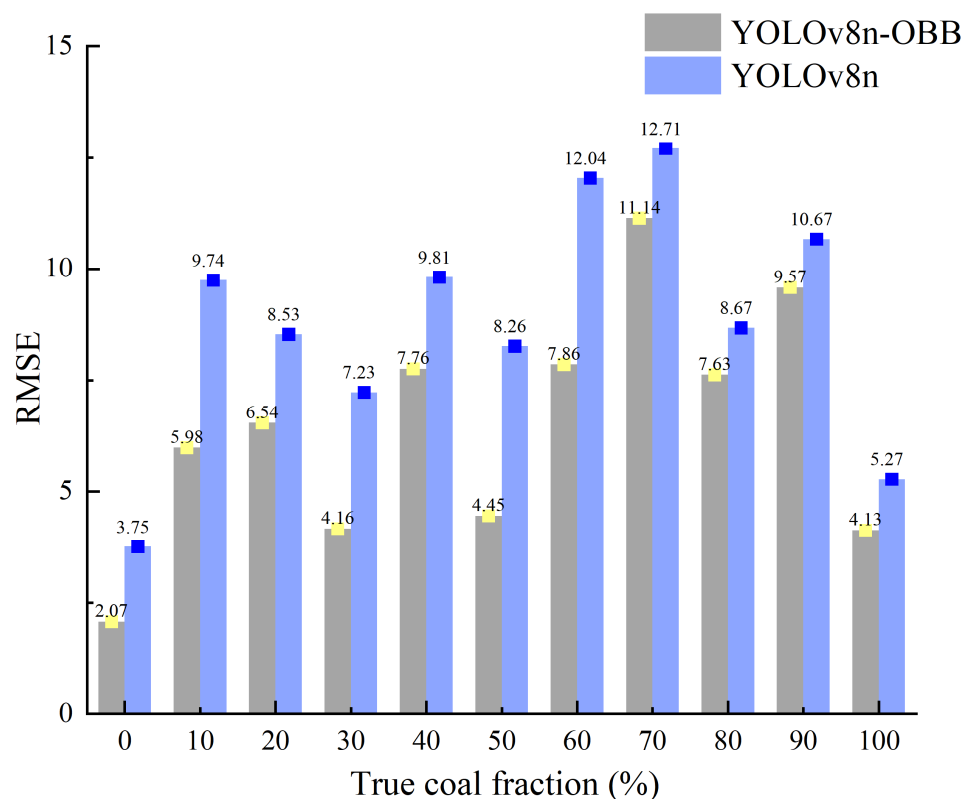


Figure 7. RMSE comparison between the two models under different true coal mass-fraction conditions.

3.3. Effects of Illumination Conditions and Particle-Size Variation

To evaluate the stability of coal–biomass blend-ratio recognition under complex operating conditions, prediction results under morning natural light, nighttime infrared supplementary illumination, and different particle-size conditions were further analyzed. Illumination conditions mainly affect image gray-level distribution, local contrast, and texture visibility, whereas particle size changes the number of particles per unit area, local boundary density, and occlusion morphology. This section therefore analyzes the error characteristics of the model under different operating conditions in terms of predicted coal fraction, Bias, RMSE, and predicted-label distribution.

Figure 8a shows that, under both morning natural light and nighttime infrared supplementary illumination, the predicted coal fraction generally increased with the true coal fraction, indicating that the model maintained a basic response to changes in blend ratio under different illumination conditions. Under morning natural light, the prediction points were generally closer to the ideal reference line $y = x$, suggesting better consistency between the model predictions and true coal fractions. By contrast, under nighttime infrared supplementary illumination, the prediction points deviated more clearly from the ideal reference line: the predicted values were slightly higher than the true values in the low-coal-fraction range, whereas they gradually became lower than the true values in the medium-to-high coal-fraction range, indicating an underestimation trend.

Figure 8b further reveals the direction of Bias under different illumination conditions. Under morning natural light, Bias fluctuated within a relatively small range and included both positive and negative deviations, suggesting that model errors were mainly random fluctuations at local ratio points. Under nighttime infrared supplementary illumination, Bias was positive in the low-coal-fraction range, indicating a degree of overestimation. As the true coal fraction increased, Bias gradually became negative, with obvious underestimation in the 70–90% range. This indicates that nighttime infrared supplementary illumination

not only increased error magnitude but also introduced a coal-fraction-related systematic Bias. This phenomenon may be related to changes in the brightness difference, texture contrast, and shadow distribution between coal and biomass particles under infrared illumination, which weakened the model response to coal fraction under medium-to-high coal-fraction conditions.

