Journal of Computer Technology and Software
ISSN:2998-2383
Vol. 1, No. 3, 2022

Enhanced Character Detection Using Coordinate Attention

Orion Kessler

Department of Computer Science, University of Colorado Boulder, USA

orion.kessler47@colorado.edu

Abstract: Angle steel is crucial in constructing the national power grid, particularly in transmission line towers. Accurate identification and recording of angle steel types are vital to prevent errors in construction processes. Traditional manual methods are prone to mistakes, necessitating an automated solution. This paper presents an improved angle steel character detection algorithm utilizing DBNet and MobileNetV3, enhanced by Coordinate Attention (CA) to replace the original attention mechanism. By collecting a specialized angle steel character dataset and implementing data augmentation and learning rate strategies, the proposed method addresses the challenges of diverse character shapes and complex backgrounds. Experimental results demonstrate that the improved algorithm significantly outperforms existing methods, achieving a remarkable increase in the F1-Score from 81.42% to 99.06%. The study validates the effectiveness of the proposed enhancements in industrial settings and lays the groundwork for further advancements in angle steel character detection.

Keywords: Angle Steel Characters, Attention Mechanisms, Multiscale-Scale Feature Fusion, Data Augmentation.

1. Introduction

Angle steel is one of the very important materials in the construction of the national power grid, and is widely used in the structure of transmission line towers [1]. The steel stamp character of the angle steel is the identity mark of the angle steel, and its main function is to distinguish the model of the angle steel material. A complete transmission tower needs to be composed of different types of angle steel, and the required quantity of different types of angle steel is also different, and in the process, the operator needs to manually identify and record and combine the characters, which is easy to cause mistakes.

With the rapid development of image recognition technology, text detection algorithms based on deep learning have been widely proposed. The realization principle of the text detection algorithm is to first use the target detection framework to detect part of the text area, and then use the post-processing method to obtain the complete text area [2]. Text detection algorithms have been widely used in finance, transportation, logistics and many other fields, but in the industrial field, they have not been fully applied. The main reason is that there is no public dataset of industrial characters, the characters are diverse in shape, and the background is complex [3]. Based on the above problems, this paper collects the angle steel character data set, combines the currently widely used DBNet to study the angle steel characters, uses MobileNetV3[4] as the feature extraction network, and introduces Coordinate Attention (CA) [5] to replace the original attention mechanism , improve multi-scale feature fusion, and finally use data augmentation and learning rate strategy to optimize the algorithm. In order to prove the effectiveness of the above improvements, the experiment compares the improvement points through the ablation method, and compares the common detection algorithms also based on MobileNetV3. The final index and effect observation show that the proposed improved method is significantly improved compared with the original algorithm, which meets the requirements of industrial angle steel character detection accuracy and provides a reference for industrial angle steel character detection methods.

2. DBNet Network Architecture

At present, there are three main versions of DBNet algorithm, DBNet_Res50, DBNet_Res18 and DBNet_Mov3. The main difference between the three versions is that the backbone network is different, and ResNet50, ResNet18 and MobileNetV3 are used as the backbone network respectively. The algorithm in this paper is improved based on MobileNetV3. The network structure of DBNet algorithm consists of three parts: Backbone, Neck and Head. The DBNet network structure is shown in Fig. 1.

The Backbone part adopts the image classification network, which is responsible for extracting the multi-scale features of the image, mainly including the CBH convolution module (convolution layer Conv, normalization layer BN and activation function Hard-Swish), BRC deconvolution module (normalization layer BN, activation function ReLU and deconvolution layer ConvTranspose), Res residual module and Res_SE residual module with SE attention mechanism (Squeeze-and-Excitation) [9]. The Neck part is located between Backbone and Head. The feature pyramid structure FPN is used to enhance the image features through multi-scale feature fusion. The four feature maps output by Backbone are enhanced and then spliced (Concat), and finally the features output by Neck The height and width of the image are one-fourth of the original image. The head part first goes through a convolutional layer, then deconvolutions twice, maps the FPN feature output by Neck from the quarter size of the original image to the original

