An Algorithm for Detecting the Integrity of Outer Frame Protection Net on Construction Site Based on Improved SSD

Yu Fan ¹,²,a, Xianqiao Chen¹, Jinguang Xie¹,ᵇ* and Zhenjie Fu¹

¹ Wuhan University of Technology, Wuhan, Hubei, China
² Hubei Key Laboratory of Transportation Internet of Things, Wuhan, Hubei, China
ᵃ email: 237047@whut.edu.cn
ᵇ Corresponding author’s email: 2535407807@qq.com

Abstract: In recent years, the continuous development of the basic construction industry has promoted the progress of urban social economy, and the safety issues in construction have attracted more and more attention. The integrity of the external protection net is of great significance to the safety of the construction personnel. Most of the traditional methods of detecting damage to the protection net have problems such as poor real-time performance and low detection accuracy. In this paper, we present a novel approach based on SSD for solving the problem of the integrity detection of the external frame protection net in the construction site. The improved network merges adjacent feature layers to enrich the semantics of shallow feature maps, thereby improving the detection accuracy of the network, and simultaneously introduces a shallower feature layer and a more advanced receptive field module to improve the detection ability of small targets. The experimental results show that compared with the SSD algorithm, the improved algorithm has higher detection accuracy in the detection task of the external protective net on the construction site, and can meet the real-time detection requirements.

1. Introduction
With the development of the construction industry, the complex construction environment, the dense labor force, and the frequent occurrence of large-scale machinery have greatly affected the safety of construction. The non-standard construction site protective equipment is an important cause of safety accidents. The integrity of the external protective net is undoubtedly very critical to the life safety of construction personnel. Therefore, it is necessary to test the integrity of the outer frame protection net, rectify the protection facilities that do not meet the requirements, and reduce the safety accidents caused by the incomplete outer frame protection net.

The detection of the integrity of the outer frame protection net is also a kind of problem in the target detection. The detection of the damage of the protection net that is easily overlooked in the vision is of great significance to safe production. At present, many scholars have conducted corresponding research on similar equipment damage detection technology. Pan Yingli [1] used machine vision related technology to conduct in-depth research on contact line insulator damage detection; Liang Zhiyong et al. [2] detected the damage of asphalt pavement based on digital image processing technology, and analyzed and compared various processing technologies; Based on visual image detection technology, Niu Lijuan et al. [3] used the OSTU method to perform image binarization to detect the damage of the underwater net clothing of the breeding cage. Most of the existing methods
for detecting damage to similar protective nets use traditional image processing and target detection methods, which have problems such as insufficient accuracy or slow detection speed, and strict requirements for the detection environment, which cannot meet the real-time requirements of the construction safety inspection process, including carrying out inspection problems and ensuring the quality of inspection needs.

In this paper, we use the image and video data from the construction site of the construction project, and after corresponding processing, finally make it into the integrity test data set of the outer frame protection net. According to the comprehensive requirements of detection accuracy, main detection target size and real-time performance, the SSD\(^4\) model is improved accordingly. We first remove the last layer of feature maps, and use channel splicing technology to perform feature fusion on adjacent feature maps to improve the semantic capabilities of shallow feature maps. Then, a shallower layer is added to the backbone network and a more advanced receptive field module is introduced to make full use of multi-scale information for small target damage detection. Experiments have proved that in the target detection task of the external protection net integrity detection data set, compared with the SSD algorithm, the accuracy is increased by 3.4\%, and the speed can meet the requirements of real-time detection on the construction site.

