

Research Article

Cite this article: An N, Huang L, Hu M, Zhu J and Wang C (2024). Improved basic elements detection algorithm for bridge engineering design drawings based on YOLOv5. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **38**, e22, 1–11
<https://doi.org/10.1017/S089006042400026X>

Received: 23 November 2023

Revised: 17 September 2024

Accepted: 03 October 2024


Keywords:

bridge engineering design drawings; basic elements detection; improved YOLOv5; attention mechanism; atrous convolution

Corresponding authors:

Junan Zhu and Chuanjian Wang;
Emails: zhujunan@ahu.edu.cn; wcj_si@ahu.edu.cn

Improved basic elements detection algorithm for bridge engineering design drawings based on YOLOv5

Ning An¹ , Linsheng Huang², Mengnan Hu², Junan Zhu² and Chuanjian Wang²

¹School of Electronic Information Engineering, Anhui University, Hefei, Anhui, China and ²School of Internet, Anhui University, Hefei, Anhui, China

Abstract

Bridge engineering design drawings basic elements contain a large amount of important information such as structural dimensions and material indexes. Basic element detection is seen as the basis for digitizing drawings. Aiming at the problem of low detection accuracy of existing drawing basic elements, an improved basic elements detection algorithm for bridge engineering design drawings based on YOLOv5 is proposed. Firstly, coordinate attention is introduced into the feature extraction network to enhance the feature extraction capability of the algorithm and alleviate the problem of difficult recognition of texture features inside grayscale images. Then, targeting objectives across different scales, the standard 3×3 convolution in the feature pyramid network is replaced with switchable atrous convolution, and the atrous rate is adaptively selected for convolution computation to expand the sensory field. Finally, experiments are conducted on the bridge engineering design drawings basic elements detection dataset, and the experimental results show that when the Intersection over Union is 0.5, the proposed algorithm achieves a mean average precision of 93.6%, which is 3.4% higher compared to the original YOLOv5 algorithm, and it can satisfy the accuracy requirement of bridge engineering design drawings basic elements detection.

Introduction

The widespread application of artificial intelligence in various industries is profoundly changing our lives. In the medical field, AI can be used for medical image recognition and disease diagnosis, thus helping doctors make better treatment decisions (Esteva et al., 2017). In the financial sector, AI enables more accurate risk assessment and fraud detection, thereby improving the security of financial services (Nabipour et al., 2020). In the field of e-commerce, AI can provide personalized recommendations and inventory management, consequently improving sales efficiency and increasing user experience (Zhang et al., 2019). Despite the significant progress of AI in many fields, there are still some challenges in the application of AI in the field of bridge engineering design, especially in the recognition of bridge engineering drawings. In bridge engineering design, safety is crucial to the life and property security of the public. Bridges, as vital infrastructure, serve an essential function in transportation, and any safety accidents could have serious repercussions on society. Engineering drawings serve as the primary execution reference for construction projects. Therefore, accurately identifying bridge engineering drawings is paramount to ensuring the safety of bridge designs.

Bridge engineering drawings usually contain a large amount of technical details and complex structural information, which poses difficulties in the recognition of drawings. Specifically, the specific textual information contained in bridge engineering drawings only accounts for a small part, while the majority is geometric graphics. Therefore, the existing text-based detection methods are not applicable. Moreover, due to the lack of color semantic information in grayscale bridge engineering design drawings, it is difficult to effectively distinguish basic elements through texture features. Additionally, the text box is a long target, which differs greatly in size from other basic elements, thus posing further difficulties in the detection of basic elements. AI models need to have a high degree of recognition accuracy and comprehension in order to accurately interpret the content of the drawings. Additionally, different designers may have varied styles, so AI models must possess strong generalization and adaptability to handle various types of drawing recognition tasks. Therefore, while artificial intelligence can save a significant amount of time and labor costs compared to the traditional manual process of reviewing and analyzing drawings, further research and technological breakthroughs are still needed to achieve more widespread applications.

As shown in Figure 1, bridge engineering design drawings are usually composed of basic elements such as plane, elevation, side, table of quantities, annotation, and text boxes. Different basic elements carry different design information. The plane, elevation, and side contain the size, shape, and topological information of the bridge structure; the quantity table contains the

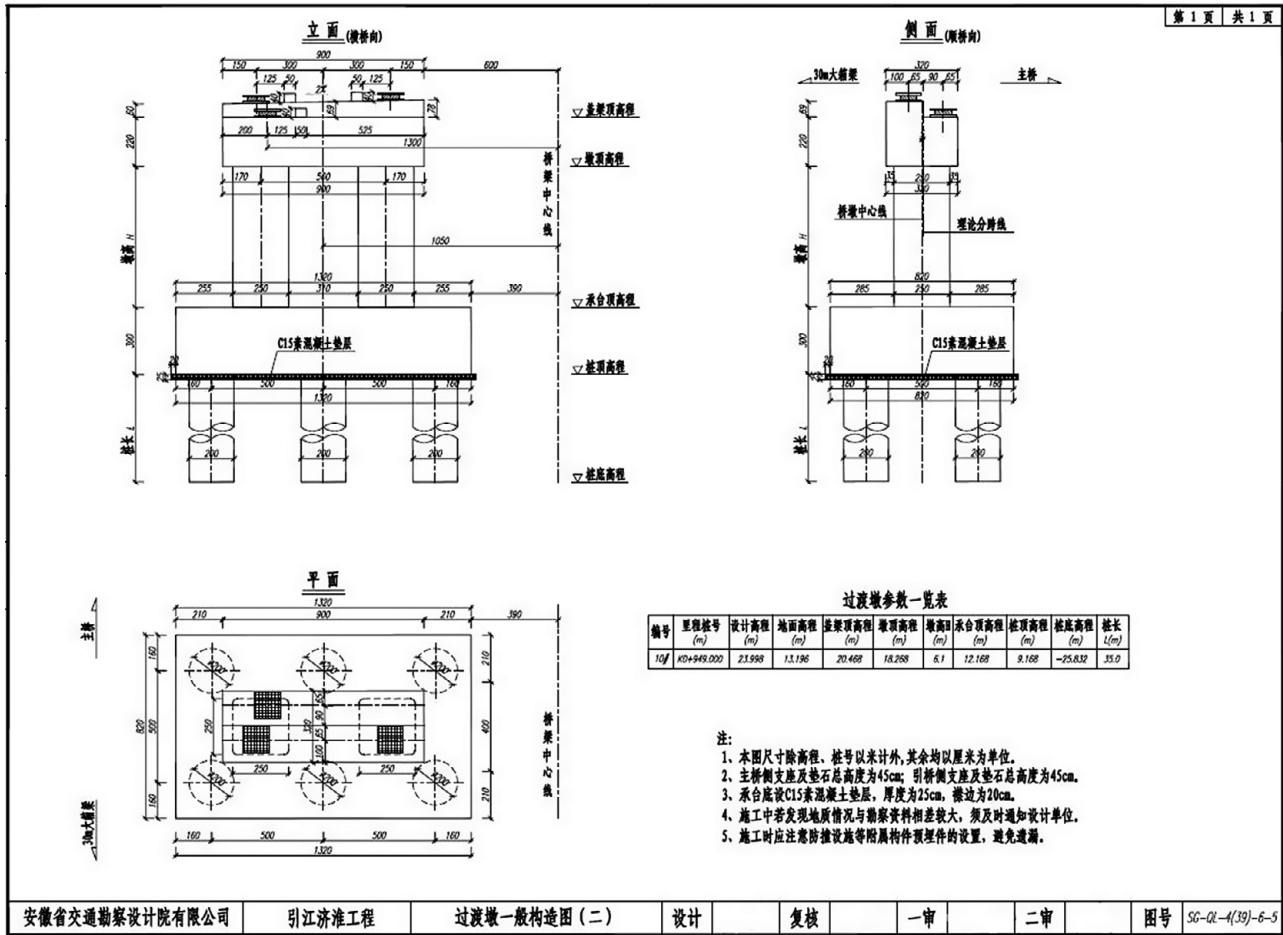


Figure 1. Example of bridge engineering design drawings.

