electronics

MDPI

Article

Substation Personnel Fall Detection Based on Improved YOLOX

Xinnan Fan [1,2], Qian Gong [2], Rong Fan [2], Jin Qian [3], Jie Zhu [4], Yuanxue Xin [2] and Pengfei Shi [1,5,*]

[1] Jiangsu Key Laboratory of Power Transmission Distribution Equipment Technology, Hohai University, Changzhou 213022, China
[2] College of Information Science and Engineering, Hohai University, Changzhou 213022, China
[3] Jiangsu Province Hydrology and Water Resources Investigation Bureau, Nanjing 210011, China
[4] Jiangsu Hydraulic Research Institute, Nanjing 210017, China
[5] College of Intelligence and Automation, Hohai University, Changzhou 213022, China
* Correspondence: shipf@hhu.edu.cn

Abstract: With the continuous promotion of smart substations, staff fall detection has become a key issue in automatic detection of substations. The injuries and safety hazards caused by falls among substation personnel are numerous. If a timely response can be made in the event of a fall, the injuries caused by falls can be reduced. In order to address the issues of low accuracy and poor real-time performance in detecting human falls in complex substation scenarios, this paper proposes an improved algorithm based on YOLOX. A customized feature extraction module is introduced to the YOLOX feature fusion network to extract diverse multiscale features. A recursive gated convolutional module is added to the head to enhance the expressive power of the features. Meanwhile, the SIoU(Soft Intersection over Union) loss function is utilized to provide more accurate position information for bounding boxes, thereby improving the model accuracy. Experimental results show that the improved algorithm achieves an mAP value of 78.45%, which is a 1.31% improvement over the original YOLOX. Compared to other similar algorithms, the proposed algorithm achieves high accuracy prediction of human falls with fewer parameters, demonstrating its effectiveness.

Keywords: deep learning; object detection; human falls; intelligent substation; YOLOX; gated non-local convolution (gnConv)

check for updates

Citation: Fan, X.; Gong, Q.; Fan, R.; Qian, J.; Zhu, J.; Xin, Y.; Shi, P. Substation Personnel Fall Detection Based on Improved YOLOX. *Electronics* **2023**, *12*, 4328. https://doi.org/10.3390/electronics12204328

Academic Editor: Andrea Asperti

Received: 15 September 2023
Revised: 10 October 2023
Accepted: 15 October 2023
Published: 18 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Substations play a vital role in the stable operation of the power system, as the hub connecting the transmission and distribution networks. Intelligent detection has been applied in multiple fields [1–3]. With the continuous promotion of smart grids, intelligent monitoring of substations has become a trend [4]. In an industrial environment like a substation, personnel are required to independently perform various tasks, including equipment maintenance, inspection, and troubleshooting. Due to the presence of complex power systems, high-voltage equipment, and various facilities in substations, personnel may face the risk of falling during their operations, which can result in personal injury, equipment failure, or even power outages. Intelligent monitoring of substations should not only include the detection of equipment operation but also ensure the personal safety of the workers, as it is an essential part of ensuring the safe operation of substations [5]. The study of personnel fall detection in substations can not only improve the safety of workers by enabling timely rescue measures but also provide important support for evaluation of risks, improving the work environment and cultivating safety awareness through the analysis of fall event data. This can reduce accident risks and provide a more reliable means for responding quickly to potential hazards.

Currently, there are two main types of fall detection methods: sensor-based methods [6–8] and computer-vision-based methods [9–11]. Sensor-based methods were first applied in the field of fall detection and have been widely researched and applied due

to their low cost, scalability, and flexibility. Sensors can be classified as wearable sensors and environmental sensors. Wearable sensors refer to devices carried on the body that detect fall events by monitoring changes in body motion and posture. This approach requires individuals to continuously wear devices, which can lead to poor user experience. Environmental sensors, on the other hand, are installed in the surroundings and detect falls by monitoring physical changes in the environment. However, this approach has limitations, such as installation position restrictions and complex data interpretation. With the rapid development of artificial intelligence technology, deep learning-based methods such as convolutional neural networks have made significant progress in image and video analysis tasks. These methods provide better real-time performance, do not interfere with the daily activities of workers, and offer higher accuracy and reliability for fall detection. Visually based algorithms first capture images through cameras, then extract relevant human features using object detection models to determine whether a person has fallen [12].

Chen et al. utilized the Mask R-CNN method to detect moving objects on complex backgrounds and proposed an attention-guided bidirectional LSTM model for final fall event detection [13]. Cai et al. designed a vision-based multitask mechanism, achieving accurate fall detection by assigning the secondary task of frame reconstruction and the primary task of fall detection [14]. García et al. employed an LSTM model for time series classification combined with data augmentation and developed a robust and accurate fall detection model [15].

