IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 19, 2020, accepted May 11, 2020, date of publication May 19, 2020, date of current version June 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995608

A Deep Learning Method to Detect Foreign Objects for Inspecting Power Transmission Lines

JINGUO ZHU[1], YUE GUO[1], FANDING YUE[1], HUAN YUAN[1,2], AIJUN YANG[1,2], (Senior Member, IEEE), XIAOHUA WANG[1,2], (Senior Member, IEEE), AND MINGZHE RONG[1,2], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Xi'an Jiaotong University, Xian 710049, China
[2]State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding authors: Xiaohua Wang (xhw@mail.xjtu.edu.cn) and Mingzhe Rong (mzrong@mail.xjtu.edu.cn)

ABSTRACT Image online monitoring technology has been widely used in transmission lines inspection, but the intelligent and efficient foreign object detection still has a gap with the ideal. In this paper, we propose a deep learning method to detect invading foreign objects for power transmission line inspection. Specifically, we design our network based on the regression strategy with oriented bounding boxes to accurately predict spatial location and orientation angle of foreign objects, as well as their categories in cluttered backgrounds. Moreover, an easy yet effective Scale Histogram Matching method is proposed to be applied to the publicly available dataset, allowing useful patterns to be exploited to detect tiny foreign objects during the pretraining procedure and boosting detection performance even with limited annotated samples. Besides, we construct an image dataset that contains common foreign objects in transmission line scenarios to evaluate proposed methods, on which experiment results show our full model achieves accuracy with 88.1% mean Average Precision (mAP). Additionally, the efficient and compact network structure allows our network to run in real-time, which provides possibilities for practical use.

INDEX TERMS Image online monitoring technology, power transmission line inspection, deep learning, oriented bounding boxes, scale histogram matching.

I. INTRODUCTION

With the rapid development of urbanization, foreign object invasion has become one of the biggest safety hazards to power transmission lines, especially high-voltage transmission lines [1]–[3]. Due to this reason, the electric power companies invest significantly in the inspection and maintenance of power transmission lines [4], [5]. Image online monitoring, as one of the most widely used technologies, plays an important role in securing the safety of power transmission lines, by which the hidden dangers can be detected to prevent unplanned power outage [6]. However, traditional diagnoses rely mainly on manual monitoring, which is not only extremely time consuming and expensive, but also very prone to human errors [4]. Moreover, as the dramatic increase of image and video data, it becomes more and more necessary to detect foreign objects automatically and efficiently [7], [8].

In recent years, deep learning has made significant progress in computer vision [9]–[11]. Owing to the large-scale datasets in generic fields and high-performance computing hardware like GPU, the convolutional neural network (CNN) advances the object classification and object detection to a new level [12], [13]. CNN also has demonstrated its strong capability and adaptability in many industrial inspection fields [14]–[16]. However, due to the particularity of the application scenario, only a few studies have been undertaken using CNN in detecting foreign objects to monitor transmission lines [4]. There are some difficulties in simply applying the generic CNN to detect foreign objects around transmission lines [17]. Because most of those collected images do not contain foreign objects, which cannot provide useful characteristics to the CNN model during training. The lack of adequate images samples makes this detection task more challenging [5]. How a CNN based model can rapidly generalize from limited datasets to perform the task of detecting foreign objects is still a huge problem [18].

A general way to solve the task with only a few labeled samples available is by adapting a pretrained model from another task [19]. Instead of starting the learning process from scratch, visual tasks usually pretrain a model on a large benchmark dataset like ImageNet [12], and then fine-tune the pretrained model on a task-specific dataset. However,