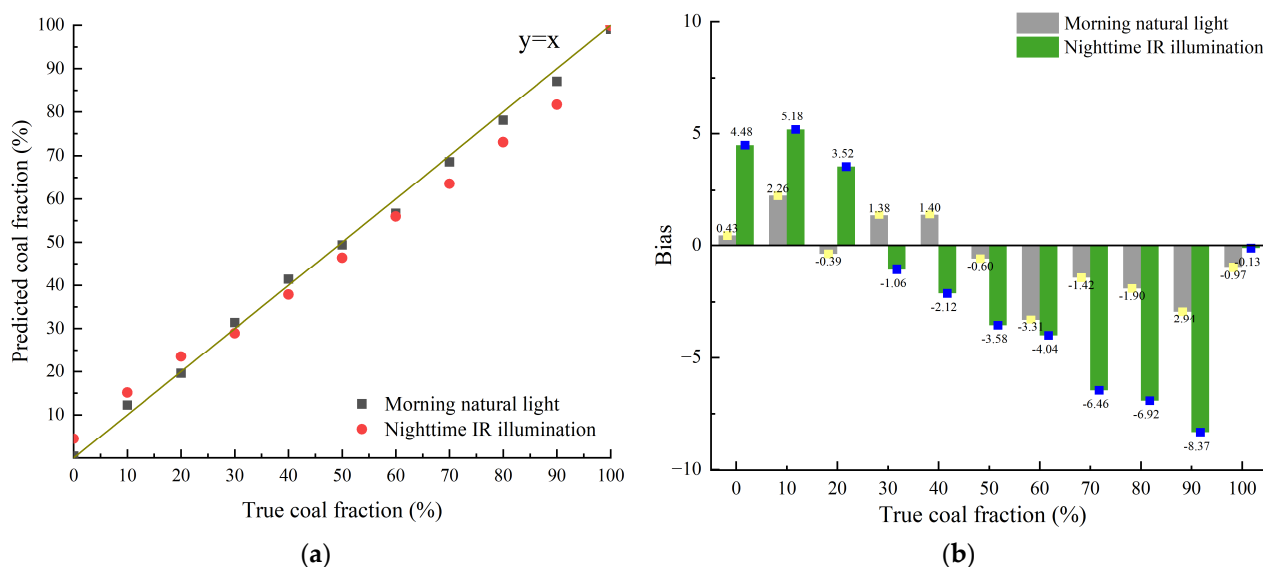


Figure 8. Effects of illumination conditions on coal-fraction estimation: (a) relationship between true and predicted coal mass fractions; (b) Bias as a function of true coal mass fraction.

Figure 9 shows clear differences in RMSE under different illumination conditions. Except for the 100% coal-fraction condition, RMSE values at most ratio points were higher under nighttime infrared supplementary illumination than under morning natural light, indicating that infrared supplementary illumination generally reduced the stability of ratio estimation. At the low coal fraction of 0%, RMSE under nighttime illumination increased substantially, consistent with the overestimation observed in Figure 8. For the medium-to-high coal-fraction range of 70–90%, RMSE values under nighttime illumination reached 11.99, 13.74, and 17.01 pp, respectively, higher than the corresponding values under morning natural light (11.14, 7.63, and 9.57 pp). This agrees with the persistent negative Bias in this range in Figure 8b and indicates that infrared supplementary illumination exacerbated the estimation error for medium-to-high coal-fraction samples. Notably, under the 100% coal-fraction condition, the RMSE under nighttime illumination was lower than that under morning natural light, suggesting that pure-coal samples could still form relatively stable visual features under this condition. Overall, nighttime infrared supplementary illumination had a more pronounced influence on mixed-ratio samples, possibly because it changed brightness differences, texture contrast, and shadow distribution between coal and biomass particles.

To further quantify the feature-distribution discrepancy between morning natural-light images and nighttime infrared supplemental-illumination images, ROI-level gray-level histogram Jensen–Shannon divergence and GLCM texture features were calculated as shown in Figure 10 and Table 12. The analysis indicates that infrared supplementary illumination introduces measurable shifts in gray-level distributions and texture patterns compared with natural-light imaging. Relatively higher Jensen–Shannon divergence values were observed at 0%, 90%, 95%, and 100% coal mass fractions, suggesting that the illumination-induced feature discrepancies are related to material composition.

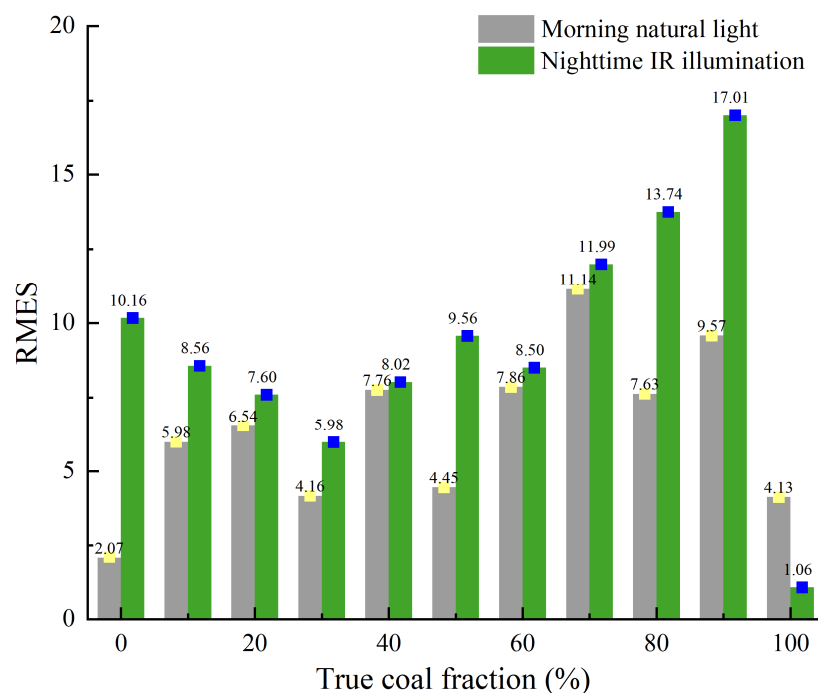


Figure 9. RMSE comparison for each true coal mass fraction under different illumination conditions.

