EasyChair Preprint

№ 11404

YOLO-Based Multi-Scale Lightweight Insulator String Defect Detection Network

Xiaosong He, Xiao Wu, Jun Peng, Yuanmin He, Haojun Dai and Xinkai Ma

November 28, 2023

YOLO-based Multi-scale Lightweight Insulator String Defect Detection Network

Xiaosong He*[††][§], Xiao Wu[§], Jun Peng[‡][¶][§], Yuanmin He[§], Haojun Dai[§], Xinkai Ma[§]
*College of Intelligent Technology and Engineering
Chongqing University of Science and Technology, Chongqing 401331, China
E-mail: xiaosh@cqust.edu.cn
[††]Corresponding author
[‡]College of Mathematics, Physics and Data Science
Chongqing University of Science and Technology, Chongqing 401331, China
[¶]Chongqing Sino-German Future Factory Research Institute, Chongqing 401331, China
E-mail: jpeng@cqust.edu.cn
[§]College of Intelligent Technology and Engineering
Chongqing University of Science and Technology, Chongqing 401331, China
E-mail: {1291383707, 523031150, 1589264219}@qq.com
irvingmxk@163.com

***Abstract*—When the defect detection model of transmission line insulator strings is deployed on edge devices such as drones, it is essential to condense the model as much as possible while increasing computation speed while maintaining accuracy. This paper proposes a new lightweight target detection algorithm based on YOLOv5, improves the C3 network structure by introducing multi-scale feature information interaction and reducing redundant channel information, and generates two modules that can consider both speed and accuracy, namely FasterC3 and Res2C3. Experiments have shown that combining these two modules can cut the number of model parameters and operations per second by 12%. Furthermore, the computation performance is faster than specific standard lightweight networks with fewer layers.**

***Index Terms*—Object detection, Lightweight, YOLOv5, Power Systems**

I. Introduction

Insulator strings are essential in transmission lines because they offer isolation and mechanical support for transmission wires. Transmission lines often operate in a natural setting and will unavoidably be impacted by hostile surroundings, resulting in partial loss of insulator strings, affecting regular transmission line performance, and, in extreme situations, paralyzing the whole power system. The power grid system's power line inspection job is mostly manual, yet this is a time-consuming, labor-intensive, and dangerous inspection approach. Because of its low cost, convenience and flexibility, high efficiency, and high safety factor, the UAV inspection system integrating artificial intelligence such as target detection, semantic segmentation, and other algorithms has become the standard of today's power inspection with the proposal and development of an intelligent grid. The study on insulator string defect detection is underway as a crucial component of the UAV power inspection system. The invention of a deep learning-based intelligent, efficient, and real-time insulator string defect detection algorithm has significant practical and research implications for smart grid transmission line inspection systems. The insulator string and its defect detection work may be classed as small target real-time detection due to the difficulty and uniqueness of its application circumstances. Deep learning approaches provide several great solutions for target identification challenges in computer vision. Convolutional neural network-based target identification techniques have long been a research focus. According to the recognition stage, target detection algorithms are divided into two categories.

One is a two-stage deep learning object detection system based on candidate frames represented by R-CNN [1], Fast R-CNN [2], and Faster R-CNN [3] deep convolutional neural networks. On aerial insulator photos, Liu et al. [4] employed the Faster R-CN network to obtain a detection accuracy of 94% and a detection speed of 10 frames per second. Liang et al. [5] modified the Faster R-CNN technique. They utilized ResNet-101 as the backbone network to build a multi-category defect detection model for insulator strings, voltage-equalizing rings, and shock-proof hammers. The method's average accuracy (mAP) is 91.1%, 11% greater than the Faster R-CNN algorithm with VGG-16 as the backbone network. Yang et al. [6] used Mask R-CNN to detect self-explosion defects and locate insulators. Gao et al. [7-8] integrated the target detection method with the semantic segmentation approach to construct a cascade detection model in order to address the tiny target detection nature of the insulator string defect detection problem. To begin, a faster R-CNN was utilized to precisely find the insulator string, and then the semantic segmentation technique was employed to discover the defect site of the identified insulator string. Although the two-stage detection network has intrinsic accuracy benefits, its network structure features generally result in a large number of parameters and sluggish detection speed, making real-time detection jobs intolerable.

