International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 23, No. 2, 2026

Research on Detecting Violations through Fences based on Improved YOLOv11

Hongfei Huang[1], Qian Chen[1], Wanying Ou[1], Yang Xu[1], and Yujie Zhang[2,*]

[1] Chengdu Technological University, Chengdu, 611730, China
[2] School of Artificial Intelligence, Chengdu Technological University, Yibin, 644000, China
[*] **Corresponding Author:** Yujie Zhang

Abstract: With the continuous advancement of intelligent security and intelligent video surveillance technologies, deep learning-based abnormal behavior detection has gradually emerged as a critical research direction in the security field. Traditional manual patrols and conventional video surveillance systems struggle to achieve rapid and accurate detection and early warning. To address the issues of missed detections, false detections, and low detection accuracy in pedestrian fence-climbing detection under complex scenarios such as occlusion, dense multi-target conditions, and simultaneous multi-person climbing, an improved YOLOv11-based illegal wall-climbing detection algorithm named ban-YOLOv11 is proposed. In the backbone module, GhostNet modules are combined with convolutional modules to effectively reduce computational cost and parameter count, enabling the model to be deployed on edge devices. Furthermore, the Convolutional Block Attention Module (CBAM), a hybrid-domain attention mechanism, is introduced at the end of the backbone network to mitigate interference from irrelevant targets in complex backgrounds, fuse multi-scale information, and achieve more comprehensive feature aggregation. Experimental results demonstrate that the proposed algorithm achieves a mean average precision (mAP) of 98.0%, which is 2.4% higher than the original model, and a detection speed of 80.6 frames per second (FPS) while reducing the parameter count. It effectively alleviates the problems of missed detections, false detections, and low detection accuracy, while maintaining high real-time performance, making it more suitable for real-time pedestrian fence-climbing detection.

Keywords: YOLOV11, Object Detection, Crossing the Wall in Violation of Rules, CBAM.

1. Introduction

Isolation facilities such as walls and fences undertake critical safety protection and regional management functions in campuses, construction sites, power construction sites, traffic isolation zones, and other closed management scenarios[1]. However, in practical applications, illegal behaviors including climbing, crossing, and crawling under walls or fences occur frequently. Such behaviors not only undermine management order but also may lead to personal injuries, property damage, and even more severe safety accidents.

For a long time, safety supervision in wall-enclosed areas has mainly relied on manual patrols, gate post management, and conventional video surveillance[2]. Although these methods can achieve basic prevention to a certain extent, they still suffer from inherent limitations: limited supervision scope, slow response speed, high rates of false and missed alarms, and elevated labor costs. Particularly under conditions of wide surveillance coverage, short duration of illegal behaviors, and rapid action changes, it is extremely difficult to achieve all-weather, full-area, and high-efficiency supervision solely through manual means. With the rapid development of deep learning, object detection, and video behavior recognition technologies, automatic recognition of abnormal behaviors using existing surveillance videos has become feasible and has demonstrated significant application value in the construction of intelligent security systems.

Deep learning-based object detection algorithms are primarily categorized into two families: two-stage and one-stage approaches. Two-stage algorithms, represented by the R-CNN series [3-5] and R-FCN [6], achieve superior detection accuracy but suffer from high computational costs due to their two independent processing stages, making them unsuitable for real-time detection requirements. In contrast, one-stage algorithms skip the candidate region generation step and perform object detection directly on the input image, offering significant advantages in inference speed. Prominent examples include the YOLO series [7-10], SSD [11], RetinaNet [12], and EfficientDet [13]. Given the stringent real-time requirements of applications such as road traffic detection and abnormal behavior detection, one-stage algorithms have become the primary focus of improvement for researchers worldwide.