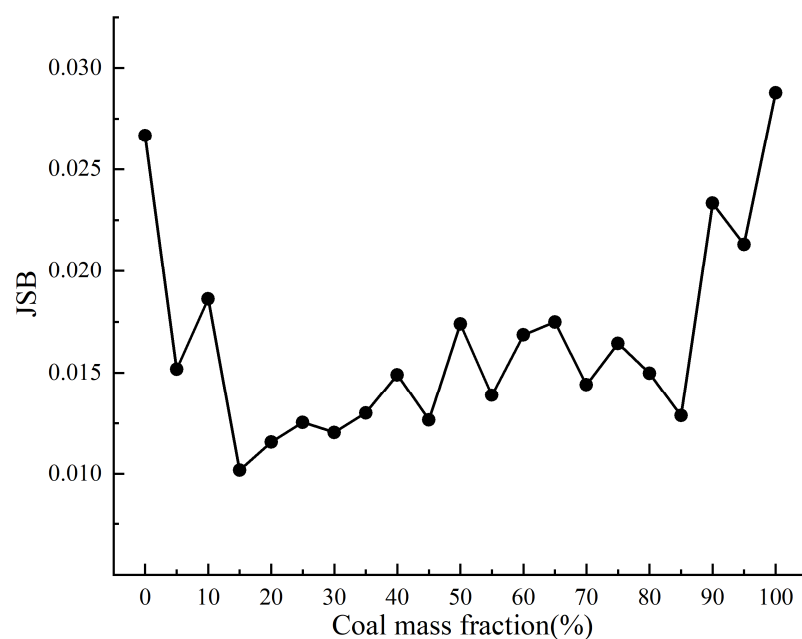


Figure 10. Jensen–Shannon divergence of gray-level histograms between morning natural light and nighttime infrared supplemental-illumination ROI images under different coal mass fractions.

Table 12. Statistical comparison of GLCM texture features between morning natural light and nighttime infrared supplemental-illumination ROI images.

GLCM Feature	am Mean $\pm$ SD	pm Mean $\pm$ SD	Relative Change (%)	Significance
Contrast	27.7941 $\pm$ 3.5097	19.1279 $\pm$ 1.7566	−31.18	$p < 0.001$
Homogeneity	0.3931 $\pm$ 0.0275	0.4640 $\pm$ 0.0226	18.03	$p < 0.001$
Energy	0.0700 $\pm$ 0.0104	0.0810 $\pm$ 0.0054	15.62	$p < 0.001$
Correlation	0.7716 $\pm$ 0.0482	0.8442 $\pm$ 0.0208	9.41	$p < 0.001$

These feature-distribution shifts correspond to the trends observed in prediction error under different illumination conditions in Figures 8 and 9. Locations with larger divergence and texture differences generally coincide with higher RMSE and Bias values, suggesting that illumination-induced domain shifts reduce cross-illumination recognition stability. This effect is especially pronounced for medium-to-high coal fractions, where local visual complexity and interlaced coal–biomass textures amplify the impact on model predictions.

Overall, these results demonstrate that the increased Bias and root-mean-square error under infrared illumination are not caused solely by brightness variation but are also associated with systematic shifts in gray-level distributions and texture features. This highlights the importance of considering domain differences when evaluating model robustness under varying illumination conditions.

As shown in Figure 11, under the three representative true coal-fraction conditions of 30%, 50%, and 70%, the RMSE values of small-particle samples were higher than those of large-particle samples. In particular, under the 30% condition, the RMSE of small-particle samples was markedly higher than that of large-particle samples, indicating that particle-scale variation had a more pronounced effect on ratio-estimation stability under this ratio condition. These results show that, under the experimental conditions of this study, smaller particle size did not necessarily make recognition easier. Instead, small-particle samples contained more particles per unit area, denser local boundaries, and more complex interlacing, occlusion, and stacking between coal and biomass particles. These factors caused model predictions to disperse among multiple adjacent ratio labels, thereby increasing RMSE.