2. Materials and methods

The improved network architecture proposed in this paper is shown in Figure 1. Compared with the traditional SSD300, a shallower feature layer is introduced followed by an improved receptive field module. Because the receptive field module \(^5\) simulates the relationship between the size of the receptive field and the eccentricity in the human visual system, it strengthens the deep features learned by the convolutional neural network model make the detection model faster and more accurate. Therefore, the improved network obtains a shallower feature layer that combines position and semantic information. Then, considering the integrity of the outer frame protection net, the main detection task is the damage of small targets. In order to ensure that the detection speed is further improved, the last layer of the original feature layer is deleted. The adjacent feature layers in the original network are merged. Since the feature layers Conv9_2 and Conv10_2 already have rich semantic information, there is no need to enhance their features. Finally, new shallow layers will be added by improving the output and phase of the receptive field module. The output after the fusion of the adjacent layers and Conv9_2 and Conv10_2 are used as the final feature layer for prediction.

![Feature fusion module](image)

Figure 1. The overall framework of the improved SSD algorithm
2.1. Feature Fusion
The main purpose of feature fusion is to combine the characteristics of the shallow feature layer and the deep feature layer to make the positioning and classification capabilities of the model stronger. Among the target detection algorithms with feature fusion, one type is to perform prediction on a single fused feature layer, such as HyperNet \cite{6} and ION \cite{7}, which merge feature layers with inconsistent sizes through channel splicing; Another way is to predict on multiple fused feature layers. FPN \cite{8} and YOLO V3 \cite{9} are fused by pixel-by-pixel addition from top to bottom, while DSSD \cite{10} and FSSD \cite{11} pass The fusion is performed by pixel-by-pixel multiplication. However, this layer-by-layer downward fusion requires multiple consecutive up-sampling operations, which increases noise and reduces detection speed. Therefore, this article only merges adjacent feature layers.

The specific feature fusion method uses feature layer n and feature layer n+1 to represent adjacent feature layers as an example. The feature layer n+1 is compressed with 1×1conv first, so that the number of channels is half of that before compression, followed by the BatchNorm and Relu layers. At the same time, the feature layer is followed by 3×3 conv, BatchNorm, and Relu layers, which not only reduces the number of channels, but also ensures that the feature layer n and the feature layer n+1 have consistent learning capabilities. Then use bilinear interpolation to extend the feature layer n+1 to the same size as the feature layer n. Finally, channel splicing is used to fuse feature layer n and feature layer n+1.

2.2. Introduction of shallower feature layer and improved receptive field module
Since the shallow information loss of the multi-feature layer in the SSD algorithm is more serious, in this article, the shallow information is very important for detecting most of the damage information of the outer frame protection network, so the SSD backbone network VGG16 is modified to introduce Conv3_3 Features are used for detection. Considering that the semantic information of Conv3_3 is relatively small, it will be very serious if it is directly used as the feature layer. Therefore, the improved receptive field module is used to enrich the shallow feature information after Conv3_3.

The improved receptive field module refers to the practice of Shang et al. \cite{12}, and makes corresponding adjustments on the basis of the original RFB. The final RFB structure adopted is shown in Figure 2. Use small convolution kernels such as 1×1, 1×3, 3×1 to replace large convolution kernels such as 3×3 and 5×5. This alternative method can effectively reduce the parameter amount and time complexity of the model. At the same time, this configuration of RFB is conducive to extracting detailed features, especially line features, so that it will play a great role in the extraction of damaged line features of the outer frame protection net.

![Figure 2. Improved receptive field module](image-url)
3. Results & Discussion

3.1. Experimental data set production and processing
Since there is no public data set available for the inspection of the integrity of the external frame protection net, in this paper, we have made a self-made data set of the construction site external frame protection net. The data sources of the external protection net data set are the video data of the construction site monitoring and the photos collected by the safety inspection construction site. The data is provided by a construction engineering company in Hubei Province, China.

Preprocessing the acquired data is mainly divided into two steps: data format conversion and normalization. Use OpenCV to convert the site surveillance video file into an image format file, and cut, zoom, and normalize the images of different resolutions obtained from different channels to 300×300 for training. Finally, annotate the pictures to complete the creation of the data set.