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

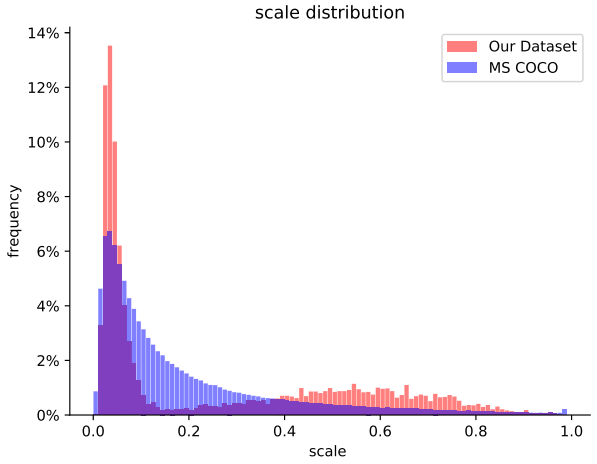


FIGURE 1. Scale Distribution of MS COCO dataset [13] and foreign object detection dataset constructed by ourselves. The scale of the objects has been normalized from 0.0 to 1.0.

the boosted performance will be greatly reduced when the domain of the task-specified dataset is different from that of the additional dataset used for pretraining [20]. Unfortunately, the object scale distribution of the image dataset used for our foreign object detection is quite different from that of the public datasets such as MS COCO [13] for generic object detection, as shown in Figure 1. Although both their scale of objects exhibits a long-tail distribution, the scale distribution of foreign objects around transmission lines is more uneven.

To make matters worse, generic object detectors only focus on detecting objects in terms of horizontal bounding boxes, which is accurate enough for most objects such as human or cars, etc [21], [22]. However, most of the foreign objects (like construction machinery such as tower cranes) in transmission line scenes, as shown in Figure 2, are typical instances of multi-oriented long and thin objects, which are better covered by oriented bounding boxes. Concretely, monitoring scenes often contain much background noise, which greatly affects the performance of the foreign object detection by interfering with the feature information of the targeted object when using the regression strategy with horizontal bounding boxes.

Given the above problems, we propose a novel CNN network with oriented bounding box regression to detect foreign objects for the inspection of transmission lines. The network, we call it *DFB-NN*, can predict the location, orientation angle, and category of each foreign object which may bring potential dangers to transmission lines. Additionally, our proposed Scale Histogram Matching technique also can be applied to the large-scale dataset to boost the pretraining performance of our CNN network. As a consequence, our *DFB-NN* can learn to extract powerful and representative features of foreign objects with limited supervised samples. More specifically, the main technique contributions made in this work can be summarized as follows:

- A multi-scale CNN network called *DFB-NN* is proposed, which can effectively integrate the low-level resolution information and high-level semantic information to provide more powerful features for foreign object detection. To the best of our knowledge, the proposed model is the first one embedding oriented bounding-boxes regression strategy into a CNN model for facilitating foreign object detection, at least in the field of transmission line inspection.
- Most importantly, an easy yet effective dataset preprocessing method, Scale Histogram Matching, is proposed to be applied to the pretrained dataset for a specified task, which allows the compact and effective network structure learning to extract useful and representative features from limited samples.
- We construct a monitoring image dataset containing five major types of foreign objects which are most prone to power transmission failure. Additionally, several experiments are carried out on the dataset to demonstrate the effectiveness of our compact network and our proposed methods.

II. RELATED WORK

A. IMAGE ONLINE MONITORING

Transmission lines against invading foreign objects is the most basic errand in the assurance of power system [23], [24]. Overhead line failures caused by invading foreign objects are relatively common anomalies, which can be caused by climatic conditions, human mistakes, flames and smoke, and construction machinery, etc [5], [25]. With the dramatic increase of sensing data, it becomes more and more desirable to make image online monitoring. [8], [26], [27] develop effective methods to detect power lines based on the images captured by automated UAV (unmanned aerial vehicle). Besides, [28] builds a model to select the best thresholds for changing scenarios to detect and track power lines when using the unmanned aerial vehicle inspection, which is proved to be robust to the complexity of the real world. However, these methods focus only on how to detect power lines from the cluttered background, which is not enough for automatic diagnosis. And [29] proposes an image processing method to measure icing thickness based on monitoring image. But this algorithm can not detect the potential dangers around the power lines. [17] uses a CNN model based on RCNN to detect foreign objects. However, the algorithm cannot achieve real-time running and only can detect wire-wound foreign objects, while it is powerless to the potential harm caused by construction machinery. In contrast to general object detection in RGB images, there are also some methods to inspect electrical equipment by using infrared thermography [30], [31]. Moreover, [7] also presents a real-time deep learning approach to detect oriented electrical equipment detection in thermal images.