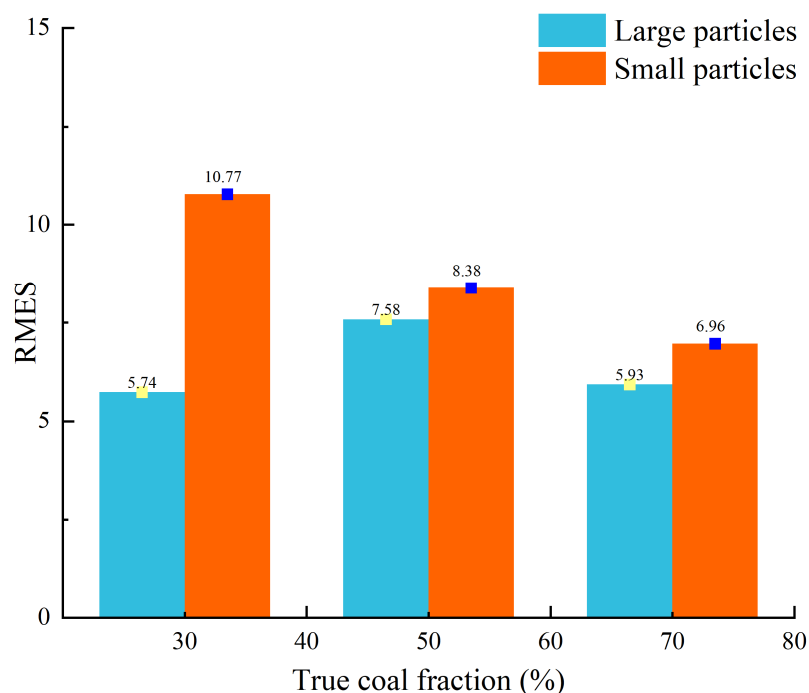


Figure 11. RMSE comparison of representative ratio samples under different particle-size conditions. Note: Small particles were defined as $6\text{ mm} < d \leq 12.5\text{ mm}$, and large particles were defined as $12.5\text{ mm} < d < 100\text{ mm}$.

Figure 12 further shows the predicted-label distributions of samples with different particle sizes under a true coal fraction of 30%. The small-particle sample formed a main peak near 30%, but also showed obvious secondary peaks near 40%, 50%, and 60%, indicating strong label dispersion in the model output. The predictions for the large-particle sample were mainly concentrated near 30% and 40%. Although adjacent-ratio shifts still existed,

the overall distribution range was narrower. Combined with Figure 11, this indicates that the higher RMSE of small-particle samples mainly originated from the dispersion of predicted-label distribution rather than from a unidirectional systematic Bias.

The illumination and particle-size experiments show that complex operating conditions reduce the stability of coal–biomass blend-ratio estimation, but the mechanisms differ. Illumination variation mainly causes imaging-domain differences by changing gray-level distribution, texture contrast, and shadow morphology, making systematic coal-fraction-related Bias more likely under nighttime infrared supplementary illumination. Particle-size variation mainly changes the number of particles per unit area, local boundary density, and occlusion-overlap morphology, causing the predicted-label distribution of small-particle high-density samples to become more dispersed. Because coal and biomass reflectance properties under visible and infrared light were not independently measured in this study, these mechanistic explanations require further verification in future research using image-statistical features and material optical properties.

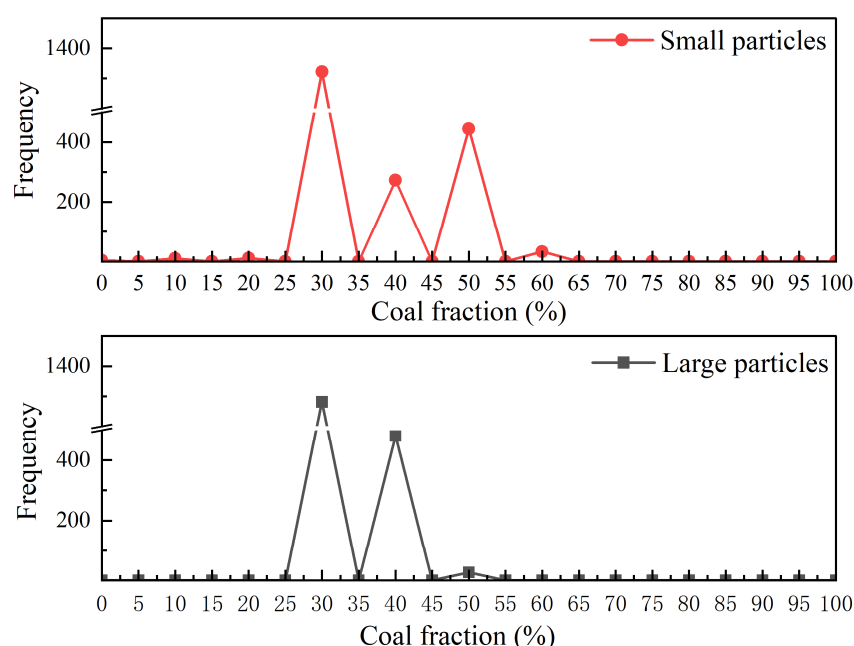


Figure 12. Ratio-label distribution for different particle-size samples at a true coal mass fraction of 30%. Note: Small particles were defined as $6\text{ mm} < d \leq 12.5\text{ mm}$, and large particles were defined as $12.5\text{ mm} < d < 100\text{ mm}$.

3.4. Typical Error-Case Analysis

Figure 13 shows representative error scenes in coal–biomass blend-ratio recognition. To facilitate analysis of error sources, each subfigure is annotated with the true coal fraction of the sample, the coal fraction corresponding to the single-frame predicted ratio-label, and the prediction confidence. It should be noted that the predicted coal fraction shown in the figure is not the result obtained by pooling extracted frames from multiple videos under the same true coal-fraction condition; it is the ratio label output by a single-frame detection result.

