Lightweight YOLOv7 Real-Time Insulator Power Equipment Defect Detection Based on Attention Improvement

Tao Wang*[1a], Jingfeng Xiao[2b], Xinxin Meng [1c], Wenzhong Yang[3d]

[1] Information and Communication Company of State Grid Xinjiang Electric Power Co,Xinjiang, Urumqi 830000;
[2] Science and Technology Digitalization Department of State Grid Xinjiang Electric Power Co,Xinjiang, Urumqi 830000;
[3] Xinjiang University, College of Information Science and Engineering,Xinjiang, Urumqi 830046;

*[a]Corresponding author: shi514121@163.com
[b]Contributing authors: 286491379@qq.com;
[c]1731224941@qq.com; [d]yangwenzhong@xju.edu.cn;

Abstract. efective faults in insulator power equipment can affect transmission equipment's normal operation and electricity consumption in the service area. To reduce or avoid transmission faults caused by defective insulator power equipment failures, structural defects of insula-tors need to be detected. In contrast, different insulator defects have significant differences in style and size. The defective parts in the UAV scenario have problems such as blurring, obscuration, and environmental factors, which lead to challenging insulator power equipment defect detection. Therefore, we propose a model specifically for detecting insulator defects in power equipment - GSNA-YOLOv7. We added GSConv to improve the New-Neck module to better balance the inference speed of the model with the detection accuracy of defective targets of power equipment, reduce the redundant information of the model, and better achieve the effect of real-time detection; improve the DownNAM module, introduce the attention mechanism, apply the weight sparsity penalty, stabilize the performance and computational efficiency, and make the model pay more attention to the defective small target information. The SFID insulator dataset and Visdrone2021 UAV dataset are trained and validated. The experimental results are analyzed, concluding that GSNA-YOLOv7 has a better detection effect for power equipment defect detection in the UAV shooting scenario and is more adaptable to detecting small targets in insulator fault defect datasets. The method is better than many existing insulator defect detections. The method outperforms numerous current insulator defect detection methods, with mAP improved by 0. 9% and 0.6% and parameter volume reduced by 0.2G, compared with the base model YOLOv7.

Keywords: Defect detection; insulator defects; small target detection; attention mechanism; power equipment

G. Guan et al. (eds.), *Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023)*, Atlantis Highlights in Computer Sciences 16,
https://doi.org/10.2991/978-94-6463-264-4_82

1 Introduction

Insulators in power equipment generally exist outdoors, making the equipment suffer from environmental influences such as temperature rain, which leads to problems such as electrical deflagration and material aging, producing defects such as self-detonation, contamination, and insulator cracks. The staff cannot go through the data of the aggregated insulator power equipment to discover the defects of the power equipment insulators in the first place, which will cause severe accidents and substantial economic losses, so the defect detection of insulators on transmission lines is critical. Defect detection is a special sub-problem and an essential process in the defect management of power equipment. Among the defect detection data sets, visually similar normal and abnormal samples exist that are only slightly different. Although traditional anomaly detection methods are well suited for data with high intraclass variance, they cannot capture subtle differences. We address this problem by identifying image features extracted in convolutional neural networks using network detection. With the growth in the use of insulator power equipment, the application of image recognition techniques with deep learning in insulator defect recognition has become one of the important research directions for defect detection. In industrial production, the quality of products is constantly monitored and improved. Therefore, there is a need to detect minor defects using insulators reliably and to take out defects in images individually for identification to achieve classification and detection of various types of defects [7][14].

To cope with the above problems related to defect detection, we propose a one-stage attention mechanism-based lightweight real-time defect detection network, GSNA-YOLOv7. In which we design a New-Neck novel lightweight structure, we replace the original convolution with GSConv 17, which preserves the original connection as much as possible, in terms of the model's insulator defect detection accuracy and inference speed are balanced to ensure real-time detection of the insulator defect part. Moreover, we add the NAMAttention 18 attention mechanism in the multi-scale fusion stage to improve the detection of insulator power equipment defects by improving the global information interaction of feature information, reconstructing the overall components of the network, and reducing redundant information, thus improving the problem of defect ambiguity, occlusion, and environmental factors that affect the detection of insulator power equipment defects in power equipment.

In summary, the contributions made in this paper are as follows.

- We designed a lightweight attentional real-time defect detection model, improving insulator defect detection's effectiveness and accuracy.
- We designed a lightweight New-Neck structure that reduces the computational cost and improves the inference speed of the network by replacing the original convolution module and adding deep convolution and dense convolution to generate feature information without model compression.
- We designed the attention pooling module to increase the global feature interaction fusion in the multi-scale fusion stage by adding the global information interaction

attention mechanism to enhance the model's attention to insulator defect sites and address the influence of environmental factors on defect detection.

- Tested on the SFID insulator dataset and the UAV Visdrone2021 dataset, our model reduces the model computation by 0.2G, achieves an FPS of 165, and improves the mAP values by 0.9% and 0.6%, respectively, compared to the YOLOv7 26 model.

2 Related Work

2.1 Target Detection

In computer vision, good progress has been made in target detection. The model based on a single-stage strategy directly regresses the classification and localization of targets, which initially reaches the requirements of real-time compatibility, etc.