The other is a one-stage detection network based on the YOLO (You Only Look Once) [9-12] and SSD [13] (Single Shot Multibox Detector) series. The main difference from the two-stage network structure is that it eliminates the candidate region creation step, conducts classification and bounding box regression right after the feature extraction network, and outputs the projected target's position and category. To detect insulator pictures, Han et al. [14] employed ResNet50 as the backbone network of YOLOv3. When compared to the original YOLOv3 method, the upgraded network model used 14.5% less RAM. Leamsaard et al. [15] introduced the attention mechanism into the Darknet-53 feature extraction network and proposed the YOLO-AFB structure for insulator string detection based on the attention mechanism and feature balance, which solved the problem of low accuracy of multi-target detection in complex backgrounds. Li et al. [16] improved the YOLOv5 model using USRNet to handle the problem of low detection accuracy of complex background objects and minor faults in transmission line inspection pictures. The two-stage and single-stage approaches described above are primarily concerned with improving the detection accuracy of the insulator string and its faulty portions in the insulator detection task. The most popular strategy is to increase the number of neural network layers. Although the accuracy of some data sets has increased, it appears that a practical crucial aspect is being overlooked, namely, when deploying the insulator string defect detection technique on the UAV, its virtually rigid model size, and computing complexity constraints. Even though the computational complexity of the single-stage technique is substantially lower than that of the two-stage algorithm, it does not fulfill the requirements for deployment on UAVs and other terminal devices with limited processing power and storage resources. As a result, research into the accuracy and speed of detecting insulator strings and their defective portions, as well as the balance of model weight reduction, is a challenging challenge that must be solved as soon as possible.

YOLOv5 is an outstanding and mature approach for computer vision target identification jobs that is simple to implement and go online. It also has a corresponding version that can help with tasks with lightweight requirements, although its lightweight version n has only 1.9 M parameters, Floating-point Of Operations or FLOPs is only 4.5G, but its detection accuracy is also greatly reduced. According to the most recent research, the key cause for the model's low Giga Floating-point Operations Per Second or GFLOPS and high FLOPs is frequent memory access. Chen et al. [17] presented a local convolution (PConv) method to overcome this problem, which may decrease redundant computations and memory access times, make better use of the device's CPU capacity, and is also particularly successful for spatial feature extraction. The purpose of this study is to develop the C3 network structure based on the YOLOv5 lightweight model, decrease model parameters, FLOPs, and further increase detection accuracy to fulfill the high precision and lightweight requirements of the insulator string and its defect detection assignment. In summary, the following are the paper's contributions:

- We propose a Res2C3 network structure that may be utilized to replace the C3 network structure and provide multi-scale information interaction properties to the model.
- We propose a FasterC3 network structure to replace the C3 network structure and significantly reduce the number of parameters and GFLOPs in the YOLOv5 model.
- We utilize the picture data from a public competition and a public Chinese Power Line Insulator Dataset(CPLID) dataset as the initial dataset. Applying data augmentation strategies, we subsequently expand the dataset to include 11111 photos containing insulator strings. On this dataset, the efficacy and efficiency of our suggested two structures and their combination procedures are validated.

II. Overview of YOLOv5 algorithm

So far, the latest algorithm of the YOLO series has been developed to the YOLOv8 [18] version, but YOLOv5 [19] has become one of the most popular and efficient solutions for target detection tasks with its excellent ecological environment, standardized operating procedures, and mature and rich application experience. YOLOv5's general structure may be separated into four modules: input, backbone network, neck network, and detecting head. The input module receives images of any size and calculates the anchor frame, among other things, while the detection head predicts the detection frame and category. The backbone and neck networks are detailed below.

The backbone network, which is primarily responsible for extracting information from incoming pictures, is a key component of object detection. In addition to the standard convolution block, the current version of YOLOv5's backbone network structure includes a C3 module including residuals to improve the model's stability and accuracy by increasing the variety of features and information interaction. It provides benefits in tiny object detection as well as dense object detection. Fig.1 depicts its network structure diagram.

