

## Electric Equipment Inspection on High Voltage Transmission Line Via Mobile Net-SSD

Wei Zhan\*, MD SUHEL RANA, Chenfan Sun, Yangyang Zhang  
School of Computer Science, Yangtze University, Jingzhou, Hubei, China

\*Corresponding Author.

### Abstract

Daily check and inspection of electrical utilities on the transmission line to find out faults or malfunction data and analyze, it's to ensure normal state of electrical equipment really difficult in any situation. Machine-controlled inspections by like robots or drones for power transmission infrastructures is an indispensable way to assure the safety of power transmission. Targeted object detection and classification of the power transmission infrastructure is the prerequisite for automatic inspection. In our experiment we have create the dedicated datasets of the electric equipment on power transmission line for multi-object detection, including our data collection, preprocessing and annotation. This work has been done multiple experiments to solve our functional problem and compare novel state of art deep learning methods such as Faster R-CNN, Mask R-CNN, YOLO, and SSD with MobileNet is a base feature extractor, to realize the electric equipment on power transmission line detection. For Condition monitoring and diagnosis identification of the importance of electric equipment on the electric transfer line, in the proposed deep detection approach, the Single-Shot Multi-box Detector (SSD) is a powerful deepmeta-architecture. The results show that our method can automatically detect electric equipment on high voltage transfer defects more accurately and rapidly than lightweight network methods and traditional deep learning methods. Results shed new light on defect detection in actual in progressive scenarios. In our research the main goal to show the implementation of the object detection on electric equipment's inspections on high voltage electric transfer lines on drone video using MobileNet-SSD object detection and recognition.

**Keywords:** object detection, convolution neural network, electric equipment detection, deep learning, caffe, mobileNet, SSD.

### I. Introduction

Electricity is one of the essential things in our modern life. The recent few year's days by day increase industrialization, technology development [1], especially after the 4th industrial revolution electricity has become one of the most essential parts of human activity. China is one of the tops developed an industrial country, on-demand of power transmission in china continues to increase and expand the scale of power systems[2] Normally two type of the electricity transmission line, one is underground and another is overhead, the overhead line being more power also needs more security for smooth service. Because almost all electric equipment is exposed outside the field and affected by the lead weight, lighting wind, snow, ice, and dust pollution, it may prone to fault easily. The safety and stability of power systems are becoming more significant [3]. So equipment of overhead lines is most important to maintains and inspections continuously. However, due to the severe geographical location of some high-voltage transmission towers, such as standing on the top of a mountain or on a river, few lines pass through inaccessible geographical areas, it is very difficult and costly to conduct regular manual inspection, for the reason, sometimes need to use helicopters with UV technology. The overall inspections are very costly and risky. In the modern world, all electric power industries have demands to Inspections the electric equipment on the high voltage

transmission lines and solve the above problem by low cost with high safety and save time. Inspections cost and safety risk can reduce by Unmanned Aerial vehicle (UAVs) or Drone [4,5] with deep learning methods.

The purpose of this paper is to study the application of existing target detection models to image data acquired by drones.

Therefore, this paper adopts deep learning [6] and computer vision methods to realize the detection and classification of electrical equipment on high-voltage transmission lines. In this paper, we will work for inspections on the specific importance of electric equipment and models is used for the image data from a drone in the power system in China. At the same technology can use anywhere any electric transmission line to get the same results.

Our work has been divided five sections into the below sections, we describe related work in section II, the proposed framework in section III, the experiment, result and discuss of the different task are presented in section IV, and in section V we draw conclusions from the results and also discuss our future work.

## II. Related work

With the recent advancements and rise of autonomous vehicles, video surveillance, and real-time face or object detection there is a huge demand for developing fast and accurate object detection systems. The developed systems are not only recognizing and classifying the class of those objects, but it also localizes the object by drawing a box around the object which is recognized. Computer vision community last few years successfully has been developed many deep learning detection frameworks to detect the object with good accuracy and results, such as SSD, YOLO, R-CNN, Fast R-CNN, Faster R-CNN and Mask R-CNN, here is YOLO and SSD are one-stage and Faster R-CNN is the two-stage detector [5]. In this work, they observed some experiments to get the best iteration epochs, batch size, and learning rate for the model to converge. In the recent works was the identification of people, airplanes, vehicles such as cars, buses, trucks, boats, and electric equipment, etc from the real-world video footage taken by drones using the deep-learning Caffe framework.

Mainly focuses on rising accuracy of object detection, lightweight and computational capable network use for the on mobile platform embedded devices experience increasing interest. One of the most popular models TensorFlow object detection API has been developed by Google, Single Shot Multi-box Detector (SSD) with MobileNets, R-FCN, and Faster R-CNN, Mask R-CNN and YOLO, etc. are the other models also. We will, in brief, discuss a few of these models in this below section.

### 2.1 Establishment of our datasets

#### 2.1.1 Data Classification/Data Collection

The safety and stability of power equipment is extremely important for overhead transmission lines, because different types of power equipment towers can be classified according to their shapes, uses, materials and structures. As shown in Fig.1. In this paper, four common electric equipment (insulator-chain tower-point, conductor-grip, ground-hpoint) on overhead transmission lines in different position are collected as the research object.

#### 2.1.2 Data Preprocessing and Augmentation

The proposed electric equipment inspections method in an aerial image high voltage transmission line inspection in Fig.1, The specifics of our architecture based deep learning is the stand of three parts

First, the image is collected physically from the drones for electric equipment inspection on the high voltage electric transmission line. Secondly Resizing, cropping labeling overall preprocessing the massive aerial image are transformed into a standard data format for train the model, and the third part is the two-stage one-tuning procedure is applied to raise the detection accuracy and characteristic. We trained a specific electric equipment data model that can be straight used to raise the electric equipment detection image that was taken by drone for the high voltage transmissions line inspections method.