One-stage detectors can be designed based on anchoring.YOLOv3 27, SSD [28], and RetinaNet 33 place anchor boxes densely on the feature map and then predict object classes and anchor box offsets.VFNet 29 and RepPoints 30 are anchorless detectors that indicate critical points, such as corner or center points, to form enclosing boxes. Compared with single-stage networks, two-stage methods such as Faster RCNN 4, Cascade R-CNN 31, R-FCN [16], and Dynamic RCNN 32, first generate region suggestions to distinguish foreground from background, and thus inference is slower. However, the improved design improves the detection performance and makes it more suitable for high-precision scenes. The R-CNN [1] model abandons the traditional violence detection method by finding the marquee regions of the input feature picture for the target to be detected, extracting the feature vectors, fine-tuning the boxes, and then determining the class of the candidate regions by the classifier. The Fast R-CNN [3] model does not need to refine the characteristics of each participating region separately. FasterR-CNN adds a region generation network to the Fast R-CNN model, improving the overall computational inference speed and the accuracy of detecting targets. The end-to-end YOLO and single-point multi-frame detector models represent the regression-based single-stage detection model.YOLO detects target objects faster than classical models of the same period and only needs to input images into the neural network to complete detection through a single stage. The SSD model incorporates the above models' regression plus candidate region mechanism and uses multi-scale feature maps for detection and convolution. For detection, set up a priori boxes in three steps. The disadvantage is that it requires artificial adjustment of the parameters of the box, the recall rate for small targets to be detected could be more effective, and there needs to be more feature extraction. Through the improvement and development of YOLO, its network performance is gradually applicable to most of the target detection, and it can also complete the identification and detection of the target to be detected well, completing the two-way improvement of real-time and accuracy.

2.2 Power Equipment Defect Detection

The defect detection of power equipment has been integrated into deep learning methods in recent years by using convolutional neural networks to identify different types of defects in power equipment. The YOLO algorithm has a faster inference speed and recognition speed compared to other baseline network models of the same period, which can better reflect the real-time requirements of defect detection in power equipment, and the network also has a better effect on defect detection accuracy. yolov1 of 23 is to input the originally captured image directly into the YOLO network and output the information of the frame directly. In YOLOv2 5, new features such as batch normalization, high-resolution classifier, anchor box convolution, dimensional clustering, and Dark-net-19 network are added. yolov3 [27] uses a fully convolutional network by dividing image regions, predicting the probability of bounding boxes, and then predicting the probability of each region. The main innovations of YOLOv4 24 are mosaic data enhancement, self-adversarial training, and normalization across small batches. However, the computational volume of the model and the inference speed required are not up to real-time, making CNNs unavailable for use in industrial environments such as power systems. To alleviate this problem, many lightweight networks have been proposed. PP-YOLOE 15 is a lightweight network structure to improve the first stage of YOLO proposed by Baidu in 2022. SqueezeNet 8 proposes a new Fire module.MobileNet 192021 series of networks have model compression with deeply separable convolution. An inexpensive operation was introduced in GhostNet 6, which has fewer parameters while obtaining the same number of feature mappings.

The detection performance of YOLOv7 26 is higher than any other known classical target detector, and it improves both the inference speed of the model network and the detection accuracy of the target features.YOLOv7 improves the different connections of the branching streams through model compression, which improves the detection performance of the defective targets to be detected in power equipment without increasing the inference speed, network depth, and computational complexity. The detection performance of the defect targets detected in power equipment is greatly improved by real-time detection. By using the YOLOv7 base model, major and minor defects in power equipment can be better detected, and real-time detection can be better achieved.

2.3 Multi-scale feature fusion

In the past, target detection development and multi-scale fusion were particularly good ways to improve model performance. The feature maps in the lower feature layers in the network model have stronger semantic information, lower resolution, and very poor attention to target feature details compared to the lower feature layers; the opposite is true for the higher layers. According to this problem, multi-scale fusion methods follow. The image pyramid model is the original method of multi-scale fusion, which creates different scales of slabs by convolution, with the disadvantage that it cannot satisfy real-time and that the inference speed and computational complexity are too large. On this basis, ASPP [9][13] performs extended convolution operations at different rates on the input feature map, consistent with the ordinary pooling layer, to extract as many features as possible. Based on ASPP, RFBNet 10 adds different sizes of convolution to enhance the network model's scale variability and learn features. FPN networks are fused by connecting different feature layer branches and are not computationally intensive.PANet 25 adds a bottom-up modular structure to the network structure of FPNs, which further improves the characteristics of FPNs and enriches the information for feature fusion. And AF-FPN [22] adds Adaptive Attention Module (AAM) and Feature Enhancement Module (FEM) to the traditional feature pyramid network. In 2020, the DRFPN [34] of CAS designed a new parameter-free feature pyramid network from an attentional perspective. This dual refinement feature pyramid network consists of two modules: the spatial refinement block (SRB) and the channel refinement block (CRB). The SRB is based on the context between adjacent levels. The CRB learns an adaptive channel merging method based on an attention mechanism.

In this paper, by adding the NAMAttention 18 mechanism to the neck structure, adding weighting factors from both channel and space dimensions, focusing on contextual information while increasing the weighting ratio of small target information for different feature layers, and combining feature fusion, the network's detection of small target defect features present in power equipment is greatly improved.