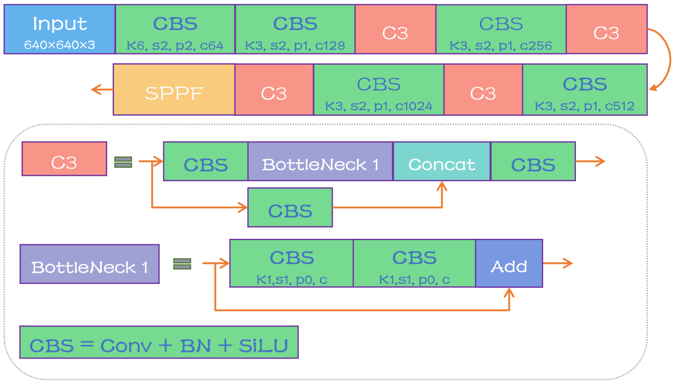


Fig. 1. Backbone structure of YOLOv5.

The neck network serves in the construction of a feature pyramid, the fusion of features, and the enhancement of feature extraction. The current version of YOLOv5 employs the PANet network, which is composed of four components: FPN, bottom-up path augmentation, adaptive feature pooling, and fully-connected fusion. FPN primarily improves the effect of target detection by fusing high-level and low-level features, particularly the detection effect of small-sized targets; Bottom-up path augmentation can shorten the information propagation path, allowing for more precise positioning of low-level features; Adaptive feature pooling can make each proposal more effective; Fully-connected fusion allows for the addition of information sources to mask prediction. Fig.2 depicts the neck network's structural diagram.

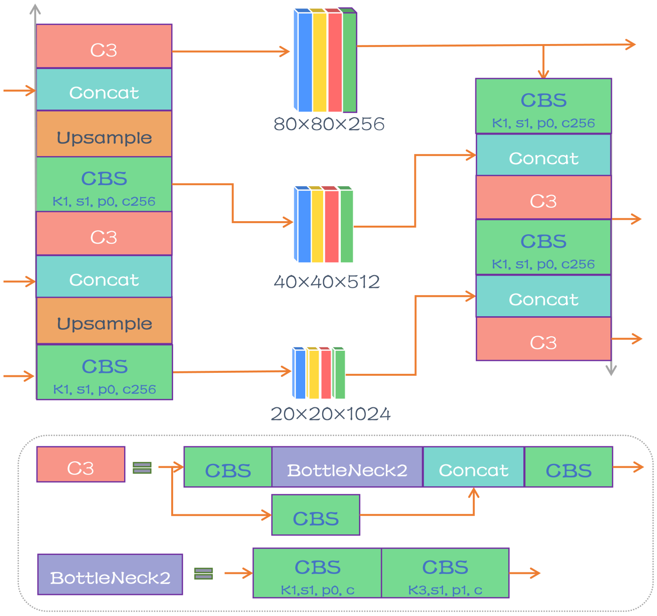


Fig. 2. Neck network structure of YOLOv5.

III. PROPOSED METHOD

The method in this paper is to improve on the foundation of YOLOv5, making it more appropriate for applications like tiny target image identification and drone and another terminal equipment deployment.

A. FasterC3 Module

To address the original model's high computational cost and vast number of parameters. We created the FasterC3 network topology to drastically reduce the amount of parameters and floating-point operations GFLOPs. The FasterC3 structure was inspired by FasterNet's FasterNet Block module. In the FasterNet essay, the author completed an in-depth investigation of the popular operator DWConv in order to produce a quicker network and demonstrated that the inefficient FLOPs are attributable to the operator's frequent memory access. The authors point out that DWConv's FLOPs are h × w × k^2 × c, whereas standard convolution's FLOPs are h × w × k^2 × c^2. Although DWConv [20] is successful in reducing FLOPs, it is frequently followed by PWConv, which cannot be used in place of conventional convolution because it will result in significant accuracy loss. Even in the inverse residual block, the width of DWConv is increased by a factor of 6 to compensate for the loss of accuracy; Nevertheless, this results in greater memory accesses, which results in non-negligible latency and slows down calculation performance. Memory accesses can now reach:

$$h \times w \times 2c' + k^2 \times c' \approx h \times w \times 2c' \qquad (1)$$

And the number of normal convolution memory accesses is as follows:

$$h \times w \times 2c + k^2 \times c^2 \approx h \times w \times 2c \qquad (2)$$

It should be noted that h × w × 2c' is utilized for memory access I/O operations and is tough to optimize further. Furthermore, the author pointed out that the various channels of the feature map have exceptionally high redundancy, hence it is recommended to convolve certain channels while leaving the others untouched. The first or final c_p channels are generated as a representation of the complete feature map for memory access of sequential or regular convolutions. The FLOPs of PConv are now available.

$$h \times w \times k^2 \times c_p^2 \qquad (3)$$

Setting c_p to 1/4 of c results in 1/16 of conventional convolution, and memory access is limited to:

$$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \qquad (4)$$

It is 1/4 of the regular convolution. The author adds the PWConv to PConv to completely and effectively utilize the information from all channels. On the input feature map, the effective receptive field resembles a T-shaped convolution, and this structure prioritizes the center position more than the standard convolution with uniform processing.

