Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 3, 2023

Real Time Detection Algorithm for Escape Ladders based on YOLOv5s

Sheng Jin *

Tianjin University of Technology and Education, China
* **Corresponding author Email**: 875597141@qq.com

Abstract: In the detection of escape ladders in the context of smart construction sites, due to the relatively small target size of the escape ladder compared to the entire input image frame, significant environmental interference, and high missed detection and false detection rates, an improved YOLOv5s escape ladder real-time detection algorithm is proposed by combining the attention mechanism network. The model uses CSPLocknet53 as the backbone network for feature extraction, introduces the attention module CA, and integrates spatial and channel information, while increasing a small amount of computation, performance has been significantly improved. Optimize the network structure of YOLOv5s algorithm, strengthen shallow feature weights to enhance small target detection effectiveness, add attention mechanisms to increase the weight of small targets and their surrounding features, and use Mosaic methods for data augmentation to improve detection accuracy and recall. After multiple repeated experiments, these experimental results have proven that the optimized YOLOv5s algorithm for real-time detection of escape ladders has an average detection accuracy (accuracy, recall) of (81.8, 82.6). Compared with the traditional YOLOv5s algorithm, the accuracy and recall have been improved by 1.4% and 1.2%, respectively. The optimized YOLOv5s algorithm can effectively improve the detection accuracy of real-time detection of escape ladders, and improve the detection and resolution performance of small escape ladder targets.

Keywords: Escape Ladder Real-time Detection; Deep Learning; Attention Mechanism; Small Target Detection.

1. Introduction

Ensuring worker safety is crucial in complex construction site environments. However, due to various factors, such as sudden mechanical failures, adverse weather conditions, and even worker negligence, danger can occur. Among them, real-time detection of escape ladders is a crucial step.

However, traditional escape ladder detection methods are often limited by environmental conditions such as lighting, visibility, noise, etc., resulting in insufficient accuracy and real-time performance. With the development of deep learning and computer vision technology, escape ladder detection algorithms based on deep learning have emerged. These algorithms can automatically and real-time detect the position and status of escape ladders, improving the safety of construction sites.

Zhang et al. [2] used parameter calibration to generate regions of interest (IOUs) and coordinate transformations, combined with the velocity and position information returned by cameras and radar, and used the mean square error loss function to calculate the updated model. S proposed an algorithm for precise positioning and target recognition based on boundary box regression, but this algorithm is limited by various indicators of the sensor. Pablo et al. [3] assumed that the data follows a Gaussian distribution and used a Gaussian mixture model for background modeling. They proposed a new particle filter algorithm for target tracking, which has better adaptability to spatial transformations. Kachach et al. [4] proposed the Directional Gradient Histogram (HOG) algorithm, which uses sliding windows to extract directional gradient features of pixels in each window.

The development of deep learning is on the rise, and more and more deep learning technologies are being applied to construction site scenarios. For the YOLO (You Only Look Once) [6,7] series of algorithms, it is the ancestor of the One Stage [5] algorithm in object detection algorithms. YOLOv5 has excellent accuracy and robustness, and YOLOv5 uses the same fully connected layer for object detection and image classification. Unlike the Two Stage [9] algorithm represented by Region Convolutional Neural Network (R-CNN) [10,11,12], it first searches for candidate regions, prunes them, and then classifies them using neural networks and support vector machines (SVM). The significant improvement in detection speed of YOLO series algorithms makes them more suitable for applications in scenarios that require high real-time detection requirements. Because deep neural networks extract features through operations such as downsampling and pooling, when the target to be detected is small, these operations will result in very few small target pixels, and the features that can be extracted will become very limited, resulting in a high missed detection rate and low detection efficiency when detecting small targets.

In summary, based on YOLOv5s, this article adds a small target detection layer [14] in the construction of the detection layer, and adds an attention mechanism network CA [15] to solve the problem of high detection rate of small target escape ladders. When constructing the dataset, a dataset containing multiple escape ladder states and environmental conditions was constructed, and standard evaluation indicators such as accuracy, recall, and mAP were used to evaluate the algorithm proposed in this paper.

1.1. YOLOv5s Algorithm Principle

YOLOv5s algorithm is an object detection algorithm that can detect and locate objects in images in real-time. The YOLOv5s algorithm transforms the object detection problem into a regression problem, dividing the image into grids using a neural network CNN and predicting each grid to determine the position and category of objects in each grid. Then, the best candidate box is matched using Non-Maximum

Suppression (NMS) [15]. At the same time, the YOLOv5s algorithm also adopts a feature pyramid network to improve the accuracy and speed of the algorithm.

Specifically, the YOLOv5s algorithm first divides the input image into SxS grids, with each grid predicting B bounding boxes, where each bounding box predicts K coordinates and K confidence scores. Among them, S, B, and K are hyperparameters of the YOLOv5s algorithm, which can be adjusted according to specific problems.