In the field of illegal climbing behavior detection, Wang et al. [14] proposed a spatiotemporal detection algorithm combining improved YOLOv5 and SlowFast [15]. They replaced the residual structures in the Fast and Slow branches with AC residual blocks and SE residual blocks, respectively, which enhanced the model's ability to learn discriminative and fine-grained features, leading to a significant improvement in detection performance. Wang et al. [16] introduced a Transformer-encoded C3TR module at the last layer of the YOLOv5s backbone network to strengthen global information extraction and multi-scale feature fusion. Additionally, they embedded an SE channel attention module in the feature fusion network to focus on critical information and improve detection accuracy. However, neither of these methods addressed the challenge of small object detection in real-world scenarios.

To tackle small object detection issues, Luo et al. constructed a channel attention-based C3-TCA module in the feature extraction stage to enhance the model's ability to distinguish similar features. In the feature enhancement stage, they fused shallow and deep features via a dual-path enhancement structure and optimized the number of

prediction branches and downsampling ratios, effectively improving the detection accuracy of tiny traffic signs. Jing et al. proposed a small object detection algorithm based on multi-scale feature fusion. They designed a bidirectional adaptive feature pyramid network to fully exploit detail features and skip connections for enhanced multi-scale fusion, and constructed a dual-head detection structure to focus on small object features, achieving a balance between detection accuracy and model lightweighting. Nevertheless, this algorithm still suffers from false and missed detections in occlusion scenarios.

To address the detection challenges in occlusion scenarios, Xue et al. proposed an improved YOLOX-based occlusion detection algorithm. They introduced a coordinate attention mechanism to enhance the saliency of target objects, highlight effective features, and suppress redundant features. Meanwhile, they removed the bottom-up path enhancement structure in the feature fusion layer to reduce network complexity and designed an adaptive feature fusion structure to improve the representation of useful features, thereby enhancing detection accuracy in occlusion scenarios. Li et al. improved the YOLOv7 algorithm for occlusion scenarios by replacing the ELAN modules in both the backbone and head with contextual transformer (CoT) modules to enhance the capture of contextual semantic information. They also fused the Biformer bidirectional self-attention transformer module with the SPPCSPC module to strengthen the model's processing capability in complex scenarios and for occluded targets. However, this method still has limitations in dense multi-target scenarios.

Most existing illegal climbing detection algorithms rely on handcrafted feature extraction or convert actions into tracking trajectories for curve judgment against predefined intersection lines. These approaches generally cannot directly detect actions in video frames, suffering from slow detection speed, low robustness, and poor cross-scenario generalization. To address these limitations, this paper proposes a one-stage illegal climbing detection model named ban-YOLOv11. Specifically, to solve the problems of difficult deployment, slow inference speed, and high latency of the original YOLOv11 on edge devices, we integrate the GhostNet lightweight network to reduce parameter count and achieve real-time monitoring. To mitigate accuracy degradation caused by excessive detection targets and model lightweighting, as well as detection precision decline due to irrelevant object interference on walls, we incorporate the convolutional block attention module (CBAM) to guide the model to focus on illegal climbers. Furthermore, to address the poor detection performance of distant small objects, we add an additional small object detection layer to improve both accuracy and real-time performance for small targets.

2. Proposed Ban-YOLOv11 Algorithm

This paper proposes a wall-climbing detection algorithm based on YOLOv11, named Ban-YOLOv11, whose architecture is illustrated in Figure 1. Ban-YOLOv11 is built upon the YOLOv11l baseline model with three core optimizations as follows.