material information of the current bridge members, implying material indexes such as reinforcement rate and steel content; the annotation contains the applicable conditions of the drawings and specific requirements before and after the construction; and the text box contains the name of the drawings, the drawing number, the design unit, the designer, and other necessary information of the drawings. In order to fully and comprehensively understand the design information in each bridge engineering design drawing as a whole, it is often necessary to understand the information in each basic element of the drawing separately. To understand each basic element, it is first necessary to localize and classify the basic element. In this paper, we propose to investigate an automatic classification and localization method for basic elements of bridge engineering design drawings, i.e., basic elements target detection.

To deal with the above problems, this paper proposes an improved basic elements detection algorithm for bridge engineering design drawings based on YOLOv5 to realize the detection of four types of basic elements in bridge engineering design drawings, namely, Figures, Forms, Annotations, and Text boxes. In terms of algorithmic improvement, Coordinate Attention (CA) is first introduced to enhance the feature extraction capability of the algorithm and alleviate the problem of difficulty in recognizing internal texture features in grayscale images. Then, Switchable Atrous Convolution (SAC) is introduced to adaptively expand the sensory field and capture multi-scale contextual information. Taking the bridge

design drawings as an example, the dataset containing four types of basic elements, namely, Figures, Forms, Annotations, and Text boxes, is self-developed. The experimental results show that the mean Average Precision (mAP) of the improved YOLOv5 algorithm is significantly improved, which realizes the detection of basic elements of bridge engineering design drawings and provides a basis for the digital management of drawings.

Following are the main contributions of our work:

1. To address the limitations of traditional methods in dealing with the detection of image elements of bridge engineering design drawings, such as low recognition accuracy and poor adaptability to complex scenes, this study innovatively applies deep learning methods to this field, which provides valuable experience and references for the subsequent image recognition research.
2. Aiming at the challenges of missing color information and multi-scale targets in drawing detection, this study innovatively integrates the Coordinate Attention mechanism and SAC in the YOLOv5 framework, which significantly improves the detection accuracy through fine design and optimization and opens up a new path for research in related fields.
3. In order to verify the effectiveness of the algorithm, we constructed a bridge engineering design drawings dataset, which contains 3000 samples covering four key elements: figures,

tables, annotations, and text boxes. We conducted sufficient experiments on this dataset, and the results show that our algorithm has good accuracy and can meet the practical needs.

Related work

So far, researchers have proposed a variety of methods to extract textual information from drawings, and the research on character recognition of engineering design drawings has been perfected, but there are few reports on the research on the detection of basic elements of engineering design drawings.

Engineering design drawing character recognition

Fan and Guan (2012) proposed a pre-segmentation algorithm based on the knowledge of engineering design drawings to realize the recognition of strings in the form of tables as well as element labeling information in drawings. Yang et al. used a template matching algorithm to obtain a template as a criterion for recognizing characters by extracting and selecting the features of the image, calculating the similarity between the template and the character to be recognized, and then judging whether it is the same character as the corresponding template based on the results. Brock et al. (2017) proposed an object detection framework based

on convolutional neural networks to achieve the localization and classification of various optical characters in engineering design drawings. Song et al. used the form of sliced table cells to locate and analyze the key information and realized the extraction of title bar information with the help of a convolutional neural network. Jiang et al. proposed an improved character detection method for drawing image characters characterized by the presence of a high number of interference, such as labeling lines, workpiece graphics, and indication symbols, which extracts a series of desired target text lines using a concatenated domain aggregation-based approach. Elyan et al. (2020) proposed a bounding box detection method for symbol localization and recognition in engineering drawings and improved the classification of symbols in engineering drawings by targeting the class imbalance problem using deep generative adversarial neural networks. Jamieson et al. (2020) used a deep learning approach to recognize text in engineering design drawings, such as piping and instrumentation drawings, and detected 90% of the text, including vertical text strings. Dong et al. (2023a) used the VGG model for feature extraction of grid engineering drawings, output the proposed candidate frames by the region candidate algorithm, achieved the unity of candidate frame size by the pooling layer of the region of interest, and achieved the character recognition of grid engineering drawings by the Faster R-CNN algorithm.

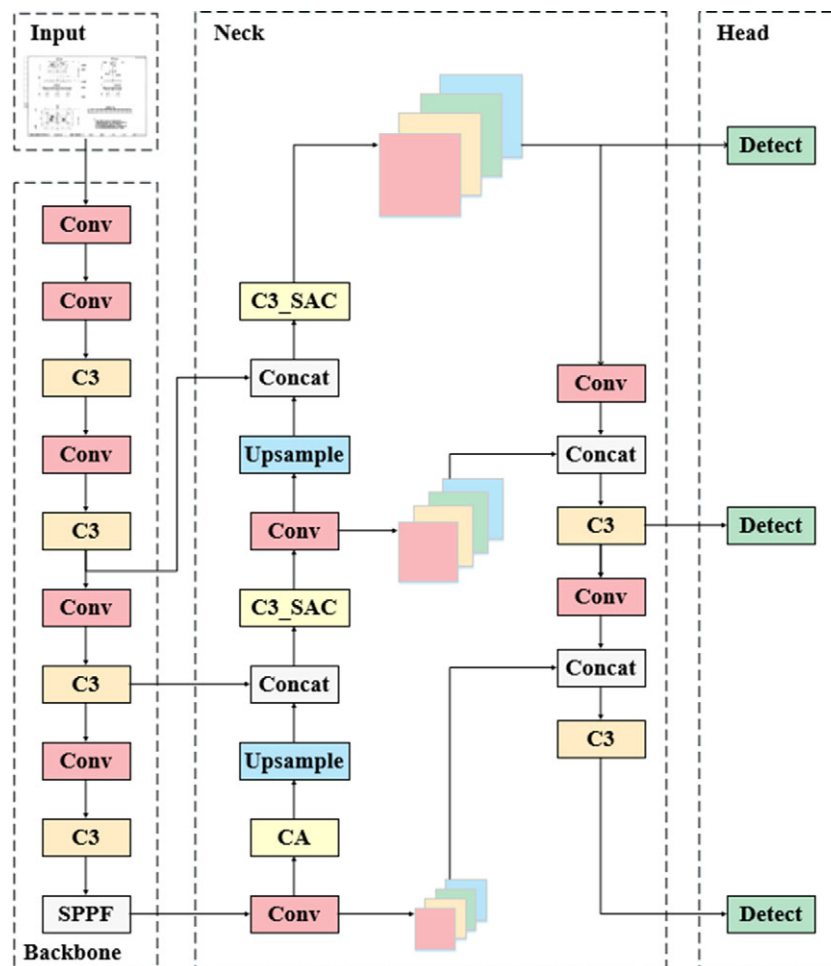


Figure 2. Improved model structure.

