www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

SMALL CONSTRUCTION MATERIALS DETECTION: AN APPROACH OF ENHANCED FEATURE EXTRACTION AND REPRESENTATION

SUBMITTED: August 2024
REVISED: December 2024
PUBLISHED: February 2025
EDITOR: Žiga Turk
DOI: 10.36680/j.itcon.2025.006

Yujie Lu, Professor (corresponding author)
College of Civil Engineering, Tongji University, Shanghai 200092, China
Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai 200092, China
Shanghai Research Institute of Intelligent Science and Technology, Tongji University, Shanghai, 200092, China
lu6@tongji.edu.cn

Yuanjun Nong, Ph.D. Candidate
College of Civil Engineering, Tongji University, Shanghai 200092, China
nyj@tongji.edu.cn

SUMMARY: *Automated construction materials detection is crucial for material lean management, such as material planning, inventory, site usage, and monitoring. However, there are numerous small materials in the construction site due to the long monitoring distances, which easily cause missed and incorrect detection owing to their indistinguishable features and complicated backgrounds. To improve detection accuracy for small materials, this study proposes an augmented detection method based on enhanced feature extraction and representation. In the proposed method, DenseNet is utilized as the backbone to enhance the feature extraction of small materials. Additionally, the explicit visual center is introduced to enhance the feature learning of small materials. Finally, the multi-scale detection structure is optimized by adding a scale to improve feature representation. Experimental results demonstrate that the average precision for small objects (APs) have improved by 5.3%, and the mean average precision (mAP) has reached 84.3%, surpassing other state-of-the-art methods. The proposed method also exhibits strong adaptability to various conditions such as shadows, blurriness, and cluttered backgrounds. Additionally, the impacts of different backbone networks and detection scales on accuracy are discussed. This research provides theoretical and practical references for material lean management and facilitates the application of digital twin in materials management.*

KEYWORDS: *Materials detection, Small object detection, Deep learning, DenseNet, Multi-scale feature representation.*

REFERENCE: *Yujie Lu & Yuanjun Nong (2025). Small construction materials detection: an approach of enhanced feature extraction and representation. Journal of Information Technology in Construction (ITcon), Vol. 30, pg. 113-132, DOI: 10.36680/j.itcon.2025.006*

1. INTRODUCTION

Material management is an essential part of construction project management. In a typical industrial facility, the expense of equipment and materials constitutes 50%-60% of the total cost (Kini, 1999). Effective material management directly impacts the project's success. Inefficient utilization of materials can result in reduced productivity and increased costs in construction projects (National Research Council, 2009). Therefore, it is vital to accurately track and manage the turnover and utilization efficiency of materials, which requires accurate identification and detection of materials. Furthermore, precise material detection also facilitates automated construction progress monitoring and the generation of 3D as-built models.

Traditional detection methods for construction materials have primarily relied on manual inspection, which is time-consuming, subjective, and error-prone. In recent years, benefiting from the development of computer technology, deep learning has been applied to automatic detection in the industry, such as construction safety inspection (Khan et al. 2021), tower crane productivity monitoring and analysis (Elgendi et al. 2023), construction activity recognition (Bhokare et al. 2022), resident quantity and distance detection (Huang et al. 2020), crack detection (He et al. 2023). Deep learning also offers a promising solution for the automatic detection of construction materials.

However, there are still challenges related to construction materials detection, particularly with small objects. On construction sites, the surveillance cameras are often mounted high up to obtain a wide viewing angle. This results in small materials being captured by surveillance cameras, as shown in Figure 1. Small objects typically occupy a tiny portion of the image and easily lead to missed or incorrect detection. This will result in inaccurate monitoring, tracking, and statistics of materials at construction site entrances and exits, on-site, and in warehouses, posing challenges to material lean management. Inadequate material management can ultimately lead to reduced productivity, cost overruns, and construction delays. The difficulty in detecting small objects can be attributed to two main reasons. First, small objects have limited pixel and visual information, making feature extraction a challenging task. And, after multiple convolution and down-sampling operations in CNN, the feature representation of small objects can become weakened or lost, further hindering effective feature learning. Second, the complex construction environment often causes small materials to be indistinguishable from the background, easily causing incorrect detection.

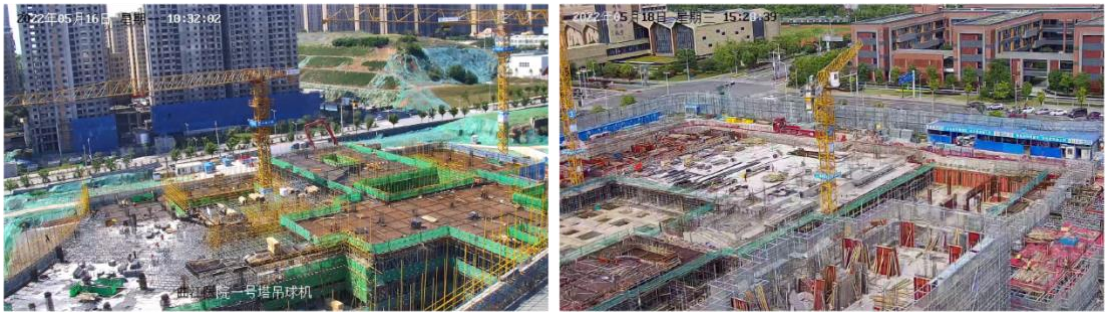


Figure 1: Small construction materials in construction sites. Due to the surveillance camera being positioned far from the construction work surface, the captured construction materials appear small, posing challenges for material detection.

To address these issues, this study proposes an augmented detection method to improve the accuracy of small material detection through enhanced feature extraction and representation. The main contributions of this study are as follows. First, an enhanced feature extraction network based on DenseNet is established, improving feature extraction for small materials and promoting feature fusion and reuse. Second, an optimized feature pyramid network is designed by introducing the explicit visual center module, enhancing feature learning for small objects. Finally, an improved multi-scale prediction network is proposed by adding one prediction scale, enhancing the utilization and representation of small material features, and strengthening the aggregation of multi-level features.

In summary, construction materials detection is crucial for material lean management. However, challenges remain in detecting small construction materials. With the significant advancements in deep learning, this study proposes a deep learning-based method to enhance the detection accuracy for small construction materials.

2. RELATED WORKS

2.1 Construction Materials detection

The development of materials detection methods has undergone a transition from manual inspection to automatic detection. Manual inspection relies on visual inspection by personnel and manual recording, which is time-consuming, labor-intensive, and error-prone. With the advancement of technology, automatic detection methods have become mainstream, which include Radio Frequency Identification (RFID) system-based, machine learning-based, and deep learning-based.

RFID-based methods use radio frequency signals for non-contact two-way communication to achieve automatic identification and tracking of materials (Song et al. 2006; Ren et al. 2011). While RFID technology improves the efficiency of material detection and tracking and reduces the identification error rate, it has limitations of high cost and insufficient security.

Machine learning uses visual features of materials for non-contact, non-destructive automatic identification and classification. The machine learning method involves manually extracting the visual features of materials, such as color, texture, and roughness, and using different classifiers like support vector machines, artificial neural networks, and logistic regression to automatically classify the materials (Dimitrov et al. 2014; Son et al. 2014; Zhu et al. 2010; Son et al. 2012; Yuan et al. 2020). However, manual feature design and extraction heavily rely on prior knowledge and are affected by lighting conditions, resulting in poor robustness.

In contrast, deep learning methods utilize convolutional neural networks (CNN) to automatically extract visual features of materials, enabling automatic and accurate material detection. Deep learning methods offer advantages such as end-to-end learning, strong robustness and generalization, and strong feature expression ability. For instance, Duan et al. (2022) utilized YOLOv3 (Redmon et al. 2018) and YOLOv4 (Bochkovskiy et al. 2020) methods to achieve automatic and accurate detection of workers, materials, machines, and layouts on construction sites. Despite the fact that deep learning methods have achieved impressive detection accuracy on large and medium size objects, they still face challenges when detecting small objects due to the limited resolution, indistinguishable features, and complicated backgrounds of small objects.

2.2 Small Objects Detection

Small object detection has always been a challenging task, even with the advancements in deep learning-based object detection. The detection accuracy of small objects remains unsatisfactory, as shown in Table 1, which illustrates the detection accuracy of deep learning-based object detection methods on the public COCO dataset. Notably, the detection accuracy of small objects (APs) is considerably lower than that of medium and large objects (APm and AP_l).

To solve this problem, several methods have been developed to improve the detection accuracy of small objects, focusing on data improvement and feature fusion. Data improvement includes data enhancement and image super-resolution, and feature fusion is multi-scale feature fusion.