PConv and FasterNet Block perform efficiently in terms of computation speed. Based on this, we apply the upgraded Faster Block to replace the bottleneck structure of the C3 module in the YOLOv5 model's neck network and realize the number of floating-point operations while guaranteeing accuracy and stability, as well as a significant decrease in the number of parameters. Fig.3 depicts FasterC3's particular network topology.

B. Res2C3 Module

For most vision tasks, multi-scale feature representation is essential. To represent multi-scale characteristics, most known approaches use a hierarchical approach. The author proposes a new method of representing multi-scale features in Res2Net [21], namely the Res2Net CNN building block, which builds hierarchical class residuals in a single residual block connection and increases the receptive field range of each network layer. In contrast to [22][23][24], the author enhances multi-scale capabilities by employing features with varying resolutions. In Res2Net, multi-scale refers to more

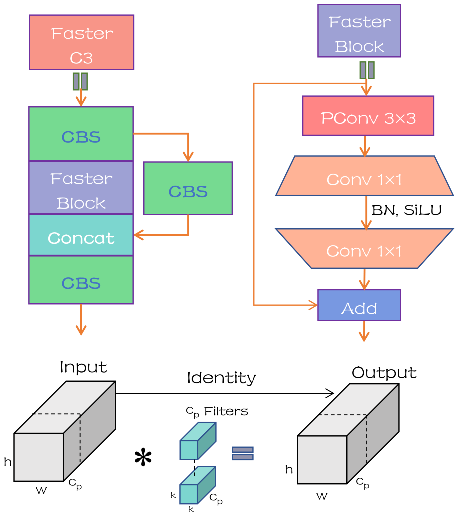


Fig. 3. The network structure of FasterC3 and PConv. The FasterC3 module consists of a 3×3 PConv and two 1×1 PWConv, using batch normalization (BN) and SiLU activation functions in the middle.

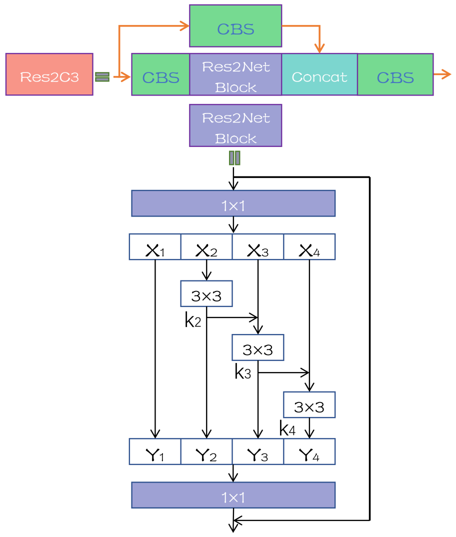


Fig. 4. The network structure of Res2C3 and Res2Net Block.

fine-grained numerous accessible receptive fields. Specifically, Res2Net divides the input features into several groups, and a group of filters first extracts features from a group of input feature maps, and then sends the output feature maps of the previous group, along with another group of input feature maps, to the next group filter, and so on until all input feature maps have been processed. Finally, the feature maps of all groups are concatenated and sent into another 1×1 filter to completely fuse the data. When passing through the 3×3 filter along all feasible path from the input feature map to the output feature map, the equivalent receptive field increases, and multiple similar feature scales are created owing to the combination effect.