image size, and finally activates the output through Sigmoid

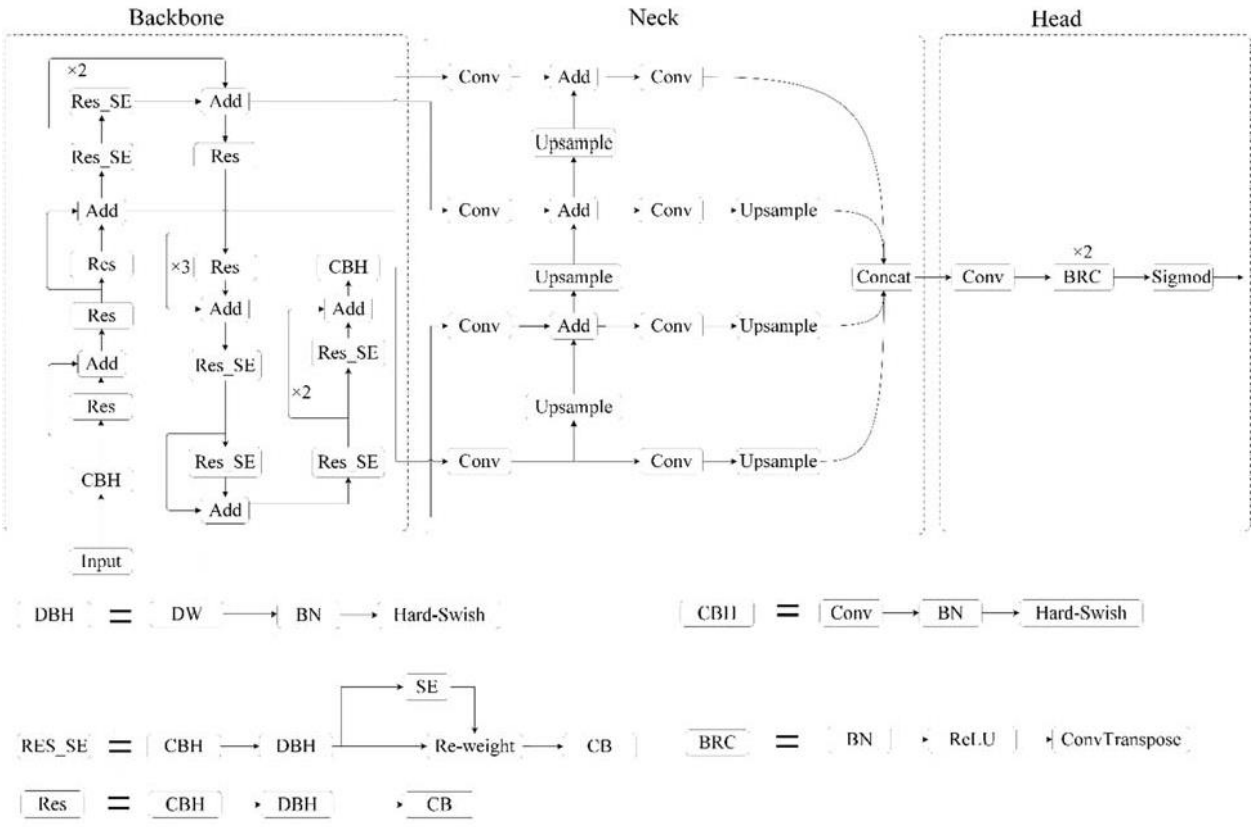


Figure 1. DBNet network structure

3. Improved DBNet Network

This paper makes three improvements to the DBNet algorithm: the Backbone part replaces the SE attention mechanism in the residual module with the new CA attention mechanism; the Neck part uses the weighted bidirectional feature pyramid (BiFPN) to improve the FPN; except for the network structure part, other non- The network structure has also been optimized. The data part is introduced into Copy-Paste [10] for data augmentation to improve the generalization of the model. The hyperparameter part adopts the Cosine learning rate reduction strategy and the learning rate warm-up strategy to ensure training. Stability in the initial stages and when the model converges. The improved DBNet network structure is shown in Fig.2.