B. OBJECT DETECTION

The traditional generic object detectors are based on the sliding window paradigms or region proposal classification using hand-crafted features [32]–[34]. With the development of deep learning, object detectors based on CNN have become a predominant trend in the field of generic object detection

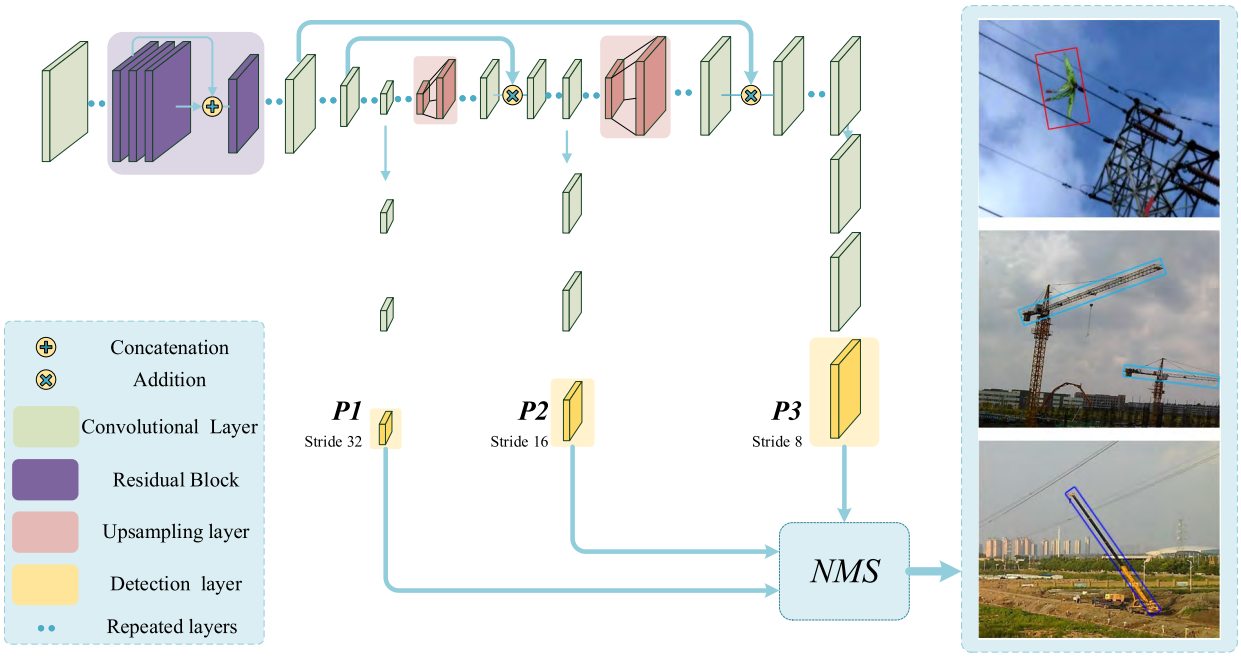


FIGURE 2. **An illustration of our proposed network architecture. Our model is based on the multi-scale feature pyramid structure to detect the category, spatial location and oriented angle of foreign objects. The backbone network is specifically designed for the transmission line inspection task. The basic CNN modules are illustrated in the legend at the bottom left.**

and have led to remarkable breakthroughs in the fields of detection applications [30], [35]–[37]. As the state-of-art object detectors, two-stage detectors such as R-CNN [38] and its descendants [39], [40] first generate class agnostic region proposals and then predict the specific class label and refine the location regression. On the other hand, the single-stage detectors regress the default anchors into detection on the feature maps directly, which can be of high computational efficiency but sacrifices partial accuracy [41]–[43]. However, the specified task like transmission line inspection is different from the issue of generic object detection. The targets around transmission lines are usually tiny, occupying only a small area in the monitoring image, and these objects are usually tilted or even dense in the images, which is a distinguishing characteristic from the MS COCO benchmark [13]. Feature pyramid network (FPN) [44] that uses the top-down architecture with lateral connections is proposed to detect objects with drastic changes in scale. Additionally, [21], [22] also present methods to detect ship targets, which also can be exhibited at any orientation, in the remote sensing images. These methods will make huge reference work for our work.