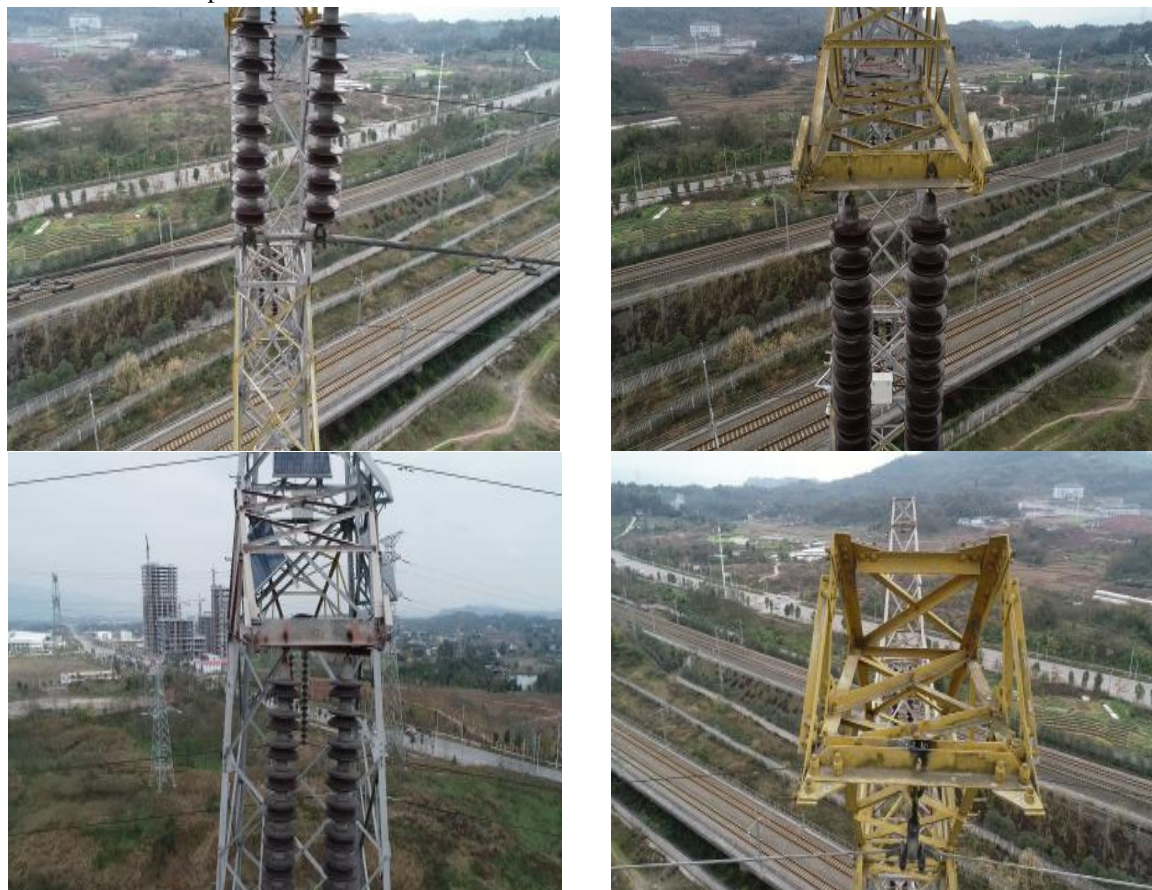


Figure 1: Insulator-chain, Tower-hpoint, Conductor-grip, Ground-hpoint.

These images are mainly from jingzhou, Hubei province of China. The other fine-tuning stage aims to train the specific model for electric equipment detection in a fixed inspection area that its typical types of equipment types and views. In this Operation, the particular electric equipment dataset is used. Different from the basic electric equipment dataset, the peculiar electric equipment dataset only carries the drones images relating to the specific high voltage electric transmission line. Direct inputting the specific equipment dataset, the parametric quantities of the basic data models are one-tuned and more meaning features of electric equipment are learned. Compared with the orthodox operation of electric equipment detection, the two-stage one-tuning will not change the workload to the electricity organization. As shown in Fig. 4, the planned acting can be easily lengthy to apply to additive assumption. If the detect project is changed to a different location, the training the process only demands to regenerate the peculiar electric equipment dataset according to the peculiar scene. while the basic dataset model remains the same.



Figure 2: Image data pre-processing pipeline

### III. Object Detector/Proposal Method

The model we designed to be able to classify and detect multiple objects at same time, Currently the best performing and most widely used one-stage and two-stage object detection networks include Fast R-CNN series, Mask R-CNN, YOLO and MobileNet-SSD. We describe a lightweight object detector network using the SSD framework with MobileNet, Faster-RCNN, Mask R-CNN, and YOLO. [7].

#### 3.1 R-CNN

Region-based Convolutions Neural Network (R-CNN) comprises 3 steps

1. Scans the input image with the selective algorithm.
2. CNN model will work on the top appropriate with 2000 selective regions
3. Consider the output of every Convolutions Neural Network and feed them into by SVM or LR (Linear regression) methods for tightening the bounding box of objects.

Firstly collects the region, classify those regions on the features and give it to classifications algorithm.

#### 3.2 Fast R-CNN

In the follow-up work, the author solved some of the drawbacks of R-CNN algorithm and rebuilt a faster object detection algorithm, which is called fast R-CNN. This method solves the problem of repeated calculations when R-CNN performs feature extraction on all region proposals, and only extracts features once for the entire image, thus significantly improving network performance

#### 3.3 Faster R-CNN

The biggest difference between Faster Region-based Convolutional Neural Networks (Faster R-CNN) and FastR-CNN is that it abandons the time-consuming selective search method, inputs the entire image into the Convolutional neural network (CNN), and uses a more efficient regional proposal network (RPN) to find all proposal regions. For the extracted features, the classifier is used to determine whether they belong to a specific class, and for the proposalbox belonging to a certain feature, the position of the proposalbox is further adjusted by the regressor.