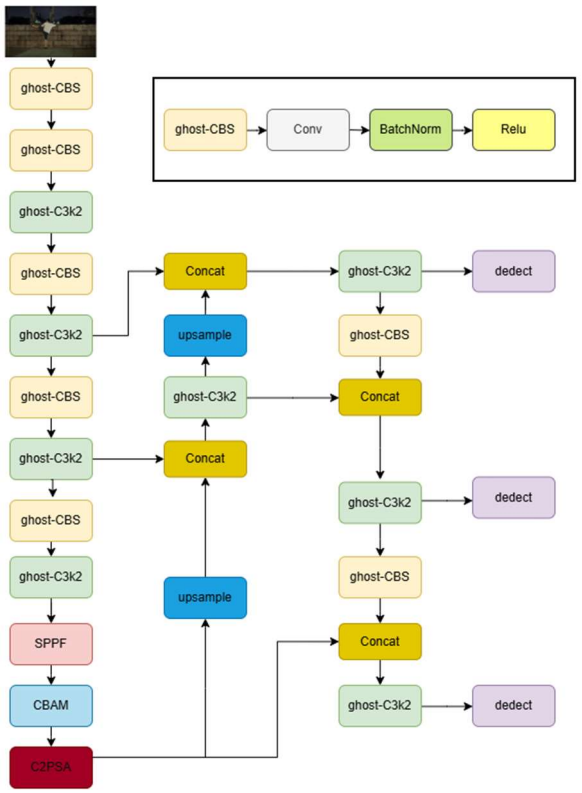

Figure 1. Ban- yolov11 model

First, we integrate the lightweight GhostNet network into the backbone to reduce computational complexity. Specifically, we combine the standard convolutional (Conv) module with GhostNet to form a GhostNet-Conv module, which is further assembled into a GhostCBS block. Additionally, we replace the original C3K2 modules in the backbone with Ghost-C3K2 modules, which incorporate channel-wise linear transformation layers on the basis of GhostNet. These layers perform a series of linear transformations on the input feature maps to enhance feature representation capability while preserving the spatial structural information of the input features. This design enables better adaptation to diverse object detection tasks and further improves feature expressiveness and accuracy, thereby boosting the overall detection performance of the model. Most importantly, the Ghost-based modules significantly reduce the number of model parameters and computational cost, enabling efficient deployment on resource-constrained edge devices.

Second, we introduce the Convolutional Block Attention Module (CBAM), a hybrid-domain attention mechanism, at the end of the backbone network. CBAM adaptively recalibrates channel-wise and spatial-wise feature responses, allowing the algorithm to allocate more computational resources to critical channels and regions of interest. This effectively enhances the model's focus on wall-climbing regions, suppresses interference from irrelevant background features, and further improves the performance of pedestrian fence-climbing detection.

2.1. Ghostnet Lightweight Network Module

Our research targets deployment on edge devices, which have inherently limited computational resources and cannot support large-scale deep learning models. Therefore, we perform comprehensive lightweight optimization on the algorithm to enable its deployment on edge detection devices,

while ensuring high inference speed, low latency, improved detection accuracy, and enhanced capability for small object detection.

Among all optimization strategies, the Ghost-C3K2 module is the core component of our lightweight design. Ghost-C3K2 is a convolutional neural network structure derived from GhostNet, which incorporates additional channel-wise linear transformation layers on the basis of the original GhostNet architecture. These layers apply a series of linear operations to the input feature maps to enhance feature representation capability. Meanwhile, Ghost-C3K2 preserves the spatial structural information of the input features, enabling it to better adapt to various object detection tasks and further improve feature expressiveness and accuracy, ultimately boosting the model's overall detection performance. Structurally, Ghost-C3K2 consists of two core components: the original C3K2 layer and GhostConv, which work synergistically to achieve a balance between model capacity and computational efficiency.

2.1.1. C3k2 Module

The original C3K2 module is designed to increase the receptive field of the network and enhance its adaptability to object scales. Structurally, a C3K2 layer consists of three convolutional layers: two 1×1 convolutional layers and one 3×3 convolutional layer. This design enables the C3K2 layer to effectively capture features at different scales and fuse them together, thereby improving overall detection performance.

2.1.2. GhostConv Module

The core principle of GhostConv is to introduce an additional ghost channel in each convolutional layer, followed by convolution operations on this channel using small convolutional kernels. This approach projects a portion of the features into a low-dimensional space through partial convolutional kernels, thereby reducing the number of parameters and computational complexity while improving the model's generalization ability to novel and unseen data. A comparison between conventional convolution and Ghost convolution is illustrated in Figure 2.