III. APPROACH

We propose a deep learning model to detect foreign objects that may bring potential hazards to transmission lines. By predicting a set of oriented bounding boxes parameterized by their center locations, widths, heights, and orientation angles, timely and accurate detection can effectively prevent accidents of transmission lines. The overall framework of our model is illustrated in Figure 2, whose backbone network is specifically designed according to characteristics of the foreign objects in transmission line environments.

A. NETWORK ARCHITECTURE

As we all know, both the low-level and high-level features are very important to object detection performance. Just like in the transmission line scenarios, the scale of foreign objects can vary greatly, depending on the distance of these objects from the monitoring cameras. To keep the completeness of the semantic and spatial information, we adopt a multi-scale feature pyramid connection to fuse multi-level information in our CNN feature extractor.

The multi-scale feature pyramid structure is proved to be effective [44], in which the bottom-up pathway and the top-down pathway are connected by lateral connection. Specifically, we choose the last 3 layers of the backbone network as reference features of the bottom-up path in our network. These 3 feature layers are represented as $P1$, $P2$, and $P3$ in the order of feature scale from small to large, which will be used in subsequent experiments. A features hierarchy is computed at different scales with a downsampling scale-step of 2. Additionally, the top-down path enhances those bottom-up layers which have stronger semantic information but have weaker or even missing spatial information. Spatial information is supplemented by lateral connections from the feature with the same spatial size via lateral connection, while semantic information is well preserved.

B. ORIENTED BOUNDING-BOX REGRESSION

In the regression strategy of ordinary object detectors like [38], [41], [43], a horizontal bounding box is determined

as (x_c, y_c, w, h), where (x_c, y_c) specifies the coordinate of the center point, w is the width, h is the height of the object box. These four parameters can better denote the spatial location and extent of horizontal targets in close proximity. However, it is difficult for 4-parameter bounding boxes to distinguish those inclined targets while choosing rotated bounding boxes to regress might be an effective way to detect targets.

In order to avoid the inherent drawbacks mentioned above, our network uses the multi-scale rotated bounding boxes to detect the foreign objects precisely. Based on these four parameters, we add an extra parameter θ to uniquely determine a rotated bounding box (x_c, y_c, w, h, θ), where the specifications of the first four parameters are the same as the horizontal box but the last parameter θ represents the angle of the rotated object. The predicted angle is in the range of $(-90°, 90°]$ since most of our detected targets are centrally symmetric.

Consistent with other anchor-based generic object detectors, our network also places prior anchors densely on the final feature maps. Let $A = (x_0, y_0, w_0, h_0, \theta_0)$ denotes a rotated anchor, while a nearby ground truth can be denoted as $G = (x, y, w, h, \theta)$. Instead of predicting the coordinates of the ground truth box G directly, the network regresses the scale-invariant translation of the center and the log-space translation of size, as well as the tangent of angle bias, related to the anchor's parameters. Specifically, the regression targets $T = \left(t_x, t_y, t_w, t_h, t_\theta\right)$ in this cell placing anchor A should be

$$\begin{aligned} t_x &= \frac{x - x_0}{w_0} \\ t_y &= \frac{y - y_0}{h_0} \\ t_w &= \log\left(\frac{w}{w_0}\right) \\ t_h &= \log\left(\frac{h}{h_0}\right) \\ t_\theta &= arctan\left(\theta - \theta_0\right). \end{aligned} \tag{1}$$

The prior sizes and angles can be obtained by clustering the ground truth boxes in the training dataset via K-means algorithm. By this means, the network can detect the foreign objects with the prior distribution of the specific dataset, thus making the training more stable and the optimization easier to converge.