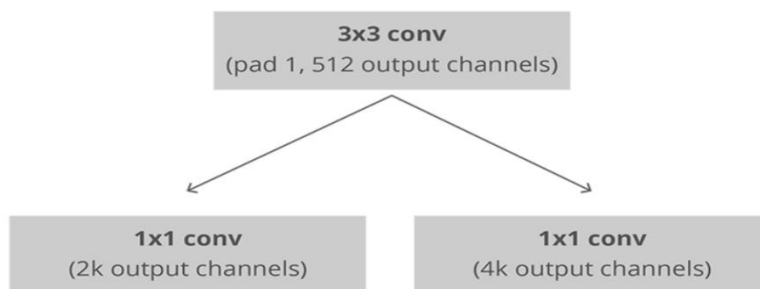


Figure 3: Convolutional implementation of an RPN architecture. (k=9 in here)

The Table1 is showing the comparison of the three region based models

Table 1 Comparison of 3 region-based Convolution Models

Model	R-CNN	Fast R-CNN	Faster R-CNN
Test time	50s	2 s	0.2 s
Speed	1x	25x	250x
mAP	66.0%	66.9%	66.9%

### 3.4 YOLO

Unlike the the two-stage object detection algorithm, YOLO's end-to-end object detection algorithm integrates region-proposal prediction and classification prediction into a neural network model. YOLO regards object detection as a regression problem. After inferring the input image, we can directly get the position and class of all objects in the image, as well as the corresponding confidence probability, and objects of different scales can be predicted through the output feature maps of different scales. Among them, the feature maps with resolutions of 52x52, 26x26, and 13x13 detect small, medium and large objects respectively.

### 3.5 Mask R-CNN

Mask R-CNN can be regarded as a general instance segmentation architecture. Mask RCNN uses the Faster RCNN prototype and adds a branch for segmentation tasks. For each proposal box of Faster R-CNN, FCN is used for semantic segmentation. The positioning and classification tasks and segmentation task are executed simultaneously. ROI pooling used in Faster-RCNN is replaced by ROI align, which greatly increases the accuracy of mask. With the introduction of semantic segmentation branch, the relationship between mask and class prediction is decoupled. There is a mask branch for semantic segmentation, and another branch for object recognition and classification.

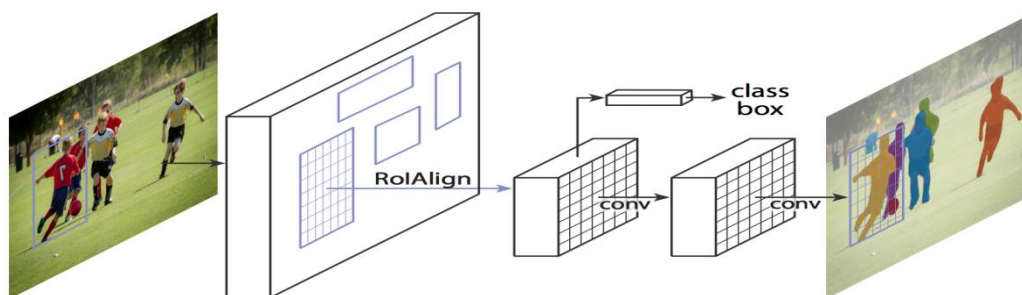


Figure 4: Mask R-CNN Framework

### 3.6 MobileNet-SSD

The SSD algorithm combines the advantages of Faster R-CNN and YOLO, and uses a regression-based method to directly return the category and position of the object in a network, so the detection speed is very fast; at the same time, it also uses the concept based on the region proposal (similar to In Faster R-CNN), in the detection process, many anchor boxes are used to select ROIs, which increases the proportion of positive samples, followed by a non-maximum suppression (NMS) step to produce the final detection [6,7].In the below mobileNet-SSD architecture.

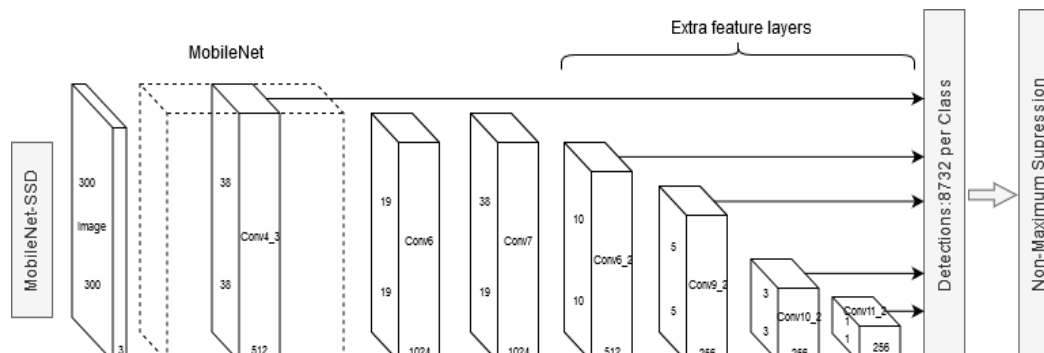


Figure 5: MobileNet-SSD Architecture

SSDs are used for object detection and classification together [9]. We are an interested predictor to detect all objects even faster than MobileNet that's why we combined SSD and MobileNet has satisfactorily predicted the bounding boxes of the objects [10]. MobileNet as backbone architecture for SSD and achieve comparable detection accuracy on our data-set and in this paper, MobileNet is the base network of SSD [7]. It can also be extended as a competent mean Reticulum in recent object detection systems [11] With small anchor networks, lightweight that can be utilized for categorization, perception and separation tasks and depth-wise dissociable convolution layer as used MobileNet-SSD, also one-stage detectors [10, 12]. We custom MobileNet before Conv-6, then we disunite all the other convolution layers. Each characteristic plant is related to the last acknowledgment layer, which Approval the perception and localization of objects of distinct scales [7].

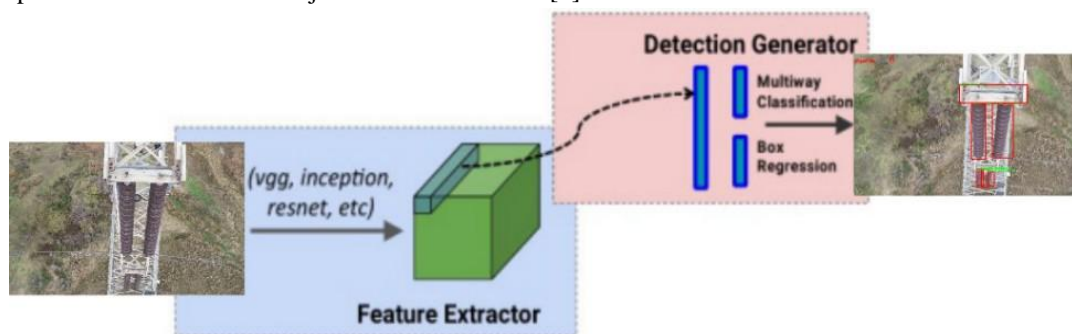


Figure 6: Processing flow of SSD object detection system

The loss of the model consists of two parts, which are the position loss and the confidence loss.

