



## Article

# An Efficient and Accurate Quality Inspection Model for Steel Scraps Based on Dense Small-Target Detection

Pengcheng Xiao <sup>1,2,3</sup> , Chao Wang <sup>1</sup> , Liguang Zhu <sup>4,\*</sup> , Wenguang Xu <sup>3</sup> , Yuxin Jin <sup>1</sup>  and Rong Zhu <sup>3</sup> 

<sup>1</sup> College of Metallurgical and Energy, North China University of Science and Technology, Tangshan 063210, China; xiaopc@ncst.edu.cn (P.X.); chaowang202@126.com (C.W.); jyx00824@126.com (Y.J.)

<sup>2</sup> Iron and Steel Laboratory of Hebei Province, Tangshan 063210, China

<sup>3</sup> Metallurgical and Ecological Engineering School, University of Science and Technology Beijing, Beijing 100083, China; xuwenguang@xs.ustb.edu.cn (W.X.); zhurong12001@126.com (R.Z.)

<sup>4</sup> College of Materials Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

\* Correspondence: zhulgts@126.com

**Abstract:** Scrap steel serves as the primary alternative raw material to iron ore, exerting a significant impact on production costs for steel enterprises. With the annual growth in scrap resources, concerns regarding traditional manual inspection methods, including issues of fairness and safety, gain increasing prominence. Enhancing scrap inspection processes through digital technology is imperative. In response to these concerns, we developed CNIL-Net, a scrap-quality inspection network model based on object detection, and trained and validated it using images obtained during the scrap inspection process. Initially, we deployed a multi-camera integrated system at a steel plant for acquiring scrap images of diverse types, which were subsequently annotated and employed for constructing an enhanced scrap dataset. Then, we enhanced the YOLOv5 model to improve the detection of small-target scraps in inspection scenarios. This was achieved by adding a small-object detection layer (P2) and streamlining the model through the removal of detection layer P5, resulting in the development of a novel three-layer detection network structure termed the Improved Layer (IL) model. A Coordinate Attention mechanism was incorporated into the network to dynamically learn feature weights from various positions, thereby improving the discernment of scrap features. Substituting the traditional non-maximum suppression algorithm (NMS) with Soft-NMS enhanced detection accuracy in dense and overlapping scrap scenarios, thereby mitigating instances of missed detections. Finally, the model underwent training and validation utilizing the augmented dataset of scraps. Throughout this phase, assessments encompassed metrics like mAP, number of network layers, parameters, and inference duration. Experimental findings illustrate that the developed CNIL-Net scrap-quality inspection network model boosted the average precision across all categories from 88.8% to 96.5%. Compared to manual inspection, it demonstrates notable advantages in accuracy and detection speed, rendering it well suited for real-world deployment and addressing issues in scrap inspection like real-time processing and fairness.

**Keywords:** steel scrap; classification; deep learning; target detection



**Citation:** Xiao, P.; Wang, C.; Zhu, L.; Xu, W.; Jin, Y.; Zhu, R. An Efficient and Accurate Quality Inspection Model for Steel Scraps Based on Dense Small-Target Detection. *Processes* **2024**, *12*, 1700. <https://doi.org/10.3390/pr12081700>

Academic Editor: Yo-Ping Huang

Received: 10 July 2024

Revised: 9 August 2024

Accepted: 12 August 2024

Published: 14 August 2024



**Copyright:** © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In today's global economic context, the steel industry occupies a pivotal role but confronts challenges in environmental protection and sustainable development. With global crude steel production reaching 1.8882 billion tons in 2023, meeting emission reduction targets set by the Paris Agreement emerges as a pressing concern for the steel industry's long-term sustainability [1,2]. Enhancing the utilization of scrap, acknowledged as the most environmentally friendly raw material in steel production [3], is crucial to achieving this objective. Electric arc furnace steelmaking and the adoption of scrap over converter steelmaking are critical strategies advancing the industry toward low-emission

steelmaking, leveraging benefits in resource conservation and reduced CO<sub>2</sub> emissions [4,5]. The high-quality inspection of scrap forms a critical link in this process, impacting not only steelmaking efficiency but also directly affecting enterprise economic benefits and environmental responsibilities [6]. Nevertheless, despite the mechanization achieved in transportation, loading, and unloading processes at numerous scrap yards [7,8], a considerable scope remains for enhancing the intelligent recognition of various grades of scrap material. The prevalence of traditional manual inspection methods introduces subjective biases, impacting inspection fairness and accuracy, besides being inefficient and hazardous, thereby falling short of modern steelmaking requirements [9]. With the rapid growth of scrap resources and the constraints of traditional inspection methods, there is an urgent need to develop AI-based algorithms for scrap classification and quality assessment systems. These systems will markedly enhance the intelligence level of detection, facilitating more precise assessments of scrap quality and costs, enhancing operational efficiency, and mitigating risks. This advancement offers robust technological support for low-emission and sustainable development within the steel industry. Hence, the development and deployment of AI-driven scrap classification and quality assessment systems are pivotal in fostering high-quality development within the steel industry.

The development of intelligent classification methods for scrap types based on deep learning algorithms commenced later compared to other fields. Nevertheless, due to the rapid advancement and extensive application of deep learning technologies in recent years, many scholars have conducted research and achieved notable outcomes in this domain. Xu et al. [9] developed a deep learning-based model for classifying and grading scrap, integrating the Squeeze-and-Excitation Network (SENet) attention mechanism with feature extraction from cross-stage local networks. Tu et al. [10] proposed a novel framework to tackle challenges in scrap grading, encompassing complex background handling, the precise detection of scrap, and grading assessment. This framework significantly enhances grading accuracy while maintaining speed and has been successfully implemented across various steel plants. Ichiro DAIGO et al. [11] employed transfer learning to classify heavy scrap using pyramid pooling semantic segmentation on a small dataset, yielding favorable classification outcomes for thickness or diameter. Xiao et al. [12] developed a deep learning network model for scrap classification and grading using SENet, incorporating cross-stage local networks for feature extraction and a spatial pyramid structure to handle cross-scale challenges in scrap images, and integrating SENet into the feature extraction network. Mei [13] investigated aspects including rust degree identification, coating recognition and detection, and the rapid detection of alloy elements using machine vision technology and LIBS technology. Qiu [14] optimized the image fogging convergence algorithm and the non-maximum suppression algorithm using the YOLOv3 model, successfully detecting scrap raw materials in input images. These studies offer new avenues for advancing the intelligent classification of scrap types.

Due to the high recycling value of non-ferrous metals, current research on AI-based algorithms for classifying scrap metal primarily concentrates on the non-ferrous metal sector to align with industrial needs and market demands. Gao et al. [15] proposed a method for sorting copper impurities in scrap material using deep learning technology, integrating optical recognition and shape detection. They achieved automatic recognition of copper (Cu) with an accuracy as high as 90.6% using separable convolutional neural networks. Penumuru et al. [16] successfully implemented automatic material identification in an Industry 4.0 environment by integrating machine vision and machine learning technologies, accurately identifying and classifying aluminum, copper, medium-density fiberboard, and low-carbon steel. Diaz-Romero et al. [17] utilized transfer learning methods including fine-tuning and feature extraction for the real-time classification of casting and forging (C&W) alloys in conveyor belt systems. Koyanaka and Kobayashi [18] incorporated neural network analysis into a waste identification algorithm for the automatic separation of lightweight metal scrap, achieving an average separation accuracy of 85% for mixed alloys including cast aluminum, forged aluminum, and magnesium across three different

ELV shredding facilities. Diaz-Romero et al. [19] proposed a deep learning model that integrates dense convolutional networks, back-propagation neural networks, and principal component analysis for predicting and evaluating the quality of composite castings, forgings, and stainless steel datasets, obtaining satisfactory outcomes. Chen et al. [20] investigated the small-sample identification and separation of non-ferrous metals based on deep learning, using data augmentation, adjusting focal loss functions, and adjusting Intersection over Union (IOU) thresholds with the YOLOv3 algorithm to achieve accuracies of 95.3% for aluminum scrap recognition and 91.4% for copper scrap recognition. These studies make substantial contributions to the advancement of intelligent classification and the separation of metal scrap using deep learning techniques.