Figure 13a presents an underestimation case for a high-coal-fraction sample under nighttime infrared supplementary illumination. The true coal fraction of the sample was 90%, whereas the model predicted a single-frame ratio label corresponding to 50%, with a confidence of 0.33. This single-frame result is consistent with the negative Bias trend in the medium-to-high coal-fraction range under nighttime infrared supplementary illumination discussed above, suggesting that infrared illumination may weaken the brightness and

texture differences between coal and biomass particles. The low prediction confidence also reflects insufficient discrimination stability under this imaging condition.

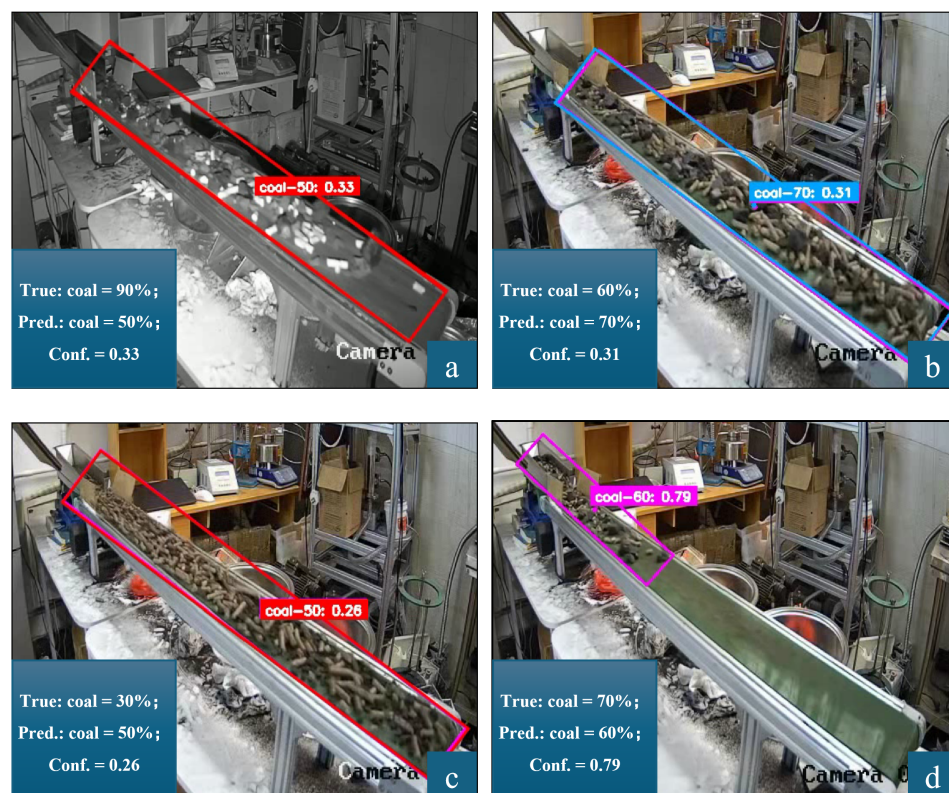


Figure 13. Typical error cases: (a) high-coal-fraction underestimation under nighttime infrared supplementary illumination; (b) intermediate-ratio confusion; (c) small-particle, high-density sample error; (d) misclassification caused by local stacking and occlusion.

Figure 13b shows an intermediate-ratio confusion case. The true coal fraction of the sample was 60%, while the model predicted a single-frame ratio label corresponding to 70%. This phenomenon is consistent with the higher RMSE in the intermediate ratio range described above. Under intermediate coal fractions, the visual boundaries between adjacent ratio labels are ambiguous, making the model prone to shifts between adjacent or near-adjacent ratio intervals. This type of error is not entirely equivalent to a semantic-class error in ordinary multi-class detection tasks; rather, it reflects the transitional and uncertain nature of adjacent labels after discretizing a continuous coal-fraction variable.

Figure 13c presents an error case for a small-particle high-density sample. The true coal fraction of the sample was 30%, whereas the model predicted a single-frame ratio label corresponding to 50%. Under small-particle conditions, dense particle boundaries, interlacing, and occlusion may cause predicted labels to spread toward adjacent or near-adjacent ratios, thereby increasing ratio-estimation error.

Figure 13d shows an underestimation case caused by local stacking and occlusion. The true coal fraction of the sample was 70%, but the model predicted a single-frame ratio label corresponding to 60%. Local stacking, occlusion, and uneven material spreading can be observed inside the detection box, and the material composition visible in the local region may not fully match the overall true ratio. When the model mainly relies on local visible regions for judgment, it can be affected by local spatial distribution, causing the prediction to shift toward a lower coal-fraction label.

These representative error cases indicate that coal-biomass blend-ratio recognition errors are not caused only by inaccurate bounding-box localization. They are jointly

formed by illumination-domain differences, particle-scale variation, local stacking and occlusion, and discretization of a continuous ratio variable. For practical online-monitoring applications, model evaluation should not focus only on mAP but should also consider Bias, RMSE, cross-illumination stability, and predicted-label distributions. For nighttime infrared illumination and small-particle high-density samples, future studies can improve prediction stability through illumination augmentation, domain adaptation, hard-sample reweighting, temporal smoothing, and continuous-ratio regression.