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

Here is “x=indicates the prediction box matches the real box (0, 1) and  
N=the number of paired default boxes then set the loss to 0.  
c=the value confidence level, and  
g=is the real box value”

### 3.7 Caffe

Caffe contains different algorithms and supports numerous deep learning frameworks, In use for the implementation and entree property in an object detection scheme. It is also proper for feed forward and fine-tuning existing system models to detect new objects. Caffe model consists of “bvlc\_reference\_caffe-net, Caffe made with aspect speed and modularity for the research and industry with massive data. Its the fundamental unit of computation is implemented as layer Data access, Convolution, Pooling, Activation Functions, Loss Functions, Drop out and learns a model from scratch and then resume learning from the saved models. Caffe can procedure over the millions of images in a day with the standardized media like a GPU. Caffe also can be existing with network models without any hard code. For the reason of open-source large number of new research is coming powered by Caffe every single moment of Caff e.All of our experiments are based on the Caffe version ofSSD.

### 3.8 Our Work

In our work, the main goal is to detect and recognize the objects in our present data-set image. We use a pre-trained CNN image model with the help of deep neural networks from Caffe. The other module contains different algorithms and supports several deep learning frameworks includes TensorFlow, Keras, PyTorch, etc. In this paper, our work has been done with three-step at the first read image input from disk, Secondly, pre-process the image for improving the quality of the image. Pass the image on the network and get the classified image is the third step. All images are categorized by this model and to find out the locations of the object position in the image. We applied the object detection algorithm to detect where is the object actually is located in the image, By combinations of Single Shot detectors (SSD) and MobileNet in the image, we gained the bounding box co-ordinates pretend (X, Y). The deep learning-based object detection has several processes to detect an object such as YOLO, Faster R-CNN, and SSDs. The Faster Region-based Convolutional Neural Network (Faster R-CNN) is one of the most popular methods to detect the object. the time cost of generating region proposals is much smaller in RPN than selective search. It is hard to understand for the beginners to implement. The task of the training phase is also challenging in Faster R-CNN. The faster implementation of the regions in faster R-CNN is a little slow on the order of 7 FPS. YOLO is a regression-based object detection algorithm. Its tasks include rapid localization and classification of specific objects. It’s much faster than Faster R-CNN, on the order of 40-90 Frames Per Second (FPS) on Graphics Processing Unit (GPU).The YOLO latest version YOLOv3 can be able also 155 FPS. But YOLO is less accurate. SSD (Single Shot Detector) is developed by Google Corporation and deployed on embedded platforms. SSD-an end-to-end object detection framework whose purpose is to locate and classify objects only at once. The SSD is a regression-based algorithm. The network generates the collection of bounding boxes with three different aspect ratios and confidence score for each bounding boxes, which are used to filter the anchor boxes with object class instances, and then a non-maximum suppression step to generate the final detection [10]. Single shot means that in the recognition process, the localization and classification of the object are carried out in one pass, and the network only makes one “look” of the image. The initial parameters of the model’s weight are pre-trained on a large dataset (such as ImageNet), whichmakes theweight to learn the richerfeatures, and retain the weight parameters similar to the features to be learned. SSD starts from the backbone network, which is used for transfer training, propagates the input image to a predetermined layer, obtains the feature map, and then moves the feature map forward to the object detection layer. We opted for SSD for the reason easy to train and directly incorporate the system for the component of the detection.

## IV. Experiments and results

Actually, At propitious, there is no openly ready importance Electric equipment (like an insulator-chain, tower-hpoint (hanging point), conductor-grip,ground-hpoint (hanging point) four items data-set. In our data-set, The genuine data-set of encasing pictures are taken a picture physically from the modern field function with drone inspection And then annotation and location by hand. The likelihood of over-fitting, we utilize the information growth 300 pictures will be utilized in the model, and afterward separate the data-set into the training set and the test set with an extent around 4:1 and a shade technique.

### 4.1 Hardware and Software geographic area

In our experiment hardware lab geographic area is Intel(R) core™ i9-9900k CPU @ 3.60GHz 3.6 GHz v3 @ 1545MHz GPU, NVIDIA GeForce RTX 2080TI, RAM 32 GB, and Software experiment geographic area are Linux, Ubuntu 16.04 LTS, python 2.7, CUDA 10.0, GCC 5.4.0.

### 4.2 Lab experiment

We set the initial learning rate to 0.001, the learning momentum to 0.9, the batch size to 24, and iterate 120k epochs. Then train multiple times and fine-tune the training parameters. The hyperparameters parameters of the model in the training process have been fine-tuned multiple times. We first adjust the momentum to 0.92, the weight decay to 0.00005, and iterate for 10k epochs. Based on the model obtained in the previous training, we continue to reduce the learning rate by 10%, change the momentum to 0.91 and weight decay to 0.000048 and change the iteration to 120k.

### 4.3 Datasets

The data-set consists of picture samples of electric equipments taken by drones. The data-set has four classes include: insulator-chain, tower-hpoint (hanging point), conductor-grip, and ground-hpoint (hanging point) as show in figure 7. The trainset contains 240 image samples, 1152 labels; the testset contains 60 sample images, 218 labels.