We replaced the bottleneck structure with C3 and combined the upgraded Res2Net Block into the C3 structure for developing the Res2C3 module because multi-scale information interaction is favorable to the target identification job. Fig4 depicts a schematic representation of its construction. After a standard convolution, the C3 structure enters the Bottleneck (Res2Net Block) structure, and the feature map is separated into four sections using 1×1 convolution. The first part of X_1 is not processed and is directly passed to Y_1; the second part of X_2 is divided into two branches after 3×3 convolution; one part continues to propagate forward to Y_2, and the other branch is passed to X_3 so that the third branch obtains the first The second branch's information; the third branch's information; the fourth branch's information, and so on. Each branch has an n/s channel number. Assuming X_i, where i ∈ {1, 2,..., s}, s = 4 in the Fig.4, and Ki() means 3×3 convolution, the output Y_i is:

$$Y_i = \begin{cases} X_i & i = 1 \\ K_i(X_i) & i = 2 \\ K_i(X_i + Y_{i-1}) & 2 \leq i < s \end{cases} \tag{5}$$

The split applies a multi-scale strategy that allows for the extraction of global and local information. Concatenate all branches and transmit them through 1×1 convolution to better integrate information at various sizes. Convolutions can process features more effectively by dividing and concatenating processes. To decrease the number of parameters, the convolution of the first division is eliminated, which could also lead to feature reuse. Using s as the scale size control parameter, higher s will provide a richer receptive field size to extract information for learning.

C. Combination Module

Because aerial images frequently contain multiple targets of insulator strings of varying distances, and there are multi-scale information interactions between the defective parts of insulator strings and the entire insulator string, the insulator strings built on the UAV inspection system Detection tasks exploit finer-grained multi-scale feature interactions to be very effective. In the YOLOv5(n) backbone network, we replace the four C3 modules with four Res2C3 modules to extract and fuse feature maps with multi-scale information; in the neck network, we use four FasterC3 modules to replace four C3 modules, so that the detection task of insulators and their defect parts can achieve a balance of speed, parameter quantity, GFLOPs, and precision.

IV. EXPERIMENTS

A. Data Preprocessing

Part of the data comes from Zhang et al. [25] CPLID. The dataset comprises regular insulators acquired by drones, with a total of 600 images. It also contains defective insulators, with a total of 248 images, which have been edited by applying data such as segmentation, affine transformation, and splicing. The approach created the image of defective insulator strings. The remaining data comes from the data set supplied by the insulator-detecting competition's organizer. This data collection includes glass and ceramic strings as well as mixed insulator strings. Because the resolution of this data is often great, 7360×4912 is the highest. As a result, we first downsize the original image of the datasets. Then, separate the primary image into numerous smaller ones by segmenting it. Finally, two datasets were combined for offline Mosaic, MixUp, and random data improvement, bringing the total number of images in the dataset to 11,111. The training, validation, and test sets are divided 8:1:1.

B. Training Strategy

We improved on the YOLOv5 model's version 7.0. The experiment employed the YOLOv5n model's pre-training weight, a learning rate of 0.01, 150 epochs, a batch size of 90, SGD as the optimization function, momentum of 0.937, label smoothing of 0.0005, and 3 epochs of warmup epochs. After training, compare your model against models like MobileNet v3 [26], Shufflenet [27], and PP-LCNet [28].

C. Evaluating Indicator

The main purpose of this paper is to propose a lightweight new network structure to adapt to the deployment of UAV terminal equipment under the premise of ensuring the accuracy of the model, that is, to improve the accuracy of the model and reduce the amount of model parameters and FLOPs. The FPS represents the frame rate per second, or the number of images that can be processed per second, and is used to evaluate the speed of detecting insulator strings, the GFLOPs, and the number of as the space complexity and time complexity indicator. The accuracy (P) and recall (R) of the model training samples must be multiplied to determine the average precision (AP). The average precision of the entire class (mAP) is the average value of the detection AP of different targets, and it is expressed as:

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

The number of properly recognized defective insulator strings is represented by TP, the number of incorrectly identified normal insulator strings as defective insulator strings is represented by FP, and the number of incorrectly predicted defective insulator strings as normal insulator strings is represented by FN.

$$AP = \int_0^1 PdR \tag{8}$$

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i \tag{9}$$

Among them, n represents the total number of categories in the training sample set, and i represents the current category's number.

D. Result Analysis

Table I illustrates the detection results of the insulator string and defective insulator string categories of the method utilized in the present paper. The precision measurement accuracy of all categories also reached 94.1%. Among them, the detection accuracy of the defect category reached 96.4%. As mentioned above, the defect category belongs to the small target category. The multi-scale feature extraction structure (Res2C3) was designed to improve the fine detection accuracy of the defect small target. After a lot of experimental analysis, when the Res2C3 structure is used in the backbone and the FasterC3 structure is used in the neck network, the model will achieve the most stable and ideal results.