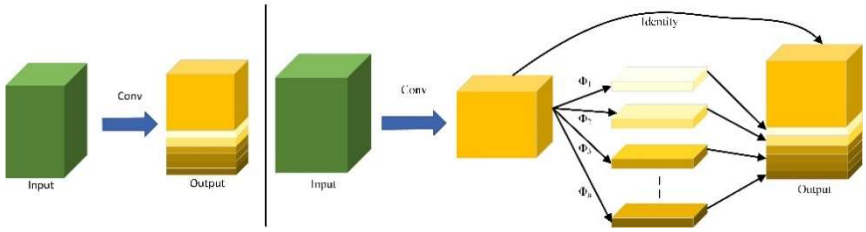

Figure 2. Comparison of convolution methods

2.2. CBAM Mixed-domain Attention Mechanism

The hybrid-domain attention mechanism integrates the advantages of both spatial attention and channel attention, enabling superior performance in both channel-wise and spatial-wise feature recalibration. The Convolutional Block Attention Module (CBAM) is the most representative implementation of this mechanism.

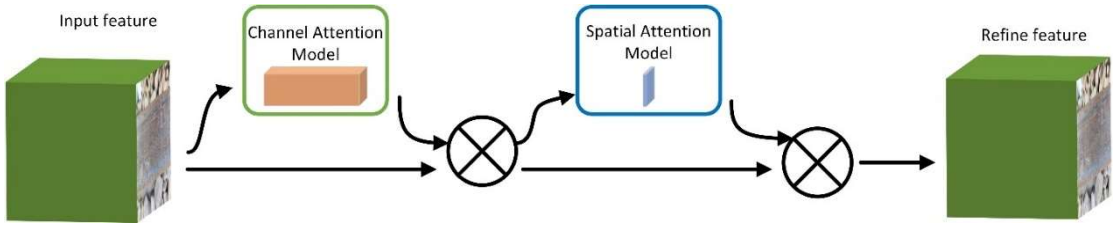

Figure 3. CBAM

The input feature map is first processed by the channel-wise attention module, and the resulting feature map is then fed into the spatial-wise attention module to obtain the final refined feature map. In the squeeze operation of the channel-wise attention module, both global average pooling and global max pooling are employed simultaneously, which enriches the extracted feature representations by capturing complementary statistical information from different perspectives. In the excitation operation, the outputs of the average pooling and max pooling branches are processed through a shared fully connected layer. The two processed results are summed element-wise and passed through a sigmoid activation function to generate the channel-wise attention weights. These weights are then multiplied with the original feature map to complete the channel-wise feature recalibration.

3. Experiment and Result Analysis

3.1. Experimental Platform and Equipment

This study focuses on the research of illegal wall-climbing detection technology. To address the challenges in illegal climbing detection, we propose the ban-YOLOv11 algorithm by improving the baseline YOLOv11 model, which achieves the goal of illegal wall-climbing detection by automatically identifying suspicious personnel climbing over walls. The parameters of the training and testing platform for the ban-YOLOv11 algorithm are summarized in Table 1.

Table 1. Test platform hardware specifications

Device Name	Configuration Name	quantity
CPU	Intel Xeon Platinum 8280	1
GPU	RTX6000 24GB	2
memory	32G	4

The software parameters of the experimental platform built for this study are shown in Table 2.

Table 2. software parameters

Software type	Version
system	ubuntu 18.05
CUDA	11.0
cuDNN	8.0.5
IDE	Pycharm
Deep learning framework	Pytorch

3.2. Introduction to Experimental Dataset

Since there is no publicly available dataset specifically dedicated to illegal wall-crossing detection, we constructed a custom illegal wall-crossing dataset named Through for this study. This dataset is compiled from publicly available abnormal behavior detection data and surveillance video frames collected from real-world scenarios, with two detection classes annotated: person and wall. It contains over 10,160 images of illegal wall-climbing samples covering various lighting conditions, weather scenarios, and occlusion levels, with a uniform resolution of 720×1280 pixels.