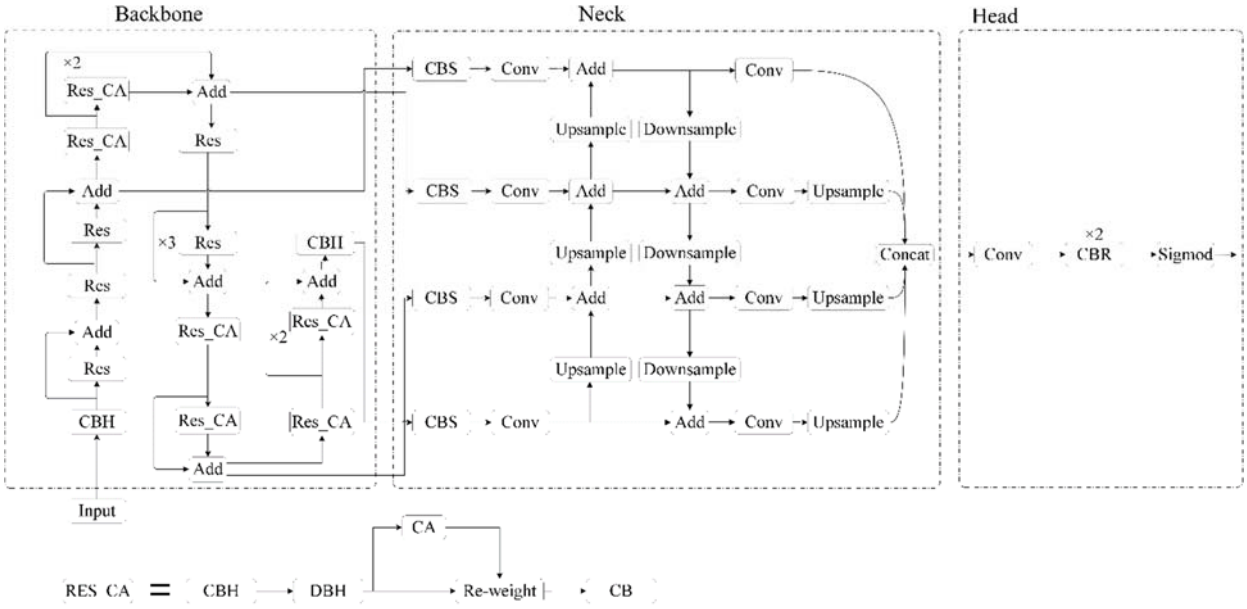
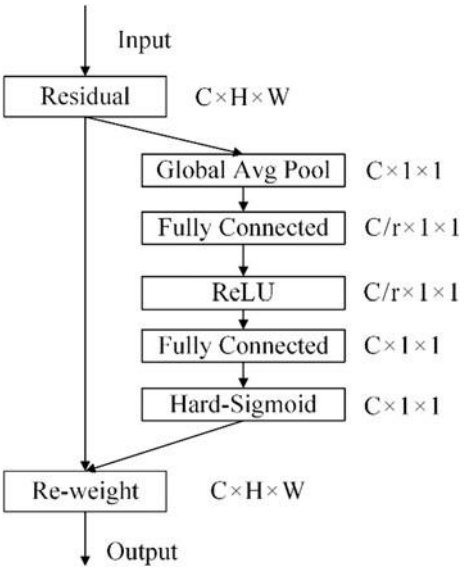