3.5. Occlusion-Based Interpretability Analysis

To further examine the basis of model prediction, an occlusion-based interpretability analysis was conducted on representative test samples. As shown in Figure 14, the visible mixed-material region, conveyor-belt background, and local highlight regions were separately masked, and the changes in detection confidence were compared. As summarized in Table 13, masking the target region reduced the detection confidence to zero for all three representative samples, with an average confidence drop of 0.8887. In contrast, masking the background and highlight regions resulted in much smaller average confidence drops of 0.0915 and 0.0683, respectively. These results indicate that the model mainly relied on the visible coal–biomass material region rather than the conveyor-belt background or local illumination highlights, supporting the effectiveness of explicit material-region localization for blend-ratio recognition.

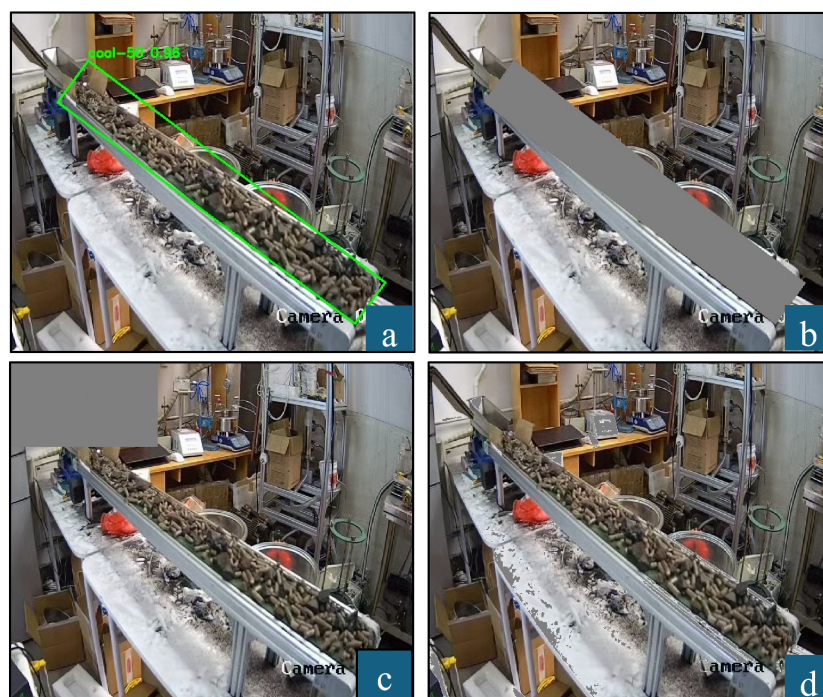


Figure 14. Representative occlusion results for model interpretability analysis: (a) original detection result; (b) target-masked image; (c) background-masked image; (d) highlight-masked image.

Table 13. Effects of target, background, and highlight occlusion on detection confidence.

Image	Class	Original	Target Masked	Background Masked	Highlight Masked	Target Drop	Background Drop	Highlight Drop
1	coal-50	0.9638	0	0.9495	0.8618	0.9638	0.0143	0.102
2	coal-30	0.8452	0	0.7374	0.8172	0.8452	0.1078	0.028
3	coal-50	0.8572	0	0.7049	0.7823	0.8572	0.1523	0.0748
Average	—	0.8887	0	0.7973	0.8204	0.8887	0.0915	0.0683

4. Discussion

Previous studies have shown that color, texture, and particle-morphology information in fuel images can be used for biomass-fuel characterization, feeding-state estimation, and coal-gangue recognition [1–7]. Recent visual studies in the coal industry have also increasingly focused on complex illumination, lightweight detection, and engineering deployment [8–10,13–16]. Compared with these studies, the coal–biomass blend-ratio recognition problem investigated here is not merely fuel-type identification or target localization; rather, it estimates coal mass fraction, a process variable, from continuous conveyor-belt images. Therefore, the main significance of this study is not to prove that YOLO models can perform general visual detection, but to demonstrate that object-detection outputs can be further transformed into blend-ratio trend information, providing a non-contact and continuous vision-based soft-sensing approach for fuel conveying processes.

The results indicate that the key to coal–biomass blend-ratio recognition is not only the selection of a larger detection network but also how the model is guided to focus on effective material regions related to ratio discrimination. Although a whole-image classification model can learn features from an entire image, it is susceptible to interference from non-material regions. By contrast, a detection model explicitly restricts the visible mixed-material region, making ratio discrimination closer to the actual physical object. The role of OBBs should also be understood from this perspective: their advantage is not simply improved geometric fitting accuracy of the bounding box, but reduced background redundancy through a more compact regional representation, enabling lightweight models to obtain more stable ratio estimates under a limited parameter scale.