However, most current research is based on experiments conducted in ideal environments, and the robustness of models for complex backgrounds like substations is generally poor. Moreover, these models have high model weights and complex network structures, which fail to meet real-time requirements. Therefore, this paper proposes an improved fall detection model based on YOLOX [16] to address the issues of low detection accuracy and poor real-time performance in the complex scenarios of substations. In the feature fusion part of YOLOX, a custom feature extraction module is implemented to enhance neck feature extraction capability, and a convolutional module is added to the head to improve detection speed, achieving accurate detection of falls in substation environments.

2. YOLOX

YOLOX is a new generation of object detection algorithm proposed by Megvii Technology in 2021. It shows significant improvements in performance compared to its predecessors, YOLOv3 [17], YOLOv4 [18], and YOLOv5. Compared to YOLOv7 [19], which further improves the target regression rate by introducing an anchor box mechanism, YOLOX uses an anchor-free box mechanism to improve the model's computational speed while maintaining detection accuracy. The overall network structure of YOLOX is depicted in Figure 1 [20], consisting of three parts: the backbone network, the feature fusion network, and the prediction heads.

2.1. The Backbone Network

The backbone network of YOLOX adopts the CSPDarknet53 architecture, which is responsible for extracting features from the input image and utilizing these features for subsequent object detection tasks. The basic idea behind CSPDarknet53 is to split the input features into two parts, where one part is processed directly through a series of convolutional layers and the other part is processed after being connected through a CSP block. This approach helps alleviate the gradient-vanishing problem and improves the efficiency of feature propagation.

The input image for detection is resized to a uniform size of 640 × 640 × 3 and fed into the Focus network structure. In this structure, every alternate pixel is selected to obtain one value, which divides the input feature map into four subfeature maps. These four subfeature maps are transposed and concatenated following certain rules to obtain a 320 × 320 × 12 feature map, which is then input into the backbone network for feature

extraction. Finally, three effective feature layers with sizes of 20 × 20 × 512, 40 × 40 × 256, and 80 × 80 × 128 are obtained as inputs for the feature fusion network.

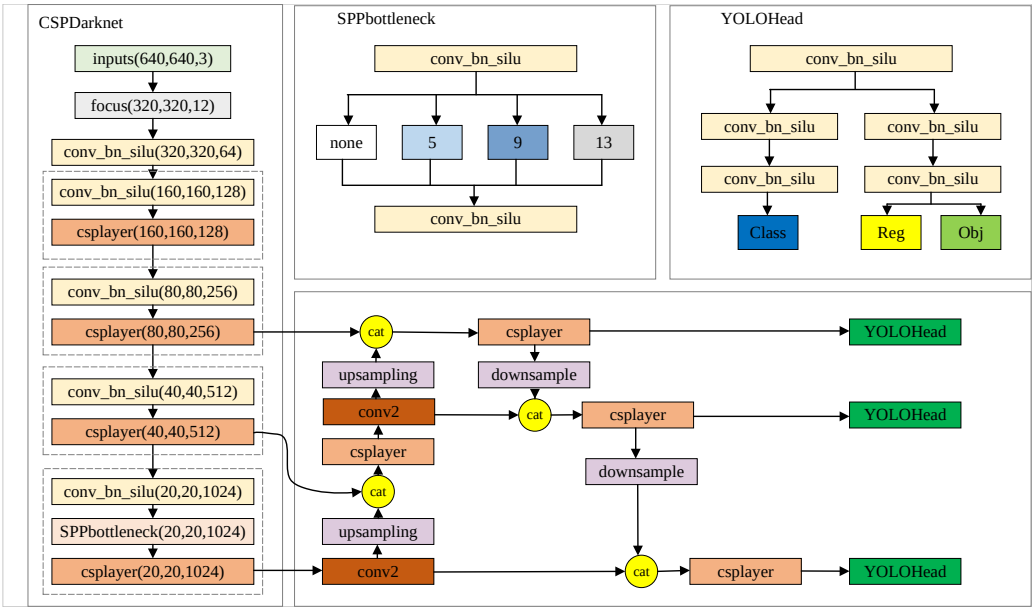


Figure 1. The network structure of YOLOX.

2.2. The Feature Fusion Network

To better fuse multiscale feature information, YOLOX incorporates the FPN (Feature Pyramid Network) [21] algorithm and the PAN (Path Aggregation Network) [22] algorithms as the upsampling and downsampling paths, respectively, in the feature fusion network. In the upsampling path, high-level feature maps extracted from the backbone network are upsampled and added element-wise to adjacent low-level feature maps to achieve cross-level feature fusion. In the downsampling path, low-level feature maps obtained from the upsampling path are downsampled and added element-wise to adjacent high-level feature maps. After passing through the feature fusion network, feature maps of different resolutions obtain rich semantic and positional information, enabling better object detection and localization.

2.3. The Prediction Head

To address the issue of conflicting objectives between classification, regression, and evaluation criteria in traditional object detection algorithms, YOLOX introduces a decoupled head structure. The decoupled head in YOLOX consists of two subheads: a classification subhead and a regression subhead. The classification subhead is responsible for predicting the class probabilities of the objects, while the regression subhead is responsible for predicting the bounding box positions and sizes of the objects. By separating the tasks of object classification and bounding box regression into independent subhead networks, the decoupled head allows them to be learned and optimized independently. Finally, the information is fused and output through concatenation. This design of the decoupled head in YOLOX facilitates more effective information exchange between the two tasks, thus improving the convergence performance and detection accuracy of the model.