The feature pyramid network is an important component of the YOLOv5s algorithm, which includes multiple convolutional and pooling layers and can extract features of different scales from input images. These features are transmitted to different detection layers, each of which outputs a set of bounding boxes and corresponding category probabilities.

YOLOv5 (You Only Look Once version 5) is a single stage object detection algorithm. It has added some new improvement ideas on the basis of YOLOv4, which has greatly improved its speed and accuracy. Compared to the previous YOLO version, YOLOv5 adopts more data augmentation methods during the training process, such as Mosaic data augmentation, adaptive anchor box calculation, adaptive image scaling, etc. In addition, YOLOv5 also incorporates some new ideas from other detection algorithms, such as Focus structure, CSP structure, FPN+PAN structure, etc. The anchor box mechanism of its output layer is the same as YOLOv4, but it improves the loss function GIOU during training_ Loss, as well as DIOU for prediction box filtering_ Nms.

During the algorithm's detection, the images input to the YOLOv5s network are preprocessed, including scaling, zeroing, and other operations, so that the network can process the images normally. After input to the network, multiple predicted boundary boxes and type results are output. At this time, YOLOv5 performs NMS (non maximum suppression) post-processing on the output results, selecting the most matching option box as the final detection result output.

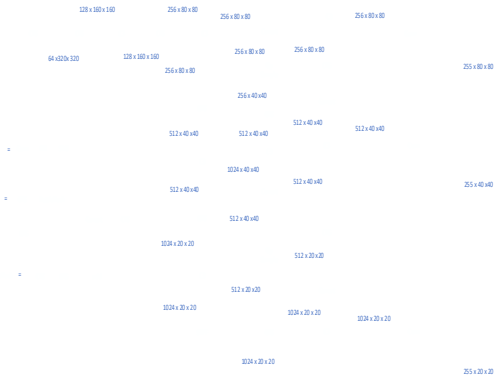


Fig 1. YOLOv5s network structure

2. Algorithm Improvement

This study introduced the CA attention network module and made some adjustments to the design and output structure of the backbone network, making the network's perception of low-level features more apparent without increasing computational parameters. The overall adaptation improvement is shown in the following figure.

2.1. Network Structure Optimization

The YOLOv5s algorithm performs multiple downsampling during feature extraction, and each downsampling results in the loss of pixel information. This can lead to poor feature extraction performance when detecting small targets, thereby affecting the detection effect. Learning small target information becomes more difficult when the number of feature layers is deep. The network model designed in this article is based on the YOLOv5s algorithm and adds a CA attention mechanism layer. This attention mechanism layer can embed position information into new attention, which can expand the network's attention range. In addition, it can also avoid the increase in computational complexity caused by large-scale attention.

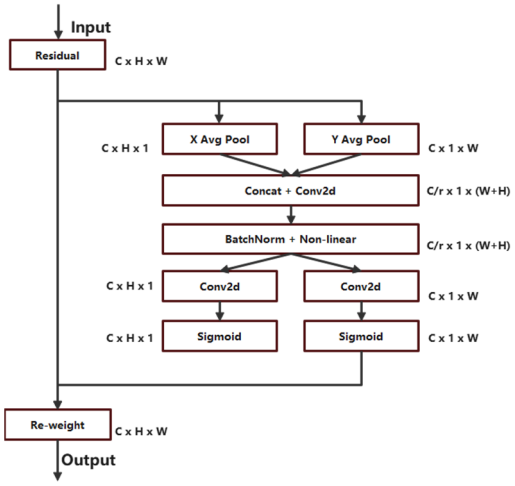


Fig 2. Network architecture flowchart of CA attention mechanism

2.2. Dataset Engineering

We constructed a dataset containing multiple escape ladder states and environmental conditions. The dataset consists of on-site collected videos, with 1000 sample images extracted every 15 frames. The environmental conditions include dim and bright environments. When annotating the dataset, visible escape ladder objects are labeled. This paper uses high pass filtering and median filtering to optimize the dataset by removing image impurities and sharpening image details, The YOLOv5s algorithm, which added a CA attention detection layer, was trained using the optimized data.

2.3. Algorithm Process

1. Prepare a dataset: Collect and annotate the image data of the escape ladder, and divide it into training, validation, and testing sets.

2. Build YOLOv5s model: Based on the YOLOv5s network structure, build an object detection model.

3. Add attention mechanism CA: Add a CA attention mechanism layer to the YOLOv5s model to improve the model's attention to small targets and optimize the neural network's detection ability for small size escape ladder images.

4. Training Model: Train the model using a training set and optimize the model parameters using a backpropagation algorithm.

5. Validation Model: Use the validation set to evaluate the model during the training process and select the model with the best performance.