Engineering design drawing basic elements detection

Song et al. (2011) stored the topological relationships and geometric constraints information between basic graph elements obtained after the vectorization process in the structure of the graph as a knowledge representation of such graphical objects and used the knowledge representation of graphical objects stored in the structure of the graph for comparison and recognition when recognizing engineering drawings. Liu et al. (2019) proposed a convolutional neural network-based detection architecture to realize the classification of three categories of engineering drawings: electrical engineering drawings, mechanical engineering drawings, and textual drawings. Zhao et al. (2021) used Faster R-CNN to recognize and classify columns and beams in frame maps and proposed an information matching method to enrich the attribute and location information of targets. Zhao et al. (2022) introduced the classical Hough transform technique into deep learning representations and proposed an end-to-end learning framework for line detection, which performs the Hough transform by parametrizing lines with slopes and deviations, transforms the deep representation into the parameter domain, and performs line detection in the parameter domain. Yang et al. (2022) proposed to use the improved Cascade RCNN algorithm combined with digital image processing technology to recognize duct plan drawings, extracting equipment categories and location information in the images, with a recognition accuracy of 80.8%, which provides a data foundation for reconstructing BIM. These previously proposed methods have low recognition accuracy, which makes it difficult to meet practical needs and deal with a single type of graphical element with weak generalization capabilities.

Methodology

Overview of the proposed method

As shown in Figure 2, YOLOv5 consists of four parts: Input, Backbone, Neck, and Head. Input preprocesses the image using Mosaic data enhancement, adaptive initial anchor frame calculation, and image scaling; Backbone uses Focus downsampling, improved Cross Stage Partial network (C.-Y. Wang et al., 2020), and Spatial Pyramid Pooling (He et al., 2015) to extract image feature information; Neck uses Feature Pyramid Network (Lin et al., 2017) and Path Aggregation Network (Liu et al., 2018) to realize the transfer of feature information between targets of different sizes; Head uses Binary Cross Entropy Loss (Zheng et al., 2020) and Complete IoU Loss to compute the classification, localization, and confidence losses, and improve the accuracy of network prediction by Non-Maximum Suppression.

Although YOLOv5 has achieved some research results in general-purpose target detection (J. Wang et al., 2023), its detection accuracy needs to be improved in bridge engineering design drawing basic elements detection scenes. In this paper, two improvements are made on the basis of YOLOv5: firstly, the CA is introduced into the feature extraction network to enhance the feature extraction capability of the algorithm and alleviate the problem of difficult identification of texture features inside gray-scale images; secondly, targeting objectives across different scales, the standard 3×3 convolution in the feature pyramid network is replaced with SAC, which adaptively enlarges the sensory field to capture multi-scale contextual information and improve the anchor frame.

Algorithm Improved Basic Elements Detection Algorithm for Bridge Engineering Design Drawings based on YOLOv5.

Input: Images (I); Labels (L); Learning Rate (η); Number of Iterations (T).

Output: Trained Model Weights (W).

- 1: Preprocess/to obtain normalized and augmented images I' .
- 2: Build the improved YOLOv5 network model by inserting the CA module and the SAC module.
- 3: Define the loss function by weighted summation of the bounding box regression loss L_{bbox} , category loss L_{cls} , and object loss L_{obj} to obtain the total loss L_{total} .
- 4: **for** $t \in [1, T]$ **do**.
- 5: Forward Propagation: Compute predictions \hat{b} , $\hat{\gamma}$, \hat{p} by passing I'_b through the network.
- 6: Compute Loss: Evaluate L_{total} using L_b , \hat{b} , $\hat{\gamma}$, \hat{p} .
- 7: Back Propagation: Compute gradients ∇L_{total} and update W using an optimizer.
- 8: Weight Update.
- 9: **end for**.

Coordinate attention

The core idea of the attention mechanism is to focus on the information that is more critical to the task at hand among the many inputs, ignoring other irrelevant information, and thus acquiring more details relevant to the goal and improving the accuracy of task processing. As shown in Figure 3, in the field of computer vision, depending on the attention focus domain, it can be categorized into channel attention mechanisms represented by Squeeze-and-Excitation Networks (SENet) (Hu et al., 2018), spatial attention mechanism represented by Spatial Transformation Neural Networks (STNs) (Jaderberg et al., 2015), and hybrid attention mechanism represented by the Convolutional Attention Module (CBAM) (Woo et al., 2018).

SENet only considers encoded inter-channel information, ignoring positional information that is critical for capturing the target structure. CBAM attempts to exploit positional information by reducing the channel dimension of the input tensor and then computing spatial attention using convolution, which can only capture local relationships and cannot model the remote dependencies necessary for visual tasks. CA (Hou et al., 2021) is a new attention mechanism module proposed after SENet and CBAM, which embeds position information into channel attention to capture not only cross-channel information but also orientation-aware and position-sensitive information, which helps the model to more accurately localize and identify objects of interest.

The key of CA lies in generating two attention maps, one emphasizing the horizontal direction and the other emphasizing the vertical direction. These two attention maps are typically generated through global average pooling and convolution operations. Subsequently, these two attention maps are multiplied with the original feature map to produce a new weighted feature map. The output of CA can be expressed as:

$$X_{new} = X \cdot \sigma(\text{Conv}(\text{GAP}_h(X))) \otimes \sigma(\text{Conv}(\text{GAP}_v(X))) \quad (1)$$

In this context, X_{new} represents the new feature map after being processed by the CA module, while X is the original feature map. The symbol \cdot denotes the element-wise multiplication operation, which is used to multiply the attention maps with the original feature

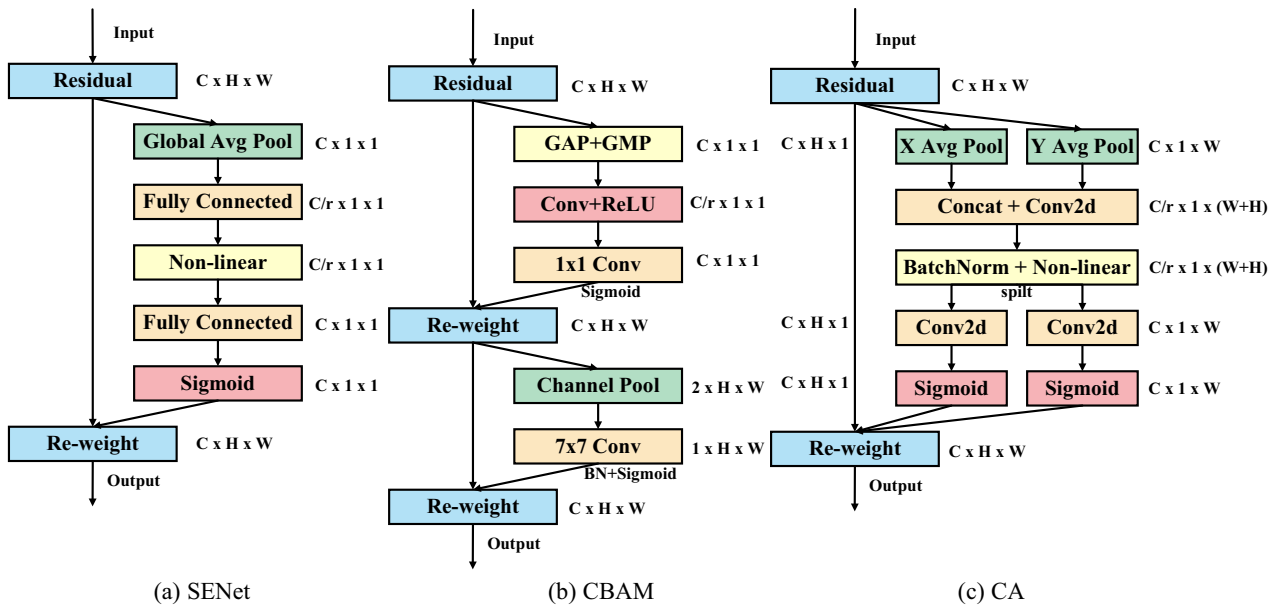


Figure 3. Attention mechanisms.

