PAPER • OPEN ACCESS

Research on Image Recognition of Power Inspection Robot Based on Improved YOLOv3 Model

To cite this article: Wei Xiong *et al* 2020 *J. Phys.: Conf. Ser.* **1486** 042034

Research on Image Recognition of Power Inspection Robot Based on Improved YOLOv3 Model

Wei Xiong[1,2], Sha Yang[1], Zhao Zhang[1,2], Liang Chen[1,2], Shuxin Huang[1,2]

[1]Nari Group Corporation, State Grid Electric Power Research Institute, Nanjing 211106, Jiangsu, China

[2]Nari Technology Co. Ltd, Nanjing 211106, China

xiongwei@sgepri.sgcc.com.cn; yangsha@sgepri.sgcc.com.cn;

Abstract. YOLO series models are widely used in power inspections, but they are prone to miss inspections for small targets and diverse targets. In response to this defect, an improved network model of YOLOv3-g suitable for GPU cores and an improved network model of YOLOv3 mini suitable for CPU cores are proposed. By reducing the number of small and medium-scale targets, the number of convolution kernels is increased, and the small and medium-scale targets are increased. The intensity of feature extraction reduces the amount of network calculations and better solves the problem of inaccurate detection and recognition of small targets in electric robot scenes.

1. Background and significance

With the advent of the intelligent era, traditional power operation and maintenance and management models have not been able to meet the needs of the current rapid development of smart grids. Integrating robotics and power technology, the comprehensive unmanned operation and maintenance inspection of power transmission, transformation and distribution through the power inspection robot has become the development trend of smart grid operation inspection in China.

As the "engine" technology in the current artificial intelligence technology, deep learning technology has efficient automatic learning and classification capabilities, and has achieved effects close to or even beyond human level in image processing problems in multiple fields. In the operation and maintenance of power scenes, deep learning technology is used to achieve high-precision, multi-layered, and comprehensive image collection for power equipment and specific parts. Through the research and application of machine learning, artificial intelligence, deep learning, data of supporting for equipment status analysis and diagnosis is provided.

In actual engineering applications, the robot operation and maintenance of the power industry requires very high equipment detection accuracy, but the reality is that there are many types of equipment in the power industry, and the same equipment has various forms, and the environment is complex. The accuracy of target recognition is not very optimistic.

YOLO is a real-time target detection algorithm that was updated to the third generation YOLOv3 in March 2018. It regards target detection as a type of regression problem, predicting the coordinates and class probability of the target's bounding box directly from the input image, thereby achieving end-to-end recognition, so its detection speed is very fast, and it can maintain a high accuracy real-time target detection.

In order to meet the actual production needs, the existing YOLO models was studied in depth. Through the improvement of the existing network structure and the design and application of the new network structure to achieve the level of actual production requirements, it can be flexibly applied to the visual recognition of inspection robots with different hardware cores.

2. Robot Vision Recognition Method Based On Improved YOLO Model

2.1Robot application environment

Deep learning is a mathematical network model established by simulating the human brain's nervous system. The most significant feature is that it requires large amounts data for training, and requires the processor to have a parallel and repetitive computing capabilities. The GPU's (Graphic Processing Unit) parallel computing structure, high memory access speed, and high floating-point computing capabilities are well suited for deep learning. GPU’s are good at performing simple and repeated operations on big data. The advantage of CPU is that it has strong scheduling, management, and coordination capabilities, and is good at complex operations such as commanding the whole world.