From an engineering application perspective, the average RMSE of approximately 6–7% obtained in this study is more suitable for online trend monitoring, anomaly warning, and operational assistance than for independent high-precision mass metering or closed-loop control. When visual estimates continuously deviate from the target co-firing ratio within a time window, they can provide early indications of feeding abnormalities, material stratification, non-uniform mixing, or fuel-supply fluctuations and can assist operators in checking weighing, feeding, and combustion-feedback information. Compared with offline sampling, the visual method provides higher temporal resolution; compared with relying only on weighing or feeding setpoints, it directly reflects the visible state of the mixed material on the conveyor-belt surface. Thus, this study extends coal–biomass fuel image recognition from static classification or object detection to online visual soft sensing for the co-firing fuel-conveying process.

Illumination condition and particle size also affected model robustness. Infrared supplementary illumination changed gray-level distributions and texture patterns, thereby contributing to systematic prediction Bias under some coal-fraction conditions. In addition, small-particle, high-density samples showed more dispersed predicted-label distributions because of denser local boundaries, stronger inter-particle occlusion, and more complex stacking morphology. These effects indicate that cross-condition robustness remains important for practical deployment and should be further improved through domain adaptation and hard-sample enhancement.

This study has several limitations. Although the experimental fuels and imaging device were consistent with practical power-plant applications, and the laboratory-scale conveyor-belt platform simulated continuous fuel conveying and multi-condition imaging, the long-term operating environment of a real power plant is more complex. Field dust deposition, equipment vibration, illumination fluctuations, lens contamination, aging of supplementary illumination devices, variations in coal and biomass types, and differences in camera installation may all affect image quality and model prediction stability. In addition, the true labels used in this study were the overall coal mass fractions at the

sample-preparation stage, whereas the model observes only local visible regions on the conveyor-belt surface in a single frame. Therefore, the two quantities are not necessarily strictly identical, which introduces inherent uncertainty in the sample-level weak labels. The multi-frame frequency-weighted strategy can reduce single-frame local fluctuations and label ambiguity, but it remains an indirect estimation method based on discrete ratio labels. Future work should collect long-term operating data on a pilot-scale platform or at a power-plant site, establish calibration relationships between visual estimates and feeding rates, belt-scale measurements, or offline fuel-analysis results, and combine cross-illumination-domain adaptation, hard-sample enhancement, continuous-ratio modeling, model compression, and edge-deployment optimization to improve robustness, physical interpretability, and engineering reliability in complex industrial environments.

5. Conclusions

This study addressed the need for online perception of blending states during coal–biomass co-firing fuel transport. A laboratory-scale conveyor-belt image dataset covering different coal fractions, illumination conditions, and particle sizes was constructed, and coal mass-fraction estimation was investigated using YOLO-series models. By establishing a whole-image classification baseline, constructing HBB and OBB annotation systems, and comparing the effects of model scale and bounding-box representation, the following conclusions were obtained:

- (1) Coal–biomass blend-ratio recognition should not be treated simply as a whole-image classification task. It is more appropriately formulated as a vision-based soft-sensing problem based on material-region detection. YOLOv8-cls achieved an average RMSE of 13.98 pp, higher than that of detection-based models, indicating that explicit localization of the visible mixed-material region improves the reliability of ratio estimation.
- (2) Increasing model scale reduced ratio-estimation error but substantially increased parameter size and inference cost. YOLOv8m achieved the lowest average RMSE, whereas YOLOv8n-OBB achieved estimation performance close to that of YOLOv8m at a smaller model scale, showing a better accuracy–efficiency balance.
- (3) OBB representation improved the ratio-estimation stability of lightweight models by more tightly covering visible material regions and reducing background interference. Compared with YOLOv8n, YOLOv8n-OBB substantially reduced the average RMSE, indicating enhanced robustness under lightweight deployment.
- (4) Illumination conditions and particle-size variation affected cross-condition recognition performance. Nighttime infrared illumination and small, high-density particle samples increased estimation error, indicating that imaging conditions, particle scale, and local stacking remain key factors for model robustness and practical deployment.

Overall, this study confirms the feasibility of a YOLOv8-based visual recognition approach for coal–biomass blending on conveyor belts. The method provides a non-contact, continuous vision-based soft-sensing approach for trend monitoring, and the YOLOv8n-OBB detection model demonstrates engineering application potential under lightweight deployment. Future work should focus on cross-illumination-domain adaptation, hard-sample enhancement, and field validation to further improve industrial applicability and robustness.

Author Contributions: Y.M.: Investigation, methodology, data curation, writing—original draft preparation, conceptualization. H.Y.: Software, investigation, writing—review and editing. C.Z.: Resources. W.L.: Resources. Z.R.: Project administration. H.P.: Project administration. X.H.: Project administration. X.W.: Project administration. Z.L.: Conceptualization, funding acquisition, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2022YFB4202000.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, ChatGPT (GPT-5.5 Thinking, OpenAI) was used only for language polishing and formatting assistance. The authors reviewed and edited all AI-assisted text and take full responsibility for the scientific content of the manuscript.