(1) Data Improvement

Data augmentation strategies are used to expand the size of the dataset and increase the number of small objects, thereby improving small object detection performance. These strategies typically include geometric transformations, color transformations, and random occlusion (Bochkovskiy et al. 2020; Kisantal et al. 2019; Gao et al. 2020; Mahaur et al. 2023). For example, Bochkovskiy et al. (2020) utilize the mosaic enhancement technique, which randomly stitches four input images into one during training to improve detection performance. However, it's important to note that data augmentation strategies can increase computational costs, and designing reasonable augmentation strategies can be challenging. Improperly designed strategies would introduce noise or irrelevant information, which can negatively affect the detection of small objects.

Super-resolution methods are employed to reconstruct low-resolution images into high-resolution ones, providing more detailed information about objects and aiding in small object detection (Goodfellow et al. 2014; Li et al. 2017; Zhang et al. 2020; Bai et al. 2018; Noh et al. 2019). These methods often utilize generative adversarial networks (GAN) (Goodfellow et al. 2014) to reconstruct high-resolution images. For instance, perceptual GANs (Li et al. 2017) have been used to generate super-resolved representations for small objects. Similarly, multi-task

GANs (Zhang et al. 2020b) have been proposed to generate sharp images from blurred small ones through up-sampling operations. However, it is worth noting that GAN-based methods can be challenging to train, and maintaining a balance between generators and discriminators during training can be difficult.

(2) Feature Fusion

CNNs produce different levels of feature maps due to down-sampling operations. High-level features contain semantic information, while low-level features provide detailed spatial information. Combining these feature maps allows for the acquisition of both spatial and semantic information (Lin et al. 2017; Liu et al. 2018; Tan et al. 2019; Ghiasi et al. 2019), thereby enhancing the detection accuracy of small objects. For instance, Feature Pyramid Networks (FPN) (Lin et al. 2017) use a top-down pathway and lateral connections to merge low-level and high-level features. PANet (Liu et al. 2018), built on FPN, introduces a bottom-up path to propagate localization information in low-level features. BiFPN (Tan et al. 2019) incorporates learnable weights to determine the importance of different features and repeatedly fuses multi-scale features. While multi-scale feature fusion can enhance the detection accuracy of small objects, it can also increase model complexity and the risk of over-fitting during training. Additionally, it overlooks the semantic gaps between features at different scales.

In the field of construction, there is limited research on small object detection, with a focus on detecting small workers. For example, Park et al. (2023) proposed a detection method that utilized DIoU-NMS, Soft-SPPF, and weighted-triplet attention to optimize YOLOv5 (Jocher et al. 2022), resulting in improved detection accuracy for small workers. Kim et al. (2023) developed a YOLOv5-based small object detection system capable of detecting multi-scale objects from small workers to large construction equipment.

To conclude, existing small object detection in the construction field primarily concentrates on worker detection, and there is little research on small construction materials detection. Furthermore, the detection methods used for small workers would not be effective in detecting small construction materials due to the differing features between workers and construction materials, such as shape, spatial location distribution, and visual information. Given that small materials are challenging to extract features and are easily confused with the complex background, this study proposes an augmented detection method for small materials, which aims to enhance detection accuracy by improving feature extraction, learning, and representation for small materials.

Table 1: The detection accuracy of different deep learning-based methods on the COCO test-dev set. The APs, APm, and AP_l are the average precision of small, medium, and large objects.

Methods	APs	APm	AP_l
SSD (Liu et al. 2016)	0.109	0.318	0.435
Cascade RCNN (Ren et al. 2017)	0.237	0.455	0.552
TridentNet (Li et al. 2019)	0.239	0.466	0.566
FCOS (Tian et al. 2019)	0.260	0.468	0.550
ATSS (Zhang et al. 2020a)	0.261	0.470	0.536
PAA (Kim et al. 2020)	0.265	0.488	0.563
TOOD (Feng et al. 2021)	0.289	0.496	0.570

3. METHOD

3.1 Overall Network Architecture

This study proposes a small material detection method based on enhanced feature extraction and representation. To improve the detection accuracy of small materials, three optimization measures were adopted: 1) Optimize the backbone network by introducing DenseNet to enhance feature extraction of small materials. 2) Optimize the feature pyramid networks by introducing an explicit visual center module to improve the network's feature learning ability for small materials. 3) Optimize the multi-scale prediction structure by adding a detection scale to improve the network's representation ability of small materials. The overall network architecture of the proposed method is shown in Figure 2.

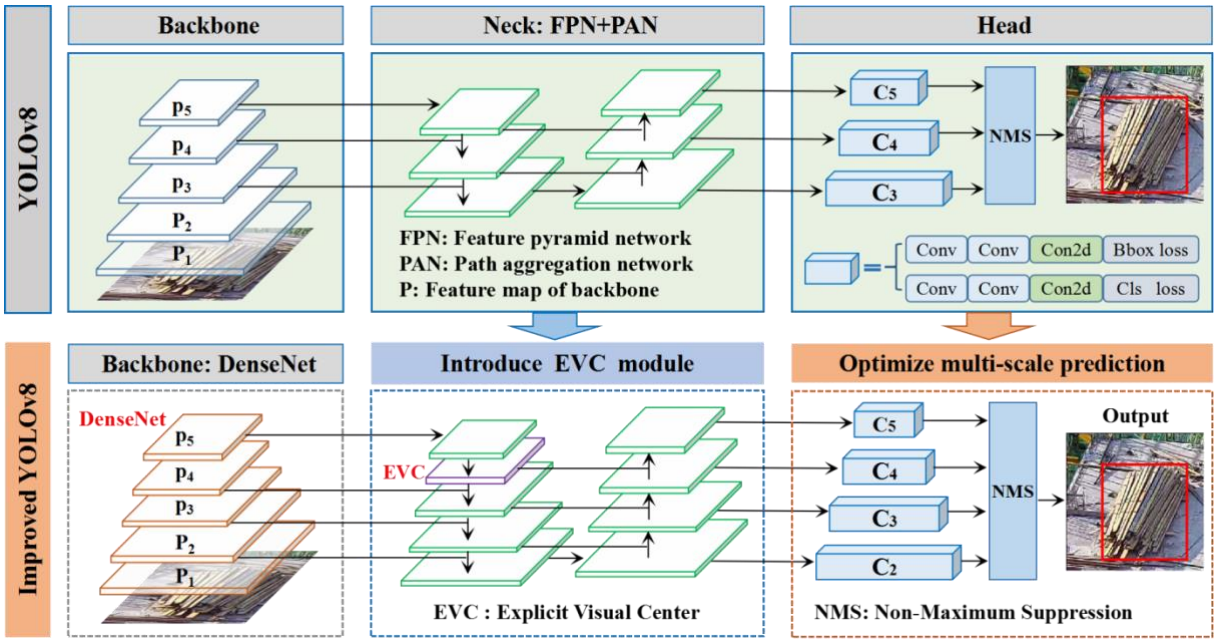


Figure 2: The overall network architecture of the proposed method.

The deep learning-based object detection methods can be divided into two main categories: one-stage methods and two-stage methods. Generally, the two-stage methods can achieve higher accuracy, while the one-stage methods have advantages in detection speed. Recently, efforts have improved one-stage methods, narrowing the accuracy gap between one-stage and two-stage methods. As one of the representative one-stage methods, YOLOv8 (Ultralytics, 2023) is widely used due to its excellent detection accuracy and speed, making it appropriate for material detection. Nonetheless, the detection accuracy of YOLOv8 is still unsatisfactory when dealing with small materials. To improve the detection accuracy of small materials, we optimize YOLOv8 in three aspects in this study, including introducing DenseNet, introducing the explicit visual center module, and Optimizing multi-scale prediction structures.

3.2 Enhance Feature Extraction By Introducing DenseNet

In the backbone network, YOLOv8 performs five down-sampling operations, which can reduce the size of the feature maps but would lead to the loss of detailed feature information, especially for small objects, and reduce the backward transmission of feature information. As the network deepens, the loss of feature information for small objects becomes more severe. Additionally, deep neural networks often face the issue of gradient vanishing, which hinders feature learning, especially for small objects.

To address these challenges and improve feature extraction for small materials, we introduce DenseNet (Huang et al. 2017) as the backbone network for feature extraction, as shown in Figure 3. DenseNet establishes dense connections between layers, where each layer receives feature maps from all preceding layers and maps its own feature maps to all subsequent layers. This strengthens the propagation of features and promotes feature fusion and reuse, allowing the network to obtain more feature information, especially from low-level feature maps.

Low-level feature maps exhibit strong representation for small objects and contain rich feature information of small objects due to their high resolution and small receptive field. By enhancing the reuse of low-level feature maps, DenseNet reduces the loss of feature information for small objects during transmission and enables the network to capture more features of small objects. Moreover, DenseNet improves the flow of gradients between layers, alleviating gradient vanishing and enhancing feature learning for small objects.