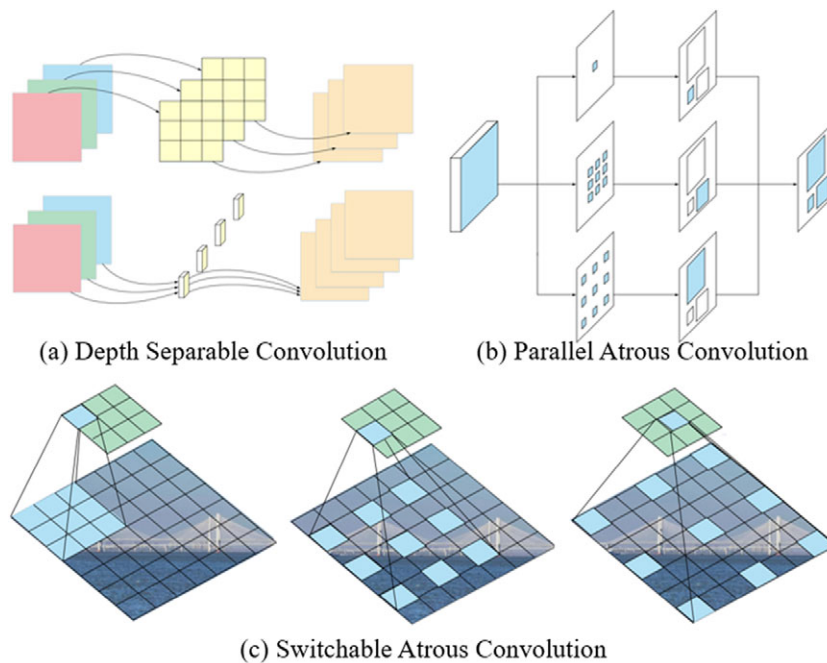


Figure 4. Atrous convolution.

map element by element. The symbol \otimes stands for the tensor multiplication operation, which combines the attention maps in the horizontal and vertical directions. The σ represents the activation function, typically using the Sigmoid function to map the values of the attention maps to a range between 0 and 1. *Conv* indicates the convolution operation, which further processes the results after global average pooling to generate the final attention maps. $GAP_h(X)$ and $GAP_v(X)$ refer to the global average pooling operations applied to the feature map X in the horizontal and vertical directions, respectively, to obtain global statistical information.

In network improvements, we place the CA module between the Conv and Upsample operations, that is, before the feature map is

upsampled and fused. Through this approach, CA is able to directly recalibrate the original feature map, enabling the model to focus more on the positional information crucial to the detection task. For the bridge engineering drawing detection task, as the basic elements in drawings usually have clear positions and directionality, CA can significantly enhance the model's attention to these key areas, thereby improving the accuracy of detection.

Switchable atrous convolution

Atrous convolution refers to the process of expanding the convolution kernel by adding some zeros between the elements of the

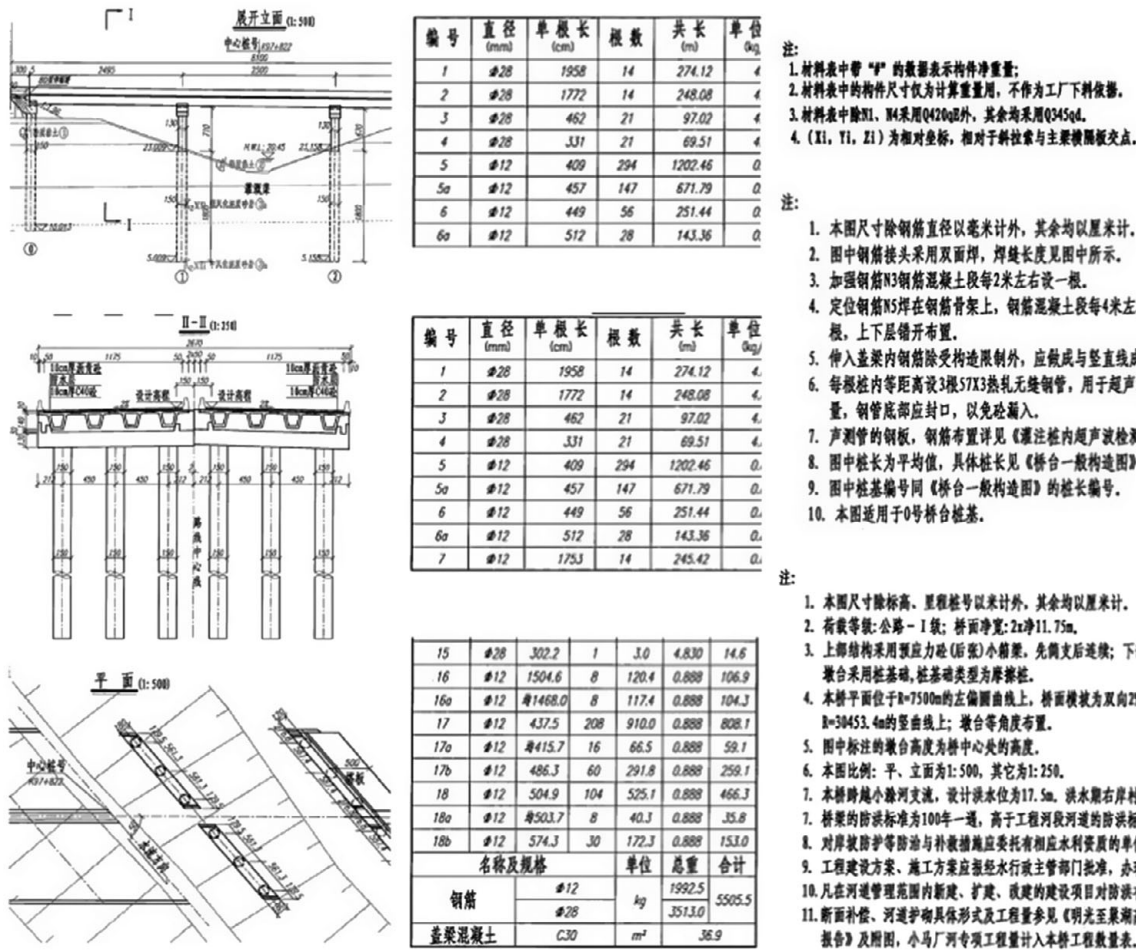


Figure 5. Example of a sample data set.

convolution kernel, which is mainly used to expand the sensory field so that the output of each convolution contains a larger range of information. Meanwhile, atrous convolution suffers from two shortcomings: on the one hand, there is the Gridding Effect, which means that when stacking multiple 3×3 convolution kernels with an atrous rate of 2, the convolution kernels are not continuous and not all input pixels are computed; on the other hand, it is effective for large targets, but not so effective for small targets.