C. TRAINING

1) LOSS FUNCTION

Consistent with generic object detectors, the loss function of our network should also be a multi-task loss that includes a classification loss L_{cls} and a location loss L_{loc}. The full multi-task loss L can be represented as

$$L = L_{cls} + \lambda_{loc} L_{loc} \tag{2}$$

where the hyper-para λ_{loc} is used to balance the two losses, which is set to 1 empirically.

Detecting foreign objects in transmission line monitoring images is faced with an extreme foreground-background class imbalance encountered during the training of anchor-based detectors. To make our detector more focused on these sparse but difficult foreground samples and prevent a large number of easy background samples from dominating the loss during training, we design our classification loss based on focal-loss [42]. Specifically, let x_i be the indicator whether the ith anchor is matched to a ground truth box, where 1 means matched at least one ground truth is matched to anchor, and 0 means no ground truth is matched to the ith anchor, then the anchor will be assigned into the background class. The classification loss is calculated as

$$L_{cls}\left(x_i, c_i, y_i\right) = \begin{cases} (1 - c_i)^\alpha \, CE\left(y_i, c_i\right) & x_i = 1 \\ (c_i - 0)^\alpha \, CE\left(y_i, c_i\right) & x_i = 0 \end{cases} \tag{3}$$

where c_i is the class prediction, y is the class label matched to the ith anchor. The cross-entropy (CE) loss for the class prediction is added a modulating factor $(1 - c_i)^\alpha$ or $(c_i - 0)^\alpha$ with tunable parameter $\alpha \geq 0$.

Considering that the ground truths for the regression task vary within a wide range, we use a more robust loss function for L_{loc}. Therefore, the loss of location and angle offset can be calculated as:

$$L_{loc}\left(x_{ij}, T_i, R_j\right) = \begin{cases} \left|\left(T_i - R_j\right)\right|^\gamma & \left|T_i - R_j\right| \leq 1 \\ \gamma\left|\left(T_i - R_j\right)\right| - \gamma + 1 & otherwise. \end{cases} \tag{4}$$

where the x_{ij} denotes that the ith anchor is assigned to the jth ground truth. T_i and R_i are the regression targets calculate by Equation 1 and the regression outputs by the network respectively. The parameter γ can adjust the sensitivity of the loss function to the outliers which can improve the robustness of the L_{loc}. When $\gamma = 2$, L_{loc} is equivalent to smoothL1 in [39].

Through cross-experiments, our network can achieve the best performance when focal loss modulating factor $\alpha = 2$ and the parameter $\gamma = 3$ for robust L_{loc}. The setting of parameters α and γ will be used in our subsequent experiments.

2) MATCHING STRATEGY

During training, if an anchor is close enough to a ground truth box, the anchor should be assigned to the prediction task of regressing to this ground truth [41]. The commonly used algorithm to evaluate the distance of two boxes is Intersection over Union (IOU), which is defined as:

$$IOU\left(A, G\right) = \frac{Area\left(A \bigcap G\right)}{Area\left(A \bigcup G\right)}, \tag{5}$$

where the $\bigcap$ and $\bigcup$ are boolean operations between an anchor A and a ground truth G. The criterion IOU can reflect the proximity of location, aspect and scale between anchors and ground truths. However, the calculation of IOU between two rotated boxes with different angles is more complicated.

Inspired by [45], we choose a simplified IOU, named IOU_{RB}, as our distance criterion for training, which is calculated by

$$IOU_{RB}(A, G) = \frac{Area\left(\hat{A} \bigcap G\right)}{Area\left(\hat{A} \bigcup G\right)} |\cos(\theta - \theta_0)|, \quad (6)$$

where $\hat{A} = (x_0, y_0, w_0, h_0, \theta)$. The temporary rotated box $\hat{A}$ will make the intersection part easier to calculate. Moreover, the IOU_{RB} decouples the distance of geometric size and rotated angle between boxes. The previous one ensures that the two boxes are close in position and scale, while the latter one ensures that the difference between angles is not too large by the *cosine* function.

It should be noted that IOU is still used to filter out the redundant detections for non-maximum suppression when post-processing, while IOU_{RB} is only used for training when selecting positive anchors.