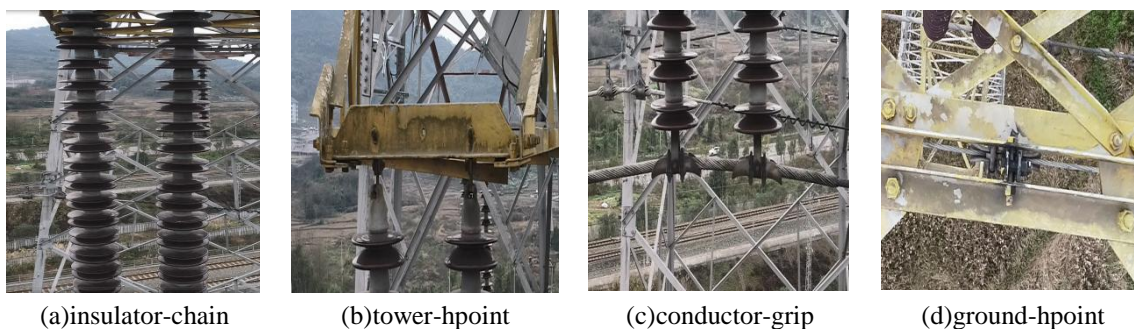


Figure 7: Dataset classes

### 4.4 Results and Discussion

As per our idea for a combination of MobileNet and SSD framework gives fast and better accurate results based on deep learning object detection. Our work is implemented by a throw-away MobileNet-SSD offer in all directions a deep neural network library of the Caffe-SSD module. Our process begins to input the image and the image pre-processed than will be given to the trained model. At first, we have Mobile-Net has trained with the COCO dataset, the COCO dataset contains approximately 2000 numbers of classes of numbers such as people, airplane, cars,



horses, etc. This input will be processed through our trained model and the objects present in the image are classified with their class name along with bounding boxes different images as an input practically applied in the SSD network and audited results.

Precision and time. The scatter diagram shown in Figure 8 visualizes the mAP and GPU time of the two meta-architectures. By analyzing the distribution intervals of the scatter points on the vertical and horizontal axes, we find that the SSD meta-architecture is faster but has relatively low precision, Faster R-CNN meta-architecture has higher precision, but the inference speed of a single image is much slower than SSD.

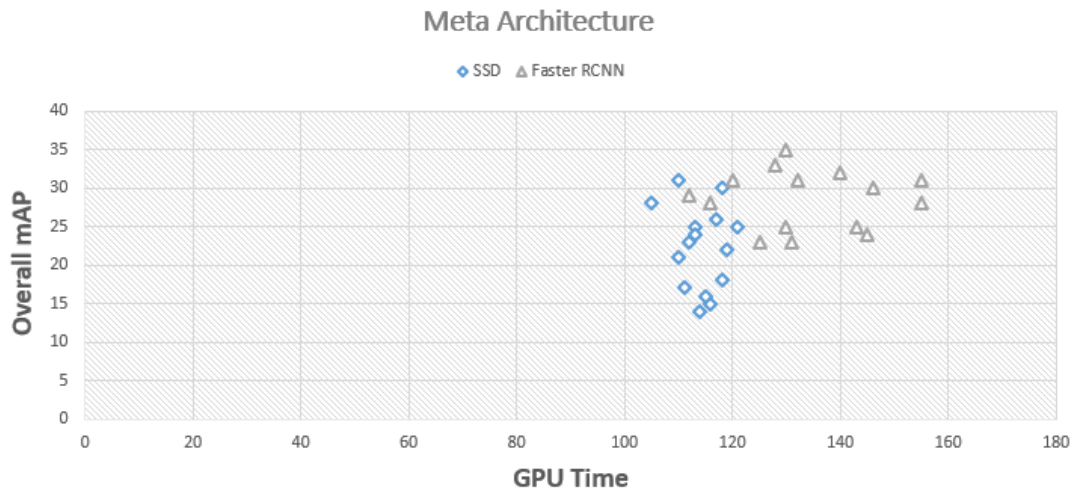


Figure 8: Accuracy and time. Overall mAP and average GPU time

The effect of the backbone (or feature extractor). In order to explore the effect of the backbone network, Figure 9 compares the three backbone networks of MobileNet, ResNet50 and GoogleNet. Generally, the stronger the detection performance of the backbone network, the stronger the object recognition effect of the corresponding meta-architecture, but experiments show that the lower accuracy MobileNet backbone is more suitable for SSD meta-architecture, and the model's detection effect will be faster, at the same time the accuracy is also higher.

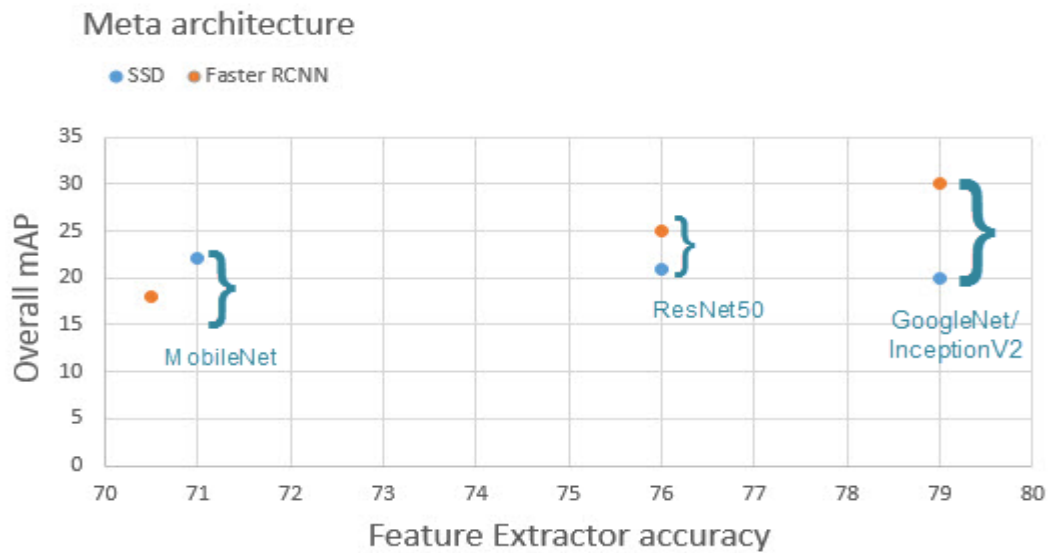


Figure 9: The effect of feature extractors for the overall mAP

The effect of image resolution size. It has been proved in many advanced experiments that the resolution of the image will have a greater impact on the accuracy of object detection. As shown in Fig 10, our experiments show that reducing the resolution of the input image reduces the accuracy of the model by an average of 18%, and also reduces the average running time by 23%. One of the reasons for this impact is that high-resolution images let for small objects to be solved.