Recent years have seen substantial advancements in object detection technology, especially in handling complex and dense scenes, with a notable focus on detecting small-scale targets. Innovations such as multi-feature extraction methods and multi-scale fusion techniques have effectively tackled the challenges of identifying dense and small-scale objects in scenarios such as aerial imaging, space exploration, and underwater target detection. Xu et al. [21] proposed enhancing the feature extraction capability of the YOLOv3 model by introducing a “replicated” Backbone network to construct an auxiliary network, thereby enhancing the detection of distant small targets such as vehicles, pedestrians, and traffic signs during driving. Ming et al. [22] addressed aerial image object detection by employing the Position-Sensitive Feature Pyramid Network (PS-FPN) to precisely extract location-sensitive features of small, densely arranged objects, and introduced a distance-rotated IoU loss to mitigate discrepancies between training and evaluation metrics. Wang et al. [23] proposed a multi-scale feature fusion pyramid network for space exploration tasks, enhancing target extraction capability with a CNN-CST module based on Swin Transformer, refining the SE attention mechanism, and introducing enhanced spatial pyramid pooling to optimize performance in detecting small targets. Fang et al. [24] developed the YOLO-RAD algorithm for dense scenes, integrating the Reception Field Attention (RFA) mechanism, the Adaptive Spatial Feature Fusion module (ASFF), and the dynamic head structure for small targets (DyHead-S), thereby substantially enhancing pedestrian detection accuracy in crowded scenarios. Zhao et al. [25] enhanced the YOLOv7 network for underwater target detection by integrating SE attention and RFE modules and incorporating Wasserstein distance as a novel metric to supplant traditional loss functions. Liu et al. [26] introduced CFNet and CBAM modules, presenting the YOLOv8-CB algorithm for enhanced lightweight multi-scale pedestrian detection via a Bidirectional Feature Pyramid Network (BIFPN) architecture.

Object detection algorithms based on deep learning are categorized into two main types: two-stage and single-stage detection algorithms [27]. Two-stage detection algorithms such as Faster R-CNN (Region-Based Convolutional Neural Networks) and Mask R-CNN (Mask Region-Based Convolutional Neural Network) are noted for their high accuracy but exhibit slower speeds compared to single-stage algorithms. Single-stage detection algorithms like the YOLO (You Only Look Once) series and SSD (Single-Shot MultiBox Detector) provide excellent speed and accuracy and have been widely adopted, despite potential challenges in detecting small targets and missed detections [28–31].

Previously, our team explored an intelligent classification of scrap in laboratory simulations, developing object detection models for recognizing scrap types [9,12,32]. However, due to factors such as high-altitude shooting in industrial settings and densely packed scrap, the collected data included small and densely distributed targets, resulting in the poor performance of the laboratory-established models. To tackle the challenge of scrap detection in industrial scenarios, this paper proposes a novel scrap-quality inspection model named CNIL-Net (CA+Soft-NMS+Improved Layer) structured on the YOLOv5 model. Through enhancements in detection layers and the integration of the CA (Coordinate Attention) mechanism and Soft-NMS (Soft non-maximum suppression algorithm), CNIL-Net classifies overlapping small-target scrap across multiple categories and scales.

Comparative evaluations with models such as YOLOv7 and YOLOv8 demonstrate the superior performance of the proposed model.

The innovative contributions of this paper are as follows:

1. First item; The detection performance of traditional object detection algorithms is enhanced for densely packed small targets, thereby achieving accurate classification in scrap images and reducing missed detections of overlapping targets;
2. Second item; Compared to manual classification and grading, the CNIL-Net model exhibits significant advantages in terms of accuracy and efficiency, paving the way for the development of an intelligent unmanned scrap acceptance system.

The structure of this paper is outlined as follows: Section 2 offers an in-depth description of the dataset creation process, along with a detailed overview of the CNIL-Net model's architecture and its related enhancement modules and algorithms. Section 3 discusses the experimental setup and evaluation criteria used for the model, and presents the training and validation results for the scrap steel dataset. Section 4 provides a summary of the research findings.

## 2. Materials and Methods

### 2.1. Preparation of the Datasets

Three high-definition industrial cameras were installed by our team at a steel company in China, mounted on brackets positioned between 10.2 m and 10.5 m high. This positioning ensures comprehensive coverage of a 5.4 m long and 2.3 m wide area used by dump trucks. Whenever a truck reaches the designated area in the scrap yard, a claw crane automatically unloads the scrap. The system automatically detects the crane's grabbing motion and captures images according to its frequency, thereby minimizing redundant detections and storage. Figure 1 illustrates a schematic diagram showing the camera positions relative to the scrap truck, with a total of 934 images retained.

In alignment with the core requirements of steel companies, where various scrap types are associated with distinct recycling prices, plates were classified based on three thickness categories, while non-plate scrap was categorized according to its type and whether it exceeded standard lengths. The data were annotated using LabelImg software and categorized into 9 labels, namely plate thickness <3 mm; plate thickness 3–6 mm; plate thickness >6 mm; airtight; inclusion; overlength (1.2–1.5 m), overlength (1.5–2 m); scattered; and ungraded. However, due to the complexity of acceptance scenarios, class data imbalance, and inadequate training on small-target samples, the model's generalization capability is compromised [33–35]. Moreover, deep learning algorithms necessitate extensive datasets to support their parameter-intensive operations. To enhance dataset robustness, we employed five data augmentation methods, namely adding noise, adjusting brightness, cropping, mirroring or rotating, and randomly combining these strategies to further diversify the dataset, as depicted in Figure 2. Following data augmentation, the dataset expanded to 3736 images, encompassing a total of 298,260 labels. For optimal model training and validation, the dataset was divided into training and validation sets at a ratio of 9:1, consisting of 3362 images for training and 374 for validation. This dataset was named ESD (Enhanced Scrap Datasets). Table 1 details the statistics of the labels for each category, and this dataset was used for training and validation for both the comparison and ablation experiments in this paper.



Figure 1. Schematic site layout.

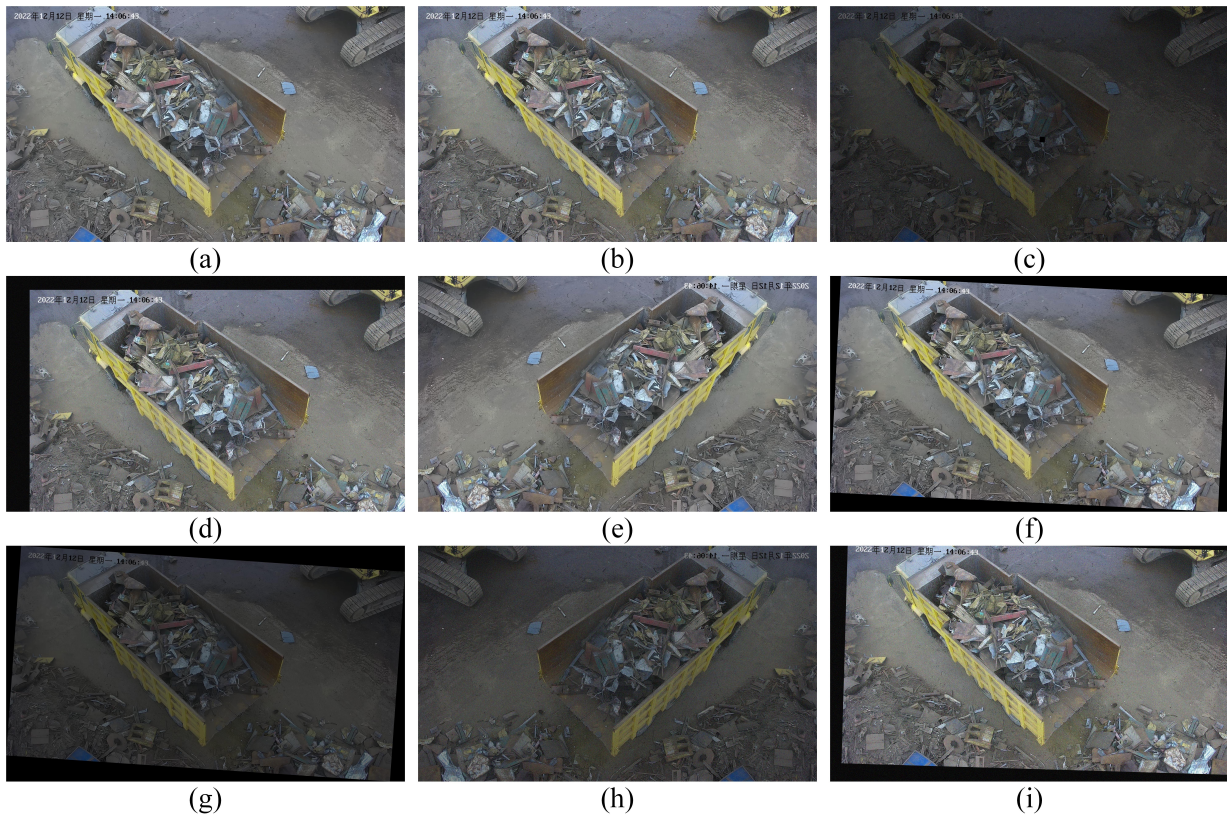


Figure 2. Image enhancement effects: (a) original image; (b) with added noise; (c) brightness adjustment; (d) cropping; (e) mirroring; (f) rotating; (g) random combination; (h) random combination; (i) random combination.