3 Methods

3.1 GSNA-YOLOv7 Network Structure

In this section, to solve the problem of small target defects in the insulator power equipment structure, we design the insulator defect detection model GSNA-YOLOv7, as shown in Figure 1, through the original input power equipment defect feature map, after the backbone network uses a multi-branch stacking module, four feature layers in the stack will again be a convolutional normalized activation function to feature integration, for the initial The New-Neck structure, as shown in Figure 2, first extracts multi-scale features in one direction by replacing part of the original convolution, and fuses them; the DownNAM module adds a feature global information attention mechanism to input the fused multi-scale features into the detection head for classification and regression operations. The New-Neck structure, shown in Fig. 2, utilizes GSConv to replace part of the original convolution and generates two new feature layers after two fusions so that the model better meets the requirements of real-time electronic equipment defect detection while maintaining the target detection accuracy of the electrical equipment; the DownNAM module is a modified PAFPN multi-scale pyramid fusion in Neck structure, fusing two feature maps of different scales Ni and Fi+1 to get a new Ni+1 feature map, and its structure is shown in Figure 3. In Part II and Part III of this section, we will offer comprehensive insights into the GSConv and DownNAM modules.

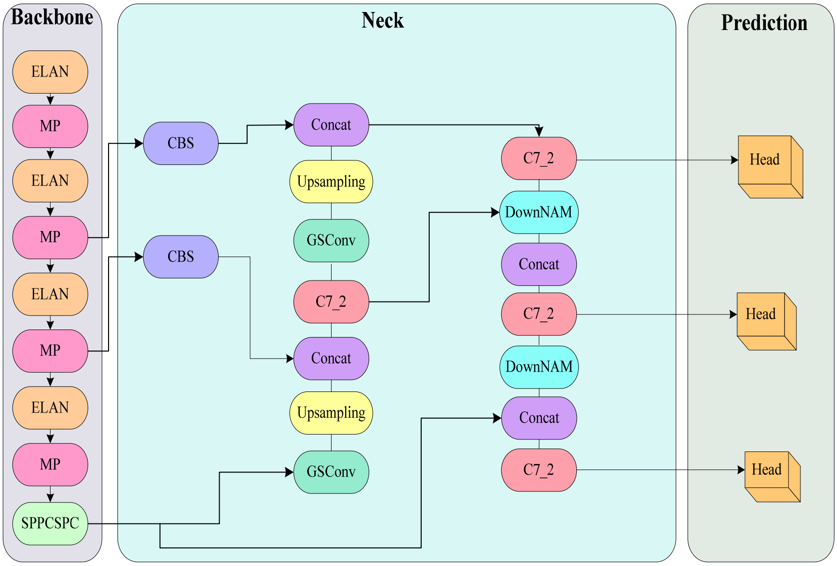


Fig. 1. GSNA-YOLOv7 model structure diagram.It includes the YOLOv7 baseline model, the New-Neck structure, and the DownNAM module.

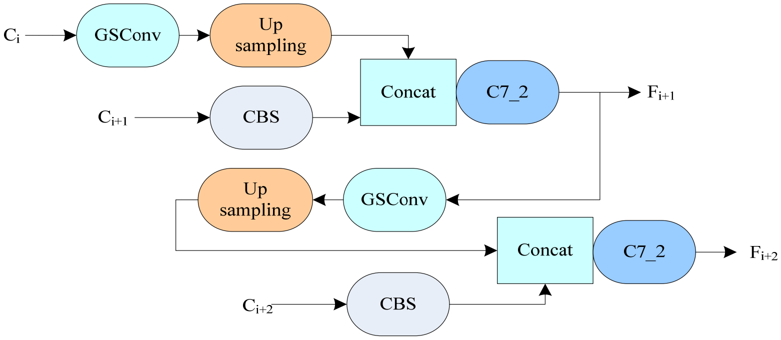


Fig. 2. New-neck detail diagram. First extracts multi-scale features Ci (i=0, 1, 2) in one direction, replaces part of the original convolution using GSConv, and the new feature layer obtained after the up-sampling operation is fused with the Ci+1 feature layer of the extracted features of the backbone network in a fusion operation, and outputs a new feature layer, Fi+1, through a multilayer convolution, and after two fusions to generate two new feature layer.

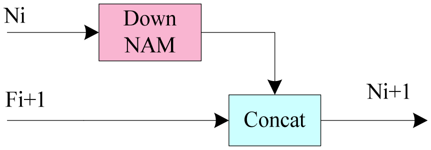


Fig. 3. Detailed diagram of the process of adding YOLOv7multi-scale features fusion by DownNAM module. Ni indicates the shallow feature map generated by the top-down path of the PAFPN structure, Fi+1 indicates the deep feature map generated by the bottom-up path of the PAFPN structure, and Ni+1 indicates the new feature map after multi-scale feature fusion.