To assess the efficacy of the improved C3 module, we designed ablation experiments on defect class detection., and the results are summarized in Table II. Model 1 refers to the baseline model; Model 2 is the backbone network using the improved Res2C3 structure; Model 3 represents the use of the FasterC3 structure in the neck network; and Model 4 is the combined use of the Res2C3 and FasterC3 structures in the backbone and neck networks. With the implementation of each improved module, both its GFLOPs and the number of parameters gradually decrease, demonstrating the effectiveness of our method in reducing model weight. The combination of these two modules yields the highest performance results. The detection accuracy of insulator string defect may reach 96.4% using only 1.5M parameters and 3.6 billion floating-point operations per second, demonstrating the viability of our improved methods.

TABLE I
PRECISION, RECALL AND MAP RESULT OF IMPROVED.

	Precision	**Recall**	**mAP**
All	**0.941**	**0.854**	**0.912**
Defect	**0.964**	**0.845**	**0.901**
Insulator	**0.917**	**0.864**	**0.923**

TABLE II
ABLATION EXPERIMENTAL.

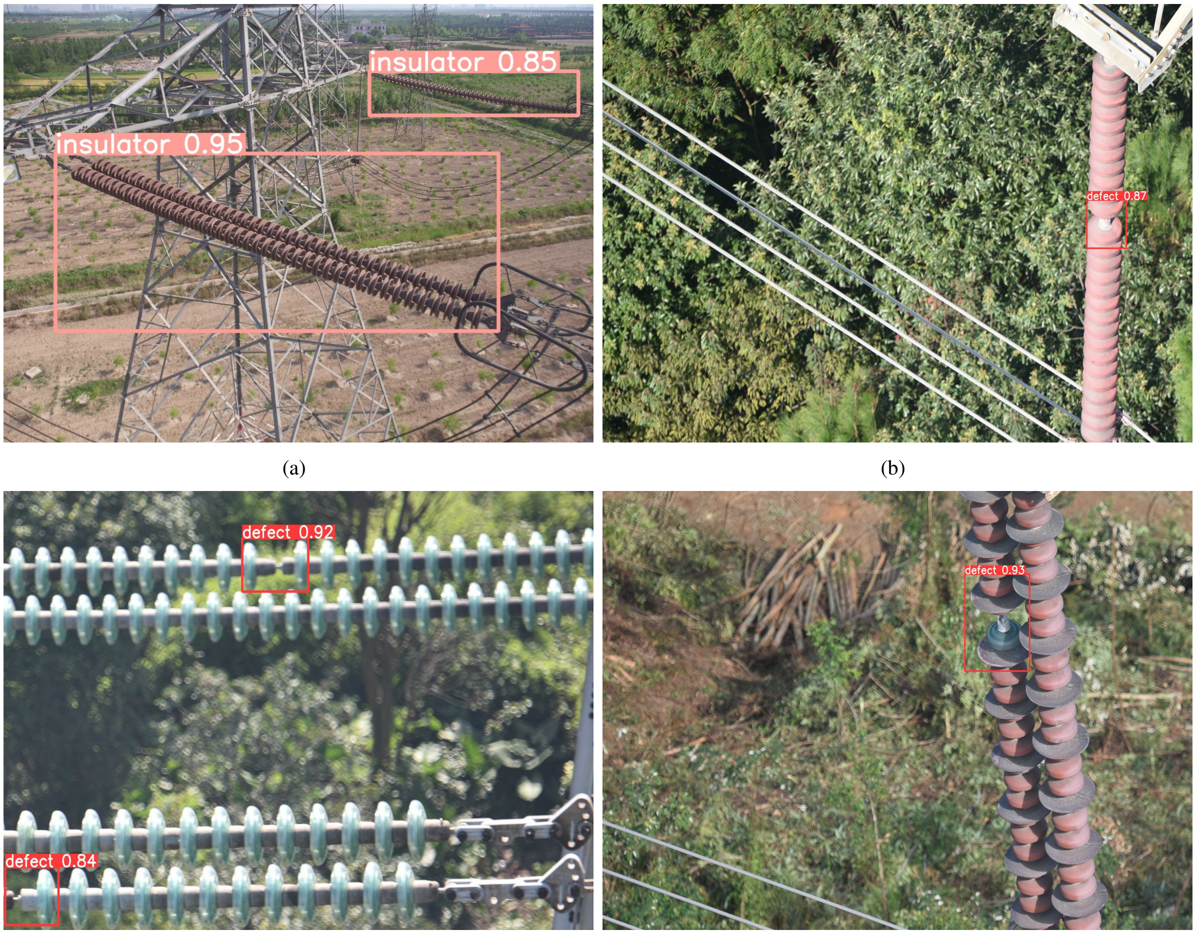
	Res2C3	**FasterC3**	**Precision**	**GFLOPs**	**Parameters**
Model1	×	×	**0.955**	**4.1**	**1.8M**
Model2	√	×	**0.969**	**3.9**	**1.7M**
Model3	×	√	**0.960**	**3.8**	**1.6M**
Model4	√	√	**0.964**	**3.6**	**1.5M**