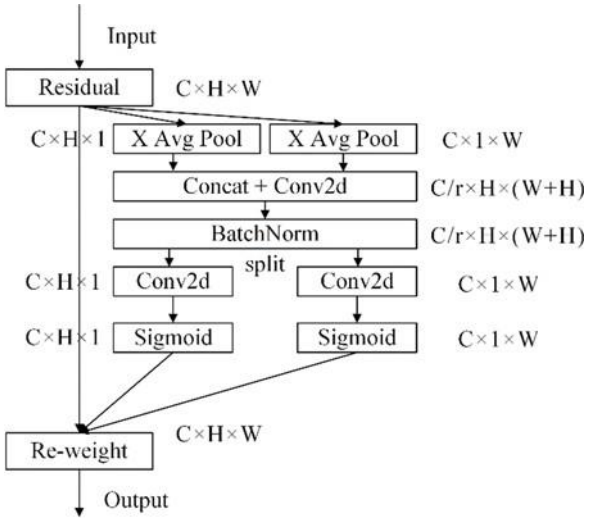
Figure 2. Improved DBNet network structure

3.1. Residual module introduces a new attention mechanism

In this paper, the SE attention mechanism of the residual module in the backbone network is changed to a novel CA attention mechanism. The CA attention mechanism can make the lightweight network MobileNetV3 obtain information of a larger area without introducing large overhead by embedding the position information into the channel attention. The structure comparison of the attention mechanism is shown in Fig. 3. The SE module in Fig. 3 is divided into two steps in structure, compression and excitation, which are respectively designed for global information embedding and adaptive recalibration of channel relationships. The specific structure is shown in Fig. 3 (a), using global average pooling (Global Avg Pool) as a compression operation, followed by a fully connected layer (FC) layer to reduce the feature dimension by r times, and then activated by the ReLU [14] function and then raised back to the original feature dimension through an FC layer, Then it is converted into a normalized weight of 0~1 through the Hard-Sigmoid function, which is an excitation operation. Finally, the newly generated feature map and the original feature map are multiplied and calibrated, that is, the Re-weight operation; the CA module encodes the channel relationship and long-range dependency through the spatial domain. The specific structure is shown in Fig. 3 (b). First, along the X The average pooling (Avg Pool) operation is performed in the direction and the Y direction respectively, and two feature maps based on X and Y are obtained, and then the concatenation (Concat) and convolution operations are performed to establish a long-range dependency, and then the BN operation is performed, and then The feature map with global information is re-divided into feature maps based on X and Y, and convolution and sigmoid normalization are performed respectively, and finally the position information along X and Y is weighted on the original feature map. In Fig. 3, "C" represents the channel, "r" represents the multiple of channel reduction, and "H" and "W" represent height and width, respectively.



(a) SE attention mechanism structure



(b) CA attention mechanism structure

Figure 3. Attention mechanism structure comparison

3.2. Introducing BiFPN structure to improve multi-scale feature fusion

The BiFPN introduced in this paper is borrowed from the feature enhancement network in EfficientDet [15]. Compared with the FPN structure, in addition to the bottom-up feature fusion, the top-down feature fusion is added, and the introduction between the input node and the output node of the same scale is introduced. A skip connection can fuse more angle steel character features and improve the ability of network feature enhancement. The BiFPN network structure is shown in Fig. 4. This structure is different from the original BiFPN structure, and is changed to a four-layer structure in order to adapt to the output of the backbone network.