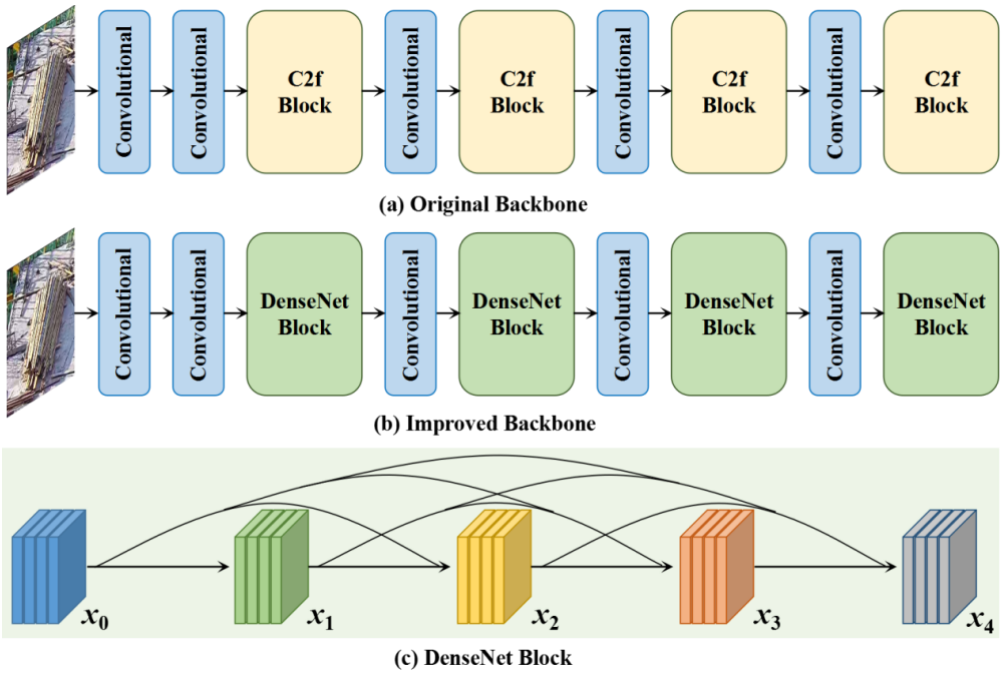


Figure 3: Backbone network and DenseNet. (a) original backbone, (b) improved backbone, (c) DenseNet block.

In traditional convolutional networks, the feature map x_{l-1} output from layer (l-1) is fed as input to layer l to obtain the output feature map x_l of layer l, which is transformed as follows:

$$x_l = \mathrm{H}_l(x_{l-1}) \quad (1)$$

Where $\mathrm{H}_l(\ )$ represents a nonlinear transformation function composed of batch normalization, corrected linear units, and 3×3 convolutional operations.

In DenseNet networks, the input of layer l comes from the features maps (x_0, x_1,...., x_{l-1}) output from all preceding layers, and its transformation is as follows:

$$x_l = \mathrm{H}_l([x_0, x_1, ..., x_{l-1}]) \quad (2)$$

where [x_0, x_1,...., x_{l-1}] represents the concatenation of the feature maps (x_0, x_1,...., x_{l-1}) output from layers 0, 1,...., (l-1).

3.3 Improve Feature Learning By Introducing the Explicit Visual Center Module

The feature pyramid network can effectively capture and fuse multi-scale features, thereby improving detection accuracy. However, existing feature pyramid networks often overly emphasize the interaction of features between different layers while neglecting the critical intra-layer features, especially those of small objects. This oversight hinders the detection of small objects. To address this issue, the explicit visual center (EVC) (Quan et al. 2023) module is introduced into YOLOv8, as shown in Figure 2. The EVC mainly consists of two parallel modules: a lightweight MLP and a learnable visual center (LVC) mechanism, as shown in Figure 4. The lightweight MLP captures the global long-term dependencies of features, while the LVC mechanism aggregates local regional features within the layer to obtain local small object information. The feature maps generated by these two modules are then concatenated to obtain both global information and refined feature representation of small objects.

(1) Lightweight MLP

In the lightweight MLP, features X are first fed into a depthwise convolution-based module to enhance feature representation capability. This module is primarily composed of group normalization (GN), depthwise convolution (DConv), channel scaling, and droppath. The processing of features in this module can be expressed as:

$$\overline{X} = \mathrm{DConv}(\mathrm{GN}(X)) + X \quad (3)$$

Subsequently, the output features $\overline{X}$ are fed into a channel MLP-based module to capture global feature information, as shown in the following equation:

$$X_1 = \mathrm{MLP}(\mathrm{GN}(\overline{X})) + \overline{X} \quad (4)$$

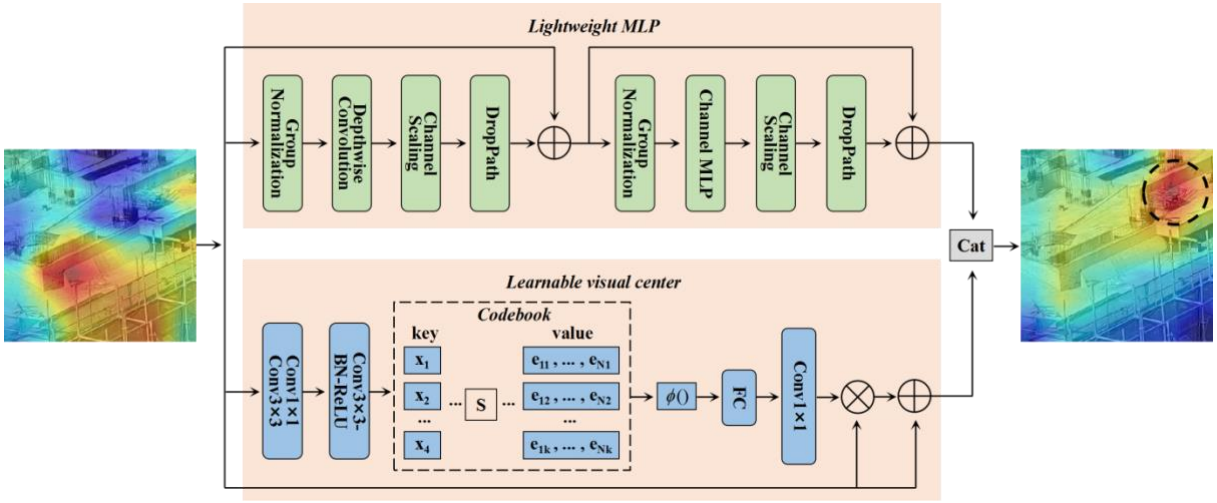


Figure 4: The structure of the explicit visual center (EVC) module.

(2) LVC

In the LVC module, features X are first encoded by a series of convolution operations. The encoded features x are then fed into a Codebook, which uses scaling factors s_k to map the position information between each feature point x_i and each learnable visual code-word b_k, as shown in the following equation:

$$\mathrm{e}_k = \sum_{i=1}^{N} \frac{e^{-s_k \|x_i - b_k\|^2}}{\sum_{j=1}^{K} e^{-s_k \|x_i - b_k\|^2}} (x_i - b_k) \quad (5)$$

Then, all e_k values are fused using the ϕ function, which consists of a batch normalization and mean layer.

$$\mathrm{e} = \sum_{k=1}^{K} \phi(\mathrm{e}_k) \quad (6)$$

Next, e is fed into a fully connected layer and a convolution layer. The output is then channel-multiplied $\otimes$ and channel-added $\oplus$ with the original features X to obtain salient feature information of small objects.

$$X_2 = X \oplus (X \otimes (\delta(\mathrm{Conv}_{1\times1}(\mathrm{e})))) \quad (7)$$

Where δ is the sigmoid function.

Finally, the features X_1 produced by the lightweight MLP and the feature X_2 generated by the LVC are concatenated along the channel dimension.

$$\mathrm{EVC}(X) = \mathrm{cat}(X_1; X_2) \quad (8)$$

3.4 Improve Feature Representation By Optimizing Multi-Scale Prediction Structure

After feature extraction from the input images, YOLOv8 generates prediction results on the feature maps at C3, C4, and C5 scales, which dose not fully exploit low-level feature information, as shown in Figure 2. High-level feature maps contain rich semantic information, which is beneficial for object classification. However, high-level

feature maps lack spatial and contextual information after multiple down-sampling operations and show weak representation ability for small objects. In contrast, low-level feature maps, with high resolution and small receptive fields, contain rich fine-grained and location information and demonstrate strong representation power for small objects, which is beneficial to the detection of small objects, as shown in Figure 5.