**Table 1.** Labeling quantities for each category of the dataset.

Category	Number of Dataset Labels	Number of Training Set Labels
<3 mm	11,248	10,166
3–6 mm	12,2812	109,766
>6 mm	10,7476	97,407
airtight	716	645
inclusion	2260	2033
overlength (1.2–1.5 m)	10,952	9940
overlength (1.5–2 m)	6720	6067
scattered	9164	8214
ungraded	26,912	24,224

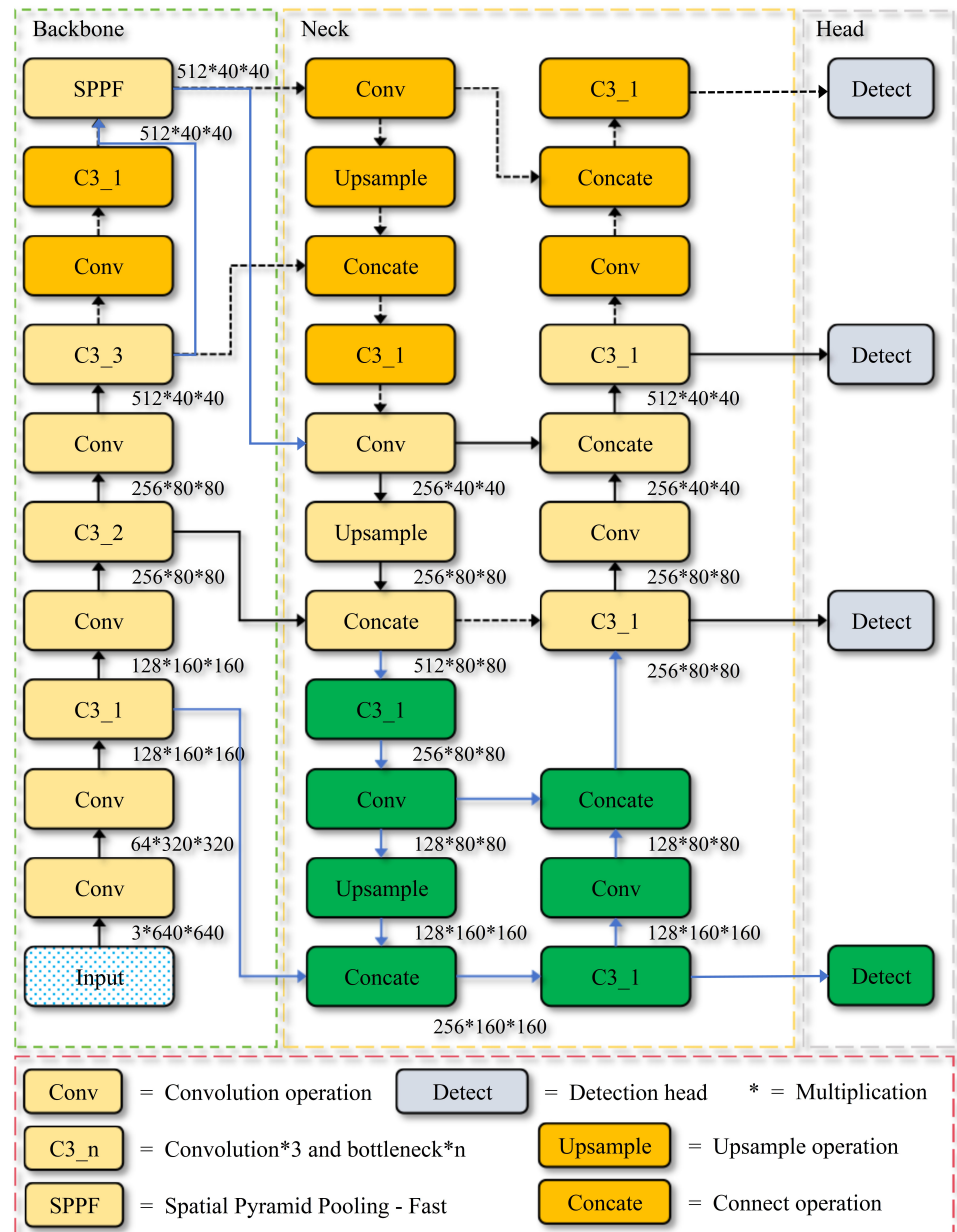
## 2.2. Model Improvement

### 2.2.1. Detection Layer Structure

In its original design, the YOLOv5 model divides the initial detection layers into P3 to P5, allowing it to detect objects at various scales. Through hierarchical downsampling operations, the model handles  $640 \times 640$ -pixel images, producing feature maps of  $20 \times 20$ ,  $40 \times 40$ , and  $80 \times 80$  at scales of  $32\times$ ,  $16\times$ , and  $8\times$ , which correspond to objects ranging from large to small sizes. However, the small size and low resolution of small-target scrap restrict the extraction of feature information and introduce considerable noise. Because of YOLOv5's significant downsampling factor, deep feature maps struggle to capture the characteristics of small-target scrap effectively, potentially resulting in missed detections and affecting overall model performance [36].

To tackle these challenges, a proposed solution involves introducing an additional P2 layer (scaled to  $160 \times 160$ ) [37]. This enhancement aims to enhance the model's ability to focus on small targets and improve the detection of distant scrap targets. Specifically, after the 17th layer in the architecture, upsampling operations are employed on the feature maps to augment their size. At the 20th layer, the  $160 \times 160$  feature map obtained is fused with the second-layer feature map from the Backbone to create a larger feature map suitable for detecting small targets.

It is important to note that as the number of detection layers in the model increases, so do parameters, network layers, and other metrics. Given the rarity of large-sized targets in aerial images of scrap, in this study, we chose to eliminate the larger detection layer P5 (corresponding to a scale of  $20 \times 20$ ), thus optimizing the model for swift deployment in industrial environments. Figure 3 illustrates the network structure after the addition of the P2 layer and the removal of the P5 layer, with dashed and blue lines indicating removed and added data flow components, while orange and green blocks represent removed and added modules. This revised network configuration is henceforth referred to as the Improved Layer (IL).



**Figure 3.** IL network structure.

### 2.2.2. Coordinate Attention (CA)

The inspection scenarios present complex backgrounds with considerable interference, and the scrap inside the carriage holds limited informative content. Furthermore, the distinctions in characteristics among different types of scrap are not clear, presenting substantial challenges for model training. To improve detection accuracy and mitigate irrelevant interference, this study introduces an attention mechanism. CA mechanisms like the well-established SENet [38] have proven to greatly enhance model performance. However, they frequently neglect crucial positional information required for selectively generating spatial features. Spatial attention mechanisms, in contrast, concentrate exclusively on identifying spatially significant regions within the network, conserving resources for critical areas while disregarding inter-channel relationships [39].

The CA mechanism integrates both channel and spatial information to enhance feature representation. It decomposes Coordinate Attention into two one-dimensional global pooling processes along each spatial direction, aggregating channel features to capture long-range dependencies and preserve precise positional information. This method produces

two distinct feature maps with directional awareness. Moreover, embedding positional information from input feature maps into aggregated feature vectors of Coordinate Attention enhances the representation of regions of interest across larger areas while mitigating the excessive computational overhead. The specific process is depicted in Figure 4. Introducing the CA mechanism enables the model to accurately detect scrap inside carriages in complex inspection scenarios, thus enhancing detection accuracy.