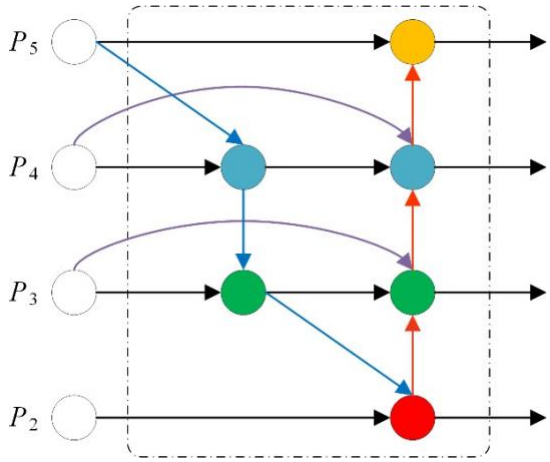


Figure 4. BiFPN network structure

3.3. Non-network structure part algorithm optimization

3.3.1. Copy-Paste data enhancement

Due to the single position of the text area in the collected angle steel data set images, this paper introduces the Copy-Paste method to optimize the model for data enhancement, by randomly selecting two training images, such as (a) and (b) in Fig. 5, to perform random jitter scaling and random Flip horizontally, select one of the images at random, and paste the text area in the image at a random location in the other image. It can better improve the richness of the training set and

enhance the robustness of the model to the background. The Copy-Paste method demonstration is shown in Fig. 5.

(a)Copy-Paste (b) Img2 (c)Img1

Figure 5. Copy-Paste method demonstration

3.3.2. Cosine learning rate descent strategy

In this paper, the Cosine learning rate drop strategy and the learning rate warm-up strategy are used to control the learning rate parameters in training. In the early stage of training, a smaller learning rate is used to train the model through the learning rate warm-up strategy, and the learning rate is gradually increased to the specified Epoch; as the model gradually converges, the Cosine learning rate reduction strategy is used in the later stage of training to gradually increase the learning rate. Falling to 0, the falling state is a Cosine curve. This strategy can better improve the stability in the early and late stages of training, and effectively improve the convergence accuracy of the final model.

4. Experiment and Result Analysis

4.1. Data set creation

To realize the improved angle steel character detection algorithm based on DBNet, the real industrial angle steel image is required to be used as a data set by marking the character area for algorithm training and verification. The data set in this paper is collected by using industrial cameras to take pictures at fixed points in an industrial environment, and it is marked by Baidu's open source tool PPOCRLabel. The angle steel dataset consists of 500 images in total, of which 400 are used for training and 100 are used for algorithm performance evaluation. Each angle steel character area frame is manually marked, and there is an average of about 1 character area frame in each angle steel image. The information statistics of angle steel character detection dataset are shown in Table 1.

Table 1. Angle steel character detection data set information statistics

Dataset classification	Number of images	Number of character area boxes
training set	400	471
validation set	100	105
total	500	576

4.2. Experimental platform and training details

The experimental environment is Ubuntu SMP 16.04.1 operating system, 1 4-core Intel(R) Xeon(R) Gold 6148 @ 2.40GHz processor, 1 32G NVIDIA V100 Tensor Core graphics card, deep learning framework PaddlePaddle 2.2.2 version, The code running environment is Python 3.7.The angle steel data set collected in this experiment is relatively small, so this paper uses the transfer learning method [17], and uses the corresponding classification network trained on ImageNet as a pre-training model to speed up the training. This experiment is trained for 24,000 iterations, using the Adam optimizer, the parameters bata1 is 0.9, bata2 is 0.999, and the learning rate is set to 0.001. The training image input is adjusted to a size of 960 × 960 pixel.

4.3. Backbone network comparison experiment

For this angle steel data set, in order to select a suitable classification network as the backbone network, this experiment is based on the DBNet model to train three classification networks, namely MobileNetV3, ResNet18 and ResNet50. The comparison of the training loss of different backbone networks is shown in Fig. 6. It can be seen that MobileNetV3 and ResNet18 with shallow networks converge faster in DBNet, and the convergence is not much different.