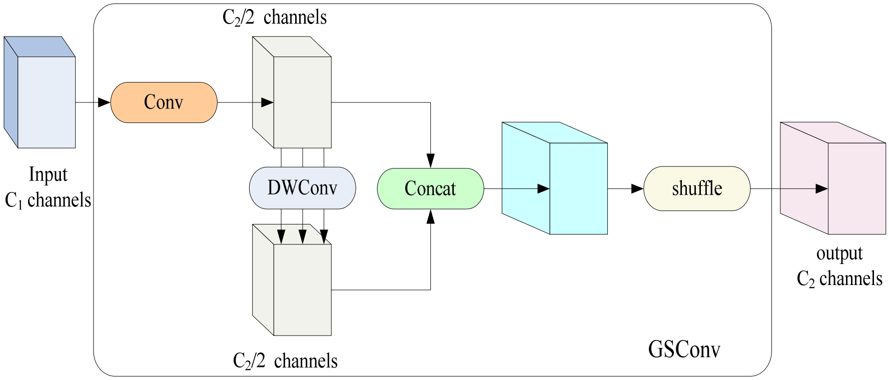


Fig. 4. Detail diagram of GSConv. The DWConv (deep convolution) [11] further reduces the computational cost and the parameter size of the large core, which facilitates the network to run on mobile devices. The original channel is fused with the channel after deep convolution. The new channel is fused with the feature map obtained from standard convolution (SC) and DSC by using shuffle so that the output of a dense convolution is fused into the deep separate convolution and finally output.

3.2 New-neck structure

The original conv is replaced by GSConv 17 in YOLOv7 26. The GSConv is introduced into YOLOv7 to get the new-Neck structure to complete the defect detection of insulator power equipment in real-time and to keep the model detection effect stable. The model's accuracy in detecting the target is guaranteed with decreased computational complexity and inference speed. GSConv exists as part of the gradient flow branch of the original convolution; if GSConv replaces the original convolution in both the backbone network and the neck layer of the model, the depth of the network will be significantly increased, which will cause an increase in the resistance of data transmission, thus making the speed of inference much higher. When the different feature layers' output from the backbone network is passed to the neck layer, the convolutional size has reached its maximum and does not need to be changed. Therefore, we chose to use GSConv only in the neck layer. At this stage, it is good to use GSConv to process the connected feature maps: no compression of model blocks is needed, and the amount of redundant information repetition is meager while having a better effect on the channel dimension learning to generate the weights part. This is shown in Figure 4. The addition of the shuffle module will allow the feature information generated by the dense convolution operation to be present at all different locations caused by the depth-separated convolution, which makes the output of the dense convolution operation as similar as possible to that of the normal convolution calculation and reduces the inference speed and computational cost of the model. It makes the defect detection of power equipment a real-time effect, increases the inference speed of the model, and reduces the complexity of the model.

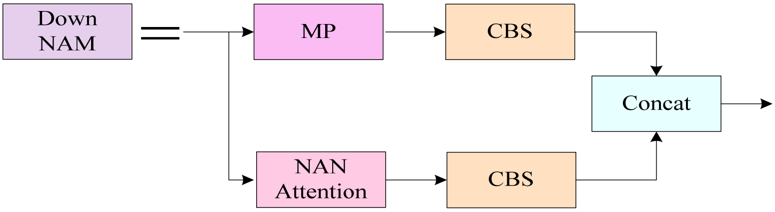


Fig. 5. Detail diagram of the DownNAM module. By adding the NAMAttention attention mechanism to the DownC module in the multi-scale fusion stage of YOLOv7, the scaling factor can show the degree of feature information in different channel dimensions and also illustrate the strength of the information representation target between channels.

3.3 DownNAM module

In the multi-scale feature fusion stage of YOLOv7, small target feature information is easily missing during transmission, and the model needs to pay more attention to small target features.The NAMAttention 18 attention mechanism, through the connection of channel information and spatial information between different submodules, enhances the information interaction between different dimensions in both ways. Therefore, we improve the DownNAM module, as shown in Figure 5, where we introduce this new module for hetero-dimensional interactions in the multi-scale fusion of the Neck structure of YOLOv7, where the global scheduling mechanism reduces the data information approximately. The global network information interaction is enhanced to improve the performance of the network model for detecting target defects in power equipment. The method can suppress insignificant weights and improve the detection effect. As shown in Figure 6. Adding light penalty weights balances the similarity degree and the detection performance in both directions. As shown in Equation (1), to calculate the weights of channel attention, which is used to represent the final output characteristics, γ is the scaling factor of each channel; to calculate the weights of spatial attention, a homogeneous normalization layer is added to calculate the pixel-by-pixel of each feature map in the spatial dimension, as shown in Equation (2).