As shown in Figure 4, it was found that researchers have made many improvements to the atrous convolution based on the above two shortcomings: Depth Separable Convolution, which mainly replaces the traditional convolution with channel-by-channel and point-by-point convolution, is not applicable to grayscale bridge engineering design drawings that lack colorful semantic information representations; Parallel Atrous Convolution, by artificially defining three branches with different atrous rates, so that the branch with a small sensory field trains a small-scale target and the branch with a large sensory field trains a large-scale target (Son et al., 2021); SAC, which is capable of obtaining different values of the switching function $S(x)$ according to the inputs and positions and adaptively choosing whether the atrous rate is 1 or 3 according to the value of $S(x)$ (Qiao et al., 2021).

The core of SAC lies in its adaptive atrous rate switching function, which allows the network to select an appropriate atrous

rate based on the characteristics of the input data. Through global average pooling and convolution operations, corresponding feature maps are generated. Subsequently, these feature maps are integrated through a fusion mechanism to form the final feature representation. The output of SAC can be expressed as:

$$Conv(x, w, 1) \rightarrow S(x) \cdot Conv(x, w, 1) + (1 - S(x)) \cdot Conv(x, w + \Delta w, r) \quad (2)$$

In this context, x is the input, w is the weight, $S()$ is the switching function, Δw denotes a weight with trainable weights, and r is the atrous rate.

In network improvement, we replaced the first two C3 modules in the Neck section with SAC modules. This allows SAC to dynamically adjust the atrous rate based on task requirements, expanding the receptive field of the convolutional layer. This means that each convolutional kernel can cover a wider input region, capturing more contextual information. At the same time, the combination of SAC with FPN and PANet in the YOLOv5 network further promotes the fusion of multiscale features. Bridge engineering drawings usually have clear structures, explicit contextual information, and different scales. The flexibility of SAC enables it to adjust the atrous rate based on different target scales, thus achieving effective detection of multi-scale targets.

安徽安文工程勘察设计有限公司
 设计
 审核
 一审
 二审
 安徽安文工程勘察设计有限公司
 设计
 审核
 一审
 二审

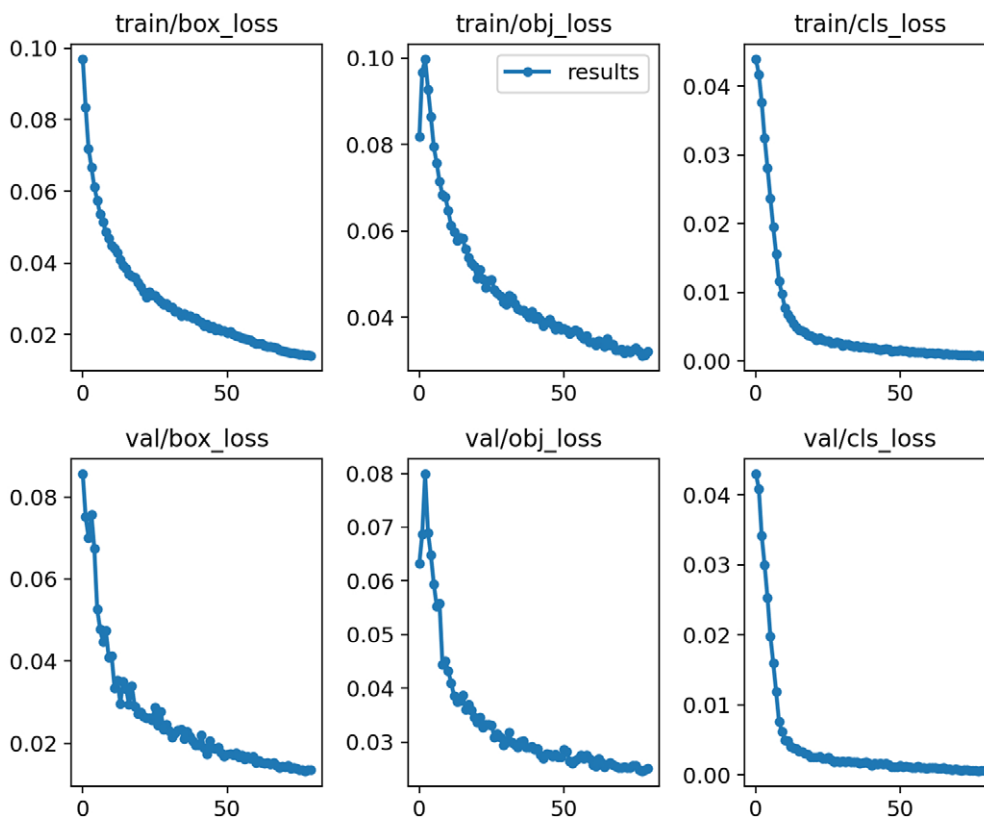


Figure 6. Convergence of improved YOLOv5 training loss.

Table 1. Comparison of results of multiple detection algorithms

Algorithm	P					R					mAP0.5	mAP0.5:0.9
	figure	ann	textbox	form	all	figure	ann	textbox	form	all		
Faster R-CNN	0.726	0.68	0.686	0.804	0.747	0.693	0.654	0.728	0.74	0.711	0.75	0.691
SSD	0.743	0.722	0.648	0.782	0.735	0.673	0.665	0.641	0.719	0.667	0.75	0.649
YOLOv5	0.853	0.92	0.929	0.925	0.909	0.876	0.95	0.95	0.94	0.926	0.902	0.812
Ours	0.863	0.87	0.94	0.925	0.894	0.933	0.88	0.95	0.995	0.939	0.936	0.823

Experiments

Experimental environment

The operating system for the experiments in this paper is 64-bit Ubuntu 20.04 LTS, the graphics card is NVIDIA GeForce RTX 3090, the deep learning framework is Pytorch 1.11.0, and the programming language is Python 3.8.

Data sets

There is a lack of publicly available datasets in the field of bridge engineering design drawing recognition due to intellectual property rights restrictions. To solve this problem, this paper creates a self-developed bridge engineering design drawing basic elements detection dataset. The data mainly comes from the scanned documents of Anhui Provincial General Research Institute of Transportation Planning and Design, covering images of different design scenes,

different design styles, and different resolutions, and data enhancement is performed by adding noise, and a total of 3,000 images are finally obtained. The collected data was labeled using LabelImg in YOLO format for four categories: Figure, Form, Annotation, and Text box. The dataset is divided into the training set and test set in a 9:1 ratio. The sample dataset is shown in Figure 5.

Network training

In the YOLOv5 model training, the smaller the loss function loss value of the model structure, the better, with an expected value of 0. To achieve the best performance of the model, the initial learning rate was set to 0.01, the momentum to 0.937, the weight decay coefficient to 0.0005, the number of iterations to 100, and the batch size to 8. After 80 iterations, the loss values stabilize, and the model reaches the optimal state. The training loss variation is shown in Figure 6.