6. Test Model: Use a test set to test the final selected model and evaluate its performance in small object detection tasks.

7. Result analysis: Analyze the test results, compare the

performance differences before and after adding attention mechanism CA, and prove the effectiveness of attention mechanism in improving the detection effect of small targets.

3. Experimental Analysis and Validation

3.1. Experimental Environment and Datasets

The experimental software and hardware settings in this article are: NVIDIA GeForce GTX 3060 GPU, 16GB memory, Inter Core i5-10200H CPU, trained using Python, Pytharm Community IDE, Python framework version 1.7.0, which can call GPU for high-speed tensor calculation, using Python 3.9 that is suitable for Python version, and operating environment is Windows 10.

Monitoring videos taken using live scenes on the construction site and videos recorded normally, including 6 types of data scenarios: day (strong light), day (weak light), dusk, in the ditch, and outside the ditch. Image data was obtained through video frame extraction. After data analysis, invalid data was deleted, fuzzy data was removed, and other operations were conducted to create a high-quality dataset. Finally, 1500 images were obtained, including 2000 samples of escape ladders. During data annotation, to avoid interference from other features, occluded parts are removed to purify the features.

3.2. Algorithm Improvement Effect

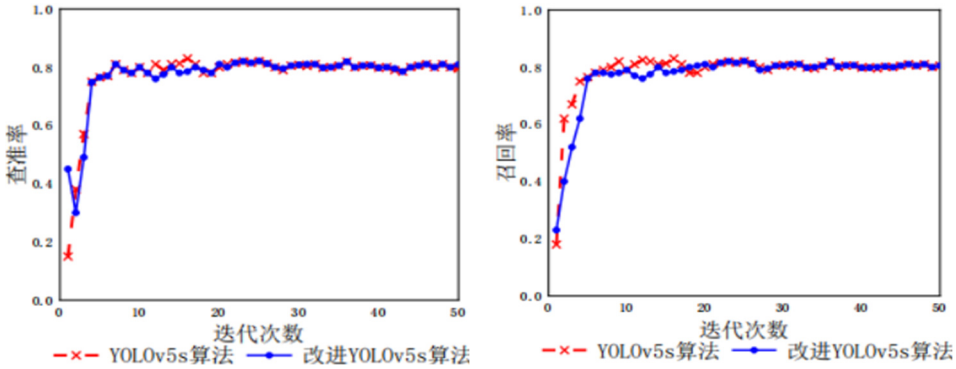
3.2.1. Performance Test Results of YOLOv5s Detection Algorithm on Self built Datasets

Improved algorithm structure, added small object detection layer CA, improved anchor box, selected anchors that are more suitable for small objects, improved YOLOv5s algorithm, added small object detection layer, easy to analyze charts, optimized and adapted YOLOv5s algorithm, improved accuracy by 1.4% and recall by 1.2% mAP@0.5 The indicator has increased by 2.9%, and the detection speed has decreased by about 0.9ms.

Table 1. Comparison of YOLOv5s and YOLOv5s with CA added

	mAP .5	mAP .5 samll	inference(ms)
YOLOv5	0.826	0.769	8
optimized YOLOv5	0.855	0.825	8.9
difference	2.9%	5.6%	0.9

3.2.2. Visualization Analysis of Iterative Process



(a) Comparison of precision rate (b) Comparison of recall rate

Fig 3. Visualization Analysis of Iteration Times and Convergence of YOLOv5s Algorithm before and after Improvement

From the comparison of the changes in iteration times, precision, and recall in the figure, it can be seen that the improved YOLOv5s algorithm has better convergence speed than the original YOLOv5s algorithm.

3.2.3. Comparison of Effects Before and after Optimization

(a) YOLOv5s Algorithm Small Target Detection Effect (b) Improvement of YOLOv5s Algorithm Small Target Detection Effect

Fig 4. Comparison of Small Target Detection Effects.

In order to demonstrate the optimization effect, the YOLOv5s algorithm before and after improvement will be analyzed and verified for the detection effect of small target escape ladders. The test video is the actual video collected on the construction site. As shown in Figure 4, in the video collected in practical engineering applications, the YOLOv5s algorithm missed the detection of small target escape ladders when detecting backlight small targets, as shown in Figure 4 (a). The improved and optimized YOLOv5s algorithm performs better in the same test, and its detection ability is stronger compared to before when targeting small targets with lower resolution.

4. Conclusion

This article investigates the method of adding CA attention mechanism layer to the YOLOv5s model and explores its optimization effect on escape ladder target detection tasks. Through experimental verification, we have demonstrated that adding a CA attention mechanism layer can improve the performance of YOLOv5s, enabling it to achieve better accuracy and robustness in object detection tasks, which helps further improve the performance and practicality of object detection. The research results of this article demonstrate the important role of attention mechanism in deep learning models, providing valuable reference for future research. We believe that in future research, we can further explore and optimize the application of attention mechanisms in the field of object detection, laying a solid foundation for achieving higher precision object detection.