To improve the detection performance for small objects, the multi-scale prediction structure is optimized by adding a C2 prediction scale, as shown in Figure 2. With this optimization, the improved network outputs prediction results on feature maps at C2, C3, C4, and C5 scales, which fully utilize low-level feature maps, retain the feature information of small materials, and enhance feature representation for small materials. Moreover, the optimization of the multi-scale prediction structure also strengthens the aggregation of multi-level features to obtain richer semantic and spatial information, thereby improving detection accuracy for small materials.

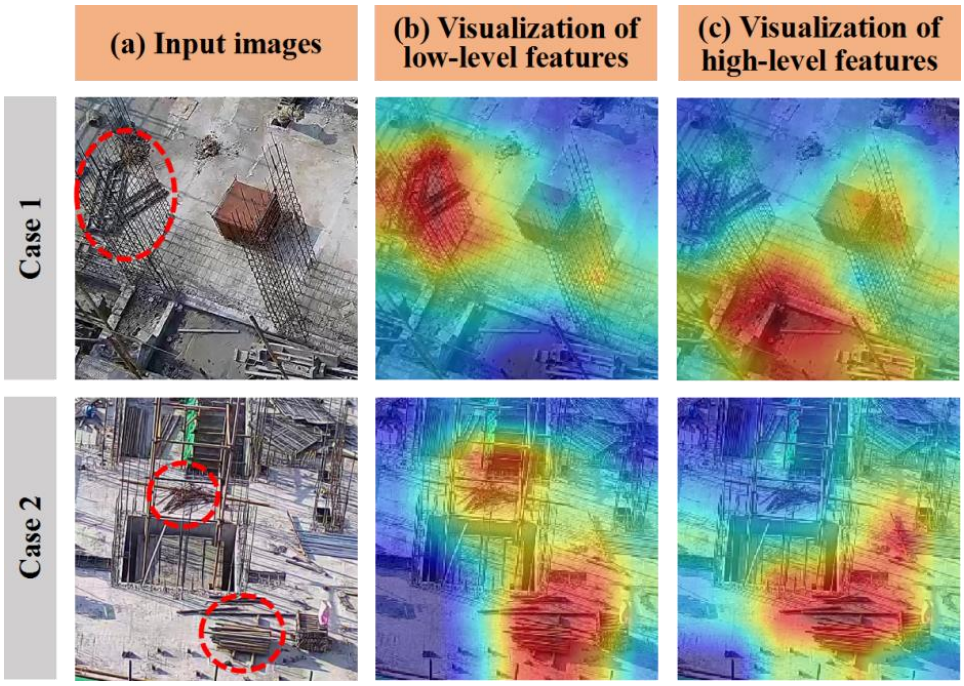


Figure 5: Visualization of low-level and high-level features. The red area in Figure 5 corresponds to the feature representation of objects.

In summary, YOLOv8 is adopted as the baseline model for material detection. To improve its detection accuracy for small materials, we optimize YOLOv8 by introducing DenseNet, leveraging the explicit visual center module, and optimizing multi-scale prediction structures.

4. DATA COLLECTION

To facilitate materials detection, a construction materials dataset was created through image acquisition and annotation. The construction images were collected using surveillance cameras deployed on construction sites, covering various viewpoints, weather conditions, and illumination conditions, as shown in Table 2. Examples of the collected images can be seen in Figure 6. The Labelme tool was utilized for data annotation. A total of 11 categories of construction materials were annotated based on common materials, as shown in Figure 7.

The constructed dataset consists of 5,005 images and 60,196 annotated objects. The distribution of objects in each category is presented in Table 3. Among them, wood_brace and rebar have the largest number of annotations, while panel and slab have fewer annotations, which aligns with the material usage and distribution on construction sites. The annotated objects were categorized into three scales based on their area sizes: small, medium, and large. Small objects occupy less than 0.2% of the entire image area, medium objects occupy from 0.2% to less than 2%, and large objects occupy more than 2%. The number of objects occupying different proportions of the image area is shown in Figure 8. It can be observed that the majority of objects are small objects. To facilitate model training and evaluation, the dataset was randomly divided into training, validation, and test sets in a ratio of 7:2:1.

Table 2: Various conditions for image acquisition.

Conditions	Detailed information
Location	9 construction sites located in 4 cities
Project type	residence, shopping mall, school
Time	day, night
Weather	sunny, rainy, foggy
Environment	shadow, occlusion

Figure 6: Examples of the collected images.

Figure 7: The annotated 11 categories of construction materials.

Table 3: Number of objects and images in each category.

Category	Number of images	Number of objects
template_aluminum	1206	7976
steel_tube	1407	3862
wood_brace	3098	18497
rebar	2287	10820
template_wood	1617	4833
hooping	759	3097
slab	161	536
climbing_scaffold	418	6007
red_brick	340	1292
panel	108	567
white_brick	611	2709

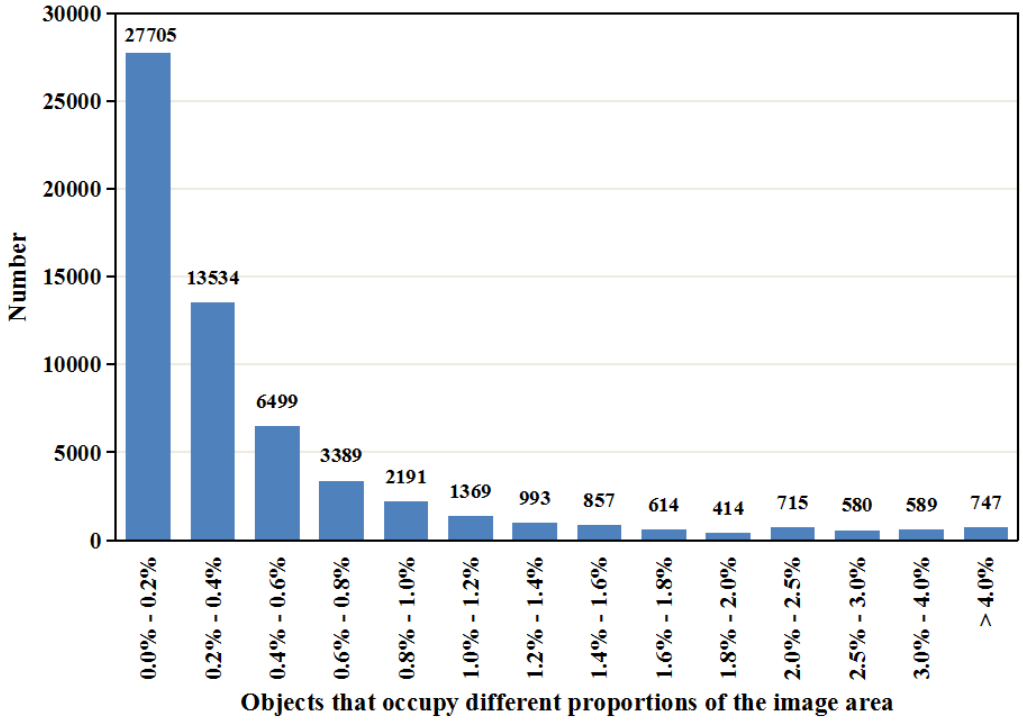


Figure 8: The number of objects that occupy different proportions of the image area.

5. RESULTS

The experiments were conducted using the following hardware and software environment: CPU Intel Core i9-11900K@ 3.50GHz, GPU Nvidia GeForce RTX 3090Ti, RAM 64 GB, Ubuntu 22.04.1, CUDA 11.7, CUDNN 8.4.1, Python 3.8.5, PyTorch 1.8.0. When training the proposed model, the initial learning rate was set to 0.001, the momentum to 0.937, the weight decay to 0.0005, the batch size to 4, the optimizer to SGD, and the epoch to 150.

The model performance was evaluated using commonly used metrics of AP (average precision) and mAP (mean average precision). The AP and mAP are derived from precision and recall, and their calculation formulas are presented in equations (9)-(12). Additionally, the APs, APm, and APl metrics were used to evaluate the model's

performance on small objects, medium objects, and large objects. The FLOPs (floating point operations) and FPS (frames per second) metrics were used to measure model complexity and detection speed.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

Where TP (True Positive) is the count of positive samples correctly identified; FP (False Positive) is the count of negative samples mistakenly identified as positive; FN (False Negative) is the count of positive samples mistakenly identified as negative.

The AP (average precision) for each category was calculated as the area under the precision-recall curve.