3. Improved YOLOX

The complexity of outdoor environments typically found in substations [23] can have a negative impact on image quality and the effectiveness of pedestrian detection algorithms. To improve the accuracy of personnel fall detection, this paper proposes an enhanced YOLOX network structure, as shown in Figure 2. Specific improvements include the addition of a custom feature extraction module, TModule, to the feature fusion network

to enhance the network's feature extraction capability; the addition of recursive gated convolution, gnConv, to the head to facilitate context information fusion and improve detection capability; and the replacement of the original IoU loss function with the SIoU loss function to enhance target localization accuracy.

The main contributions of this paper are as follows:

1. In order to extract rich multiscale features, a feature extraction module is designed in the feature fusion part of YOLOX. This module enhances the neck's feature extraction capability while reducing computational complexity and parameter count. It extracts semantic information that includes diverse characteristics of substation personnel.
2. In the YOLOX head, after the feature map undergoes convolutional normalization and activation functions, gnConv (gated non-local convolution) is introduced. This recursive convolution captures key information from the feature layers, improving the accuracy and speed of the model detection without introducing additional parameters.
3. The smoothed IoU (SIoU) loss function is used to address the problem of the IoU (intersection over union) loss function not considering the angle information of the bounding boxes. By fully considering the influence of angle on model training, the SIoU loss function allows the model to adapt better to targets with different angles and shapes. It provides more accurate position information for bounding boxes and improves the model's regression capability.

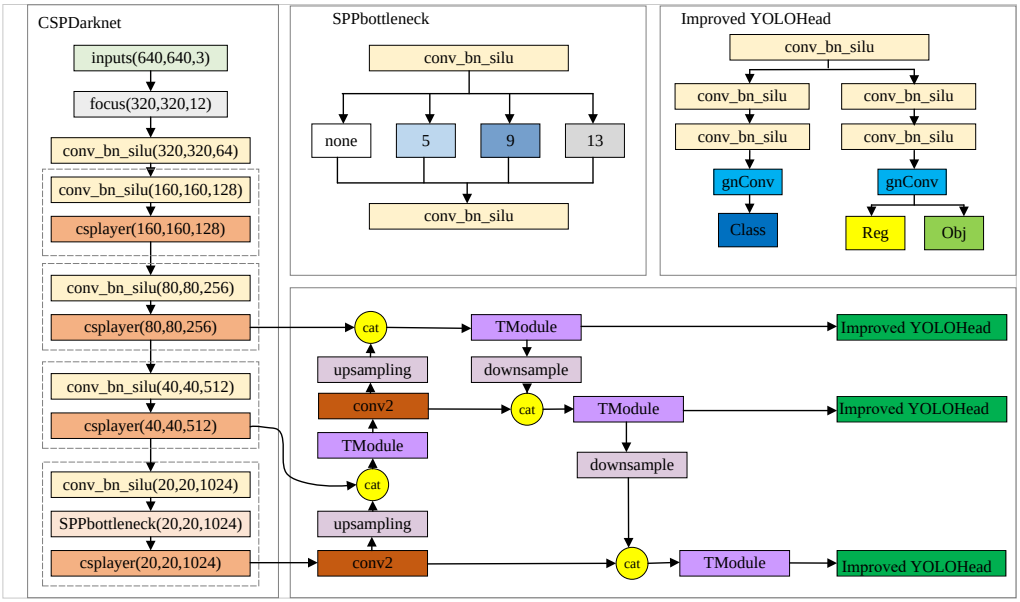


Figure 2. The network structure of Improved YOLOX.

3.1. Tmodule

In the substation scenario, where the background complexity is high, this paper proposes a redesign of the feature extraction module, as shown in Figure 3, to better capture the local features and contextual information of personnel falls. The input of the customized module is first split, and each branch compresses the number of channels by half using a 1×1 convolution. Then, the upper branch continues to split, maintaining spatial invariance of features with a 3×3 convolution and a stride of 1, then stacks with the lower branch. The features are then integrated through a 3×3 convolution and a 1×1 convolution before being stacked and merged with the original branch. Finally, the features are output through a 1×1 convolution. This module is placed in the neck of the feature extraction network, enhancing the feature extraction capability of the convolutional neural network while reducing model complexity.

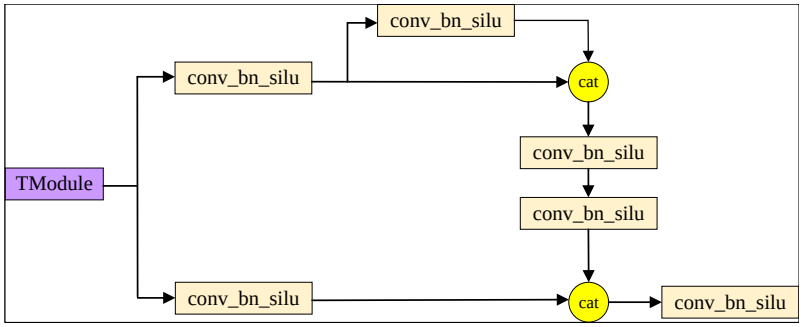


Figure 3. The Network Structure of TModule.