Conflicts of Interest: Author Yisheng Mao was employed by the company Guangdong Electric Power Development Co., Ltd. Author Xu Huang was employed by the company Guangdong Electric Power Development Co., Ltd. Author Cuihua Zhang was employed by the company Guangdong Red Bay Power Generation Co., Ltd. Author Weihui Liao was employed by the company Guangdong Red Bay Power Generation Co., Ltd. Author Zhilong Ruan was employed by the company Guangdong Red Bay Power Generation Co., Ltd. Author Haibing Pu was employed by the company Guangdong Red Bay Power Generation Co., Ltd. Author Xiaolong Wu was employed by the company Xian Thermal Power Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

YOLO	You Only Look Once
HBB	Horizontal Bounding Box
OBB	Oriented Bounding Box
RMSE	Root Mean Square Error
mAP	Mean Average Precision
NMS	Non-Maximum Suppression
GFLOPs	Giga Floating-Point Operations

References

1. Lukas, J.; Kolb, S.; Heinbuch, J.; Willig, L.; Plankenbuehler, T.; Mueller, D.; Karl, J. Image-based biomass characterization: Comparison of conventional image processing and a deep learning approach. *Fuel* **2023**, *341*, 127705. [\[CrossRef\]](#)
2. Oegren, Y.; Sepman, A.; Fooladgar, E.; Weiland, F.; Wiinikka, H. Development and evaluation of a vision driven sensor for estimating fuel feeding rates in combustion and gasification processes. *Energy AI* **2024**, *15*, 100316. [\[CrossRef\]](#)
3. Plankenbuehler, T.; Mueller, D.; Karl, J. An adaptive and flexible biomass power plant control system based on on-line fuel image analysis. *Therm. Sci. Eng. Prog.* **2023**, *40*, 101765. [\[CrossRef\]](#)
4. Gudavalli, C.; Bose, E.; Donohoe, B.S.; Sievers, D.A. Real-time biomass feedstock particle quality detection using image analysis and machine vision. *Biomass Convers. Biorefinery* **2022**, *12*, 5739–5750. [\[CrossRef\]](#)
5. 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* **2022**, *198*, 111415. [\[CrossRef\]](#)
6. Zhang, C.; Dou, D.; Sun, F.; Huang, Z. Detecting coal content in gangue via machine vision and genetic algorithm-backpropagation neural network. *Measurement* **2022**, *201*, 111739. [\[CrossRef\]](#)
7. Liu, H.; Xu, K. Recognition of gangues from color images using convolutional neural networks with attention mechanism. *Measurement* **2023**, *206*, 112273. [\[CrossRef\]](#)
8. Zeng, Q.; Zhou, G.; Wan, L.; Wang, L.; Xuan, G.; Shao, Y. Detection of coal and gangue based on improved YOLOv8. *Sensors* **2024**, *24*, 1246. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Zong, G.; Yue, Y.; Shan, W. Optimization study of coal gangue detection in intelligent coal selection systems based on the improved YOLOv8n model. *Electronics* **2024**, *13*, 4155. [\[CrossRef\]](#)
10. Gao, L.; Yu, P.; Dong, H.; Wang, W. Multi-scale fusion lightweight target detection method for coal and gangue based on EMBS-YOLOv8s. *Sensors* **2025**, *25*, 1734. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Wang, X.; Zhao, P.; Yue, S.; Gu, Y. An improved cross algorithm edge detection method and its application in coal gangue identification. *Int. J. Coal Prep. Util.* **2026**, *46*, 835–847. [\[CrossRef\]](#)
12. Chen, P.; Nie, X.; Ning, Y.; Zhang, Y. Learning efficient and adaptive cross-channel dependencies for weakly-supervised object detection. *IEEE Trans. Multimed.* **2025**, *27*, 8954–8966. [\[CrossRef\]](#)

13. Zhang, K.; Wang, T.; Yang, X.; Xu, L.; The, J.; Tan, Z.; Yu, H. STATNet: One-stage coal-gangue detector based on deep learning algorithm for real industrial application. *Energy AI* **2024**, *17*, 100388. [[CrossRef](#)]
14. Cao, Z.; Fang, L.; Li, Z.; Li, J. Lightweight target detection for coal and gangue based on improved YOLOv5s. *Processes* **2023**, *11*, 1268. [[CrossRef](#)]
15. Lv, Z.; Fan, Y.; Sha, T.; Cui, Y.; Wu, Y.; Lv, H.; Sun, M.; Tu, Y.; Xu, Z.; Wang, W. A large-scale open image dataset for deep learning-enabled intelligent sorting and analyzing of raw coal. *Sci. Data* **2025**, *12*, 403. [[CrossRef](#)] [[PubMed](#)]
16. Cao, X.; Liu, H.; Liu, Y.; Li, J.; Xu, K. Coal and gangue detection networks with compact and high-performance design. *Sensors* **2024**, *24*, 7318. [[CrossRef](#)] [[PubMed](#)]
17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: Piscataway, NJ, USA, 2016; pp. 779–788. [[CrossRef](#)]
18. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 6517–6525. [[CrossRef](#)]
19. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767. [[CrossRef](#)]
20. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning RoI Transformer for oriented object detection in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 2844–2853. [[CrossRef](#)]
21. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018*; IEEE: Piscataway, NJ, USA, 2018; pp. 3974–3983. [[CrossRef](#)]
22. Ultralytics. Oriented Bounding Boxes Object Detection. Available online: <https://docs.ultralytics.com/tasks/obb/> (accessed on 29 March 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.