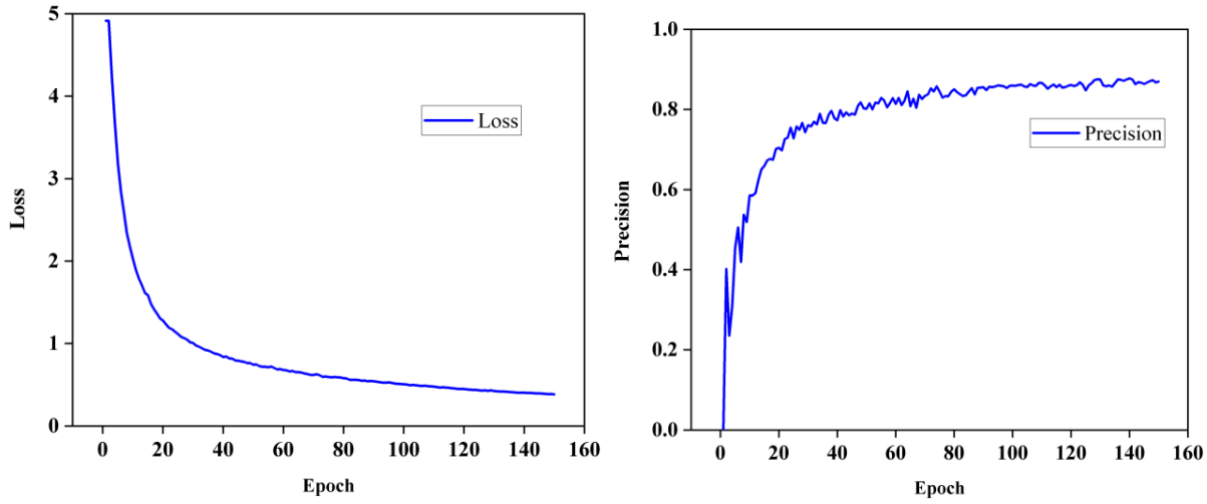
$$AP = \int_0^1 P(R)dR \quad (11)$$

The mAP (mean average precision) was obtained by averaging the AP for all categories.

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \quad (12)$$

5.1 Experiment Results

The loss and precision curves during training are shown in Figure 9. It can be seen that the loss consistently decreases, and the accuracy continuously increases until reaching a stable value as the training epoch increases, indicating that the model has converged during training.



(a) *(b)*

Figure 9: The loss and precision curves during training. (a)the loss curve, (b) the precision curve.

The Precision-Recall curve is shown in Figure 10, where the average precision for each category is the area under the curve. It can be seen that the proposed method has achieved high accuracy in each category.

5.2 Comparison with State-of-the-art Methods

Table 4 presents the detection performance of the proposed method and other state-of-the-art methods on the validation set. It can be seen that the proposed method exhibits a significant accuracy advantage in construction materials detection, achieving the best mAP of 0.843 among all methods. It also achieves the best APs of 0.768, outperforming other methods with a significant accuracy advantage. These results demonstrate the effectiveness and superiority of the proposed method in detecting construction materials, especially for small materials.

Although the proposed method is inferior to YOLOv8 in terms of FLOPS and FPS, it still shows advantages compared with other methods.

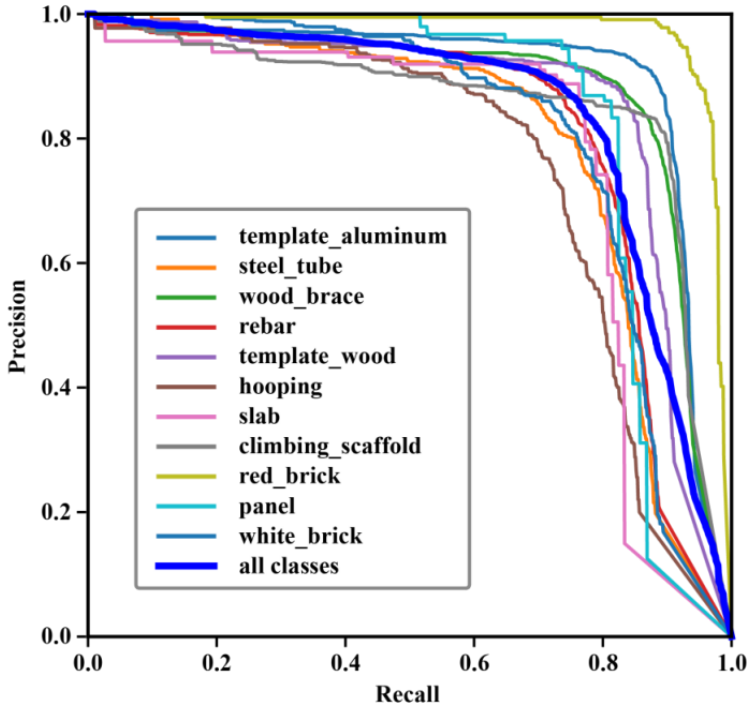


Figure 10: The Precision-Recall curve of the proposed method on the validation set.

5.3 Ablation Experiments

Ablation experiments were conducted to evaluate the effectiveness and performance of different optimization strategies. The experimental results are shown in Table 5. It can be seen that each strategy alone can improve detection accuracy, with the most significant improvement observed for small objects (APs). The reason is that the DenseNet enhances the feature extraction for small materials, the EVC module improves the feature learning for small objects, and the C_2 detection scale improves the utilization of small materials features. After integrating all strategies into YOLOv8, the accuracy is further improved, with the mAP reaching 0.843 and the APs improving by 5.3% (from 0.715 to 0.768). These results validate the effectiveness of the optimization strategies. Although using all optimization strategies caused an increase in FLOPs, the detection speed of the proposed method still meets the real-time detection criteria (FPS≥20).

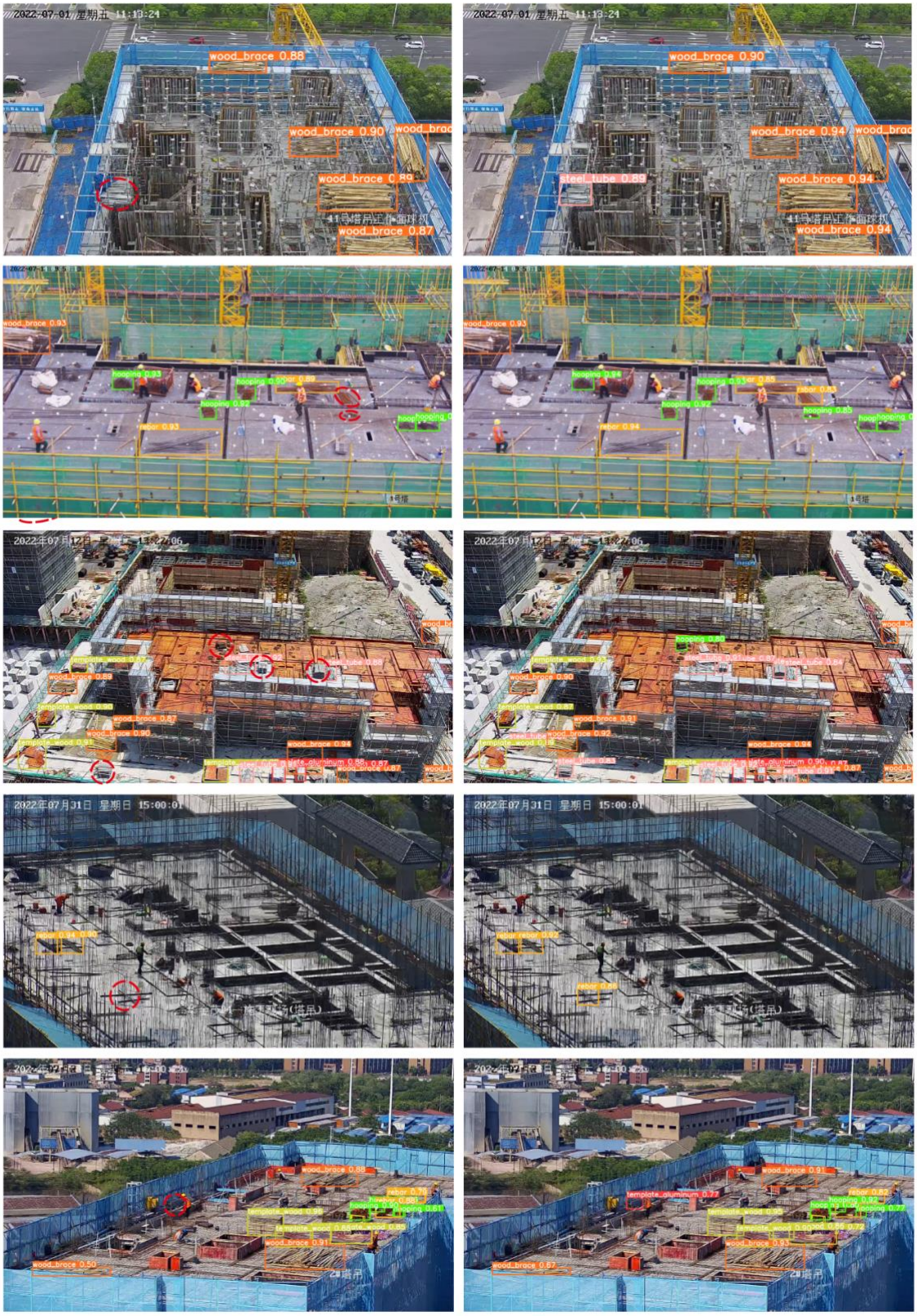
Table 4: Effects of different improved strategies on detection performance.