3.2. Gated Non-Local Convolution

gnConv (gated non-local convolution) [24] combines gated convolution and a recursive design to effectively capture the contextual relationship in image data to achieve high-order feature interactions. A schematic diagram of the gnConv structure is shown in Figure 4. The input of gnConv is a feature map with channel C, and after the first layer of convolution, the number of channels doubles. In parentheses, C represents the number of output channels, and the remaining information is represented by *. The convolutional output of the first layer is divided into two parts: the first part is used by the next layer, and the second part is fed into the deep separable convolution to output three parts as inputs for the other three layers. It enhances the feature representation without introducing additional computational complexity.

The input feature map is denoted as x (with dimensions of H × W × C). After passing through a linear layer, we obtain two feature maps: p0 (with dimensions of H × W × C) and q0 (with dimensions H × W × C). Feature map q0 undergoes a depth-wise convolution operation and is then dot-multiplied by feature map p0, resulting in feature map p1. Finally, feature map p1 is processed through a linear layer to produce the output feature map (y). The output of the recursive gated convolution can be represented as follows:

$$\left[p_0^{HW\times C}, q_0^{HW\times C}\right] = \Phi_{\text{in}}(x) \in R^{HW\times 2C} \tag{1}$$

$$p_1 = f(q_0) \odot p_0 \in R^{HW\times C} \tag{2}$$

$$y = \Phi_{\text{out}}(p_1) \in R^{HW\times C} \tag{3}$$

where f represents the depth-wise convolution, and $\odot$ denotes the dot product operation.

In the YOLOX head, after the feature map goes through convolutional normalization and activation functions, the recursive gated convolution is introduced to further extract the crucial information from the feature layers. This improves the accuracy and speed of the model detection.

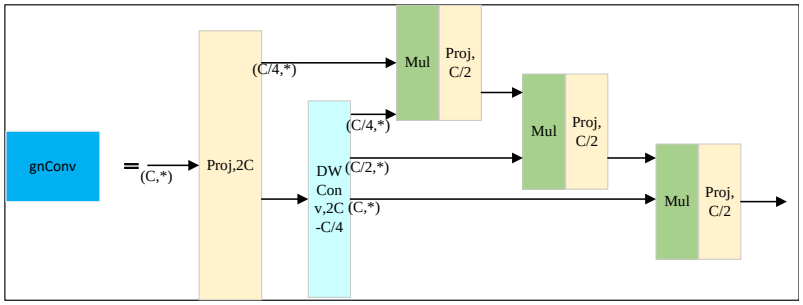


Figure 4. The network structure of gnConv.

3.3. Improvement of Loss Function

In object detection, the definition of the loss function has a significant impact on the final performance of the model [25–27]. In YOLOX, the GIoU (generalized intersection over union) [28] loss function is used as the localization loss function. However, GIoU only considers position and shape information and does not account for the angle loss between the predicted and ground truth bounding boxes. To effectively improve the regression accuracy of the predicted boxes, in this paper, we replace GIoU with SIoU (soft intersection over union) [29].

The SIoU loss function consists of two components:

1. IoU Loss: This component is used to measure the overlap between the predicted box and the ground truth box. It uses the standard IoU (intersection over union) calculation formula to compute the intersection-over-union ratio of the predicted box and the ground truth box and combines it with the target classification loss required in the object detection task.
2. Smooth L1 Loss: This component is used to smooth the process of bounding box regression. It applies the smooth L1 loss function to the difference between the coordinates of the predicted box's bounding box and the ground truth box to mitigate noise and instability during the regression process.

Given a predicted box P and a ground truth box G, the SIoU loss function can be defined as follows:

$$L_{SIoU} = 1 - IoU(P, G) + \frac{\Delta + \Omega}{2} \tag{4}$$

$$IoU(P, G) = \frac{|P \cap G|}{|P \cup G|} \tag{5}$$

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t}\right) \tag{6}$$

$$\Lambda = 1 - 2 * \sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \tag{7}$$

$$\Omega = \sum_{t=w,h} \left(1 - e^{-\omega_t}\right)^{\theta} \tag{8}$$

where $\rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{C_w}\right)^2$, $\rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{C_h}\right)^2$, $\gamma = 2 - \Lambda$, IoU(P, G) represents the intersection-over-union ratio between the predicted box P and the ground truth box G. Δ represents the distance loss, C_w represents the width of the minimum bounding rectangle for the ground truth box and the predicted box, C_h represents the height of the minimum bounding rectangle for the ground truth box and the predicted box, Λ represents the angle loss, $x = \frac{c_h}{\sigma}$, c_h represents the vertical distance between the centers of the ground truth box and the predicted box, σ represents the horizontal distance between the centers of the ground truth box and the predicted box, Ω represents the shape loss, $\omega_w = \frac{|w - w^{gt}|}{\max\left(w, w^{gt}\right)}, \omega_h = \frac{|h - h^{gt}|}{\max\left(h, h^{gt}\right)}$, and $\left(w, w^{gt}\right)$ represent the width of the predicted box and the ground truth box, respectively. Similarly, $\left(h, h^{gt}\right)$ represents the height of the predicted box and the ground truth box, respectively.