Table III demonstrates that the method we proposed in the present research outperforms the current mainstream lightweight backbone network in terms of parameter amount and actual frame number. In terms of average precision, our approach outperforms the PP-LCNet, MobileNet v3, and ShuffleNet v2 networks, as well as YOLOv5's lightweight n model on GFLOPs.

Experiment results show that when lightweight networks like MobileNet v3 and PP-LCNet are applied to train on this

TABLE III
COMPARED WITH OTHER NETWORKS.

	Precision	Recall	mAP	Parameters	GFLOPs
PP-LCNet	**0.909**	**0.788**	**0.882**	**1.616M**	**3.3**
ShuffleNets v2	**0.891**	**0.768**	**0.862**	**2.141M**	**4.3**
MobileNet v3	**0.899**	**0.787**	**0.879**	**1.625M**	**2.5**
YOLOv5n	**0.938**	**0.880**	**0.936**	**1.761M**	**4.1**
Proposed Network	**0.941**	**0.854**	**0.912**	**1.552M**	**3.6**



(a)

(b)

(c)

(d)

Fig. 5. The partial testing results.(a) represents the detection results of insulator string targets at different distances, including both far and near targets. (b), (c), and (d) represent the detection results of the model at the defect parts of insulator strings made of various materials.

dataset, the results is not as excellent as the least n version model in YOLOv5. As a consequence, rather than rebuilding the backbone network, this paper redesigns the backbone and neck networks to obtain a lightweight insulator string detection network.

Figure 5 exhibits the results of testing the training weights using the test sets. Based on the test results, our method can accurately locate small target defect sites on insulator strings. Moreover, our proposed method demonstrates strong generalization ability and can identify various types of insulator strings and their respective defect sites, including ceramic, glass, and composite insulator strings.ulator strings and their defect portions, such as ceramic, glass, and composite insulator strings.

V. Conclusion

In this paper, a lightweight and accurate detection approach for insulator strings and their defective portions is provided. To improve detection accuracy, this method mainly relies on the Res2C3 module to perform multi-scale feature information interaction. FasterC3 aims to eliminate redundant channel information, improve the number of parameters and memory accesses, and successfully accomplish model lightweighting. The accuracy and speed of the insulator string detection model can be improved by replacing the improved C3 structure on the basis of the YOLOv5 model, thereby meeting the requirements of deploying lightweight insulator string detection models on UAVs and other terminal equipment and realizing end-to-end real-time insulator string detection. After lightweight, the model's parameter is only 1.5M, which is 1.7M less than the smallest model parameter count in the latest YOLOv8. And the weight size of the model only accounts for 3.4M, and the inference speed is substantially faster when compared to some lightweight networks. Some model compression approaches, such as pruning and knowledge distillation, will be introduced into object detection tasks in future work to improve the performance of the lightweight network even more.

Acknowledgment

The presented work is partially funded by the National Science and Technology Major Project (No. 2016ZX05060), the Scientific Research Fund of Chongqing University of Science and Technology (No. ckzg201914), the Cooperation Project between Chongqing Municipal Undergraduate Universities and Institutes Affiliated to the Chinese Academy of Sciences in 2021 (No. HZ2021015). The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

References

[1] R. Girshick, J. Donahue, T. Darrell, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the *IEEE conference on computer vision and pattern recognition*, pp. 580-587, 2014.

[2] R. Girshick. "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile*, pp. 1440-1448, 2015.

[3] S. Ren, K. He, R. Girshick, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems*, pp. 28, 2015.