Method	Improved strategies			mAP	APs	APm	AP_l	FLOPs	FPS
	DenseNet	EVC	C2						
YOLOv8m				0.821	0.715	0.849	0.853	78.8	39.2
Method-A	√			0.835	0.752	0.856	0.853	95.6	31.3
Method-B		√		0.832	0.747	0.848	0.856	97.8	28.2
Method-C			√	0.834	0.751	0.852	0.854	98.1	29.1
Method-D	√	√		0.838	0.759	0.857	0.861	129.3	26.5
Method-E	√		√	0.840	0.761	0.854	0.865	114.8	25.7
Method-F		√	√	0.837	0.756	0.853	0.868	138.4	24.3
Ours	√	√	√	0.843	0.768	0.859	0.874	148.5	22.1

Table 5: The detection performance of different methods on the validation set.

TA means template_aluminum, ST means steel_tube, WB means wood_brace, TW means template_wood, CS means climbing_scaffold, RB means red_brick, WK means white_brick, HP means hooping. IOU=0.5. APs, APm, and APl are the average precision for small, medium, and large objects.

Method	AP											mAP	APs	APm	AP_l	FLOPs (G)	FPS
	TA	ST	WB	TW	CS	RB	WK	HP	rebar	slab	panel						
Faster-RCNN (Ren et al. 2017)	0.853	0.674	0.812	0.752	0.828	0.940	0.750	0.627	0.739	0.668	0.792	0.767	0.681	0.796	0.830	431.2	5.9
Cascade-RCNN (Cai et al. 2018)	0.845	0.689	0.809	0.755	0.835	0.945	0.737	0.639	0.722	0.685	0.786	0.768	0.662	0.799	0.847	433.2	13.9
FCOS (Tian et al. 2019)	0.876	0.678	0.834	0.782	0.828	0.958	0.763	0.666	0.745	0.690	0.786	0.782	0.673	0.823	0.811	437.0	16.3
NAS-FCOS (Wang et al. 2020)	0.882	0.722	0.855	0.789	0.839	0.938	0.748	0.691	0.760	0.671	0.767	0.787	0.662	0.829	0.856	330.2	12.6
ATSS (Zhang et al. 2020a)	0.887	0.712	0.849	0.788	0.880	0.944	0.747	0.699	0.774	0.652	0.796	0.793	0.691	0.834	0.857	444.6	15.6
Vfnet (Zhang et al. 2021)	0.889	0.730	0.860	0.801	0.835	0.960	0.724	0.710	0.776	0.684	0.760	0.794	0.651	0.844	0.867	424.7	13.0
Reppoints (Yang et al. 2019)	0.892	0.725	0.859	0.800	0.872	0.950	0.746	0.730	0.775	0.617	0.786	0.796	0.671	0.846	0.854	425.3	13.0
Autoassign (Zhu et al. 2020)	0.903	0.737	0.865	0.815	0.874	0.958	0.780	0.730	0.800	0.656	0.785	0.809	0.702	0.851	0.859	438.7	14.5
PAA (Kim et al. 2020)	0.903	0.773	0.878	0.808	0.864	0.958	0.748	0.735	0.808	0.705	0.807	0.817	0.715	0.859	0.873	444.6	7.3
TOOD (Feng et al. 2021)	0.912	0.764	0.884	0.831	0.855	0.970	0.746	0.733	0.798	0.750	0.766	0.819	0.709	0.860	0.878	287.7	10.0
YOLOv8m (Ultralytics, 2023)	0.907	0.784	0.870	0.838	0.860	0.946	0.784	0.735	0.794	0.731	0.785	0.821	0.715	0.849	0.853	78.8	39.2
Ours	0.917	0.793	0.879	0.865	0.850	0.967	0.823	0.749	0.817	0.773	0.836	0.843	0.768	0.859	0.874	148.5	22.1



(a) YOLOv8 *(b) Our method*

Figure 11: The detection results of YOLOv8 and the proposed method on the test set.

5.4 Visualizations of Detection Results

Visualizations of small materials detection results between the proposed method and YOLOv8 on the test set are shown in Figure 11. It can be seen that the proposed method exhibits excellent detection performance for small objects and accurately detects small construction materials. However, YOLOv8 suffers from missed detection, with the red dashed circle in Figure 11 being the missed detection objects. The visualization results demonstrate the effectiveness of the proposed method in detecting small construction materials.

6. DISCUSSION

6.1 Impact of Different Backbone Networks on Accuracy

Several experiments were conducted to compare the impact of DenseNet and other backbone networks on detection performance, such as HorNet, RepVGG, and ConvNext. The experimental results, presented in Table 6, demonstrate that YOLOv8m-DenseNet achieves the best mAP and APs, surpassing other methods by a significant accuracy advantage. The effectiveness of DenseNet stems from its ability to enhance feature propagation between layers, promote feature fusion and reuse, and enable the network to obtain more feature information. In addition, DenseNet can alleviate gradient vanishing and promote the network's feature learning for small objects by improving the flow of gradients between layers. YOLOv8m-DenseNet is inferior to other methods in FLOPs and FPS since the dense connection in DenseNet leads to more computational overhead and memory access cost.

Table 6: Effects of different backbone networks on detection performance.

Method	Backbone	mAP	APs	APm	AP_l	FLOPs	FPS
YOLOv8m	YOLOv8m (Ultralytics, 2023)	0.821	0.715	0.849	0.853	78.8	39.2
YOLOv8m-HorNet	HorNet (Rao et al. 2023)	0.820	0.726	0.844	0.850	69.7	34.1
YOLOv8m-RepVGG	RepVGG (Ding et al. 2021)	0.821	0.719	0.845	0.862	75.5	43.7
YOLOv8m-ConvNeXt	ConvNeXt (Liu et al. 2022)	0.823	0.727	0.850	0.848	63.1	44.4
YOLOv8m-ResNext	ResNext (Xie et al. 2017)	0.823	0.730	0.846	0.851	55.2	43.5
YOLOv8m-ResNet	ResNet (He et al. 2016)[57]	0.824	0.728	0.851	0.852	66.3	42.4
YOLOv8m-DenseNet	DenseNet (Huang et al. 2017)	0.835	0.752	0.856	0.853	95.6	31.3

6.2 Effect of Different Output Prediction Scales on Accuracy

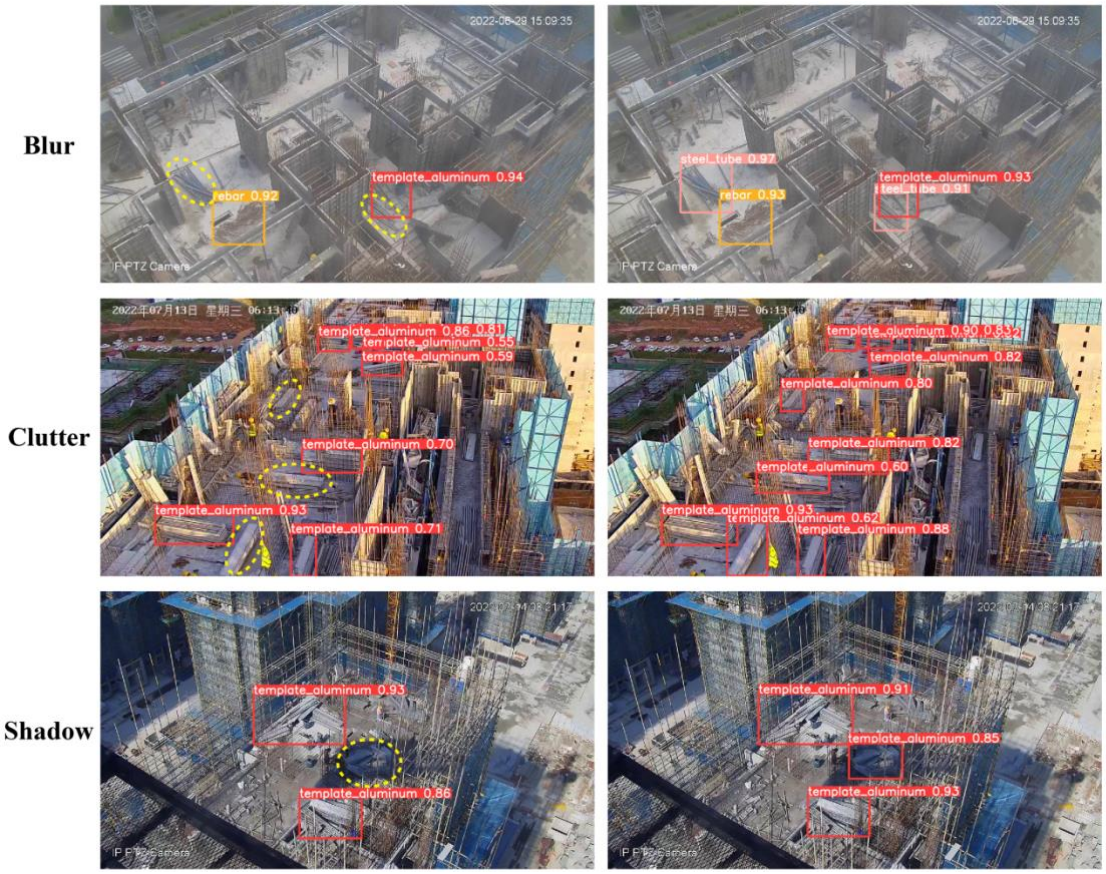
Comparative experiments were conducted to investigate the effect of different output prediction scales on detection performance. The experiment results, reported in Table 7, show an increase in accuracy with more prediction scales. This improvement can be attributed to the ability of networks with more prediction scales to fuse more feature maps at different levels, thereby improving detection accuracy. However, more output scales increase computational overhead and model complexity, resulting in a drop in FPS and an increase in FLOPs. Notably, when using five prediction scales of C1, C2, C3, C4, and C5, the improvement in accuracy is not significant, suggesting that the model's performance has reached saturation. Therefore, to improve detection accuracy while balancing the detection speed and accuracy, this study adopts the C2, C3, C4, and C5 scales as the output prediction scales.