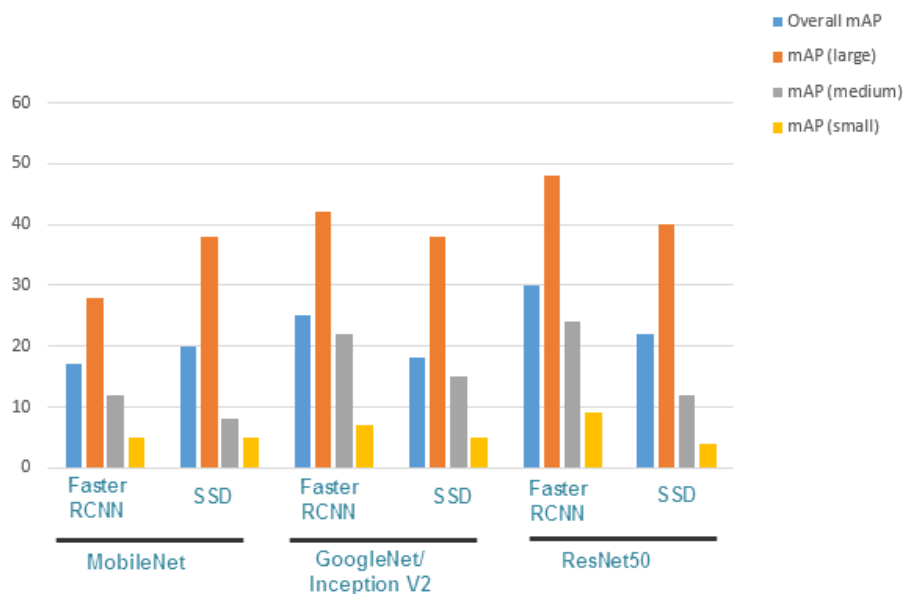


Figure 10: Accuracy stratified by object size, meta-architecture and feature extractor, we set the image resolution to 300

Memory analysis. We measured the total memory usage during the experiment as a benchmark. As shown in Figure 11, we tested 6 different combinations of backbone networks and meta-architectures. The final experimental data showed that the performance of the backbone network and memory usage with high correlation, the stronger the backbone network is, the more memory resources it takes up.

It can be seen from Figure 11 that when the backbone network is MobileNet, the least memory is required.

At the same time, we compare the Mobile-SSD model with the advanced single-stage model yolov5 and the

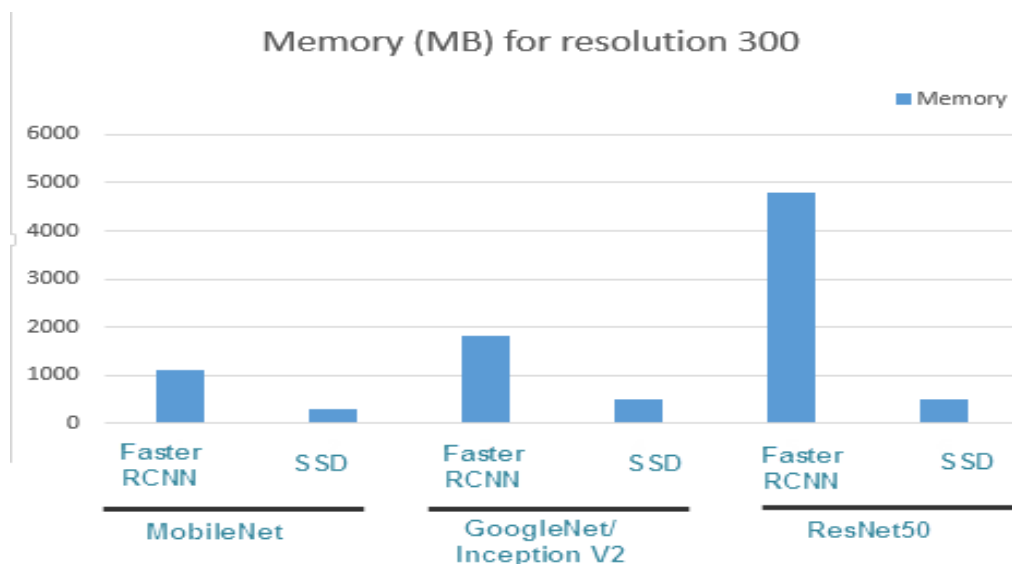


Figure 11: Memory usage for each model.

two-stage model Mask-RCNN. There are three evaluation indicators, namely mAP, Inference speed, and the size of the model. As shown in table 2. Through data comparison, it is found that the size and average detection time of a single image of the Mobile-SSD model is much shorter than Mask RCNN, and the detection mAP is much greater than that of yolov5. Compared with Mask-RCNN and yolov5, it achieves a better balance among the three evaluation indicators.

Table 2 Comparison of 3 Advanced CNN Models

Model	Model size	Inference speed	mAP[%]
Mobile-SSD	23M	0.021s	37.4
Mask RCNN[13]	240M	0.125s	52.3
YOLOv5[14]	14M	0.015s	25.6

## V. Conclusions

Our research uses Convolutional Neural Network (CNNs) to detect electric equipment from videos or images taken by drones, in this research, we allow the CNN model to recognize some of importance object types such as insulator-chain, conductor- grip, tower-hpoint, and ground-hpoint. At the same time, we used lightweight MobileNet-SSD, more accurate powerful framework object detector. Through some experiments, we compared the performance of SSD and Faster R-CNN meta-architecture under different backbone networks, taking into account GPU time, mAP, and memory usage. With the advantages of speed and precision, it is more suitable as a model for

deployment on mobile devices. We also compared it with modern models, the final experimental results show that MobileNet-SSD has achieved a better balance between this three evaluation Indices.