[4] X. Y. Liu, H. Jiang, J. Chen, et al. "Insulator detection in aerial images based on faster regions with convolutional neural network." *2018 IEEE 14th international conference on control and automation (ICCA). IEEE*, pp. 1082-1086, 2018.

[5] H. G. Liang, C. Zuo, W. M. Wei. "Detection and evaluation method of transmission line defects based on deep learning," vol. 8, pp. 8448-38458, 2020.

[6] Y. L. Yang, Y. Wang, H. Y. Jiao. "Insulator identification and self-shattering detection based on mask region with convolutional neural network." *Journal of Electronic Imaging*, vol. 28, pp. 053011, 2019.

[7] F. Gao, J. Wang, Z. Z. Kong, et al. "Recognition of insulator explosion based on deep learning." *2017 14th international computer conference on wavelet active media technology and information processing (ICCWAMTIP). IEEE*, pp. 79-82, 2017.

[8] X. F. Li, H. S. Su, G. H. Liu. "Insulator defect recognition based on global detection and local segmentation," vol. 8, pp. 59934-59946, 2020.

[9] J. Redmon, S. Divvala, R. Girshick, et al. "You only look once: Unified, real-time object detection," in Proceedings of the *IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.

[10] J. Redmon, A. Farhadi. "YOLO9000: better, faster, stronger," in Proceedings of the *IEEE conference on computer vision and pattern recognition*, pp. 7263-7271, 2017.

[11] J. Redmon, A. Farhadi. "YOlOv3: An incremental improvement." *arXiv preprint arXiv*:1804.02767, 2018.

[12] A. Bochkovskiy, C. Y. Wang, H. M. Liao. "YOlOv4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv*:2004.10934, 2020.

[13] W. Liu, D. Anguelov, D. Erhan, et al. "Ssd: Single shot multibox detector." *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing*, pp. 21-37, 2016.

[14] J. M. Han, Z. Yang, Q. Y. Zhang, et al. "A method of insulator faults detection in aerial images for high-voltage transmission lines inspection." *Applied Sciences*, vol.9, pp. 2009, 2019.

[15] J. Ieamsaard, S. N. Charoensook, S. Yammen. "Deep learning-based face mask detection using YOLOv5." *2021 9th International Electrical Engineering Congress (iEECON). IEEE*, pp. 428-431, 2021.

[16] B. Li, Y. L. Li , X. S. Zhu,S. Wang, et al. "Multi-target Detection in Substation Scence Based on Attention Mechanism and Feature Balance." *Power System Technology*, vol. 46, pp. 2122-2132, 2022.

[17] J. R. Chen, S. Kao, H. He, et al. "Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks." *arXiv preprint arXiv*:2303.03667, 2023.

[18] J. Glenn, YOLOv8, https://github.com/ultralytics/ultralytics , 2023.

[19] J. Glenn, YOLOv5, https://github.com/ultralytics/YOLOv5, 2022.

[20] F. Chollet. "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the *IEEE conference on computer vision and pattern recognition*, pp. 1251-1258, 2017.

[21] S. H. Gao, M. M. Cheng, K. Zhao, et al. "Res2net: A new multi-scale backbone architecture." *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, pp. 652-662, 2019.

[22] S. F. Chen, E. Z. Xie, C. J. Ge, et al. "Cyclemlp: A mlp-like architecture for dense prediction." *arXiv preprint arXiv*:2107.10224, 2021.

[23] B. Zoph, V. Vasudevan, J. Shlens, et al. "Learning transferable architectures for scalable image recognition," in Proceedings of the *IEEE conference on computer vision and pattern recognition*, pp. 8697-8710, 2018.

[24] X. H. Ding, X. Y. Zhang, J. Han, et al. "Scaling up your kernels to 31x31: Revisiting large kernel design in cnns," in Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11963-11975, 2022.

[25] T. X. Zhang. "Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks."

[26] A. Howard, M. Sandler, G. Chu, et al. "Searching for mobilenetv3," in Proceedings of the *IEEE/CVF international conference on computer vision*, pp. 1314-1324, 2019.

[27] N. N. Ma, X. Y. Zhang, H. T. Zheng, et al. "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in Proceedings of the *European conference on computer vision (ECCV)*, pp. 116-131, 2018.

[28] C. Cui, T. Q. Gao, S. Y. Wei, et al. "PP-LCNet: A lightweight CPU convolutional neural network." *arXiv preprint arXiv*:2109.15099, 2021.