Table 7: Effects of different output prediction scales on detection performance.

Method	Output prediction scales					mAP	APs	APm	AP_l	FLOPs	FPS
	C1	C2	C3	C4	C5						
Method-1				√	√	0.807	0.648	0.849	0.851	62.3	52.1
YOLOv8m			√	√	√	0.821	0.715	0.849	0.853	78.8	39.2
Method-2		√	√	√	√	0.834	0.751	0.852	0.854	98.1	29.1
Method-3	√	√	√	√	√	0.835	0.749	0.854	0.859	126.3	18.1

6.3 Adaptability in Various Construction Environments

To examine the adaptability of the proposed method for construction materials detection in various environments, we compared the detection results of the proposed model and YOLOv8 in various construction scenes, including blurred detection conditions, cluttered backgrounds, and shadow interference. As shown in Figure 12, the proposed method accurately detects construction materials in various environments, while YOLOv8 suffers from missed detection, as indicated by the yellow dashed circle in Figure 12. The effectiveness of the proposed method in complex environments can be attributed to the feature reuse promoted by DenseNet, which enables the network to obtain more feature information and enhances its feature representation capability in various environments. Moreover, the EVC module enables the model to obtain richer global information. Additionally, optimizing multi-scale prediction structures can obtain more feature information, which improves the network's feature learning ability in various conditions. These results demonstrate the strong adaptability and robustness of the proposed method in various construction scenarios.



(a) YOLOv8 *(b) Our method*

Figure 12: The detection results of YOLOv8 and our method in complex scenes. The yellow dashed circle represents the missed objects.

6.4 Application and Limitation

(1) Application

The proposed method can be integrated into construction site monitoring and material management systems. By analyzing video and image data from the monitoring system and connecting detection results to the material management system, this integration enables efficient tracking of material movement, usage monitoring, and inventory management. This comprehensive approach enhances material lean management, improves tracking efficiency, reduces waste, and strengthens site safety.

The proposed method offers several key benefits. At construction site entrances and exits, it can accurately detect and track the types and quantities of materials moving through the site, assisting in the management of material reception and transfer. On the construction site itself, the method can help monitor daily material usage and consumption, linking directly with warehouse inventory management. The method also enhances safety by detecting construction materials placed near work surface edges, triggering timely alerts for site managers to prevent falling hazards. In the material warehouse, the method can help track inventory levels and material changes, combining this data with construction site consumption patterns to optimize inventory management and material procurement planning.

The proposed method has been successfully implemented across several large-scale construction projects in China. Through real-time analysis of site surveillance video, it enables automatic monitoring of daily material usage, supporting efficient material management and advancing the intelligent oversight of on-site resources.

(2) Limitation

Although the proposed method can be integrated with the monitoring and material management systems, it still encounters several limitations and challenges. The proposed method depends on monitoring videos and images of the construction site for detection, which can be constrained by the availability of such visual data. Furthermore, the detection accuracy of the method is influenced by occlusion. When materials are heavily or entirely occluded, they may not be successfully detected.

7. CONCLUSION

Automated construction detection plays a crucial role in material lean management, automatic construction progress monitoring, and 3D as-built model generation. Traditional material detection relies heavily on manual inspection, which is time-consuming and error-prone. Deep learning has advanced the automatic detection of construction materials. However, there are numerous small materials in the construction site, which are difficult to detect due to their low resolution and insufficient features, posing challenges to refined materials management. To solve this problem, this study proposes an augmented detection method for small materials based on enhanced feature extraction and representation. The experiment results show that the proposed method significantly improves the detection accuracy for small materials and exhibits superior performance in detecting small materials. Furthermore, the proposed method demonstrates strong adaptability and robustness to various construction conditions such as shadows, blurriness, and cluttered backgrounds.

This study contributes to the body of knowledge by developing an improved detection method for small materials, which improves the detection accuracy for small materials. The proposed method offers three main advantages. First, it establishes an enhanced feature extraction network utilizing DenseNet, thereby improving feature extraction and utilization of small materials. Second, it designs an improved feature pyramid network by introducing the EVC module, enhancing the feature learning of small objects. Finally, it constructs an enhanced multi-scale prediction network by incorporating a C2 detection scale, enabling network to obtain richer feature information of small materials and improving multi-level feature fusion. This research facilitates material lean management and contributes to the potential application of digital twin in materials management.

Despite the aforementioned contributions, this study still has one limitation: the uneven data distribution between categories in the dataset. For instance, the number of objects in the slab and panel categories is lower compared to other categories. Future studies are anticipated to expand the dataset to balance the number of objects across different categories, reducing the impact of category imbalance on model performance. In addition, this study will further investigate the feasibility and effectiveness of network structures such as Transformer and Mamba in enhancing the detection accuracy of small construction materials. This study will also explore pruning or compressing the model to make it lightweight, enabling deployment on embedded devices.

ACKNOWLEDGEMENTS

The study was supported by the National Key Research & Development Program of China (2022YFC3801700), National Natural Science Foundation of China (52078374), Science and Technology Commission of Shanghai Municipality (22dz1207100), Chinese Academy of Engineering (2024-XZ-37), and Fundamental Research Funds for the Central Universities (2024-1-ZD-02, 22120240236).

REFERENCES

Bai Y., Zhang Y., Ding M., Ghanem B. (2018). Finding Tiny Faces in the Wild with Generative Adversarial Network, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 21-30, https://doi.org/10.1109/CVPR.2018.00010.

Bhokare S, Goyal L, Ren R, Zhang J (2022). Smart construction scheduling monitoring using YOLOv3-based activity detection and classification, ITcon Vol. 27, pg. 240-252, https://doi.org/10.36680/j.itcon.2022.012

Bochkovskiy A., Wang C., Liao H.M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv, http://arxiv.org/abs/2004.10934,2020

Cai Z., Vasconcelos N. (2018). Cascade R-CNN: Delving Into High Quality Object Detection, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6154-6162, https://doi.org/10.1109/CVPR.2018.00644

Dimitrov A., Golparvar-Fard M. (2014). Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections, Advanced Engineering Informatics, 28(1), 37-49, https://doi.org/10.1016/j.aei.2013.11.002.

Ding X., Zhang X., Ma N., Han J., Ding G., Sun J. (2021). RepVGG: Making VGG-style ConvNets Great Again, arXiv, https://doi.org/10.48550/arXiv.2101.03697

Duan R., Deng H., Tian M., Deng Y., Lin J. (2022). SODA: A large-scale open site object detection dataset for deep learning in construction, Automation in Construction, 142, 104499, https://doi.org/10.1016/j.autcon.2022.104499.