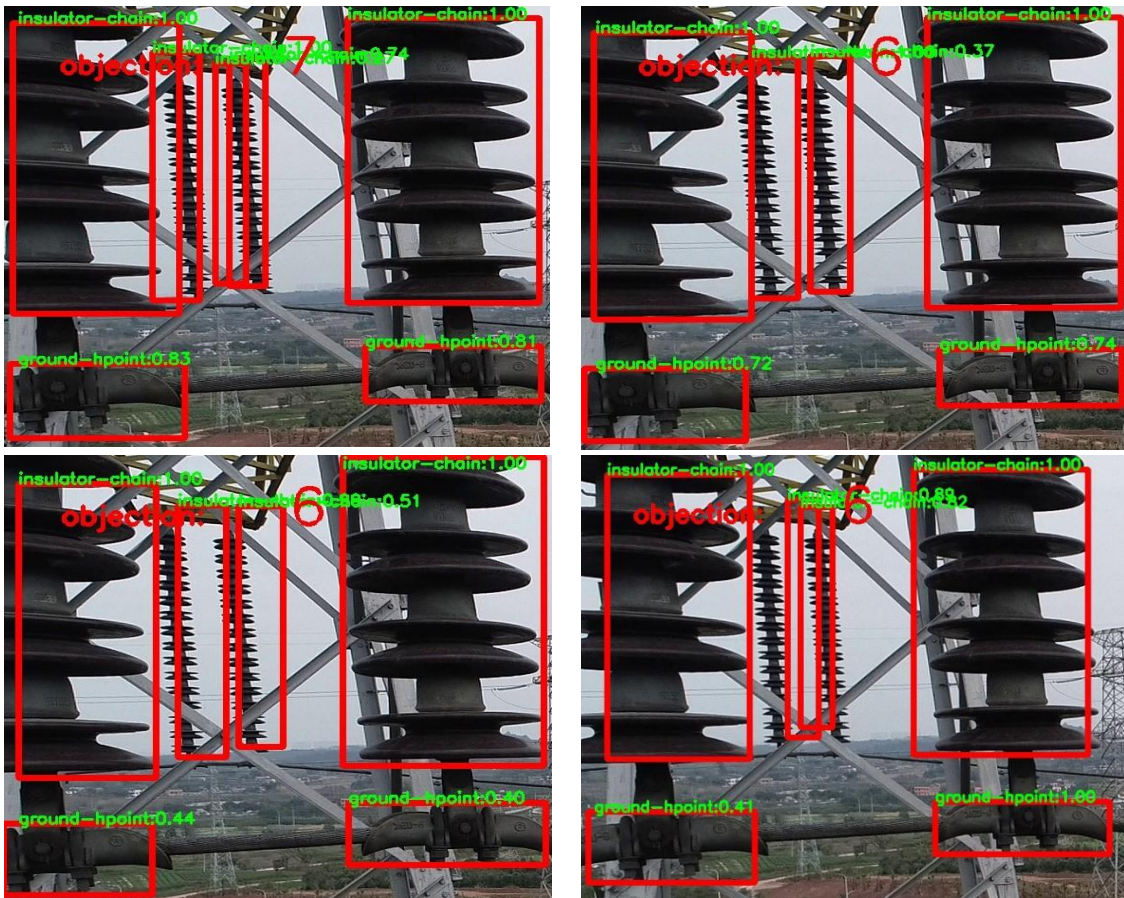
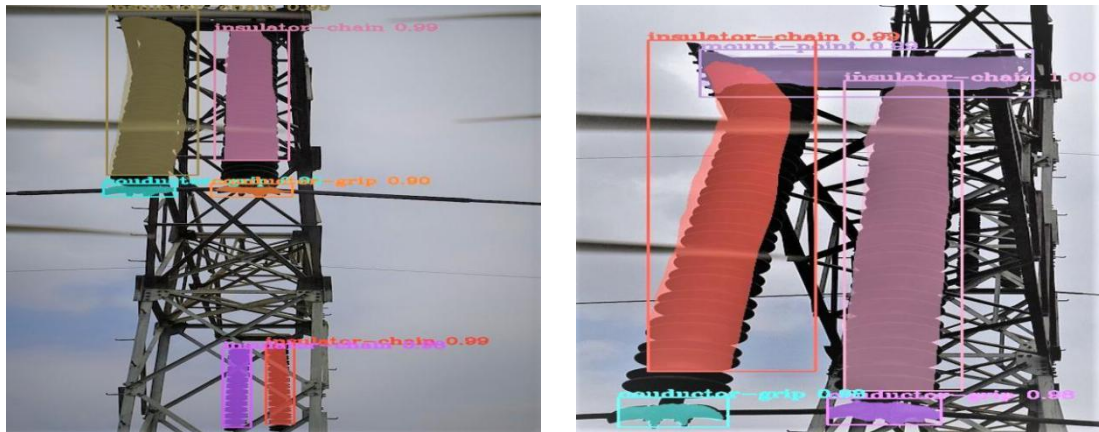