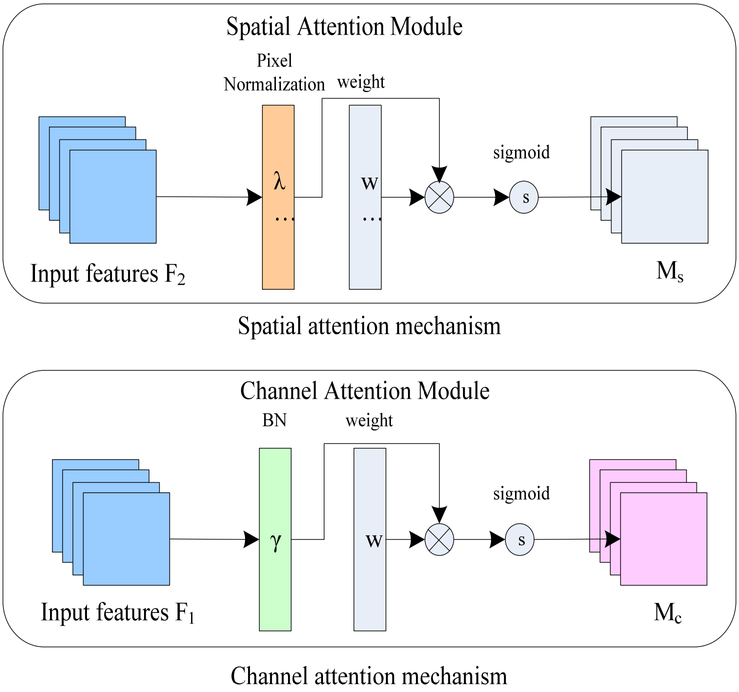
$$\mathrm{M_c} = \mathrm{sigmoid}(\mathrm{W_\gamma}(\mathrm{BN}(\mathrm{F_1}))) \tag{1}$$

Where M_c is network output features, γ represents the scaling factor of each channel.

$$M_s = \mathrm{sigmoid}(W_\lambda(BN_s(F_2))) \tag{2}$$

Where M_s represents network output features, W represents network weights; λ represents the scaling factor.

There is no redundant amount of information computation, such as full connectivity, etc., the characterization of defective features of devices that do not need to be detected is reduced by the addition of regularization terms, and the scaling in BN is directly used in the calculation of attention weights, which are missing in other attention methods. The scaling factor is the variance in BN; the larger the variance, the more the channel changes; therefore, the information in that channel will be more prosperous and critical. The NAMAttention attention mechanism is added to the multi-scale fusion module of YOLOv7 to focus on small and medium target information of electrical equipment defects in the feature layer with fewer parameters, thus enabling the network to learn critical information better.



Spatial attention mechanism

Channel attention mechanism

Fig. 6. Details of the two submodules Spatial and Channel. γ and λ in the two submodules are scaling factors, respectively. The scaling factor of the BN layer is added to the spatial dimension to handle pixel normalization, to indicate the importance of weights, and thus to measure the importance of spatial features; the weight contribution factor is added to enhance the effect of attention.

4 Experiments

4.1 Experimental dataset

We use the insulator SFID dataset to evaluate our network. The insulator SFID dataset has 13,000 insulator training images and more than 2,700 insulator validation images, containing two classes of insulators and their defective components, on whose training and validation sets we obtained detection results. At the same time, we trained and tested the detection of our model for the UAV scenario using the UAV public dataset Visdrone2021, which contains UAV Ten categories of UAVs, 6471 training images, and 548 validation images captured.

4.2 Experimental equipment and details

We perform all experiments on eight NVIDIA A40 GPUs. All models are based on the deep learning framework PyTorch 1.10.0. we set the momentum to 0.937, the initial learning rate to 0.01, and the weight decay to 0.0005, using a stochastic gradient descent (SGD) [12] optimizer. The total epoch is set to 300, and the batch size is set to 2. The image size in the insulator SFID dataset was resized to 1536 × 1536 and the image size in the Visdrone2021 dataset was resized to 640 × 640 for training and validation.

4.3 Evaluation metrics

We used average precision, average recall, AP50, mAP, AP_S, AP_M, AP_L, and FLOPs as evaluation metrics. mAP is a common measure of overall ability for all classes. It is simply the average of AP50 for all classes, while AP50 has an IOU threshold of 0.5. In addition, to compare with the computational complexity of different networks, time complexity (FLOPs) was chosen to represent the inference speed of different models.

To validate the performance of our GSNA-YOLOv7 for insulator power equipment defect detection, we compared the GSNA-YOLOv7 method with many classical state-of-the-art models, including the one-stage network YOLOv3 27 and the two-stage network Faster R-CNN 4. In addition, there are newer YOLOv5 25 and YOLOv7 26 models.

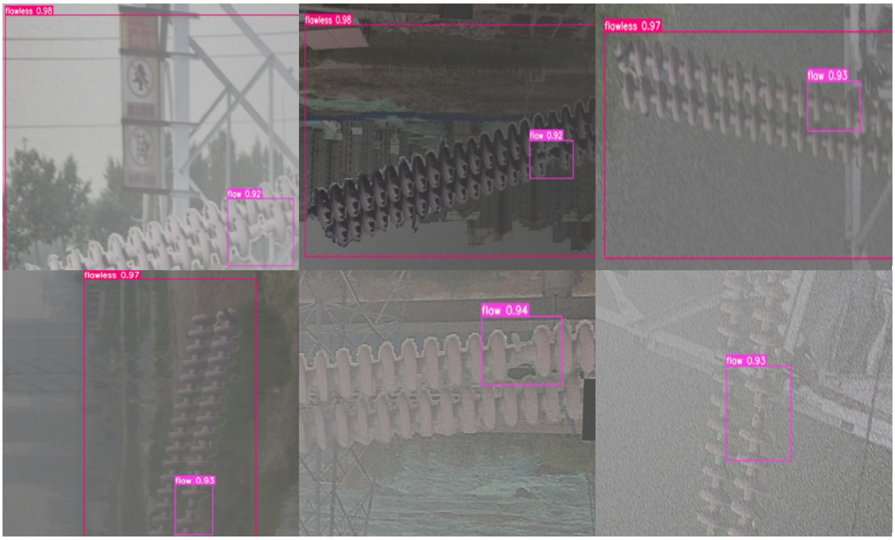


Fig. 7. Visualization of GSNA-YOLOv7 model for power equipment defect detection. The large red box indicates the insulator power equipment tested as a whole; the small pink box indicates the specific defective part of the insulator tested.