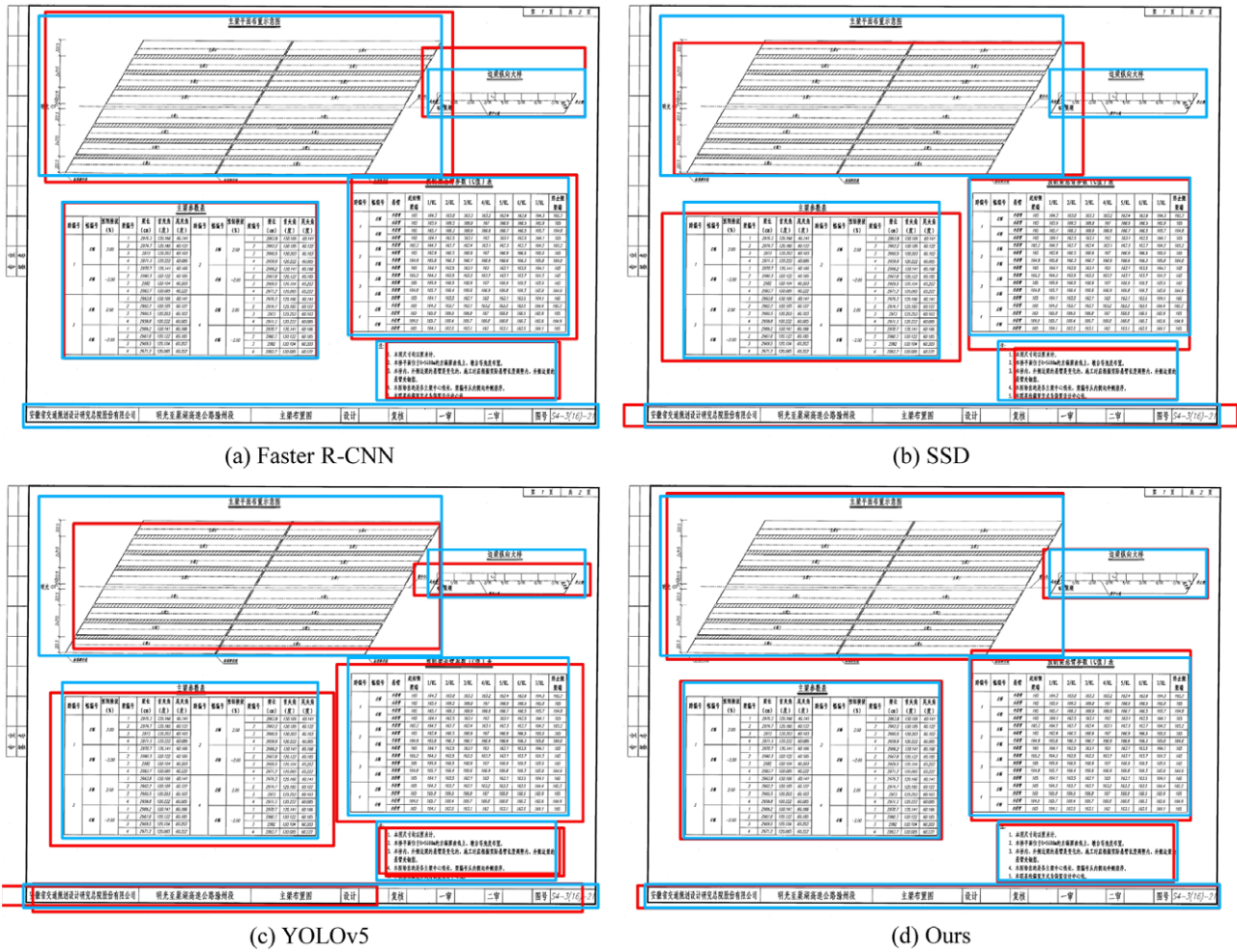


Figure 7. Comparison of the results of different detection algorithms. Red represents the detection result, and blue represents the ground truth.

Evaluation indicators

In this paper, Precision (P), Recall (R), and mean Average Precision (mAP) are used as evaluation metrics. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$AP = \frac{1}{M} \sum P(R) \tag{5}$$

$$mAP = \frac{1}{M} \sum AP_i \tag{6}$$

where TP denotes the number of detection frames that are correctly predicted, FP denotes the number of detection frames that are incorrectly predicted, FN denotes the number of detection frames that are missed, AP denotes the average precision, M denotes the number of categories, P(R) denotes the accuracy P corresponding to a different recall rate R, and AP_i denotes the average precision of the ith iteration. mAP0.5 denotes the mAP when the IoU threshold

is set to 0.5, and mAP0.5:0.9 denotes the average mAP over different IoU thresholds from 0.5 to 0.9 in steps of 0.05.

Results and discussion

In order to verify that the algorithm proposed in this paper has better results, it is experimentally compared with classical target detection algorithms such as Faster R-CNN, SSD, and YOLOv5 under the same configuration conditions. The specific experimental results are shown in Table 1.

As is shown in Table 1, the algorithm in this paper reduces the accuracy by 1.5% and improves the recall by 1.3% over the original YOLOv5 algorithm, improves the mean average precision by 3.4% when the threshold is set to 0.5, and improves the mAP0.5:0.9 by 1.1%. In particular, for the category of text boxes with long dimensions, the accuracy is improved by 1.1%, and for the categories of figures and tables, where internal texture features are difficult to distinguish, the recall is improved by 5.7% and 5.5%, respectively. The evaluation indexes of the detection results of this paper's algorithm are much higher than those of Faster R-CNN and SSD, which indicates that the performance and reliability of the algorithm proposed in this paper are stronger and can meet the accuracy requirements for the detection of basic elements of bridge engineering design drawings.