D. SCALE HISTOGRAM MATCH

Inspired by the illumination histogram matching technology which can transform an image's illumination histogram to another specified histogram, we propose an easy but efficient scale histogram matching as a detector calibration technique to keep the object scale consistency between two datasets.

The probability density function of object scale s of dataset $\mathcal{X}$ is denoted as $P_{scale}(s; \mathcal{X})$. Our method is to use a scale transformation function T to match the probability distribution of object scale in the additional pretrained dataset $\mathcal{I}$ to that in the targeted dataset $\mathcal{E}$ for the specified task, which is given in Equation 7.

$$P_{scale}(s; T(\mathcal{I})) \approx P_{scale}(s; \mathcal{E}) \quad (7)$$

The dataset $T(\mathcal{I})$ after scale histogram matching can be used to pretrain our scratch model, of which object scales are forced to be aligned to the task-specified dataset. In our paper, MS COCO is used as the additional pretrained dataset $\mathcal{I}$, while the dataset used for our transmission line inspection task is used as the task-specified dataset $\mathcal{E}$. The details of Scale Histogram Matching for the additional pretrained dataset are shown in Algorithm 1, in which $\tilde{\mathcal{I}} = T(\mathcal{I})$. We roughly assume that the scales are uniformly distributed over any scale range $R[K]$ on the dataset scale histogram. Therefore, the cumulative histogram can be calculated continuously and expressed as a continuous piece-wise function composed of multiple linear functions. Because there is maybe more than one object with different scales in one image, resizing each object in one image will destroy the image structure. We regard the mean scale s_i as the scale representative of the i-th image, which is used in [46].

IV. EXPERIMENTS

In this section, we first introduce our dataset and present our experimental setting in detail. Then we present the results of some groups of experiments to evaluate our proposed methods.

Algorithm 1 Scale Histogram Matching

Input: $\mathcal{E}$: targeted dataset
$\mathcal{I}$: additional pretrained dataset
N: number of bins in dataset scale histogram
I_i: the i-th image in dataset $\mathcal{I}$;
G_{ij}: the j-th object in I_i;
$R[k]$: the scale range of k-th histogram bin.
$S(G)$: the scale of object G
Output: $\tilde{\mathcal{I}}$: dataset after scale matching
begin
  $f_1(\mathcal{E})$,$f_2(\mathcal{I})$ ← histogram of $\mathcal{E}$,$\mathcal{I}$
  $F_1(\mathcal{E})$,$F_2(\mathcal{I})$ ← cumulative histogram of $\mathcal{E}$,$\mathcal{I}$
  $\tilde{\mathcal{I}} \leftarrow \varnothing$
  for I_i *in* $\mathcal{I}$ **do**
    // calculate mean size of objects in I_i
    $s_i \leftarrow mean\left(S\left(G_{ij}\right)\right)$
    // calculate the targeted scale
    $\tilde{s}_i == F_1^{-1}(F_2(s_i))$
    // resize image
    $\tilde{I}_i = resize\left(I_i, \sqrt{\tilde{s}_i / s_i}\right)$
    $\tilde{\mathcal{I}} \leftarrow \tilde{\mathcal{I}} \cup \tilde{I}_i$
  end
  return $\tilde{\mathcal{I}}$
end

TABLE 1. Statistics on the number of images and instances for each category in the training and testing dataset.

subsets	training set		val & testing set	
	images	instances	images	instances
WO	503	560	500	532
BC	1038	1062	1023	1096
FT	1026	1334	1035	1457
TC	1217	1875	1287	1920
WF	500	509	500	507

A. DATASET AND SETTING

We collected 8000 transmission line monitoring images captured by cameras mounted on transmission line towers. These images contain some invading foreign objects that may bring harm to transmission lines, including construction machinery like tower cranes, wildfire or smog in the surrounding environment and foreign objects wound around transmission lines such as kites. There are 10852 foreign objects with complex background in this dataset. Based on the difference in appearance and the occurrence frequency of foreign objects in the collected samples, we divide these foreign objects into five categories: on-board cranes (BC), tower cranes (TC), forklift trucks (FT), wire-wound foreign objects (WO) and wildfires (WF).