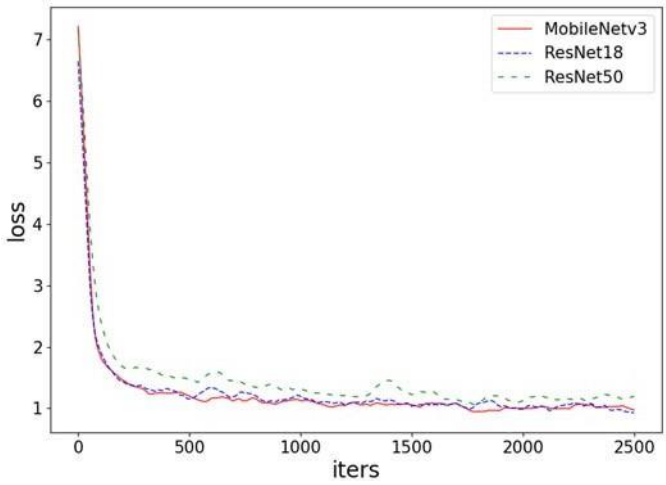


Figure 6. Comparison of training losses of different backbone networks

In addition to the analysis of the loss situation, the most important thing is the performance indicators of different classification networks in the DBNet algorithm, which are evaluated using the angle steel data set. The experimental comparison results of different backbone networks are shown in Table 2. It can be seen from Table 2 that MobileNetV3 is lighter than ResNet18 under the condition of similar convergence speed, and the evaluation index is higher, so MobileNetV3 is used as the backbone network in the experiments later in this paper.

Table 2. Comparison experiment of different backbone networks

Backbone	Precision	Rcall	F1-Score	Params(M)
MobileNetV3	76.03	87.62	81.42	5.70
ResNet18	71.85	92.38	80.83	47.20
ResNet50	77.57	79.05	78.30	97.20

4.4. Ablation experiment

In order to verify the effectiveness of the algorithm improvement and the importance of each improvement point to the algorithm, this paper conducts ablation experiments on the improvement points. The results of the algorithm improvement ablation experiments are shown in Table 3. Among them, strategy A is to replace the SE attention mechanism in the backbone network with the CA attention mechanism; strategy B is to introduce the BiFPN structure to improve the multi-scale feature network structure; strategy C is to improve the non-network structure part, and perform Copy-Paste data enhancement and learning rate optimization The strategy optimizes the algorithm.

Table 3. Ablation experiment

Strategy	Precision	Rcall	F1-Score	Params(M)
DBNet	76.03	87.62	81.42	5.70
DBNet+A+B	80.67	91.43	85.71	6.90
DBNet+A+C	92.92	100.00	96.33	6.00
DBNet+B+C	79.34	91.43	84.96	6.60
DBNet+A+B+C(Ours)	98.13	100.00	99.06	6.90

The first row of Table 3 is the reproduction result of the DBNet algorithm on the angle steel character data set in this experiment. As a comparison of the results of the improved algorithm in this paper, the last row of Table 3 is the result of the improved algorithm, which is used as the benchmark for this ablation experiment. ①A+B+C: Using all improvement strategies, the indicators have improved significantly, reaching 99.06%, and the number of parameters has reached 6.9M. ②A+B: Without strategy C, the F1-Score decreased by 13.35%, and the parameters remained unchanged. ③A+C: Without strategy B, the F1-Score decreased by 2.73%, and ④ the parameter amount decreased by 0.9M. B+C: Without strategy A, the F1-Score decreased by 14.1%, and the number of parameters decreased by 0.3M.

To sum up, it can be seen that without strategy A, the performance index declines most obviously, indicating that strategy A is the most important in algorithm improvement; without strategy B, the performance declines the least, indicating that strategy B has relatively little impact on the model; Strategy C, the performance drops more and the parameter quantity does not change, indicating that strategy C has a greater impact on the model.