References

[1] WU T, ZHOU P, LIU K, etal. Multi-agent Deep Reinforcement Learning for Urban Traffic Light Control in Vehicular Networks [J]. IEEE Transactions on Vehicular Technology, 2020, 69(8): 8243-8256.

[2] ZHANG X, ZHOU M, QIU P, etal. Radar and Vision Fusion for the Real-time Obstacle Detection and Identification [J]. Industrial Robot An International Journal, 2019, 46(3): 391-395.

[3] PABLO B, CHRISTIANO B, FABIANO, et al. A Novel Video Based System for Detecting and Counting Vehicles at User-Defined Virtual Loops [J]. Expert Systems with Applications, 2015, 42(4): 1845-1856.

[4] KACHACH R, CANAS J M. Hybrid Three-dimensional and Support Vector Machine Approachfor Automatic Vehicle Tracking and Classification Using a Single Camera [J]. Journal of Electronic Imaging, 2016, 25(3): 033021.

[5] TIAN Y N, YANG G D, WANG Z, et al. Apple Detection During Different Growth Stages in Orchards Using the Improved YOLO-v3 Model [J]. Computers and Electronics in Agriculture, 2019,157:417-426.

[6] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-time Object Detection [C] . Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 779-788.

[7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 7263-7271.

[8] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [R]. arXiv: 1804. 02767v1, 2018.

[9] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal Speed and Accuracy of Object Detection [J/OL]. arXiv: 109342020. http://arxiv.org/abs/2004.10934, 2020.

[10] GIRSHICK R, DONHUE J, Darrell T, et al. Region-based Convolutional Networks for Accurate Object Detection and Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.

[11] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-based R-CNNs for Fine-grained Category Detection [C]. European Conference on Computer Vision. ECCV, Zurich, 2014: 834-849.

[12] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[13] ZHU Maotao, XING Hao, FANG Ruihua. Vehicle Detection Based on YOLO-TridentNet [J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(11): 1-8.

[14] LIU Changyuan, WANG Qi, BI Xiaojun. Multi-target and Small-scale Vehicle Target Detection Method [J]. Control and Decision, 2021,36(11):2707-2712.

[15] YAN Shaohua, ZHANG Duo, CHU Hongyu, et al. A Review of YOLO Object Detection Based on Deep Learning [J]. Journal of Electronics & Information Technology, 2022,44 (10):3697-3708.

[16] CAI Renhao, CHENG Ning, PENG Zhiyong, et all. Lightweight Infrared Dim Vehicle Target Detection Algorithm Based on Deep Learning [J]. Infrared and Laser Engineering, 2022,51(12):357-367.

[17] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions [J]. arXiv preprint arXiv: 1610.02357, 2016.

[18] HUANG Yinglai, LI Daming, LÜ Xin, et all. A Detection Method of Lentinus Edodes Based on Improved YOLOv4 Algorithm [J]. Journal of Harbin University of Science and Technology, 2022,27(4):23-31. Journal of Harbin University of Science and Technology.

[19] Dong X, Yan S, Duan C. A Lightweight Vehicles Detection Network Model Based on YOLOv5 [J]. Engineering Applications of Artificial Intelligence, 2022, 113: 104 914.

[20] FAWZI A, SAMULOWITZ H, TURAGA D, et al. Adaptive Data Augmentation for Image Classification [C]. IEEE International Conference on Image Processing. IEEE, 2016: 3688-3692.

[21] Yin Zhang, Guiyi Zhu, Tianjun Shi, Kun Zhang, Junhua Yan. Small Object Detection in Remote Sensing Images Based on Feature Fusion and Attention[J]. Acta Optica Sinica, 2022, 42(24): 2415001.

[22] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. 14th European Conference on Computer Vision (ECCV), 2016: 21-37.

[23] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.

[24] Redmon J, Farhadi A, Ieee. YOLO9000: Better, Faster, Stronger[C]. 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.

[25] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[C]. 15th European Conference on Computer Vision (ECCV), 2018: 122-138.

[26] Zhang X, Zhou X Y, Lin M X, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]. 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 6848-6856.

[27] Am A, Id O, B B, et al. - Vision Transformers in Image Restoration: A Survey[J]. - Sensors (Basel). 2023 Feb 21; 23 (5): 2385.

[28] J H, L S, S A, et al. - Squeeze-and-Excitation Networks[J]. - IEEE Trans Pattern Anal Mach Intell. 2020 Aug;42(8):2011-2023. doi:(- 1939-3539 (Electronic)): - 2011-2023.