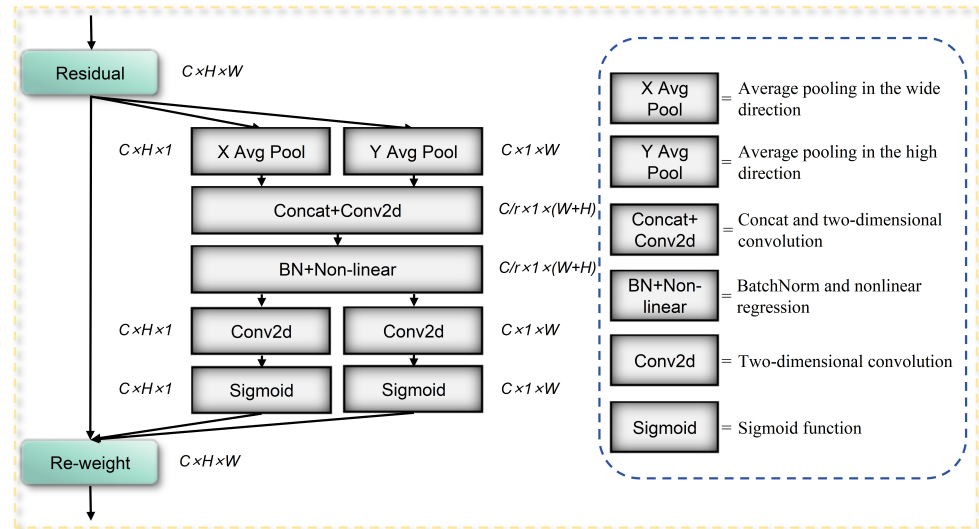


Figure 4. CA structure diagram.

### 2.2.3. Soft Non-Maximum Suppression (Soft-NMS)

During the post-processing phase of object detection, non-maximum suppression (NMS) is frequently utilized to filter detection boxes, extracting high-confidence object detection results while minimizing false positives with low confidence. Its formula is depicted in Equation (1). However, in scenarios with stacked scrap, traditional NMS frequently erroneously removes overlapping or closely adjacent scrap boxes, resulting in missed detections. To tackle this issue, the Soft-NMS is introduced. The Soft-NMS calculates overlap not with simple binary thresholds but by integrating a penalty function to adjust detection box scores. When multiple overlapping bounding boxes are detected, the Soft-NMS employs Gaussian weighting to adjust their confidence levels. It sorts bounding boxes based on their confidence scores, applying a weighting function to reduce the confidence of lower-scored boxes rather than outright removing them. This approach mitigates the issue of NMS operations erroneously deleting overlapping detection boxes, as outlined in Equation (2) [40].

$$S_i = \begin{cases} S'_i, & IoU(M, b_i) < N_t \\ 0, & IoU(M, b_i) \geq N_t \end{cases} \quad (1)$$

$$S_i = S'_i \cdot e^{-\frac{(IoU(M, b_i))^2}{\sigma}} \quad (2)$$

In this formula,  $S_i$  represents the current confidence score of the detection frame,  $S'_i$  represents the confidence score of the detection frame before any modification,  $M$  denotes the frame with the highest score,  $N_t$  stands for the preset overlap threshold, and  $\sigma$  represents the algorithm's standard deviation.

### 2.3. CNIL-Net Model Structure

In this paper, the aforementioned improved modules and algorithms are integrated to construct the CNIL-Net scrap-quality inspection network model based on YOLOv5. Specifically, the addition of the detection layer P2 enhances the model's capability to detect small-target scrap, while the removal of P5 reduces model complexity. The CA mechanism



is incorporated into the Backbone to consider both channel and spatial information in the scrap features. The non-maximum suppression algorithm in the loss function is replaced with Soft-NMS to mitigate the model's problem regarding missing dense and overlapping targets. The structure of the CNIL-Net model is illustrated in Figure 5.

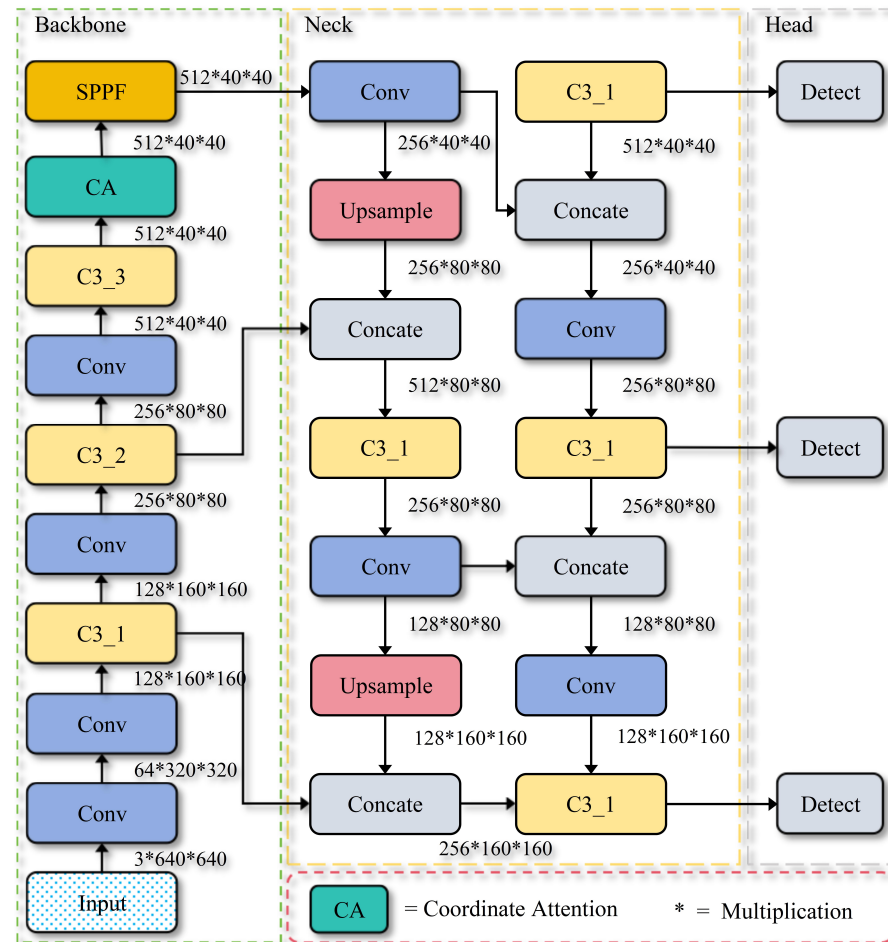


Figure 5. CNIL-Net structure.

### 3. Results and Discussion

During this experimental model training, the optimizer (SGD) was chosen, the image size (imgsz) was fixed at 960 pixels  $\times$  960 pixels, and the neural network was trained with a batch size of 16 samples per iteration. Following a comparison of activation functions like SiLU and Mish, SiLU was chosen as the activation function for this experiment. After evaluating the regression performance of loss functions like CIoU and GIoULoss on the ESD dataset, CIoU Loss was adopted as the loss function for this experiment. Both the comparison experiment and the ablation experiment were conducted using 200 epochs, with additional training sessions of 250 and 300 epochs added to the CNIL-Net network model to observe convergence.

#### 3.1. Experimental Environment and Evaluation Index

##### 3.1.1. Experimental Environment

The experimental setup in this paper included the Ubuntu 9.3.0 operating system, with the PyTorch framework employed for model training and validation. Acceleration was achieved using four NVIDIA GeForce RTX4090\*24 G graphics cards paired with Intel(R) Xeon(R) Gold 5318Y 2.1 GHz CPUs, accelerated with CUDA 11.4. Programmed in Python 3.8.10 using the Visual Studio Code.

### 3.1.2. Evaluation Metrics

Precision and recall gauge the accuracy and completeness of model checking, respectively. However, they sometimes exhibit a trade-off, which can be reconciled by introducing F1-value evaluation metrics. We evaluated the model using metrics such as mAP, F1 score, number of network layers, parameters, GFLOPs, and inference time. The formulas for several of these metrics are presented as follows:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (3)$$

$$\text{F1} = 2 \cdot \frac{P \cdot R}{P + R} \quad (4)$$

In these formulas,  $N$  represents the total number of categories,  $i$  denotes the  $i$ th category,  $\text{AP}$  stands for the average precision,  $P$  represents the precision rate, and  $R$  denotes the recall rate.