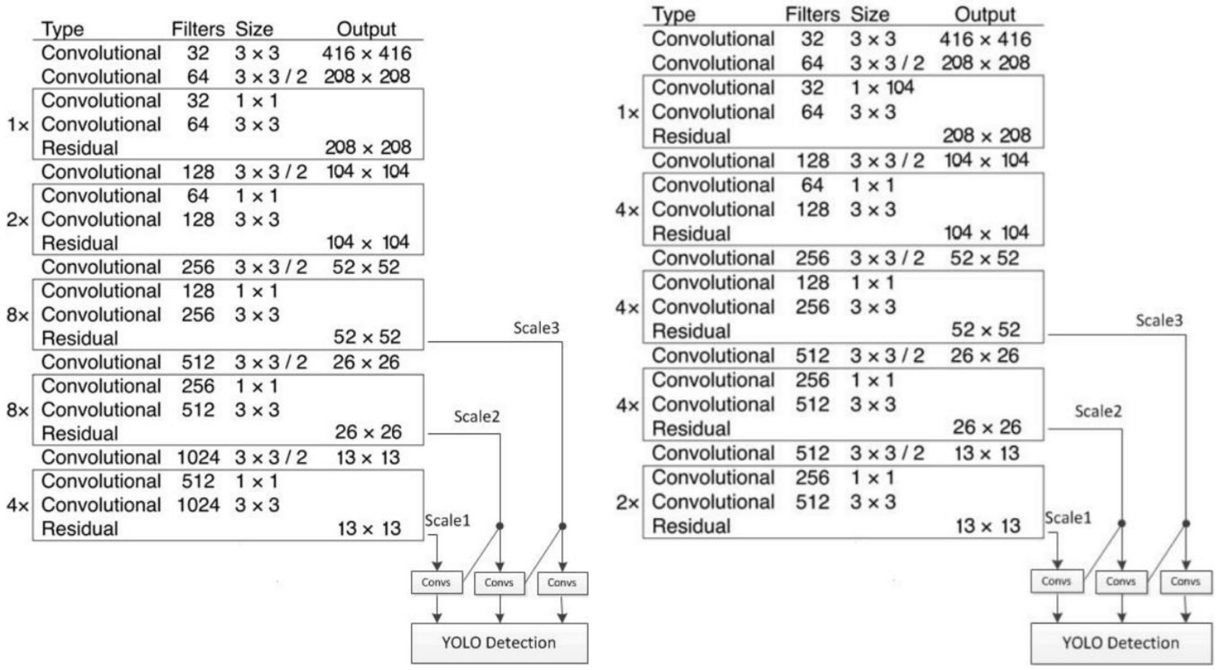
At present, deep learning-based visual recognition technology is mainly based on GPU servers to complete model training. However, the application of network models based on embedded GPU hardware platforms and server CPU hardware platforms has certain limitations, and computing power and power consumption need to be considered. The problem requires further transplantation research.

By studying the optimization technology of deep network model, the required memory and calculation can be reduced, the inspection efficiency can be improved, and the cost can be reduced. This is also a necessary consideration for robot inspection instead of manual inspection. This paper fully considers the design of the network model, carries out research on transplantation applications on portable and embedded devices, and implements deep neural networks to run on low-configuration embedded hardware platforms to meet performance requirements and promotion requirements.

2.2YOLOv3 model optimization

The original YOLOv3 structure, as shown in Figure 1, Filters represents the number of convolution kernels, Size represents the size of the convolution kernel, and Output represents the resolution of the feature map after convolution.

From the YOLOv3 three-level prediction structure in Figure 1(1), it can be seen that “Scale1” is responsible for large-scale detection, “Scale2” is responsible for medium-scale detection, and “Scale3” is responsible for small-scale detection. There are few large-scale targets in the power system inspection scene, and more devices can be positioned as small-to-medium-scale targets. This article uses the method of deleting network layers to optimize the original YOLOv3 model, and the optimized network structure YOLOv3-g is shown in the figure. 1 (2) shown.



(1) Original YOLOv3 structure (2) YOLOv3-g optimized structure

Figure 1. Comparison of the original YOLOv3 structure and YOLOv3-g optimized structure

The optimization principle is to reduce the number of convolution kernels before Scale1 output, delete the 1024 3 * 3 convolution kernels that slow down the operation speed, and replace them with two sets of 512 of 3*3 convolution kernels to increase the small and medium scale Number of cores. The optimized YOLOv3-g network structure increases the number of convolution kernels for small and medium-scale targets, increases the strength of extracting features of small and medium-scale targets, reduces the number of convolution kernels for large-scale targets, and reduces to some extent The amount of network computing.

Optimizing the network YOLOv3-g can effectively guarantee the detection efficiency and quality of the inspection robot with the GPU as the core.

2.3YOLOv3 tiny network improvement

The YOLOv3 model has low image recognition efficiency in the inspection robot based on the CPU core hardware platform, and cannot achieve the real-time detection effect. The actual test is 8 seconds per sheet. Using YOLOv3 tiny recognition speed is much faster than YOLOv3 recognition speed, can reach one second per sheet, but the accuracy is not optimistic, the mAP value is relatively low, only about 33.9%, and the recall rate for small targets is extremely low.