4. Dataset and Experimental Platform

The person falling dataset is a crucial component for training, evaluating, and improving the fall detection model. It provides the model with learning material and validates and optimizes the model during the training process [30,31], enabling the fall detection model to better learn the features of the target. To solve the problem of limited scale and inability to cover various situations and changes in the current fall detection dataset in the field of substations, we comprehensively utilized an open source fall detection dataset (https://aistudio.baidu.com/aistudio/datasetdetail/94809, accessed on 25 November 2021) in Baidu AIStudio and self-made substation scene fall data. In the experiment,

the data were expanded using methods such as horizontal flipping, random cropping, and angle rotation. There are a total of over 7000 datasets that cover various indoor and outdoor scenes, as shown in Figure 5; some typical fall scenarios are presented as references for evaluation of the performance of the model in real situations.

The dataset was annotated with targets using the LabelImg tool, and after annotation was completed, it was saved as an xml file, which was then converted into VOC2007 data format. In order to improve the generalization ability of the network model, 90% of the images were used for model training in the experiment, and the remaining 10% of the data was used to verify the model performance. The dataset used for model training was divided into a training set and a validation set in a 9:1 ratio. The function of the training set is to set the parameters of the classifier and regressor, then train the classification and regression algorithms and, finally, fit multiple classification regressors for the fall detection algorithm. The function of the validation set is to identify the algorithm weights with the highest recognition accuracy, detect the weights of each trained algorithm, record the algorithm accuracy, and select the weight parameters corresponding to the algorithm with the highest accuracy. The function of the test set is to predict the optimal algorithm obtained from the training and validation sets and measure the effectiveness of the algorithm.

We used the PyTorch framework to train the network model, with a total of 300 epochs trained. In the network model, the input size is 640 × 640. The server configuration used in this study is presented in Table 1.

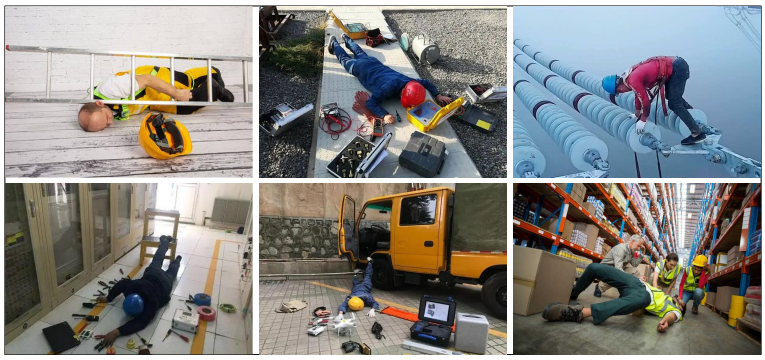


Figure 5. Partial dataset images of personnel falling in substations.

Table 1. Detailed environmental configuration.

Operating System	Ubuntu 20.04
CPU	i9-12900K CPU
GPU	NVIDIA RTX 3090
Random Access Memory	64.00 GB
Deep Learning Framework	Pytorch
Integrated Development Environment	VSCode
Programming Language	Python3.7

5. Experimental Results and Analysis

5.1. Evaluation Metrics

In order to further evaluate the detection accuracy of the model in this article, indicators such as precision (P), recall (R), average precision (AP), and mean average precision (mAP) were selected for evaluation. AP and mAP avoid the impact of unequal confidence levels in different models on evaluation and can be used for the vast majority of models in the field of object detection. The mAP is the average value of AP across all classes. We used mAP to calculate the mean accuracy of each category corresponding to the specified intersection over union in the fall detection model. The mAP value ranges from 0 to 1; a higher value indicates better performance of the object detection algorithm across multiple classes. The calculation formulas for AP and mAP are shown as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{11}$$

$$mAP = \frac{\sum_i^k AP_i}{k} \tag{12}$$

where TP (true positive) represents the targets that were originally labeled as positive samples and also predicted as positive samples by the model, FP (false positive) represents the targets that were originally labeled as negative samples but predicted as positive samples by the model, FN (false negatives) represents the targets that were originally labeled as positive samples but predicted as negative samples by the model, P denotes precision, and R represents recall. AP is obtained by calculating the area under the precision–recall curve, k represents the total number of categories, and mAP is derived by averaging the AP values across all classes.