## 3.2. Comparison Experiments

### 3.2.1. Comparison of Different Algorithms

To validate CNIL-Net's superiority in scrap detection, we conducted comparison experiments with several models, namely YOLOv7-tiny, YOLOv7, YOLOv8m, YOLOv8l, and a scrap type recognition model previously proposed by our team [9,12,33]. YOLOv7-tiny is a lightweight variant of YOLOv7, whereas YOLOv8m and YOLOv8l are distinct versions within the YOLOv8 series. The primary differences among these models lie in their specific versions within the YOLO detection framework and their respective scales. The experimental results for these models, evaluated using the ESD dataset, are presented in Table 2. The CNIL-Net model outperforms other algorithms in terms of mAP and F1 values. In contrast, the YOLOv7-tiny model, despite having a lower number of network layers and parameters, and less model complexity compared to CNIL-Net, yielded only 68.6% and 72.2% in mAP and F1 values, respectively. Compared to previously proposed models in the lab, CNIL-Net exhibited higher model complexity but fewer network layers and parameters, resulting in respective improvements of 9.8%, 8.3%, and 8.1% in mAP. These results underscore CNIL-Net's suitability for scrap detection scenarios, combining effectiveness with lightweight design for enhanced scrap category detection.

**Table 2.** Comparison of different models.

Model	mAP (%)	F1 (%)	Number of Network Layers	Parameters	GFLOPs
YOLOv7-tiny	68.6	72.2	208	6,029,244	13.1
YOLOv7	89.5	89.7	314	36,524,924	103.3
YOLOv8m	88.5	88.4	295	25,902,640	79.3
YOLOv8l	90.8	90.7	365	4,391,520	164.9
Literature 9	86.7	86.2	304	21,389,990	49.0
Literature 13	88.2	87.2	309	21,390,088	49.1
Literature 27	88.4	87.4	307	21,389,982	49.0
CNIL-Net	96.5	91.4	272	7,980,198	68.6

### 3.2.2. Comparison of Different Model Scales

YOLOv5, developed by the Ultralytics team, has proven to be highly efficient, accurate, and lightweight, demonstrating robust detection capabilities across various scenarios. The model can be categorized into five versions—n, s, m, l, and x—based on variations in network depth and width, where each subsequent version (from n to x) progressively increases both depth and width. Increasing the network depth and width generally enhances the

model's detection accuracy, albeit at the cost of increased model complexity and training inference time.

To determine the optimal network depth and width for the scrap acceptance model, we compared and evaluated the performance of these five models on the ESD dataset, presenting detailed analysis results in Table 3. As shown in Table 3, the YOLOv5x model exhibits the highest detection performance but also has significantly more network layers and parameters and higher complexity compared to other models, posing challenges for deployment in industrial settings and meeting real-time detection needs. In contrast, YOLOv5m maintains a high mAP value of 88.8% with moderate model complexity, striking a balance between lightweight design and high accuracy. Based on this comparison, the YOLOv5m model was selected as the pretraining model in this study due to its comprehensive consideration of detection performance and real-time requirements.

**Table 3.** Comparison of performance indexes of different scale models.

Model	mAP (%)	F1 (%)	Number of Network Layers	Parameters	GFLOPs
YOLOv5n	66.8	67.4	270	1,776,094	4.3
YOLOv5s	80.5	79.7	270	7,043,902	16.0
YOLOv5m	88.8	88.0	290	20,885,262	48.0
YOLOv5l	91.3	89.6	367	46,151,358	107.8
YOLOv5x	93.6	92.5	444	86,227,246	203.9

### 3.2.3. Comparison of Different Detection Layers

To tackle the challenge of inadequate feature capture of small discarded steel objects by the YOLOv5 model, we enhanced the model's attention toward these targets through the incorporation of an additional P2 layer. Experimental validation illustrated that the network architecture, enhanced with the P2 layer, excels in detecting multi-scale and small steel targets. Nevertheless, the addition of the fourth detection layer resulted in a notable increase in parameter count and complexity, thereby prolonging inference times, which is not conducive to swift deployment in industrial settings. To strike a balance between detection performance and model complexity, each detection layer underwent individual evaluation. Evaluation metrics for the various detection layers are detailed in Table 4.

Due to the scarcity of large targets in waste steel images captured by high-altitude cameras, the performance of the P5 detection layer was subpar, yielding an mAP of only 44.9%. In contrast, the smaller-scale detection layers P2, P3, and P4 yielded mAPs of 83.5%, 95.0%, and 76.3%, respectively. As a result, the P5 layer was eliminated from the model, and a lightweight network structure IL was devised using only P2 to P4 layers. Table 5 presents a comparative analysis of the performance of the two detection structures before and after enhancement. From Table 5, it is evident that while the mAP of the IL structure marginally decreased (by merely 0.8%), there were substantial reductions of 25.1%, 65.1%, and 16.9% in network depth, parameter count, and complexity, respectively, underscoring the feasibility and efficacy of the model's lightweight properties.

**Table 4.** Comparison of performance metrics of different detection layers.

Layer	mAP (%)	F1 (%)
P2	83.5	83.1
P3	95.0	92.9
P4	76.5	78.0
P5	44.9	45.2

**Table 5.** Improvement in each model.

Layers	mAP (%)	F1 (%)	Number of Network Layers	Parameters	GFLOPs
P2–P5	96.1	94.4	350	22,831,608	82.4
P2–P4	95.3	93.3	262	7,970,190	68.5

### 3.3. Ablation Experiments

Ablation experiments were conducted on the ESD dataset to assess the impact of each enhancement module on model performance. The benchmark model YOLOv5, six enhancement models, and the CNIL-Net model were utilized for this purpose. Table 6 presents the specifications of each enhanced model, while Table 7 displays the training outcomes and evaluation metrics for each model. From Table 7, it is evident that the CNIL-Net model proposed in this study achieved a substantial enhancement in mAP, achieving 96.5%, a 7.7% increase over the benchmark model. Concurrently, the number of network layers and the number of parameters in this model were reduced by 6.2% and 61.8%, respectively.

**Table 6.** Improvement details for each model.

Model	IL	CA	Soft-NMS
Improved model 1	✓		
Improved model 2		✓	
Improved model 3			✓
Improved model 4	✓	✓	
Improved model 5	✓		✓
Improved model 6		✓	✓
CNIL-Net	✓	✓	✓

**Table 7.** Evaluation indicators for different models.

Model	mAP (%)	F1 (%)	Number of Network Layers	Parameters	GFLOPs
YOLOv5	88.8	88.0	290	20,885,262	48.0
Improved model 1	95.3	91.6	262	7,970,190	68.5
Improved model 2	88.0	81.1	300	20,905,254	48.0
Improved model 3	91.3	84.2	290	20,885,262	48.0
Improved model 4	95.9	92.4	272	7,980,198	68.6
Improved model 5	96.0	90.5	262	7,970,190	68.5
Improved model 6	90.0	81.1	300	20,905,254	48.0
CNIL-Net	96.5	91.4	272	7,980,198	68.6

A comparison between the enhanced model 1 and the benchmark model reveals that restructuring to P2–P4 markedly enhances model accuracy, resulting in a 6.5% improvement in mAP and a 3.6% increase in F1 score over the YOLOv5 baseline. This enhancement stems from the capability of the P2–P4 detection layers to acquire feature maps at scales of  $40 \times 40$ ,  $80 \times 80$ , and  $160 \times 160$ . These shallow feature maps bolster the model's proficiency in detecting low-resolution targets from high-altitude captures and efficiently capturing feature details of small-scrap targets in the ESD dataset, thereby mitigating positional information loss and omission issues. It is important to note that acquiring a large-scale feature map entails increased sampling operations in enhanced model 1, thereby augmenting network complexity. Furthermore, the removal of the detection layer P5 also entails eliminating layers 7 and 8 from the Backbone section, thereby reducing the

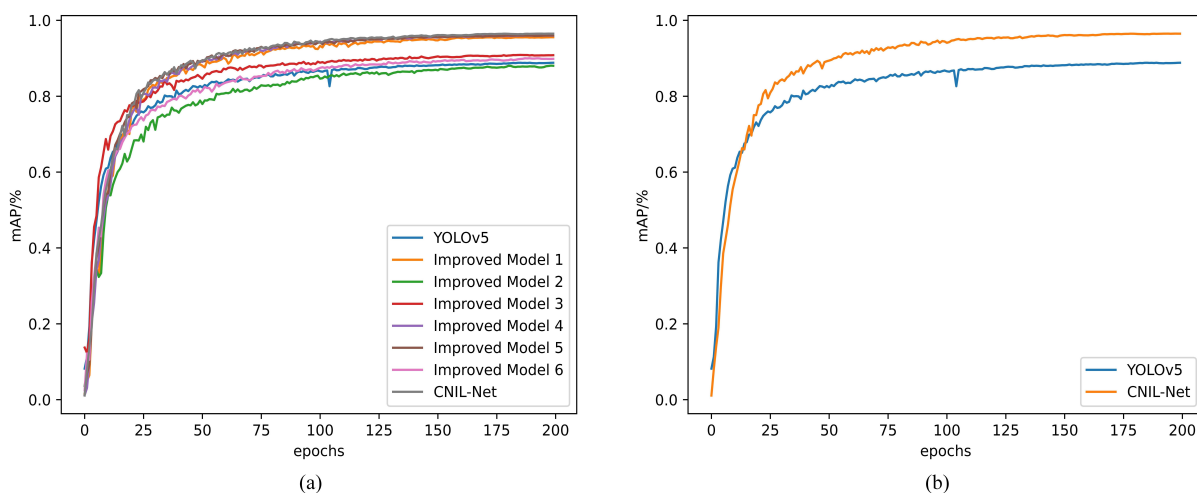
associated convolutional and pooling layers and significantly decreasing network depth and parameter count.