We also exhibit the statistical distribution of aspect ratio and oriented angle θ of foreign objects in our collected dataset in Figure 3. The aspect ratio tends to be larger than 1, indicating that the objects are usually long and thin. It is important to note that, there are quite a few samples rotated at extreme tilt angles around $\pm 70°$. As we mentioned before, it is the

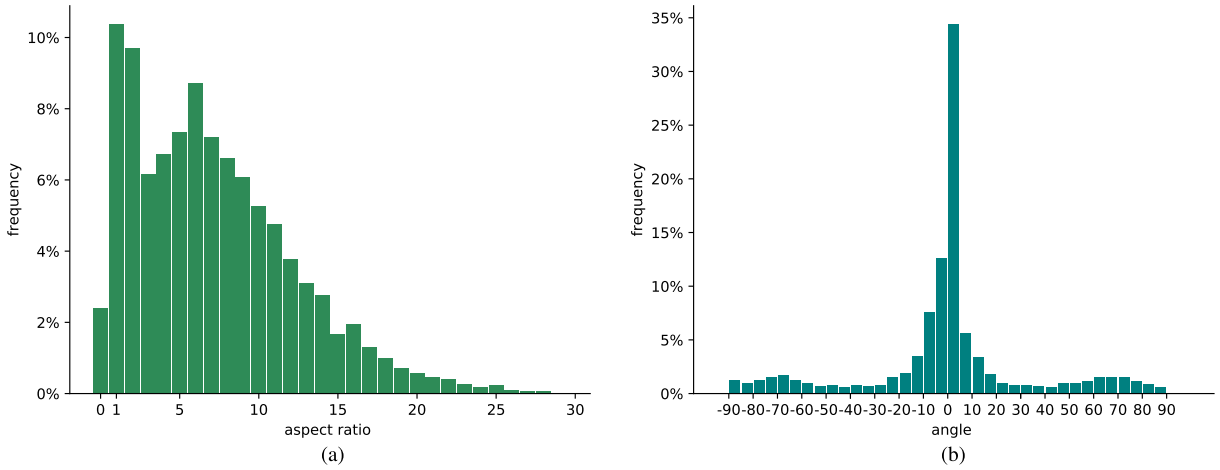


FIGURE 3. Main statistics distribution of the samples in our dataset: (a)distribution of aspect ratio, (b)distribution of angle θ.

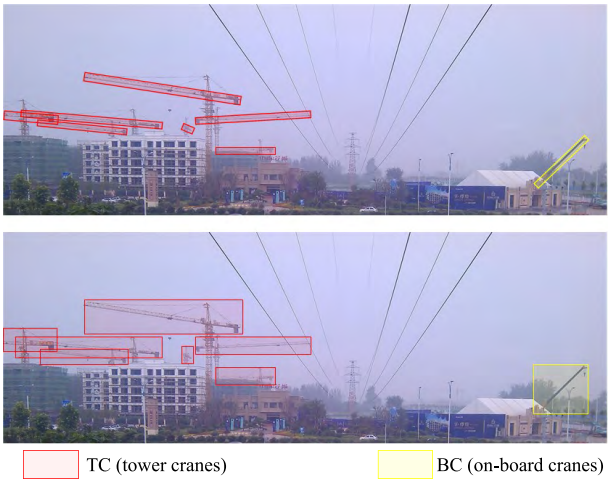


FIGURE 4. A typical transmission line scene containing invading foreign objects. Two different annotation protocols are also exhibited: instances annotated with oriented bounding boxes (the top image) or horizontal bounding boxes (the bottom images).

unbalanced distribution of foreign objects in the power transmission environment that indeed increases the difficulty of detection. Considering this, we attempt to use the oriented bounding box regression strategy to detect foreign objects in this particular application scenario.