5.2. Model Training

The process of model training is essentially the process of fitting model parameters. In this study, the model utilizes Adam (adaptive moment estimation) as the optimizer. A total of 300 epochs were trained, divided into two steps:

In the first step, the parameters of the backbone network are frozen to expedite the training process. The learning rate is set to 0.001, and the batch size is set to 32.

In the second step, the parameters of the backbone network are unfrozen to fully learn the features of the detection targets and achieve better convergence. The learning rate is set to 0.0001, and the batch size is set to 16. To prevent the model from getting stuck in local optima, a cosine annealing decay schedule is employed for learning rate adjustment.

The changes in the loss function throughout the entire training process are shown in Figure 6. It can be seen that at the 140th epoch, the network tends to converge, the loss function changes smoothly, and the fluctuation amplitude is not significant, indicating that the improved model has the best training effect.

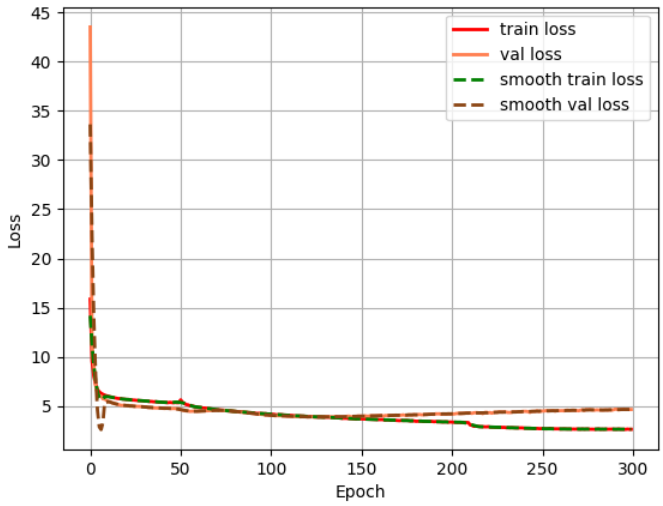


Figure 6. Changes in the total loss function.

5.3. Test Results

In order to demonstrate the effectiveness of the proposed method, it was experimentally compared with four classic object detection methods: Faster RCNN [32], YOLOv5, YOLOv7, and YOLOX. Faster RCNN is a classic two-stage detection algorithm that generates candidate boxes through a region recommendation network, then performs target classification and boundary box regression. YOLOv5, YOLOv7, and YOLOX are classic single-stage detection algorithms. The comparative experimental results are shown in Table 2. The results demonstrate that the mAP detection accuracy of the improved algorithm presented in this paper reaches 78.45%, which is an improvement of 9.32% over Faster-RCNN, 3.98% over YOLOv5, and 0.27% over YOLOv7. Moreover, the improved algorithm significantly enhances the detection speed of the model. Compared to the original baseline model, the detection accuracy is increased by 1.31%, while only adding an additional 0.1 M parameters. The effectiveness of the algorithm proposed in this paper is verified through comparisons with mainstream object detection algorithms.

Table 2. Comparative experimental results.

Model	mAP/%	Params (M)
Faster-RCNN	69.13	28.296
YOlOv5	74.47	7.06
YOLOX	77.14	8.938
YOlOv7	78.18	40.329
Ours	78.45	9.045

5.4. Ablation Experiments

To evaluate the impact of each improvement strategy adopted in this paper on the detection performance of YOLOX, ablation experiments were conducted on the dataset, as shown in Table 3. In the table, Model A represents the addition of SIoU on top of the base model, Model B represents the addition of TModule on top of Model A, and Model C represents adding SIOU and gnConv on top of the base model. From the data in the table, it can be observed that the inclusion of the SIoU loss function improvement strategy leads to a 0.16% increase in mAP. Building upon this, the addition of the designed TModule further improves mAP by 0.96%. Finally, with the inclusion of the gnConv improvement strategy, there is a further increase in mAP. The results of the ablation experiments demonstrate that the improved algorithm results in an overall mAP enhancement of 1.31% compared to the original baseline model. Additionally, they validate that the improvement strategies proposed in this paper effectively enhance the detection accuracy of pedestrian falls in substation scenarios.