The introduction of CA into the initial network, similar to Improved Model 2, resulted in a decline in model performance. This can be attributed to placing CA in the eighth layer of the Backbone, where the input feature map size is  $20 \times 20$ . Given the dataset's abundance of small-sized targets, the attention mechanism predominantly learns negative information. Conversely, Improved Model 4 incorporates CA into the higher layer of the spatial pyramid pooling, aligning it with larger-scale output feature maps. This adaptation enhances the model's ability to discern channel and positional information, leading to slight improvements in both mAP and F1 values.

Both Improved Model 3 and Improved Model 5 adopt the Soft-NMS instead of the traditional NMS. All scraps within the compartment were densely packed and overlapped with each other; the Soft-NMS employs Gaussian weighting on overlapping detection frames to adjust confidence levels, mitigating the inadvertent deletion of overlapping detections caused by NMS operations, and lowering model false-negative rates. A comparison with the benchmark model and Improved Model 1 shows that integrating the Soft-NMS algorithm significantly enhances model training accuracy. Specifically, the mAP of Improved Model 3 and Improved Model 5 increased by 2.5% and 0.7%, respectively.

Leveraging the enhancements of Improved Model 1, the CNIL-Net model integrates CA and incorporates the Soft-NMS algorithm. In comparison to Improved Model 1, the CNIL-Net model exhibited slight increments in network layers, parameters, and complexity but achieved a 1.2% improvement in mAP. When contrasted with the benchmark model YOLOv5, the CNIL-Net model enhanced its mAP and F1 scores by 7.7% and 3.4%, respectively, alongside reductions in network layers and parameters. The model excelled in accurately capturing small targets, identifying densely overlapped targets correctly, and providing superior accuracy in identifying the different types of scrap materials. Furthermore, the CNIL-Net model boasts a complexity of only 68.6 GFLOPs, making it suitable for swift deployment in industrial settings and ensuring reliability in practical applications.

Figure 6 illustrates the training outcomes of each model and the CNIL-Net model on the ESD dataset before and after enhancements. Figure 6a depicts the mAP curves during the training of each model, while Figure 6b provides a comparative analysis between the CNIL-Net model and the YOLOv5 model. Here, the curves plot the number of training epochs on the horizontal axis and mAP values on the vertical axis. The results from Figure 6 clearly indicate that upgrading to the CNIL-Net model leads to a substantial enhancement in model detection accuracy.



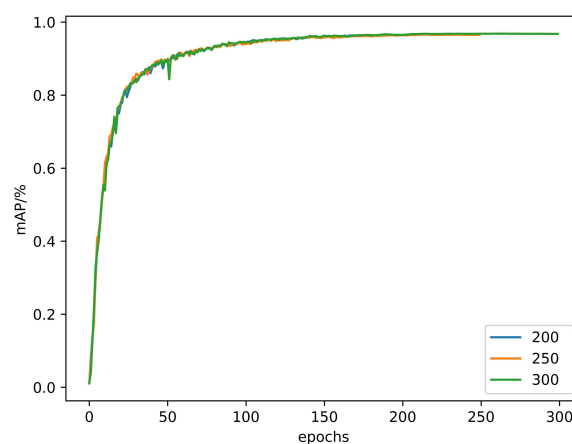
**Figure 6.** Image enhancement effects: (a) mAP curves for each model; (b) the mAP curves comparison between the CNIL-Net model and the baseline model training.

### 3.4. Analysis of Experimental Results

During the training of the CNIL-Net network model, we experimented with increasing the number of training epochs to 250 and 300, with results depicted in Figure 7. The analysis of Figure 7 reveals that the mAP value stabilizes around 96.5% after 200 training epochs, indicating model convergence. The confusion matrix, a graphical representation depicting algorithm performance, presents matrices [41] for both the benchmark model and the CNIL-Net model trained over 200 epochs, which are displayed in Figures 8 and 9, respectively. Here, the matrix portrays the true categories (True) along the horizontal axis and the predicted categories (Predicted) along the vertical axis. The main diagonal elements denote correctly identified samples (TPs) per category, while the upper and lower triangular regions signify missed and false detections, respectively. A comparison of confusion matrices between the baseline model and CNIL-Net model reveals slight reductions in correctly detected samples for overlength categories (1.2–1.5 m and 1.5–2 m) after model improvement, while accuracy improves notably across other categories. An examination of the CNIL-Net model's confusion matrix reveals relatively high misdetection rates in two scrap categories, 3–6 mm and >6 mm, primarily stemming from numerous scrap items outside compartment boundaries in the images. These categories constitute a significant portion of the overall samples, thereby contributing disproportionately to the overall misdetection rate. Overall, the model demonstrates strong performance across all categories, achieving over 86% accuracy, with six categories exceeding a 90% correct detection rate.

The ESD validation dataset consisted of 374 images and 29,798 labels. We validated the CNIL-Net model on the validation set, presenting the performance metrics for each category in Table 8. The validation results demonstrate that the detection accuracy and F1 score for each category exceed 90%. Notably, the airtight category achieves the highest AP and F1 scores at 98.8% and 97.7%, respectively. Additionally, the AP for the two most frequently labeled thickness categories—3–6 mm and >6 mm—also reaches 96.0% and 93.3%, respectively.

Figure 10 depicts the images captured at the scrap acceptance site, while Figure 11 illustrates the application of the CNIL-Net scrap-quality inspection model post-inference, detailing each scrap item with its category label and confidence level. A comparison of images before and after inspection reveals the accurate recognition of the location and category of scraps by the model. During scrap-quality inspection, the model processed a single picture in just 5.76 ms, with a 15–20 s interval between each claw machine grab. Thus, the CNIL-Net model meets the practical requirements for accuracy and processing speed in scrap inspection.



**Figure 7.** Effect of model convergence after increasing the number of rounds in the CNIL-Net network.

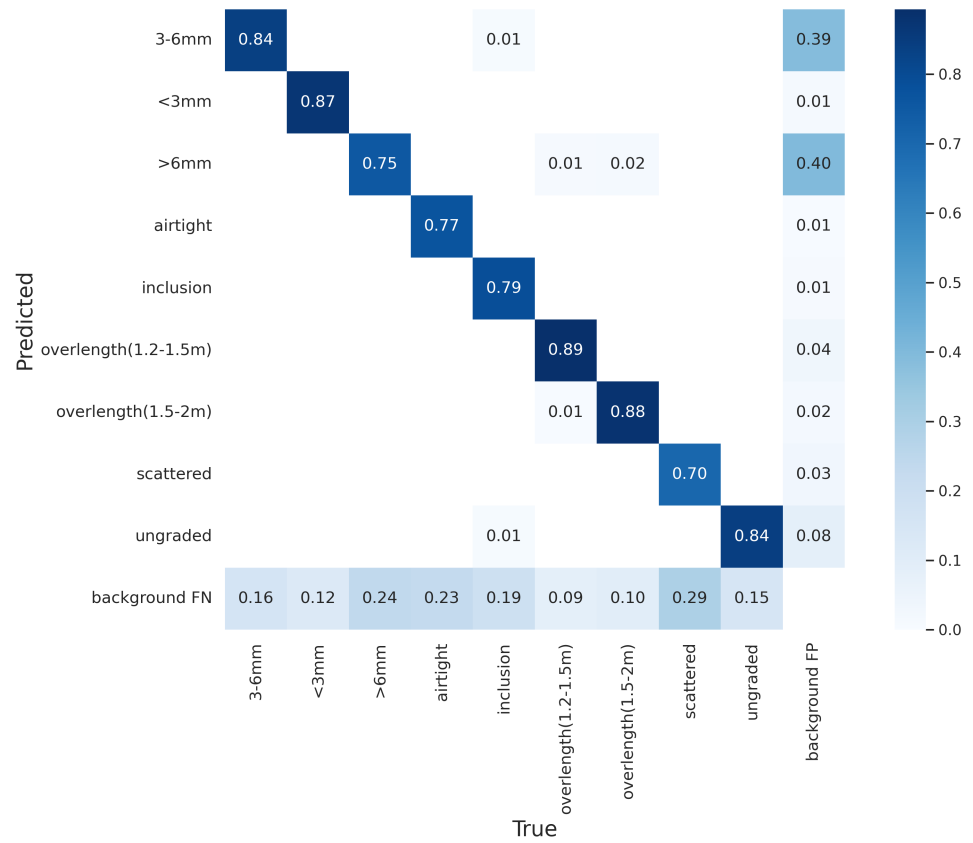


Figure 8. YOLOv5 model training confusion matrix.