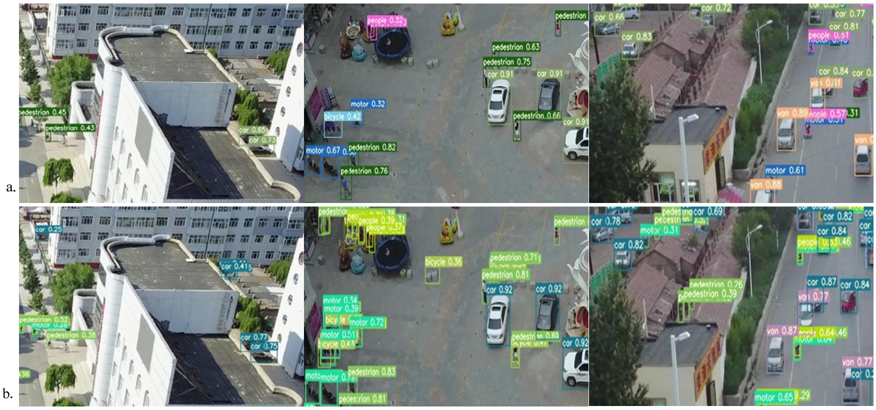


Fig. 8. Visualization of target detection on visdrone2021-val. a.Visualization of the yolov7 model; b. Visualization of GSNA-YOLOv7 model.

4.4 Ablation experiments

We studied the ablation of our designed New-Neck and DownNAM modules on the insulator SFID dataset, respectively, and the results from Table 1 can prove the theoretical basis of this paper.

First, we replace the neck structure of the YOLOv7 model with our designed New-Neck module, which reduces the number of parameters in the overall structure of the network and improves its mAP value by 0.2% relative to the YOLOv7 model. In this network design case, its FLOPs and Params are lower than those of the YOLOv7 model.

Second, the mAP value is improved by 0.7% due to the addition of the DownNAM module with NAMAttention attention mechanism, which gives the model an overall better ability to focus on defective sites, while the overall number of parameters of the model does not increase much. Finally, our model reduces about 0.2G in inference speed and improves about 0.9% in mAP. Compared to the baseline YOLOv7 model, our GSNA-YOLOv7 has better performance, computational and storage advantages.

4.5 Comparative experiments

To demonstrate the effectiveness of our GSNA-YOLOv7 method, we performed validation using the insulator power equipment SFID dataset and the UAV Visdrone 2021 dataset. The comparison with the classical and newer models, shown in Tables 2 and 3, better illustrates the real-time accuracy of our model in defect detection, highlighting the advantages of our model in power equipment defect detection and its superiority in terms of the number of parameters and inference speed, which is better than the newer model.

First, as shown in Table 2, our model was compared with the classical YOLOv3 27 single-stage network and the two-stage network Faster R-CNN 4 in the SFID dataset. Our GSNA-YOLOv7 achieves an mAP value of 97.25%, higher than YOLOv3, and a faster R-CNN, while the FLOPs are only a small percentage higher than the former, but the final detection is higher than the former. Compared with most models, GSNA-YOLOv7 performs better with fewer parameters and lower inference speed.

Secondly, for the newer models YOLOv5 25, YOLOv7 26, etc., our model GSNA-YOLOv7 has better advantages in terms of the number of parameters and inference speed, while the final mAP values obtained on the insulator SFID dataset are significantly improved relative to the other models, being 0.9% higher than those of the YOLOv7 model. mAP values for GSNA-YOLOv7 have a higher mAP value than YOLOv5 and have a lower inference speed and number of parameters than that of the YOLOV5 model.

As shown in Table 3, our model on the UAV Visdrone2021 dataset can reach better detection results when compared with YOLOv3 and Faster R-CNN; it also has a 0.6% higher mAP value than the YOLOv7 model. As shown in Table 4, it can be seen that the GSNA-YOLOv7 model has a lower number of parameters and inference speed compared with other models, and the FPS can reach 165, which meets the real-time requirements of defect detection.

We also made a visualization of the effect of detection in the two datasets separately, as shown in Figures 7 and 8. In Figure 7, we can see that the GSNA-YOLOv7 model has a good detection effect for insulators and their defective parts with high detection confidence; in Figure 8, we made a visualization effect plot comparison between GSNA-YOLOv7 model and YOLOv7 model in the UAV Visdrone2021-val dataset, a indicates the detection of YOLOv7 model effect and b indicates the detection effect of GSNA-YOLOv7 model. The comparison of the upper and lower images shows that the YOLOv7 model has the phenomenon of missing and wrong detection for the images taken by the UAV scene, and the detection confidence for the target object is lower than that of our GSNA-YOLOv7 model.