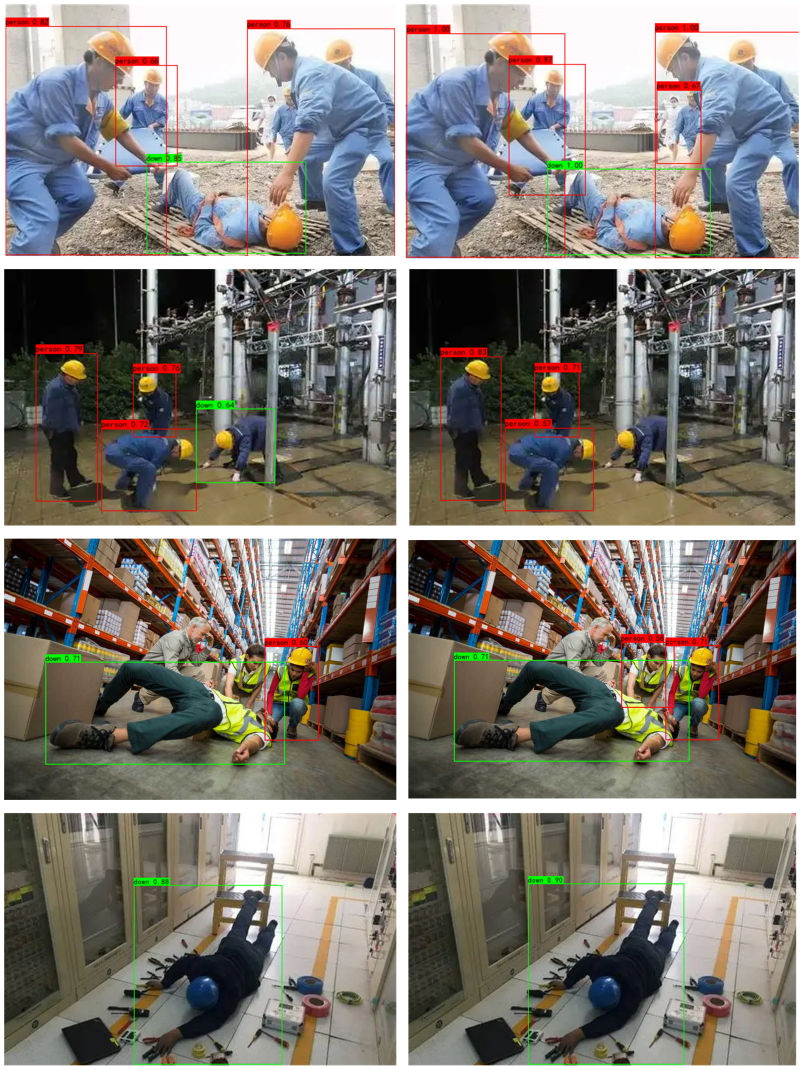
Table 3. The Results of ablation experiments.

Model	SIoU	TModule	gnConv	mAP (%)
Base Model				77.14
A	✓			77.30
B	✓	✓		78.26
C	✓		✓	77.43
Ours	✓	✓	✓	78.45

5.5. Visualization of Detection Results

In this study, the detection performance of the original YOLOX model and that of the algorithm proposed in this paper were visually compared, as shown in Figure 7. The improved algorithm performs better than the baseline model, which addresses the issues of false negatives and false positives in the original algorithm. The benchmark model often fails to detect small target personnel, such as those located at a distance, mainly due to insufficient feature extraction of the target, insufficient attention to the target in complex

backgrounds, and the inability to solve the problem of large differences in target scales among substation personnel. By improving the algorithm, the localization and feature extraction capabilities of multiscale targets in complex backgrounds can be enhanced, thereby generating more accurate detection frames and alleviating the situation of missed detections to a certain extent.



(a) Original YOLOX Algorithm (b)Ours

Figure 7. Comparison of detection results before and after improvement.

6. Conclusions

This article proposes an improved algorithm based on YOLOX to address the issue of low detection accuracy in personnel fall detection in actual substation working environments. The algorithm designs a feature extraction module in the YOLOX feature fusion section, enhancing the neck feature extraction ability. By optimizing the loss function and adding a recursive gated convolution module at the head, the detection speed is improved, resulting in better model convergence and regression performance during the training process, as well as accurate detection of personnel falling in substation scenarios. The experimental results show that compared with the original algorithm, the improved algorithm proposed in this paper is associated with an increase of 1.31% in mAP. The improved algorithm has more advantages in balancing parameter quantity and accuracy. Although it adds fewer parameter quantities, mAP is the best, indicating that it can improve the detection accuracy of multiscale targets in substations while meeting the real-time detection requirements at the expense of a certain detection speed, indicating the effectiveness of

the improved algorithm. In future research, we will attempt to prune and quantify the algorithm before transplanting it to the development board to achieve real-time detection of on-site terminals.