Figure 12: Result of detection with SSD base on MobileNet



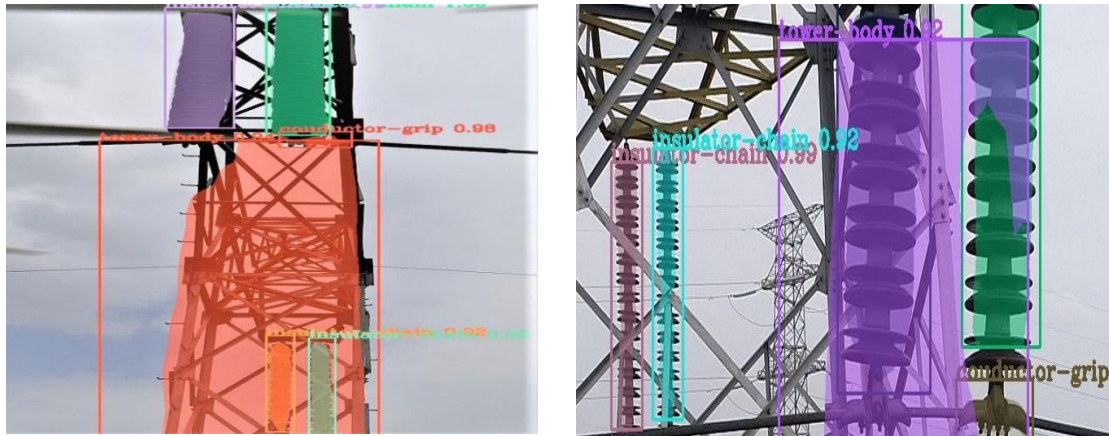


Figure 13: Result of detection with Mask R-CNN

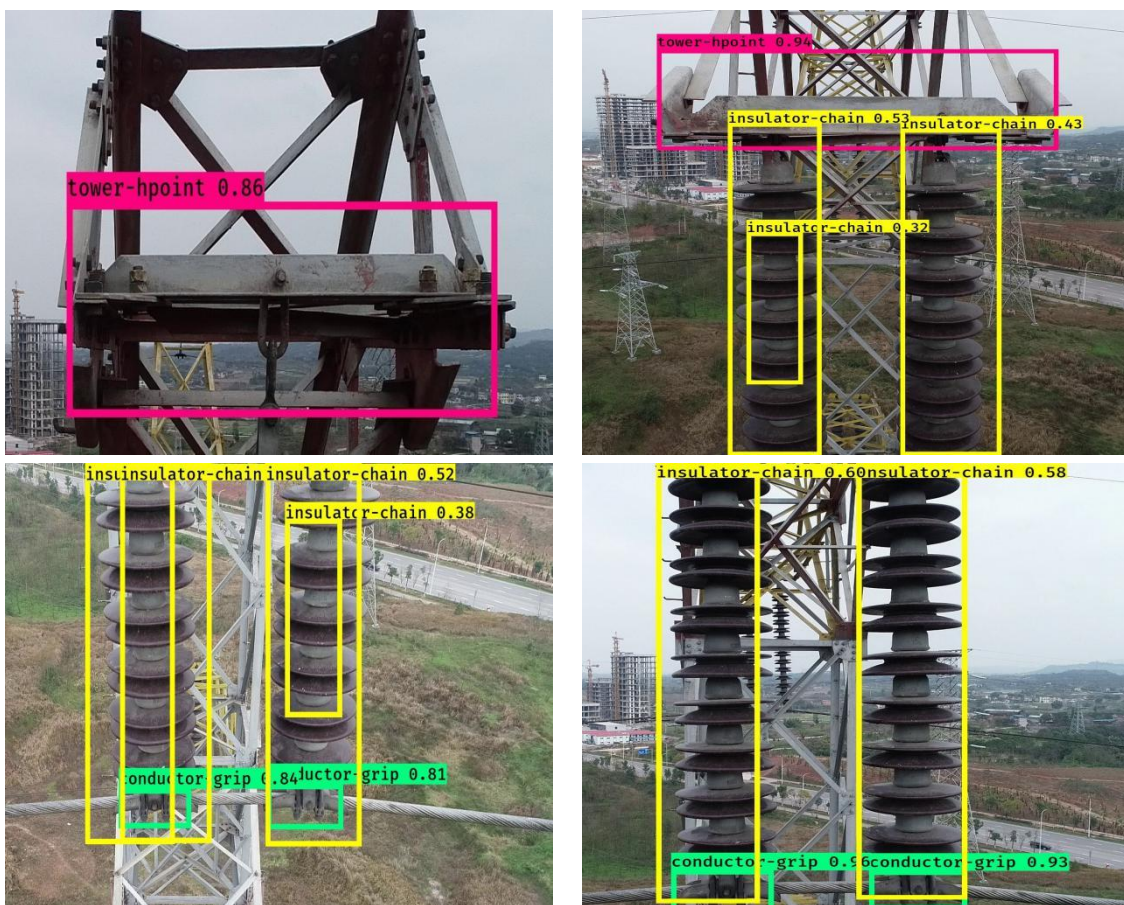


Figure 14: Result of detection with YOLOv5

### Acknowledgment

This research was funded by the China University Industry-University-Research Innovation Fund “New Generation Information Technology Innovation Project” (2019ITA03004).

## References

- [1] Chenfan Sun, Wei Zhan\*, Jinhiu She, Yangyang Zhang, "Object Detection from the Video Taken by Drone via Convolutional Neural Networks", *Mathematical Problems in Engineering*, vol. 2020, Article ID 4013647, 10 pages, 2020. <https://doi.org/10.1155/2020/4013647>.
- [2] Y. Xia, J. Lu, H. Li, and H. Xu, "A Deep Learning Based Image Recognition and Processing Model for Electric Equipment Inspection," *2nd IEEE Conf. Energy Internet Energy Syst. Integr. EI2 2018 - Proc.*, pp. 1–6, 2018, doi. 10.1109/EI2.2018.8582593.
- [3] B. Tian *et al.*, "Transmission line image defect diagnosis preprocessed parallel method based on deep learning," *Proc. - 2018 3rd Int. Conf. Mech. Control Comput. Eng. ICMCCE 2018*, vol. 1, pp. 299–303, 2018, doi. 10.1109/ICMCCE.2018.00068.
- [4] X. Miao, X. Liu, J. Chen, S. Zhuang, J. Fan, and H. Jiang, "Insulator detection in aerial images for transmission line inspection using single shot multibox detector," *IEEE Access*, vol. 7, pp. 9945–9956, 2019, doi: 10.1109/ACCESS.2019.2891123.
- [5] R. Bai, H. Cao, Y. Yu, F. Wang, W. Dang, and Z. Chu, "Insulator Fault Recognition Based on Spatial Pyramid Pooling Networks with Transfer Learning (Match 2018)," *ICARM 2018 - 2018 3rd Int. Conf. Adv. Robot. Mechatronics*, pp. 824–828, 2019, doi. 10.1109/ICARM.2018.8610720.
- [6] L. Steinmann, L. Sommer, and A. Schumann, "Fast and Lightweight Person Detector for Unmanned Aerial Vehicles," pp. 2–6.
- [7] S. Singh, "Real-Time Object Detection System using Caffe Model," pp. 5727–5732, 2019.
- [8] A. Driaba and V. Klyachin, "Recognition of Various Objects from a Certain Categorical Set in Real Time Using Deep Convolutional Neural Networks," pp. 4–9.
- [9] R. Sheri, N. Jadhav, R. Ravi, A. Shikhare, and S. Sannakki, "Object Detection and Classification for Self-Driving Cars," vol. 4, no. 3, pp. 179–183, 2018.
- [10] A. Arora, A. Grover, R. Chugh, and S. S. Reka, "Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector," *Wirel.Pers. Commun.*, no. 0123456789, 2019, doi: 10.1007/s11277-019-06294-1.
- [11] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.
- [12] He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [13] Thuan D. Evolution of yolo algorithm and yolov5: the state-of-the-art object detection algorithm. 2021.