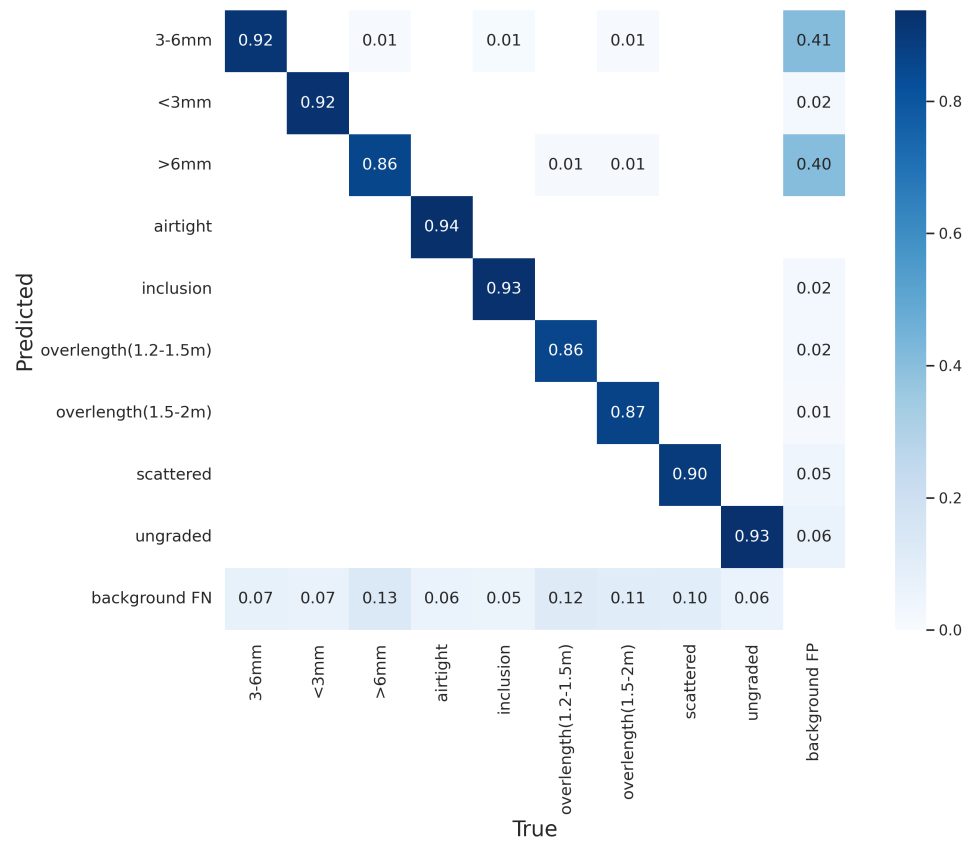


Figure 9. CNIL-Net model training confusion matrix.

**Table 8.** Evaluation indexes of the validation set for each category.

Class	Labels	AP (%)	F1 (%)
<3 mm	1082	97.1	94.3
3–6 mm	13,046	96.0	93.9
>6 mm	10,069	93.3	90.6
airtight	71	98.8	97.7
inclusion	227	97.5	93.6
overlength (1.2–1.5 m)	1012	97.1	92.4
overlength (1.5–2 m)	653	96.1	92.5
scattered	950	95.6	92.9
ungraded	2688	97.1	95.0

**Figure 10.** YOLOv5 model training confusion matrix.**Figure 11.** YOLOv5 model training confusion matrix.

#### 4. Conclusions

To address challenges in scrap-quality inspection, the CNIL-Net model was developed on the YOLOv5 architecture, enhancing the detection layers and incorporating the CA



mechanism and the Soft-NMS algorithm, followed by comparative analysis. The results show the following findings:

1. Introducing the P2 detection layer while removing P5 from the original four-layer structure reduced network layers, parameters, and complexity by 25.1%, 65.1%, and 16.9%, respectively. Although this adjustment led to a slight 0.9% decrease in mAP, it enhanced the model's ability to detect small targets and significantly reduced its computational burden.
2. Incorporating the CA mechanism and the Soft-NMS algorithm successfully addressed challenges in feature channel extraction and precise location identification without adding complexity. Moreover, it mitigated issues related to non-maximum suppression, resulting in a 1.2% increase in mAP alongside improvements in detection layer performance.
3. The CNIL-Net model achieved an average accuracy of 96.5% across all categories in scrap-quality inspection. Compared to the YOLOv5 benchmark, it demonstrated a 7.7% improvement in mAP and achieved an impressive single-image inference speed of 5.76 ms, fully meeting the requirements for industrial scrap-quality inspection.

Compared to traditional manual inspection methods, the CNIL-Net scrap-quality inspection model offers significant advantages in accuracy, safety, and fairness. Its high precision and lightweight design facilitate swift deployment in industrial settings, paving the way for intelligent, unmanned scrap acceptance systems. Moving forward, the team plans to generate additional high-quality, multi-category scrap datasets and iteratively enhance the intelligent quality inspection system to facilitate widespread industrial adoption.

**Author Contributions:** Conceptualization, P.X. and L.Z.; methodology, P.X., C.W. and R.Z.; software, P.X., C.W. and W.X.; validation, P.X., C.W. and R.Z.; formal analysis, P.X., C.W. and L.Z.; investigation, P.X., C.W., L.Z. and Y.J.; resources, L.Z., Y.J. and R.Z.; data curation, P.X., C.W., W.X. and Y.J.; writing—original draft preparation, P.X. and C.W.; writing—review and editing, L.Z. and R.Z.; visualization, P.X. and C.W.; supervision, L.Z. and R.Z.; project administration, P.X. and L.Z.; funding acquisition, P.X. and L.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Key Fund Projects of China (No. U21A20114) and the Hebei Provincial Science and Technology Programme of China (No. 23561007D).

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy restrictions.

**Acknowledgments:** The author acknowledges the HBIS Company Limited and Steel Laboratory of Hebei Province in this study. Their generous sponsorship and technical support provided crucial resources for experiments, significantly advancing the progress and outcomes of the research.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhou, S.; Tong, Q.; Pan, X.Z.; Cao, M.; Wang, H.; Gao, J. Research on Low-Carbon Energy Transformation of China Necessary to Achieve the Paris Agreement Goals: A Global Perspective. *Energy Econ.* **2021**, *95*, 105137. [[CrossRef](#)]
2. Branca, T.A.; Fornai, B.; Colla, V.; Murri, M.M.; Streppa, E.; Schröder, A.J. The challenge of digitalization in the steel sector. *Metals* **2020**, *10*, 288. [[CrossRef](#)]
3. Fan, Z.; Friedmann, S.J. Low-carbon production of iron and steel: Technology options, economic assessment, and policy. *Joule* **2021**, *5*, 829–862. [[CrossRef](#)]
4. Zhang, X.; Jiao, K.; Zhang, J.; Guo, Z. A review on low carbon emissions projects of steel industry in the World. *J. Clean. Prod.* **2021**, *306*, 127259. [[CrossRef](#)]
5. Voraberger, B.; Wimmer, G.; Dieguez Salgado, U.; Wimmer, E.; Pastucha, K.; Fleischanderl, A. Green LD (BOF) steelmaking—reduced CO<sub>2</sub> emissions via increased scrap rate. *Metals* **2022**, *12*, 466. [[CrossRef](#)]
6. Celada-Casero, C.; López, F.A.; Capdevila, C.; Castelo, R.; Oliver, S. Scrap for New Steel. *New Mater. Circ. Econ.* **2023**, *149*, 202–232. [[CrossRef](#)]
7. Van den Eynde, S.; Diaz-Romero, D.J.; Engelen, B.; Zaplana, I.; Peeters, J.R. Assessing the efficiency of Laser-Induced Breakdown Spectroscopy (LIBS) based sorting of post-consumer aluminium scrap. *Procedia Cirp* **2022**, *105*, 278–283. [[CrossRef](#)]