The YOLOv3 mini model uses the same backbone network and darknet19 framework as YOLOv3 and YOLOv3 tiny. The input picture size of the YOLOv3 mini model is 416*416. After several convolution and pooling operations, a feature map with a resolution of 13*13 is obtained. After several convolutions, the YOLO detector is used for the first time. The target result is output. At this time, the target is a large scale. After upsampling process and several convolution pooling, the feature map with a resolution of 26*26 is performed on the connection layer to obtain a feature map with a resolution of 26*26. After several convolutions, use the YOLO detector for the second target result output, at which time the target is of a medium scale; then operate in the same way to finally obtain a 52*52 feature map, and use the YOLO detector for the third target output At this time, the target is a small scale. In addition, due to the addition of a scale detection, the number of network layers increases. At this time, the problem of gradient disappearance and gradient explosion is considered, and a residual layer is added to the network. By skipping the connection, activation can be obtained from a certain layer of the network, and then rapid feedback is provided. Give another layer.

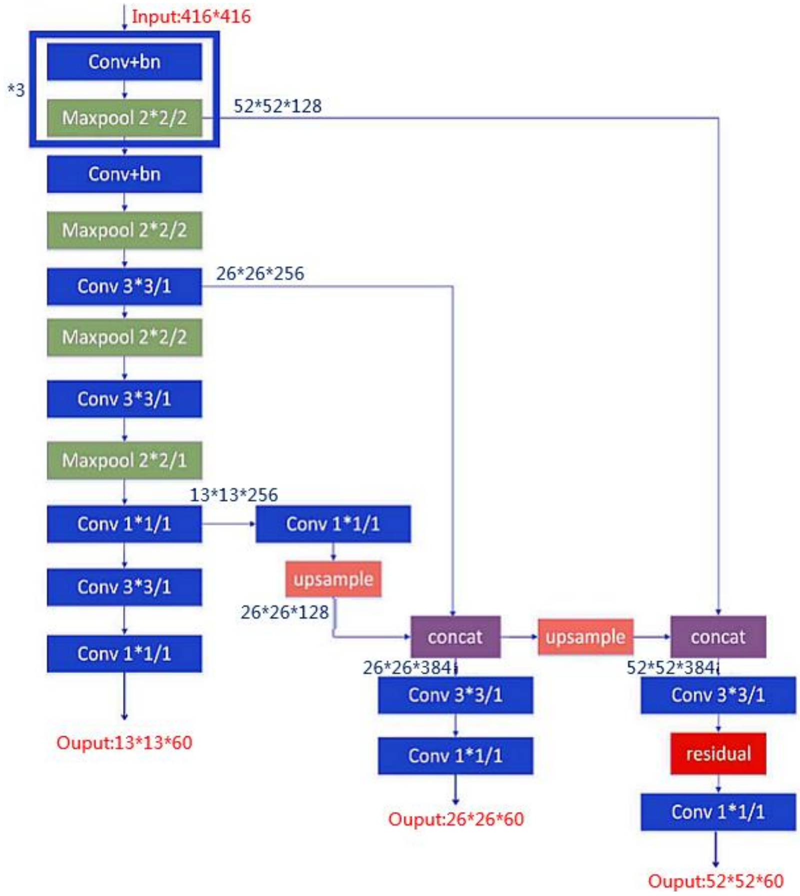


Figure 2. YOLOv3 mini structure

3. Experimental analysis

There are two types of power inspection robots: GPU-based robots and CPU-only robots. The robot with GPU platform is mainly integrated with NVIDIA JetsonTX2, which has high performance, low power consumption, small size, etc., and it uses the NVIDIA Pascal ™ architecture and has 256 CUDA cores. Robots that only use CPU cores, mainly using Intel Core i7-610E processors, dual cores and four threads, clocked at 2.53GHz. One core is used to perform inference calculations on image recognition, and the other core is used to handle other tasks of the robot.

Firstly, we collect samples of typical equipment commonly used in power systems, such as pressure plates, LED lights, digital meters, circuit breaker instructions, moisture absorbers, pointer meters, etc. Secondly, based on these samples, the above studied models are trained and experimentally verified. The sample is shown in Figure 3.