4.5. Algorithm comparison experiment

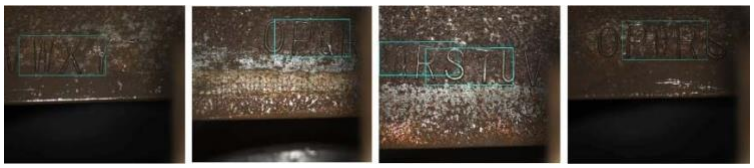
In order to verify the excellence of the improved algorithm, this paper selects the commonly used text detection algorithms based on MobileNetv3 for comparison. The experimental results of different algorithms are shown in Table 4. As can be seen from Table 4, for the character detection of angle steel data set, the performance of segmentation algorithm is better than that of regression algorithm. Compared with the original DBNet algorithm and other algorithms based on MobileNetv3 as the backbone network, the improved DBNet algorithm based on the coordinate attention mechanism has the best performance indicators, and the increase in the amount of parameters is also within the acceptable range.

Table 4. Comparison experiment of different algorithms

Algorithm	Precision	Rcall	F1-Score	Params(M)
DBNet	76.03	87.62	81.42	5.70
PSENet	78.91	96.19	86.70	5.50
EAST	51.91	64.76	57.63	6.40
Ours	98.13	100.00	99.06	6.90
Algorithm	Precision	Rcall	F1-Score	Params(M)

In order to more clearly show the improved effect of the algorithm in this paper, 4 images are randomly selected to test the performance of different algorithms, and the test comparison results are shown in Fig. 7. It can be seen that in Fig. 7 (a) the algorithm has detection errors and incomplete detection; (b) the algorithm has too large detection frames and too many detection frames; (c) the algorithm has incomplete detection and detection errors and overlap. ; (d) The algorithm is improved for this paper, the detection frame is positioned accurately, and there is no false detection and missed detection.

(a) DBNet



(b) PSENet

(c) EASTt

(d) Ours

Figure 7. Test comparison result display

5. Conclusion

This paper proposes an improved detection algorithm for angle steel characters based on DBNet, aiming at the problems of false detection and missed detection in manual recording of industrial angle steel characters. Among them, CA and BiFPN improve the feature perception ability of the model for the angle steel character area, and the Copy-Paste data enhancement and Cosine learning rate optimization strategy enhance the robustness of the model. The experimental results on the angle steel character data set show that the method in this paper has the highest index and the best effect among the common text detection algorithms based on MobileNetV3. Compared with the traditional DBNet algorithm, the F1-Score index has increased from 81.42% to 99.06%, which proves that this paper effectiveness of the method. Subsequent work will collect more angle steel character images as datasets to further strengthen the generalization of the model, thereby broadening the detection scenarios of the model.

References

[1] H.Y. Liu, Z.L. Li, Z.L. Huang: Progress in Building Steel Structures, Vol. 23 (2021) No. 12, p. 47-55.

[2] Y.H. Li, Y.Y Chen: Computer Engineering and Applications, Vol. 57 (2021) No. 06, p. 42-48.

[3] K. Wang, F. Yang, S. Jiang: Computer Application Research, Vol. 37 (2020) No. S2, p. 22-24.

[4] Q.L. Zhou, L. Ma, L. Cao, et al: Smart Agriculture, Vol. 4 (2022) No. 01, p. 47-56.

[5] C. Xie, H.Y. Zhu, Q. Fei: IET Image Processing, Vol. 16 (2022) No. 01, p. 273-284.

[6] J. Hu, L. Shen, S. Albanie, et al: IEEE Trans on Pattern Analysis and Machine Intelligence, Vol. 42 (2020) No. 08, p. 2011-2023.

[7] Q.F. Guo, L. Liu, X. Zhang, et al: Journal of Engineering Mathematics, Vol. 37 (2020) No. 05, p. 521-530.

[8] B.M. Li, R.L. Jin, Z.F. Xu, et al: Journal of Zhengzhou University (Engineering Edition), Vol. (2022) No. 01, p. 20-26.

[9] Y.X. Feng, Y.M. Li: Data Mining, Vol. 8 (2018) No. 04, p. 186-200.

[10] J. Schmidt-Hieber: Annals of Statistics, Vol. 48 (2020) No. 04, p. 1875-1897.