8. Ruan, J.; Qian, Y.; Xu, Z. Environment-friendly technology for recovering nonferrous metals from e-waste: Eddy current separation. *Resour. Conserv. Recycl.* **2014**, *87*, 109–116. [[CrossRef](#)]
9. Xu, W.; Xiao, P.; Zhu, L.; Zhang, Y.; Chang, J.; Zhu, R.; Xu, Y. Classification and rating of steel scrap using deep learning. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106241. [[CrossRef](#)]
10. Tu, Q.; Li, D.; Xie, Q.; Dai, L.; Wang, J. Automated scrap steel grading via a hierarchical learning-based framework. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–13. [[CrossRef](#)]
11. Daigo, I.; Murakami, K.; Tajima, K.; Kawakami, R. Thickness classifier on steel in heavy melting scrap by deep-learning-based image analysis. *ISIJ Int.* **2023**, *63*, 197–203. [[CrossRef](#)]
12. Xiao, P.; Xu, W.; Zhang, Y.; Zhu, L.; Zhu, R.; Xu, Y. Research on scrap classification and rating method based on SE attention mechanism. *Chin. J. Eng.* **2023**, *45*, 1342–1352. [[CrossRef](#)]
13. Mei, Y. Research on Intelligent Classification of Scrap Based on Machine Vision and LIBS Technology. Ph.D. Thesis, University of Science and Technology Beijing, Beijing, China, 21 December 2020. [[CrossRef](#)]
14. Qiu, Z. Smart Rating System of Steel Scrap Based on Improved YOLOv3 Algorithm. Ph.D. Thesis, Anhui University of Technology, Ma'anshan, China, 10 November 2020. [[CrossRef](#)]
15. Gao, Z.; Sridhar, S.; Spiller, D.E.; Taylor, P.R. Applying improved optical recognition with machine learning on sorting Cu impurities in steel scrap. *J. Sustain. Metall.* **2020**, *6*, 785–795. [[CrossRef](#)]
16. Penumuru, D.P.; Muthuswamy, S.; Karumbu, P. Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. *J. Intell. Manuf.* **2020**, *31*, 1229–1241. [[CrossRef](#)]
17. Diaz-Romero, D.; Sterkens, W.; Van den Eynde, S.; Goedeme, T.; Dewulf, W.; Peeters, J. Deep learning computer vision for the separation of Cast-and Wrought-Aluminum scrap. *Resour. Conserv. Recycl.* **2021**, *172*, 105685. [[CrossRef](#)]
18. Koyanaka, S.; Kobayashi, K. Incorporation of neural network analysis into a technique for automatically sorting lightweight metal scrap generated by ELV shredder facilities. *Resour. Conserv. Recycl.* **2011**, *55*, 515–523. [[CrossRef](#)]
19. Diaz-Romero, D.J.; Van den Eynde, S.; Sterkens, W.; Engelen, B.; Zaplana, I.; Dewulf, W.; Goedeme, T.; Peeters, J. Simultaneous mass estimation and class classification of scrap metals using deep learning. *Resour. Conserv. Recycl.* **2022**, *181*, 106272. [[CrossRef](#)]
20. Chen, S.; Hu, Z.; Wang, C.; Pang, Q.; Hua, L. Research on the process of small sample non-ferrous metal recognition and separation based on deep learning. *Waste Manag.* **2021**, *126*, 266–273. [[CrossRef](#)] [[PubMed](#)]
21. Xu, Q.; Lin, R.; Yue, H.; Huang, H.; Yang, Y.; Yao, Z. Research on Small Target Detection in Driving Scenarios Based on Improved Yolo Network. *IEEE Access* **2020**, *8*, 27574–27583. [[CrossRef](#)]
22. Ming, Q.; Miao, L.; Zhou, Z.; Song, J.; Yang, X. Sparse Label Assignment for Oriented Object Detection in Aerial Images. *Remote. Sens.* **2021**, *13*, 2664. [[CrossRef](#)]
23. Wang, X.; Liu, Y.; Xu, H.; Xue, C. A Spatial Small Target Detection Method Based on a Multi-Scale Feature Fusion Pyramid. *Appl. Sci.* **2024**, *14*, 5673. [[CrossRef](#)]
24. Fang, Y.; Pang, H. An Improved Pedestrian Detection Model Based on YOLOv8 for Dense Scenes. *Symmetry* **2024**, *16*, 716. [[CrossRef](#)]
25. Zhao, M.; Zhou, H.; Li, X. YOLOv7-SN: Underwater Target Detection Algorithm Based on Improved YOLOv7. *Symmetry* **2024**, *16*, 514. [[CrossRef](#)]
26. Liu, Q.; Ye, H.; Wang, S.; Xu, Z. YOLOv8-CB: Dense Pedestrian Detection Algorithm Based on In-Vehicle Camera. *Electronics* **2024**, *13*, 236. [[CrossRef](#)]
27. Xu, D.; Wang, L.; Li, F. Review of Typical Object Detection Algorithms for Deep Learning. *Comput. Eng. Appl.* **2021**, *57*, 10–25. [[CrossRef](#)]
28. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [[CrossRef](#)]
29. Sirisha, U.; Praveen, S.P.; Srinivasu, P.N.; Barsocchi, P.; Bhoi, A.K. Statistical analysis of design aspects of various YOLO-based deep learning models for object detection. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 126. [[CrossRef](#)]
30. Dhillon, A.; Verma, G.K. Convolutional neural network: A review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* **2020**, *9*, 85–112. [[CrossRef](#)]
31. Kaur, R.; Singh, S. A comprehensive review of object detection with deep learning. *Digit. Signal Process.* **2023**, *132*, 103812. [[CrossRef](#)]
32. Xiao, P.; Xu, W.; Zhang, Y.; Zhu, L.; Zhu, R.; Xu, Y. Classification and Rating of Scrap Steel Based on Deep Learning. *Adv. Eng. Sci.* **2023**, *55*, 184–193. [[CrossRef](#)]
33. Tong, K.; Wu, Y. Deep learning-based detection from the perspective of small or tiny objects: A survey. *Image Vis. Comput.* **2022**, *123*, 104471. [[CrossRef](#)]
34. Oksuz, K.; Cam, B.C.; Kalkan, S.; Akbas, E. Imbalance problems in object detection: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3388–3415. [[CrossRef](#)]
35. Yao, P.; Shen, S.; Xu, M.; Liu, P.; Zhang, F.; Xing, J.; Shao, P.; Kaffenberger, B.; Xu, R.X. Single model deep learning on imbalanced small datasets for skin lesion classification. *IEEE Trans. Med. Imaging* **2021**, *41*, 1242–1254. [[CrossRef](#)]
36. Liu, Z.; Gao, X.; Wan, Y.; Wang, J.; Lyu, H. An improved YOLOv5 method for small object detection in UAV capture scenes. *IEEE Access* **2023**, *11*, 14365–14374. [[CrossRef](#)]

37. Mi, Z.; Gao, Y.; Xu, X.; Tang, J. Steel strip surface defect detection based on multiscale feature sensing and adaptive feature fusion. *AIP Adv.* **2024**, *14*, 045005. [[CrossRef](#)]
38. Zhang, C.; Zhu, L.; Yu, L. Review of Attention Mechanism in Convolutional Neural Networks. *Comput. Eng. Appl.* **2021**, *57*, 64–72. [[CrossRef](#)]
39. Liu, J.; Liu, J.; Luo, X. Research progress in attention mechanism in deep learning. *Chin. J. Eng.* **2021**, *43*, 1499–1511. [[CrossRef](#)]
40. Chen, F.; Zhang, L.; Kang, S.; Chen, L.; Dong, H.; Li, D.; Wu, X. Soft-NMS-enabled YOLOv5 with SIOU for small water surface floater detection in UAV-captured images. *Sustainability* **2023**, *15*, 10751. [[CrossRef](#)]
41. Heydarian, M.; Doyle, T.E.; Samavi, R. MLCM: Multi-label confusion matrix. *IEEE Access* **2022**, *10*, 19083–19095. [[CrossRef](#)]

**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.