Data augmentation techniques including random flipping, brightness adjustment, and Gaussian blur are applied to expand sample diversity and enhance the model's generalization ability. Annotations are performed using the LabelImg tool. For pedestrian fence-climbing behavior, each image is annotated and exported as a YOLO-format text file with two labels: no-climb (non-crossing) and climb (crossing). The dataset is finally split into training, validation, and test sets in a ratio of 7:2:1. Partial samples of the dataset are shown in Figure 4.

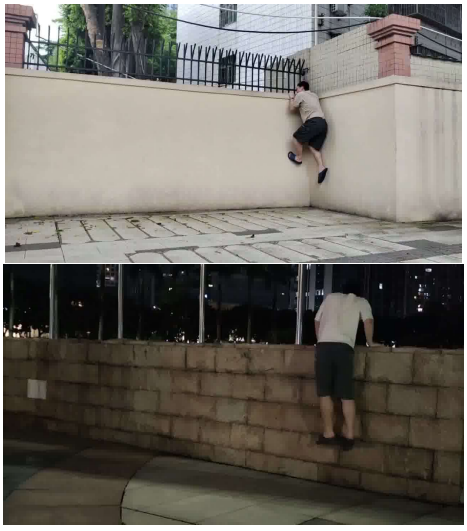

Figure 4. dataset

3.3. Performance Evaluation Metrics

To comprehensively and intuitively verify the detection performance of the proposed ban-YOLOv11 algorithm, we adopt four quantitative evaluation metrics: mean Average Precision (mAP), Recall (R), Frames Per Second (FPS), and per-image inference time. Among them, mAP is calculated as the mean of the Average Precision (AP) values across all detection classes. Precision (P), Recall (R), and mAP are defined by Equations (1) to (5) respectively.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$E = 1 - R = \frac{TN}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P(R) \tag{4}$$

$$mAP = \frac{1}{c}\sum_{j=1}^{c} AP_j \tag{5}$$

Among these metrics, TP (True Positive) denotes positive samples that are correctly classified as positive. FP (False Positive) denotes negative samples that are incorrectly classified as positive. FN (False Negative) denotes positive samples that are incorrectly classified as negative.

3.4. Experimental Results and Analysis

Table 3. Experimental results

model	mAP@0.5/%	Recall/%	FPS
Yolov8n	93.5	95	14.6
Yolov8s	95.3	93	54
Yolov8m	94.6	98	56.8
Yolov11m	94.3	96	76.2
Yolov11l	95.6	97	73.2
Ban-Yolov11l	98.0	100	80.6

Ban-YOLOv11 outperforms all comparative models across all evaluation metrics: it achieves a mean average precision (mAP) of 98.0%, representing a 2.4% improvement over the original YOLOv11; attains a frame rate of 80.6 frames per second (FPS), maintaining excellent real-time performance; and has a parameter count of 17.2M, demonstrating remarkable lightweighting effectiveness. It effectively addresses the issues of missed detections and false detections in complex scenarios and meets the deployment requirements of edge devices.

4. Over-the-wall Detection Effect

Randomly selected wall-climbing images are tested using the improved network, and the detection results are presented in Figure 13. The wall-climbing regions are marked with bounding boxes, where different colors are used to distinguish between climbing and non-climbing behaviors. The detection confidence of each prediction is displayed at the top-left corner of the corresponding bounding box. As shown in Figure 13, except for a few small targets with detection confidence around 0.9, the confidence scores of all other detections are very high, even reaching 1.0. These results demonstrate that the improved network can accurately detect the location and type of wall-climbing behaviors, exhibiting excellent detection capability with significantly reduced duplicate bounding boxes.



Figure 5. Inspection effect diagram

5. Conclusion

Addressing the issues of low detection accuracy, high rates of missed and false detections, large parameter count, and difficulty in deploying on edge devices in the existing YOLOv11 model for detecting unauthorized fence climbing, this paper proposes an improved algorithm, the ban-YOLOv11 detection algorithm. This algorithm is achieved through the improvement of a lightweight backbone network and the integration of a hybrid attention mechanism. Experimental analysis shows that the improved algorithm achieves a 2.4% increase in mAP compared to the original YOLOv5111 structure, with a detection speed of 80.6fps. It exhibits excellent adaptability and real-time performance in complex scenes, meeting the real-time detection requirements for unauthorized fence climbing behavior. In summary, the improved algorithm has stronger accuracy and real-time performance, making it more suitable for real-time detection. It achieves efficient and accurate detection of unauthorized fence climbing behavior. The main research conclusion for the next step is to further expand the dataset, optimize the model structure, and enhance the algorithm's generalization ability and scene adaptability.