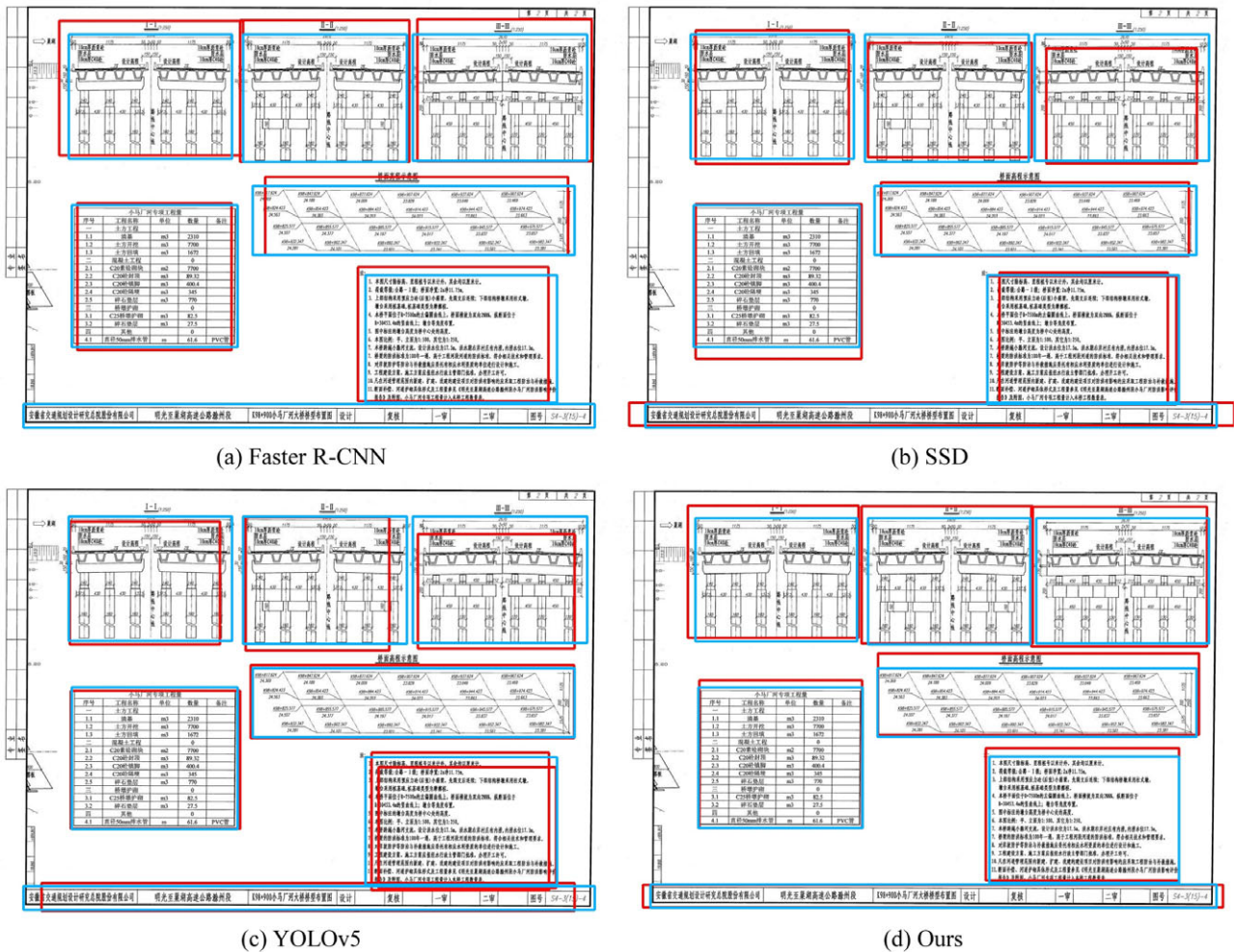


Figure 8. Comparison of the results of different detection algorithms. Red represents the detection result, and blue represents the ground truth.

Table 2. Comparison of detection results of multiple attention mechanisms

Algorithm	P	R	mAP0.5	mAP0.5:0.9
YOLOv5	0.909	0.926	0.902	0.812
YOLOv5 + SENet	0.885	0.914	0.902	0.721
YOLOv5 + CBAM	0.912	0.926	0.941	0.793
YOLOv5 + CA	0.928	0.96	0.921	0.818

Table 3. YOLOv5 ablation experiment results

YOLOv5	CA	SAC	P	R	mAP0.5	mAP0.5:0.9
✓			0.909	0.926	0.902	0.812
✓	✓		0.928	0.96	0.921	0.818
✓	✓	✓	0.894	0.939	0.936	0.823

To more intuitively see the detection differences between different algorithms, some of the detected images are selected for demonstration, and the results are shown in Figure 7. Figure 7(a) shows the detection using the Faster R-CNN algorithm, and it is observed that the text box at the bottom of the drawing, which contains a lot of design information, is not detected, and leakage occurs; Figure 7

(b) shows the detection using the SSD algorithm, and it is observed that a graph in the upper right of the drawing is missed; Figure 7 (c) shows the detection using the YOLOv5 algorithm, and it is observed that two different detection results are obtained for an annotation detection target, and misdetection occurs, which is due to the fact that compared to the RGB three-channel color image, the bridge engineering design drawings are missing an effective representation of the semantic information of color between the channels, which results in a less obvious differentiation between the external contour of the graphic and the internal texture and a lower detection accuracy. At the same time, the YOLOv5 algorithm for the text box also only recognizes a part of the text box and does not recognize the whole, this is due to the core idea of YOLOv5 is to divide the whole image into a number of grids and use a number of anchor frames within the grids for prediction, and the text box is a long target that spans across a number of grids, which makes the selection of the anchor frames and the overall recognition of the image difficult; and Figure 7(d) shows the detection using the algorithm proposed in this paper, and it is observed that for figures, forms, annotations, and text boxes are accurately recognized, and the detection results are closest to the real labels. Similarly, Figure 8(a) shows the detection using the Faster R-CNN algorithm, which misses the text box at the bottom of the drawing; Figure 8(b) shows the detection using the SSD algorithm, which mistakenly detects the category of annotations; Figure 8(c) shows the detection using the YOLOv5 algorithm, which

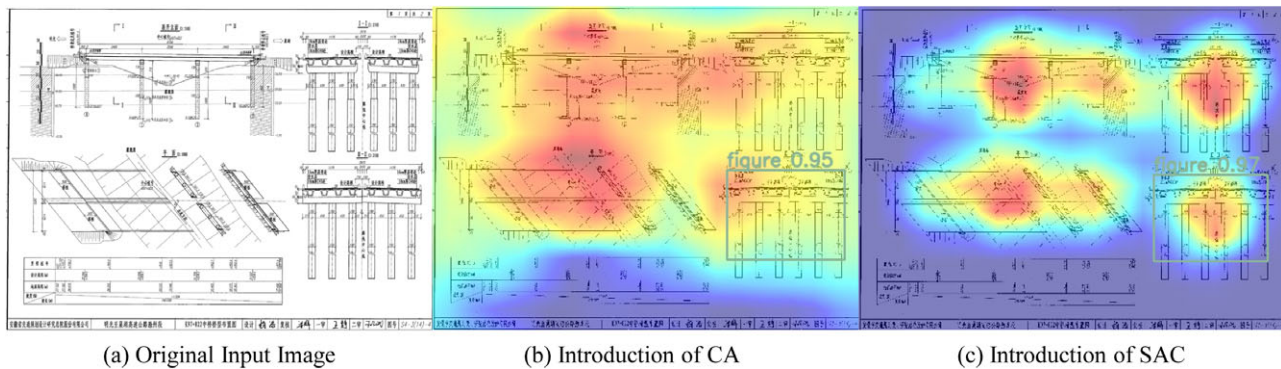


Figure 9. Feature map visualization.

does not accurately detect the figure and suffers from information loss; and Figure 8(d) shows the detection using the improved algorithm of the present paper, which is most closely aligned with the real label. From the above comparison of the detection results of different algorithms, it can be seen that the improved YOLOv5 algorithm proposed in this paper has a better detection effect in the detection scenario of basic elements of bridge engineering design drawings.

In addition, in order to verify that the CA has a better effect than the channel attention mechanism and the hybrid attention mechanism, a comparison experiment was conducted under the same configuration conditions. The specific experimental results are shown in Table 2.

As is shown in Table 2, after the introduction of SENet, the evaluation indexes of drawing recognition not only did not improve but also declined, which is due to the characteristics of bridge engineering design drawings belonging to the grayscale image, the color of which is dominated by black and white, and a lack of inter-channel color semantic information, and the channel attention mechanism only considers the encoding of inter-channel information, which leads to a reduction in the detection accuracy. The introduction of CBAM resulted in a 3.9% improvement in mAP0.5, which is related to the ability of CBAM to capture localized positional information. After the introduction of CA, all evaluation indexes are significantly improved, especially for the recall rate, which is significantly improved by 3.4%, thanks to the fact that CA can capture the characteristics of direction perception and location information. This shows that CA is more suitable for the detection of drawing basic elements than the channel attention mechanism and the hybrid attention mechanism.

In order to verify the improvement effect of YOLOv5 in this paper, CA and SAC are added sequentially to the original YOLOv5, respectively, while keeping the same experimental configuration to judge the effectiveness of each improvement point, and the specific experimental results are shown in Table 3.

As is shown in Table 3, after the introduction of CA, the overall evaluation indexes are improved, especially the P and R, which are improved by 1.9% and 3.4%, respectively, indicating that the attention mechanism improves the feature extraction capability. After the introduction of SAC, the P and R decrease, and the mAP0.5 obtains a further improvement of 1.5%, indicating that the SAC strengthens the multi-scale detection capability.