3.2. Experimental training strategy
The external protection net data set constructed in this paper has a total of 13,520 sample pictures. The sample pictures are divided into 3 sample sets. The training set contains 11520 pictures, the validation set contains 1000 pictures, and the test set contains 1000 pictures. The experimental model in this article is based on the Pytorch framework. The configuration of the experimental environment is as follows: the operating system is Ubuntu 16.04, the memory capacity is 32G, the CPU model is INTEL I7 8700K, and the 2 GPU models are GTX1080Ti. The hyperparameter Batch_size is set to 64, the number of iterations is set to 300, the learning rate is set to 0.001, the learning decay rate is set to 0.1, the weight decay rate is set to 0.0005, and the momentum factor is set to 0.9.

3.3. Analysis of results
Table 1 shows the detection results of the improved algorithm in this paper and the other 3 kinds of target detection algorithms with excellent detection performance on the outer frame protection net data set, namely SSD, and the improved DSSD [10] and FSSD [11] based on SSD. In this paper, we select the detection accuracy evaluation index and the detection speed evaluation index for comparative experiments. This proves that the improved algorithm proposed in this paper has sufficient advantages on the basis of satisfying on-site real-time detection.

| Method   | mAP(%) | FPS |
|----------|--------|-----|
| SSD      | 78.4%  | 48  |
| FSSD     | 80.3%  | 32  |
| DSSD     | 81.2%  | 9   |
| This paper’s | 81.8%  | 28  |

Figure 3 is the comparison of the specific detection effects of SSD and the improved algorithm in this paper. Since there are many small damages in the outer frame protection net, the detection ability of small targets is the key to determining the detection effect. The comparison shows that the improved network model proposed in this paper is more prominent in the detection ability of small targets, and the missed detection phenomenon is significantly less than other detection models.
4. Conclusions
In order to solve the problem of the integrity detection of the external frame protection net on the construction site, we design an improved version of the external frame protection net integrity detection network based on SSD. The improved network uses channel splicing to perform feature fusion on adjacent feature maps, thereby improving the detection accuracy of small targets, and adding more shallow layers to the backbone network while introducing more advanced receptive field modules, making full use of multi-scale information for small target detection. Compared with the improved network, the accuracy of the integrity detection of the external protective net has been improved in the improved network, and it can also guarantee the requirements of on-site collection and real-time detection.

Acknowledgments
This research was funded by the National Major Scientific and Technological Project for the Innovation of Green and Intelligent Inland Ships and the National Natural Science Foundation of China (61503289).

References
[1] Pan yingli. Research on Detection Technology of Catenary Insulator Breakage Based on Machine Vision [D]. Lanzhou Jiaotong University, 2018.
[2] Liang zhiyong, Long lifang, Huang shibin. Asphalt pavement damage detection technology based on digital image processing [J]. Chongqing Architecture, 2009(01):40-42.
[3] Niu lijuan, Du hai, Zhao yunpeng, Bi chunwei. Research on Visual Inspection Method of Underwater Net Clothing Damage in Cage Cage [A]. China Ocean Engineering Society. Proceedings of the 19th China Ocean (Coast) Engineering Symposium (Part 2) [C]. China Ocean Engineering Society: Ocean Engineering Branch of China Ocean Society, 2019:5.
[4] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016.
[5] Liu S, Huang D. Receptive field block net for accurate and fast object detection[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 385-400.
[6] Kong T, Yao A, Chen Y, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition
(CVPR). IEEE, 2016.

[7] Bell S, Zitnick C L, Bala K, et al. Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[8] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.

[9] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.

[10] Fu C Y, Liu W, Ranga A, et al. Dssd: Deconvolutional single shot detector[J]. arXiv preprint arXiv:1701.06659, 2017.

[11] Li Z, Zhou F. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv:1712.00960, 2017.

[12] Shang T, Dai Q, Zhu S, et al. Perceptual Extreme Super Resolution Network with Receptive Field Block[J]. 2020.