Thus, when validated on the insulator power equipment dataset, our approach has a fine balance between performance and lightweight.

Table 1. Comparison results of different modules on the SFID dataset

Methods	FLOPs	Size	Mean Precision	Mean Recall	mAP [%]	AP50 [%]
YOLOv7	104.7G	1536	99.87	99.75	96.01	99.71
+New-neck	101.8G	1536	99.89	99.74	96.22	99.79
+New-neck+DownNAM	102.3G	1536	99.98	99.82	96.94	99.89

Table 2. Comparison results of the effect of each model on the SFID dataset

Methods	Size	Mean Precision	Mean Recall	mAP [%]	AP50 [%]
YOLOv327	1536×1536	99.67	99.30	93.50	99.50
Faster R-CNN4	1536×1536	99.10	99.00	90.98	99.19
YOLOv525	1536×1536	99.81	99.64	95.59	99.63
YOLOv726	1536×1536	99.87	99.75	96.01	99.71
GSNA-YOLOv7(ours)	**1536×1536**	**99.98**	**99.82**	**96.94**	**99.89**

Table 3. Comparison of model effects on VisDrone2021-DET-val

Methods	Size	APval	AP50	AP75	AP_S	AP_M	AP_L
YOLOv327	640×640	15.0	27.2	14.6	6.3	21.5	36.1
Faster-RCNN +ResNeXt10110	640×640	22.6	40.2	23.1	9.6	29.3	40.3
YOLOv5-X24	640×640	22.6	38.6	21.8	13.9	32.4	42.6
YOLOV72	640×640	28.4	48.6	28.1	18.2	40.2	51.7
GSNA-YOLOv7(ours)	**640×640**	**29.0**	**49.3**	**28.9**	**19.1**	**40.9**	**52.1**

Table 4. Comparison of the operational performance of each target detection model at the same level

Methods	#Param.	FLOPs	FPS
YOLOv3	59.6M	158.0G	27
YOLOv5-L	46.5M	109.1G	99
YOLOv5-X	86.7M	205.7G	83
YOLOv7	36.9M	104.7G	161
GSNA-YOLOv7(ours)	**34.2M**	**102.3G**	**165**

5 Conclusion

In this paper, we proposed a lightweight GSNA-YOLOv7 model better to solve the problems in insulator power equipment defect detection and be more adaptable to real-time detection of insulator defect targets. First, we utilize the E-ELAN structure of YOLOv726 to enhance the learning capability of the network by scaling the design and fusing new branches while maintaining the original gradient branches to improve the model's overall performance. Secondly, the GSConv module is added to the YOLOv7model to improve the Neck structure in it to better reduce the computational cost of the model from the accuracy and speed direction so that the computational complexity of the network is reduced and the performance of the network is ensured; by introducing the NAMAttention 18 attention mechanism, compared with the original classical attention mechanism, with the deletion of fully connected layer and convolution, reduces the computational effort and inference speed, incorporates scaling factors in the attention weights, and suppresses unwanted detection target features from the inclusion of regular terms, enabling our network to better notice insulator defect targets. In many comparison and ablation experiments on the insulator SFID dataset and the Visdrone2021 dataset, our model obtains high mAP values compared to better classical target detection models. We will continue to upgrade this model structure and apply it to other related areas of power equipment detection. In addition, we will gradually improve the model's effectiveness in detecting defective targets of electrical equipment after many experiments from different electronic equipment datasets.

Author Contributions: Conceptualization, T.X. (Tao Wang) and J.X.(Jingfeng Xiao); methodology, T.X. (Tao Wang); software, T.X. (Tao Wang); validation, T.X. (Tao Wang), J.X. ; formal analysis, T.X. (Tao Wang); investigation, T.X. (Tao Wang); resources, T.X. (Tao Wang); data curation, T.X. (Tao Wang); writing—original draft preparation, T.X. (Tao Wang); visualization, T.X. (Tao Wang); supervision, T.X. (Tao Wang); project administration, T.X. (Tao Wang); writing—review and editing and funding acquisition, J.X.(Jingfeng Xiao), W.Y.(Wenzhong Yang), X.M.(Xinxin Meng). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and technology project support from State Grid Xinjiang Electric Power Co（Grant No. 5230XT220004）.

Data Availability Statement: This study did not report any data. We used public data for research.

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

References

1. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2014.