Figure 3. Typical equipment samples

Run YOLOv3 and optimized model YOLOv3-g on the JetsonTX2 platform, run YOLOv3 mini network on the CPU platform, and measure the calculation amount of the neural network with FLOPS (floating point operations per second) (this value is generated after the network structure is analyzed after running) , YOLOv3 mini, YOLOv3, YOLOv3 optimized BFLOPS, weight size and time-consuming to detect a frame of image, as shown in Table 1.

Table 1 Comparison of network structure sizes

Network model	BFLOPS	Weight Size(Mb)	Time spent (seconds)	mAP (Accuracy)	Recall
YOLOv3	65.392	240.8	0.25	98.1%	97.2%
Optimized YOLOv3	26.156	100.5	0.1	98.8%	98.5%

Running YOLOv3 tiny network and YOLOv3 mini network on Intel Core i7-610E processor, FLOPS (floating point operations per second) is used to measure the amount of computation of the neural network (this value will generate a network structure after the analysis is completed). BFLOPS for YOLOv3 tiny and YOLOv3 mini networks. Identification and comparison of typical power equipment are shown in Table 2.

Table 2. Comparison of YOLOv3 tiny and YOLOv3 mini networks

Network model	BFLOPS	Weight Size(Mb)	Time spent (seconds)	mAP (Accuracy)	Recall
YOLOv3 tiny	10.374	26.823	0.83	63.3%	92.9%
YOLOv3 mini	12.904	39.7	1．02	67.5%	94.7%

The test results show that although the optimization model YOLOv3-g cuts the number of convolution kernels responsible for extracting large-scale targets, it has little impact on the identification of the test set of typical power equipment. The optimization model YOLOv3-g has a slightly higher accuracy than YOLOv3. The detection speed is also faster than YOLOv3; the YOLOv3 mini network draws on multi-scale fusion technology to better solve the detection and recognition of small targets. The results of running on a robot with only CPU cores can well meet the time and requirements of inspection inspection standards Precision.

 1486 (2020) 042034 doi:10.1088/1742-6596/1486/4/042034

4. Conclusion

This paper proposes a network model YOLOv3-g with embedded GPU core hardware platform and a network model YOLOv3 mini with CPU core hardware platform, which are suitable for inspection robots. The optimized model YOLOv3-g is suitable for GPU inspection robots. By increasing the number of convolution kernels for small and medium-scale targets, reducing the number of convolution kernels for large-scale targets, it reduces the amount of network calculations and increases Extraction of small and medium-scale target features; the improved model YOLOv3 mini draws on the feature pyramid network, that is, it first goes through the process of upsampling to ensure grid resolution, and then uses high-level semantics and low-level semantics to improve the visual recognition of power inspection robot Efficiency and accuracy. The two models must improve the detection efficiency and detection quality, and also meet the needs of running on low-profile servers or embedded hardware platforms of inspection robots, which has a promotion and practical significance.

References

[1] Abadi M. TensorFlow: learning functions at scale[J]. Acm Sigplan Notices, 2016, 51(9):1-1.
[2] Chao W, Lan Z, Li Q, et al. Enabling Flexible Resource Allocation in Mobile Deep Learning Systems[J]. IEEE Transactions on Parallel & Distributed Systems, 2019, PP(99):1-1.
[3] Prasad S, Ramkumar B. Passive copy-move forgery detection using SIFT, HOG and SURF features[C].IEEE International Conference on Recent Trends in Electronics. 2017.
[4] Dai Weicong, Jin Longxu, Li Guoning, Zheng Zhiqiang. Improved YOLOv3 real-time detection algorithm for aircraft in remote sensing images[J]. Opto-Electronic Engineering, 2018, 45 (12): 81-89.
[5] Yang Guang, Zhou Pengju, Zhang Songbin, et al. Image recognition of substation inspection robot based on convolutional neural network [J]. Software, 2017 (12): 190-192.
[6] Zhao Jisheng. Research on Substation Monitoring Image Recognition Method Based on Convolutional Neural Networks [D]. North China Electric Power University, 2016.
[7] Li Yandong. Research on key technologies of computer vision based on convolutional neural network[D]. University of Electronic Science and Technology of China, 2017.
[8] Summary of verification work of artificial intelligence processing technology for unmanned aerial vehicle inspection images [R].Beijing: Chinese Academy of Electric Sciences.2018.
[9] Dong Leigang. Application of feature pyramid network in image detection[J]. Science and Technology Innovation, 2018(10): 90-91.