Author Contributions: X.F. designed the analysis, designed the research experiment, and wrote the original and revised manuscript; Q.G. conducted data analysis and was responsible for details of the work; R.F. verified data and conducted statistical analysis; J.Q. and J.Z. collected the data and conducted the analysis; and Y.X. and P.S. verified image data analysis and guided the direction of the work. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Changzhou Sci & Tech Program (CE20235053), in part by the Key Project of Jiangsu Provincial Key Laboratory of Transmission and Distribution Equipment Technology Team (2023JSSPD01), and in part by the Fundamental Research Funds for the Central Universities (B220202020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data can be shared up on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, Y.C.; Tsai, H.W.; Huang, C.Y.; Wu, Z.R. Based-on Computer Vision Applications for Bus Stop Passenger Detection System. In Proceedings of the 2023 IEEE 3rd International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB), Taichung, Taiwan, 15–17 April 2023; pp. 152–154.
2. Wang, X.; Wu, J.; Zhao, J.; Niu, Q. Express Carton Detection Based On Improved YOLOX. In Proceedings of the 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 16–18 December 2022; Volume 5, pp. 1267–1272.
3. Cai, X.; Ding, X. A comparative study of machine vision-based rail foreign object intrusion detection models. In Proceedings of the 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 29–31 January 2023; pp. 1304–1308.
4. Tang, W.; Chen, H. Research on intelligent substation monitoring by image recognition method. *Int. J. Emerg. Electr. Power Syst.* **2020**, *22*, 1–7. [CrossRef]
5. Wang, S. Substation Personnel Safety Detection Network Based on YOLOv4. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 26–28 May 2021; pp. 877–881.
6. Chen, H. Design of Intelligent Positioning Shoes for Elderly Fall Monitoring Based on GPS and MPU-6000 Acceleration Sensor. In Proceedings of the 2022 International Conference on Wearables, Sports and Lifestyle Management (WSLM), Kunming, China, 17–19 January 2022; pp. 43–46.
7. de Quadros, T.; Lazzaretti, A.E.; Schneider, F.K. A movement decomposition and machine learning-based fall detection system using wrist wearable device. *IEEE Sensors J.* **2018**, *18*, 5082–5089. [CrossRef]
8. Rachakonda, L.; Marchand, D.T. Fall-Sense: An Enhanced Sensor System to Predict and Detect Elderly Falls using IoMT. In Proceedings of the 2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Nicosia, Cyprus, 4–6 July 2022; pp. 448–449.
9. Feng, Y.; Wei, Y.; Li, K.; Feng, Y.; Gan, Z. Improved Pedestrian Fall Detection Model Based on YOLOv5. In Proceedings of the 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Beijing China, 3–5 October 2022; pp. 410–413.
10. Chutimawattanakul, P.; Samanpiboon, P. Fall detection for the elderly using yolov4 and lstm. In Proceedings of the 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Huahin, Thailand, 22–27 May 2022; pp. 1–5.
11. Dey, A.; Rajan, S.; Xiao, G.; Lu, J. Fall event detection using vision transformer. In Proceedings of the 2022 IEEE Sensors, Dallas, TX, USA, 30 October–2 November 2022; pp. 1–4.
12. Zhou, L.; Li, W.; Ogunbona, P.; Zhang, Z. Jointly learning visual poses and pose lexicon for semantic action recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 457–467. [CrossRef]
13. Chen, Y.; Li, W.; Wang, L.; Hu, J.; Ye, M. Vision-based fall event detection in complex background using attention guided bi-directional LSTM. *IEEE Access* **2020**, *8*, 161337–161348. [CrossRef]
14. Cai, X.; Li, S.; Liu, X.; Han, G. Vision-based fall detection with multi-task hourglass convolutional auto-encoder. *IEEE Access* **2020**, *8*, 44493–44502. [CrossRef]

15. García, E.; Villar, M.; Fáñez, M.; Villar, J.R.; de la Cal, E.; Cho, S.B. Towards effective detection of elderly falls with CNN-LSTM neural networks. *Neurocomputing* **2022**, *500*, 231–240. [CrossRef]
16. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
19. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
20. Zhou, L.; Zhong, H.; Chen, G. Improved YOLOX Pedestrian Fall Detection Method Based on Attention Mechanism. *Chin. J. Electron Devices* **2023**, *46*, 404–413.
21. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
22. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
23. Lu, S.; Zhang, Y.; Su, J. Mobile robot for power substation inspection: A survey. *IEEE/CAA J. Autom. Sin.* **2017**, *4*, 830–847. [CrossRef]
24. Rao, Y.; Zhao, W.; Tang, Y.; Zhou, J.; Lim, S.; Lu, J. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *arXiv* **2022**, arXiv:2207.14284.
25. Chen, Y.; Zhang, B.; Li, Z.; Qiao, Y. Ship Detection with Optical Image Based on Attention and Loss Improved YOLO. In Proceedings of the 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 22–24 July 2022; pp. 1–5.
26. Du, S.; Zhang, B.; Zhang, P. Scale-Sensitive IOU Loss: An Improved Regression Loss Function in Remote Sensing Object Detection. *IEEE Access* **2021**, *9*, 141258–141272. [CrossRef]
27. Zhang, C.; Xiong, A.; Luo, X.; Zhou, C.; Liang, J. Electric Bicycle Detection Based on Improved YOLOv5. In Proceedings of the 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC), Suzhou, China, 22–24 April 2022; pp. 1–5.
28. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
29. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
30. Shi, W.; Han, X.; Wang, X.; Li, J. Optimization Scheduling Strategy with Multi-Agent Training Data Rolling Enhancement for Regional Power Grid Considering Operation Risk and Reserve Availability. In Proceedings of the 2023 8th Asia Conference on Power and Electrical Engineering (ACPEE), Tianjin, China, 14–16 April 2023; pp. 1774–1781.
31. Xu, Y.; Goodacre, R. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test.* **2018**, *2*, 249–262. [CrossRef] [PubMed]
32. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.