https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html

2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In IEEE transactions on pattern analysis and machine intelligence, 37(9):1904 – 1916,2015. doi: 10.1109/TPAMI.2015.2389824.
3. Ross Girshick. Fast r-cnn. In Proceedings of the IEEE inter national conference on computer vision, pages 1440 – 1448,2015. https://doi.org/10.48550/arXiv.1504.08083
4. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN:Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems (NIPS), pages 91 – 99,2015. DOI: 10.1109/TPAMI.2016.2577031.
5. Redmon, Joseph , and A. Farhadi . "YOLO9000: Better, Faster, Stronger." IEEE Conference on Computer Vision & Pattern Recognition IEEE, 2017:6517-6525. DOI: 10.1109/CVPR.2017.690.
6. K. Han, Y . Wang, Q. Tian, J. Guo, C. Xu, and C. Xu. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1580 – 1589. DOI: 10.1109/CVPR42600.2020.00165.
7. Z. Feng, L. Guo, D. Huang and R. Li, "Electrical Insulator Defects Detection Method Based on YOLOv5," 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS), Suzhou, China, 2021, pp. 979-984, doi: 10.1109/DDCLS52934.2021.9455519.
8. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. arXiv preprint arXiv:1602.07360, 2016. https://doi.org/10.48550/arXiv.1602.07360
9. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp.834 – 848, 2018. DOI: 10.1109/TPAMI.2017.2699184.
10. S. Liu, H. Di, and Y . Wang. Receptive field block net for accurate and fast object detection. 2017. https://doi.org/10.1007/978-3-030-01252-6_24
11. Chollet, François. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. https://doi.org/10.48550/arXiv.1610.02357
12. K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. arXiv: 151203385 Cs, Dec. 2015, Accessed: Mar. 29, 2020. [Online]. Available:http://arxiv.org/abs/1512.03385. DOI: 10.1109/CVPR.2016.90.
13. L. C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017. https://doi.org/10.48550/arXiv.1706.05587
14. X. Zhang et al., "InsuDet: A Fault Detection Method for Insulators of Overhead Transmission Lines Using Convolutional Neural Networks," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 5018512, doi: 10.1109/TIM.2021.3120796.
15. Shangliang Xu, Xinxin Wang, Wenyu Lv, Qinyao Chang, Cheng Cui, Kaipeng Deng, Guanzhong Wang, Qingqing Dang, Shengyu Wei, Yuning Du, et al. Pp-yoloe: An evolved version of yolo. arXiv preprint arXiv:2203.16250, 2022. https://doi.org/10.48550/arXiv.2203.16250

16. J. Dai, Y. Li, K. He, and J. Sun. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, pages 379 – 387, 2016. https://doi.org/10.48550/arXiv.1605.06409
17. Hulin Li, Jun Li, Hanbing Wei, Zheng Liu, Zhen fei Zhan, and Qiliang Ren. 2022. Slim-neck by gsconv: A better design paradigm of detector architectures for autonomous vehicles. arXiv preprint arXiv:2206.02424. https://doi.org/10.48550/arXiv.2206.02424
18. Z. S. Yichao Liu, Yueyang Teng, Nico Hoffmann. NAM: Normalization-based Attention Module. arXiv - CS - Computer Vision and Pattern Recognition (2021). https://doi.org/10.48550/arXiv.2111.12419
19. A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y . Zhu, R. Pang, V . V asudevan et al.. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1314 – 1324. DOI: 10.1109/ICCV.2019.00140.
20. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018. DOI : 10.1109/CVPR.2018.00474.
21. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017. https://doi.org/10.48550/arXiv.1704.04861
22. Wang, J.; Chen, Y.; Gao, M.; Dong, Z. Improved YOLOV5 network for real-time multi-scale traffic sign detection. arXiv 2021, arXiv:2112.08782. https://doi.org/10.1007/s00521-022-08077-5
23. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You Only Look Once:Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016,779-788. DOI: 10.1109/CVPR.2016.91.
24. Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. YOLOV4:Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934,2020. https://doi.org/10.48550/arXiv.2004.10934
25. Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, Ayush Chaurasia, TaoXie, Liu Changyu, Abhiram V, Laugh ing, tkianai, yxNONG, Adam Hogan, lorenzomammana, AlexWang1900, Jan Hajek, Laurentiu Diaconu, Marc, Yonghye Kwon, oleg, wanghaoyang0106, Yann Defretin, Aditya Lohia, ml5ah, Ben Milanko, Benjamin Fineran, Daniel Khromov, Ding Yiwei, Doug, Durgesh, and Francisco Ingham. ultralytics/YOLOV5:v5.0 - YOLOV5-P6 1280 mod els, AWS, Supervise.ly and YouTube integrations, Apr. 2021.
26. Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real time object detectors. arXiv preprint arXiv:2207.02696 (2022). DOI: 10.1109/CVPR52729.2023.00721.
27. J. Redmon and A. Farhadi. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018. https://doi.org/10.48550/arXiv.1804.02767
28. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In European Conference on Computer Vision, 2016. https://doi.org/10.1007/978-3-319-46448-0_2
29. H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf. Varifocalnet: An iou-aware dense object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8514 – 8523. DOI: 10.1109/CVPR46437.2021.00841.
30. Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin. Reppoints: Point set representation for object detection. In The IEEE International Conference on Computer Vision (ICCV), Oct 2019. DOI: 10.1109/ICCV.2019.00975.

31. Z. Cai and N. V asconcelos. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6154 – 6162. DOI: 10.1109/CVPR.2018.00644.
32. H. Zhang, H. Chang, B. Ma, N. Wang, and X. Chen. Dynamic R-CNN: Towards high quality object detection via dynamic training. In ECCV, 2020. https://doi.org/10.1007/978-3-030-58555-6_16
33. T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. In 2017 IEEE International Conference on Computer Vision (ICCV), 2017. DOI: 10.1109/ICCV.2017.324.
34. J. Ma, and B. Chen. Dual refinement feature pyramid networks for object detection. arXiv preprint arXiv:2012.01733, 2020. DOI: 10.1109/CVPR.2017.106.