To have a more intuitive understanding of how much attention the model pays to different targets and to judge whether the network learns the right features or information, the feature map is visualized as GradCAM (Gradient-weighted Class Activation

Mapping) (Selvaraju et al., 2017), and the results are shown in Figure 9. Figure 8(a) shows the original image entered at the time of prediction, which contains four figures, one form, and one text box; Figure 8(b) shows the heat map drawn after the introduction of CA, where blue represents low attention and red represents high attention, and the darker the color, the greater the degree of correlation, it is reasonable to observe that the model mainly relies on the features of the two figures in the upper left to determine the detection target as a figure; and Figure 8(c) shows the heat map of the model after the introduction of SAC, and it is observed that the model determines the detection target to be a figure based on the features of all figures, the feature range is expanded, and the attention is more accurate and focused. This shows that the model training is able to learn the correct features, and the introduction of CA and SAC can improve the feature extraction capability.

However, this study also has certain limitations. Firstly, the calculation speed of the model needs to be improved, for example, by drawing on the parallel processing capabilities of neuromorphic computing (Ji et al., 2022; Ji et al., 2023; Dong et al., 2023b) to design deep learning models that can more efficiently utilize multi-core, GPU, and other hardware resources, thereby accelerating the processing speed of object detection tasks. Additionally, the generalization ability of the model needs to be enhanced, which may require expanding a larger dataset to meet the requirements of recognizing engineering drawings with diverse design styles.

Conclusion

This paper proposes an improved algorithm for detecting basic elements in bridge engineering design drawings based on the YOLOv5 framework. The core of this method lies in the introduction of CA, which significantly enhances the algorithm's feature extraction capabilities. Additionally, by incorporating SAC, the receptive field is adaptively expanded, enabling the capture of crucial multi-scale contextual information for precise detection.

To validate the effectiveness of our proposed method, extensive experiments were conducted on a self-constructed dataset containing 3,000 bridge engineering design drawings. The experimental results are particularly noteworthy, demonstrating substantial improvements. Specifically, our improved algorithm achieves a mAP of 93.6% when the IoU threshold is set to 0.5, representing a 3.4% improvement compared to the baseline YOLOv5 algorithm. This significant performance enhancement not only highlights our method's advantage in accurately detecting basic elements but also reveals its broad application potential in the field of engineering design.

This research not only provides more efficient, accurate, and innovative methods for bridge design and construction but also demonstrates the immense potential and broad prospects of artificial intelligence in engineering design. We firmly believe that this study represents a significant step forward for artificial intelligence in engineering design, opening new directions and possibilities for future research and applications.

Data availability statement. Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments. This research is supported by the Key Research and Development Program of Anhui Province. (202304A05020063).

References

- Brock, A, Lim, T, Ritchie, JM and Weston, N.** 2017. Convnet-based optical recognition for engineering drawings. In *37th Computers and Information in Engineering Conference*, DETC2017-68186. American Society of Mechanical Engineers.
- Dong, Z, Ji, X, Wang, X, Gu, Y, Wang, J and Qi, D.** 2023a. Icncs: Internal cascaded neuromorphic computing system for fast electric vehicle state of charge estimation. *IEEE Transactions on Consumer Electronics*.
- Dong Z, Zhao Y and Tian FF** (2023b) Character recognition and detection algorithm for power grid engineering drawings based on improved convolutional neural network. *Electronic Design Engineering* 31(13), 27–31
- Elyan E, Jamieson L and Ali-Gombe A** (2020) Deep learning for symbols detection and classification in engineering drawings. *Neural Networks* 129, 91–102
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM and Thrun S** (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639), 115–118
- Fan F and Guan JH** (2012) Engineering drawing string and labeling information extraction. *Computer Engineering and Application* 48(7), 161–164
- He K, Zhang X, Ren S and Sun J** (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(9), 1904–1916
- Hou, Q, Zhou, D and Feng, J.** 2021. Coordinate attention for efficient mobile network design. In *Proceedings of the Ieee/Cvf Conference on Computer Vision and Pattern Recognition*, 13713–13722.
- Hu, J, Shen, L and Sun, G.** 2018. Squeeze-and-excitation networks. In *Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition*, 7132–7141.
- Jaderberg, M, Simonyan, K, Zisserman, A, et al.** 2015. Spatial transformer networks. *Advances in Neural Information Processing Systems* 28.
- Jamieson, L, Moreno-Garcia, CF and Elyan, E.** 2020. Deep learning for text detection and recognition in complex engineering diagrams. In *2020 International Joint Conference on Neural Networks (Ijcnnc)*, 1–7. IEEE.
- Ji, X, Dong, Z, Han, Y, Lai, CS and Qi, D.** 2023. A brain-inspired hierarchical interactive in-memory computing system and its application in video sentiment analysis. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Ji X, Dong Z, Lai CS and Qi D** (2022) A brain-inspired in-memory computing system for neuronal communication via memristive circuits. *IEEE Communications Magazine* 60(1), 100–106
- Lin, T-Y, Dollár, P, Girshick, R, He, K, Hariharan, B and Belongie, S.** 2017. Feature pyramid networks for object detection. In *Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition*, 2117–2125.
- Liu, L, Chen, Y and Liu, X.** 2019. Engineering drawing recognition model with convolutional neural network. *Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI 2019)*, 112–116.
- Liu, S, Qi, L, Qin, H, Shi, J and Jia, J.** 2018. Path aggregation network for instance segmentation. In *Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition*, 8759–8768.
- Nabipour M, Nayyeri P, Jabani H, Mosavi A, Salwana E and Shahab S** (2020) Deep learning for stock market prediction. *Entropy* 22(8), 840
- Qiao, S, Chen, L-C and Yuille, A.** 2021. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *Proceedings of the Ieee/Cvf Conference on Computer Vision and Pattern Recognition*, 10213–10224.
- Selvaraju, RR, Cogswell, M, Das, A, Vedantam, R, Parikh, D and Batra, D.** 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the Ieee International Conference on Computer Vision*, 618–626.
- Son, H, Lee, J, Cho, S and Lee, S.** 2021. Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In *Proceedings of the Ieee/Cvf International Conference on Computer Vision*, 2642–2650.
- Song X, Li YC and Liu JF** (2011) Topology-based engineering drawing recognition method. *Journal of Shenyang University of Architecture (Natural Science Edition)* 27(4), 6
- Wang, C-Y, Mark Liao, H-Y, Wu, Y-H, Chen, P-Y, Hsieh, J-W and Yeh, I-H.** 2020. Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the Ieee/Cvf Conference on Computer Vision and Pattern Recognition Workshops*, 390–391.
- Wang J, Chen Y, Dong Z and Gao M** (2023) Improved yolov5 network for real-time multi-scale traffic sign detection. *Neural Computing and Applications* 35 (10), 7853–7865
- Woo, S, Park, J, Lee, J-Y and Kweon, I-S.** 2018. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (Eccv)*, 3–19.
- Yang M, Zhao Y and Deng X** (2022) Two-dimensional drawing recognition of duct planes based on improved cascade rcnn. *Journal of Civil Engineering and Management* 4, 39
- Zhang S, Yao L, Sun A and Tay Y** (2019) Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)* 52 (1), 1–38
- Zhao K, Han Q, Zhang C-B, Xu J and Cheng M-M** (2022) Deep hough transform for semantic line detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 4793–4806
- Zhao Y, Deng X and Lai H** (2021) Reconstructing bim from 2d structural drawings for existing buildings. *Automation in Construction* 128, 103750
- Zheng, Z, Wang, P, Liu, W, Li, Y, Ye, R and Ren, D.** 2020. Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the Aaai Conference on Artificial Intelligence*, 34:12993–13000.