Acknowledgments

"Received funding support from the Provincial College Students' Innovation and Entrepreneurship Training Program of Chengdu Technological University (Project No. 20250431610)".

References

[1] MATSUMURAS,HIGAKIH.Extensionofrh2swl for collision-free data message transmissions by subsidiary channelin wide area wireless multihop networks[C].2010IEEE WirelessCommunicationand NetworkingConference.IEEE,2010:1-6.

[2] Zhang Tai, Zhang Wei, Liu Yanyan. Detection algorithm for person climbing over in perimeter video surveillance [J]. Journal of Xi'an Jiaotong University, 2016, 50(6): 47-53

[3] GIRSHICKR,DONAHUEJ,DARRELL T,etal. Richfeaturehierarchiesforaccurateobjectdetection andsemanticsegmentation[C].Proceedingsofthe IEEE Conference on Computer Vision and Pattern Recognition,2014:580-587.

[4] RENS,HEK,GIRSHICKR,etal.FasterR-CNN: Towards real-time object detection with region proposalnetworks[J].IEEE TransactionsonPattern Analysis & Machine Intelligence,2017,39 (6): 1137-1149.

[5] HEK,GKIOXARIG,DOLLAR P,etal.MaskRCNN[J].IEEE TransactionsonPattern Analysis & MachineIntelligence,2020,42(2):386-397.

[6] DAIJF,LIY,HE K M,SUNJ.R-FCN:Object detection via region-based fully convolutional networkers [C].Proceedings Systems, Barcelona Spain,5-12Dec,2016.NewYork:CurranAssociates Inc,2016:379-387.

[7] REDMONJ,DIVVALA S,GIRSHICK R,etal.You onlylookonce:Unified,realtimeobjectdetection[C]. ProceedingsoftheIEEE Conference on Computer VisionandPatternRecognition,2016:779-788.

[8] REDMON J,FARHADI A. YOLO9000:Better, faster,stronger [C].Proceedings of the IEEE ConferenceonComputerVisionandPatternrecognition, 2017:7263-7271.

[9] REDMON J, FARHADI A. YOLOv3: An incrementalimprovement[J].ArXivpreprintarXiv: 1804.02767,2018.

[10] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv preprint arXiv: 2004. 10934,2020.

[11] LIU W,ANGUELOV D,ERHAN D,etal.SSD: Single shot multibox detector [C]. European Conferenceon Computer Vision.Springer,Cham, 2016:21-37.

[12] LINTY,GOYALP,GIRSHICKR,etal.Focalloss fordenseobjectdetection[C].ProceedingsoftheIEEE InternationalConferenceon Computer Vision,2017: 2980-2988.

[13] TAN M,PANG R,LE Q V.Efficientdet:Scalable andefficientobjectdetection[C].Proceedingsofthe IEEE/CVF Conference on Computer Vision and PatternRecognition,2020:10781-10790.

[14] WANGL, ZHAOT. Detection of climbing behavior based on deep learning [J]. Computer Systems and Applications, 2023, 32(5): 262-272

[15] FEICHTENHOFERC,FAN H Q,MALIKJ,etal. Slowfastnetworksforvideorecognition[C].2019 IEEE/CVF

International Conference on Computer Vision(ICCV).Seoul:IEEE,2019:6201-6210.

[16] WANG Yuanpeng, WAN Haibin, HUANG Kai, et al. Real-time detection algorithm for abnormal behavior of escalator passengers based on YOLOv5s [J]. Laser & Optoelectronics Progress, 2024, 61(8): 211-218