Elgendi E-B O, Shawki K M, Ashraf Mohy A (2023). Video analysis for tower crane production rate estimation, ITcon Vol. 28, pg. 138-150, https://doi.org/10.36680/j.itcon.2023.007

Feng C., Zhong Y., Gao Y., Scott M.R., Huang W. (2021). TOOD: Task-aligned One-stage Object Detection, IEEE/CVF International Conference on Computer Vision, 3490-3499, https://doi.org/10.1109/ICCV48922.2021.00349

Gao C., Tang W., Jin L., Jun Y. (2020). Exploring Effective Methods to Improve the Performance of Tiny Object Detection, European Conference on Computer Vision, vol 12539, pp 331-336, https://doi.org/10.1007/978-3-030-68238-5_25

Ghiasi G., Lin T.Y., Pang R., Le Q.V. (2019). NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection, arXiv, https://doi.org/10.48550/arXiv.1904.07392

Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2014). Generative Adversarial Networks, arXiv, https://doi.org/10.48550/arXiv.1406.2661

He K., Zhang X., Ren S., Sun J. (2016). Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, https://doi.org/10.1109/CVPR.2016.90

He X., Tang Z., Deng Y., Zhou G., Wang Y., Li L. (2023). UAV-based road crack object-detection algorithm, Automation in Construction, 154, 105014, https://doi.org/10.1016/j.autcon.2023.105014

Huang G., Liu Z., Van Der Maaten L., Weinberger K.Q. (2017). Densely connected convolutional networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708, https://doi.org/10.1109/CVPR.2017.243

Huang Q, Hao K (2020). Development of CNN-based visual recognition air conditioner for smart buildings, ITcon Vol. 25, pg. 361-373, https://doi.org/10.36680/j.itcon.2020.021

Jocher G., Chaurasia A., Stoken A., Borovec J., Chanvichet V., Kwon Y., Xie T., Michael K., Fang J. (2022). Ultralytics/yolov5: v6. 2-YOLOv5 classification models, https://github.com/ultralytics/yolov5

Khan N., Saleem M.R., Lee D., Park M.W., Park C. (2021). Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks, Computers in Industry, 129, 103448, https://doi.org/10.1016/j.compind.2021.103448

Kini D.U. (1999). Materials Management: The Key to Successful Project Management, Journal of Management in Engineering, 15(1), 30-34, https://doi.org/10.1061/(ASCE)0742-597X(1999)15:1(30)

Kisantal M., Wojna Z., Murawski J., Naruniec J., Cho K. (2019). Augmentation for small object detection, arXiv, https://doi.org/10.48550/arXiv.1902.07296

Kim K., Seok Lee H. (2020). Probabilistic Anchor Assignment with IoU Prediction for Object Detection, European Conference on Computer Vision, 12370, 355-371, https://doi.org/10.1007/978-3-030-58595-2_22

Kim S., Hong S.H., Kim H., Lee M., Hwang S. (2023). Small object detection (SOD) system for comprehensive construction site safety monitoring, Automation in Construction, 156, 105103, https://doi.org/10.1016/j.autcon.2023.105103

Li J., Liang X., Wei Y., Xu T., Feng J., Yan S. (2017). Perceptual Generative Adversarial Networks for Small Object Detection, 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1951-1959, https://doi.org/10.1109/CVPR.2017.211

Li Y., Chen Y., Wang N., Zhang Z. (2019). Scale-Aware Trident Networks for Object Detection, arXiv, https://doi.org/10.48550/arXiv.1901.01892

Lin T.Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. (2017). Feature pyramid networks for object detection, IEEE Conference on Computer Vision and Pattern Recognition, arXiv, 2117-2125, https://arxiv.org/abs/1612.03144

Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A.C. (2016). SSD: Single shot multibox detector, in: 2016 European Conference on Computer Vision (ECCV), Springer, Amsterdam, The Netherlands, pp. 21-37, https://doi.org/10.1007/978-3-319-46448-0_2

Liu S., Qi L., Qin H., Shi J., Jia J. (2018). Path aggregation network for instance segmentation, IEEE Conference on Computer Vision and Pattern Recognition, arXiv, 8759-8768, https://arxiv.org/abs/1803.01534

Liu Z., Mao H., Wu C.Y., Feichtenhofer C., Darrell T., Xie S. (2022). A ConvNet for the 2020s, arXiv, https://doi.org/10.48550/arXiv.2201.03545

Mahaur B., Mishra K.K. (2023). Small-object detection based on YOLOv5 in autonomous driving systems, Pattern Recognition Letters, 168, 115-122, https://doi.org/10.1016/j.patrec.2023.03.009

National Research Council. (2009). Advancing the competitiveness and efficiency of the U.S. construction industry. Washington, DC: National Academies Press, http://www.nap.edu/catalog/12717.html

Noh J., Bae W., Lee W., Seo J., Kim G. (2019). Better to Follow, Follow to Be Better: Towards Precise Supervision of Feature Super-Resolution for Small Object Detection, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 9724-9733, https://doi.org/10.1109/ICCV.2019.00982

Park M., Tran D.Q., Bak J., Park S. (2023). Small and overlapping worker detection at construction sites, Automation in Construction, 151, 104856, https://doi.org/10.1016/j.autcon.2023.104856

Quan Y., Zhang D., Zhang L., Tang J. (2023). Centralized Feature Pyramid for Object Detection, in IEEE Transactions on Image Processing, vol. 32, pp. 4341-4354, doi: 10.1109/TIP.2023.3297408

Ren S., He K., Girshick R., Sun J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137-1149, https://doi.org/10.1109/TPAMI.2016.2577031

Ren Z., Anumba C.J., Tah J. (2011). RFID-facilitated construction materials management (RFID-CMM)-A case study of water-supply project, Advanced Engineering Informatics, 25(2), 198-207, https://doi.org/10.1016/j.aei.2010.02.002

Rao Y., Zhao W., Tang Y., Zhou J., Lim S.N., Lu J. (2022). HorNet: Efficient High-Order Spatial Interactions with Recursive Gated Convolutions, arXiv, https://doi.org/10.48550/arXiv.2207.14284

Redmon J., Farhadi A. (2018). YOLOv3: An Incremental Improvement. arXiv, https://arxiv.org/abs/1804.02767

Son H., Kim C., Kim C.W. (2012). Automated Color Model–Based Concrete Detection in Construction-Site Images by Using Machine Learning Algorithms, Journal of Computing in Civil Engineering, 26(3), 421-433, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000141

Son H., Kim C., Hwang N., Kim C., Kang Y. (2014). Classification of major construction materials in construction environments using ensemble classifiers, Advanced Engineering Informatics, 28(1), 1-10, https://doi.org/10.1016/j.aei.2013.10.001

Song J., Haas C.T., Caldas C.H. (2006). Tracking the Location of Materials on Construction Job Sites, Journal of Construction Engineering and Management, 32(9), 911-918, https://doi.org/10.1061/(ASCE)0733-9364(2006)132:9(911)

Tan M., Pang R., Le Q.V. (2019). EfficientDet: Scalable and Efficient Object Detection, arXiv, https://doi.org/10.48550/arXiv.1911.09070

Tian Z., Shen C., Chen H., He T. (2019). FCOS: Fully convolutional one-stage object detection, in: 2019 IEEE International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea, pp. 9626-9635, https://doi.org/10.1109/ICCV.2019.00972

Ultralytics. (2023). YOLOv8, https://github.com/ultralytics/ultralytics

Wang N., Gao Y., Chen H., Wang P., Tian Z., Shen C., Zhang Y. (2020). NAS-FCOS: Fast Neural Architecture Search for Object Detection, arXiv, https://doi.org/10.48550/arXiv.1906.04423

Xie S., Girshick R., Dollár P., Tu Z., He K. (2017). Aggregated Residual Transformations for Deep Neural Networks, arXiv, https://doi.org/10.48550/arXiv.1611.05431

Yang Z., Liu S., Hu H., Wang L., Lin S. (2019). RepPoints: Point Set Representation for Object Detection, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 9656-9665, https://doi.org/10.1109/ICCV.2019.00975.

Yuan L., Guo J., Wang Q. (2020). Automatic classification of common building materials from 3D terrestrial laser scan data, Automation in Construction, 110, 103017, https://doi.org/10.1016/j.autcon.2019.103017

Zhang S., Chi C., Yao Y., Lei Z., Li S.Z. (2020a). Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9756-9765, https://doi.org/10.1109/CVPR42600.2020.00978

Zhang Y., Bai Y., Ding M., Ghanem B. (2020b). Multi-task Generative Adversarial Network for Detecting Small Objects in the Wild. Int J Comput Vis 128, 1810–1828. https://doi.org/10.1007/s11263-020-01301-6

Zhang H., Wang Y., Dayoub F., Sünderhauf N. (2021). VarifocalNet: An IoU-aware Dense Object Detector, arXiv, https://doi.org/10.48550/arXiv.2008.13367

Zhu B., Wang J., Jiang Z., Zong F., Liu S., Li Z., Sun J. (2020). AutoAssign: Differentiable Label Assignment for Dense Object Detection, arXiv, https://doi.org/10.48550/arXiv.2007.03496

Zhu Z., Brilakis I. (2010). Parameter optimization for automated concrete detection in image data, Automation in Construction, 19(7), 944-953, https://doi.org/10.1016/j.autcon.2010.06.008