To compare the pros and cons of the two different forms of bounding box regression strategy, we use two different annotated protocols to annotate all 8000 images. Just as Figure 4 shows, the spatial extent and class of each instance in the collected images can be specified by oriented and horizontal bounding boxes. These two datasets are called RBB-FO (rotated bb) and HBB-FO (horizontal bb) respectively. Each of them contains 4000, 2000 and 2000 images for training, validation, and testing.

B. EVALUATION INDICATORS

Similar to other object detection framework, mean Average Precision (mAP) is used to evaluate the detection performance of models quantitatively. The detected box which is considered to be a true positive should satisfy the following two conditions: the matching IOU between the detected box and any ground truth and should be higher than an IOU threshold $T_{IOU} = 0.5$, and the abstract difference between their angles should be smaller than the angle threshold $T_\theta = 10°$.

C. IMPLEMENTATION DETAILS

We use PyTorch to implement and train CNN detectors. For the input images, we resize them to 640 $\times$ 640. We use randomly horizontal flipping as the only data augmentation meth during the training phase. All the networks are trained for 60 epochs. The initial learning rate is set to 0.02 with a decay rate of 0.01 every 20 epochs. We optimize our CNN models using Stochastic Gradient Descent (SGD) method with 0.9 momentum and 0.0005 weight decay. For fair comparison, we set the batch size to 32 on 4 RTX 2080ti for all the networks.

For anchor settings, we use the anchor scale of 32^2 to 512^2 on features, each of which has three aspect ratios $1:2, 1:1, 2:1$. Similar to [44], anchors will be located at each pyramidal level if a network adopts a multi-scale network structure. Due to some experimental models aim to predict the angle of an object by regressing an oriented box, these models whose regression strategies are with oriented bounding boxes will add additional rotated anchors. Specifically, apart from vertical anchors of which angle of boxes is 0°, we also rotate all anchors by $\pm 70°$, making those anchors more suitable for regressing to oriented objects. The degree number 70 is chosen according to the angle distribution of our RBB-FO in Figure 3 (b). For oriented bounding box regression, anchors are assigned to ground-truth objects using an IOU_{RB} threshold of 0.5, and to background if their IOU_{RB} is smaller than 0.4. Additionally, anchors will be ignored during training if they have IOU_{RB} in $[0.4, 0.5]$. The same assignment strategy is also applicable to the regression prediction of horizontal bounding boxes, except that we use traditional IOU instead of IOU_{RB} when selecting positive anchors via threshold comparison.

TABLE 2. Performance of two networks with different pretraining strategy. Abbreviations explanation: task-specified dataset (TD); pretrained dataset (PD); HBB-FO dataset (H-F); RBB-FO dataset (R-F); M: MS COCO dataset (M); ImageNet dataset (I); Scale Histogram Matching (SHM).

network	TD	PD	SHM	mAP
Faster R-cnn	H-F	\		82.5
	H-F	I		84.9
	H-F	M		85.7
	H-F	M	√	87.4
YOLO v3	H-F	\		78.5
	H-F	I		81.2
	H-F	M		81.9
	H-F	M	√	84.3
	R-F	\		72.1
	R-F	I		80.5
	R-F	M		81.5
	R-F	M	√	83.2

D. ABLATION EXPERIMENTS

1) EVALUATION OF SCALE HISTOGRAM MATCHING

Scale Histogram Matching is proposed mainly to alleviate the problem caused by the difference of scale distribution between the additional pretrained dataset and task-specified dataset. We will evaluate its effectiveness in the following.

Results in Table 2 suggest that pretraining on MS COCO often gets better detection performance than pretraining on ImageNet dataset. However, the improvement gained from knowledge transfer from other dataset is quite limited, since the object scale of MS COCO is different from that of the foreign objects around transmission lines.

Consistent with our intention, the detection performance can be further improved by transforming Scale Histogram Matching on dataset MS COCO, which validates the effectiveness of the scale alignment strategy for different datasets. Specifically, Fater RCNN, as one of the representatives of CNN-based two-stage object detectors, can get 1.7% improvement in mAP. Also as the most representative of the one-stage CNN detector, YOLO v3 can get 2.4% improvement. We found that, with Scale Histogram Matching, the one-stage detector can get higher mAP improvement than two-stage detectors. This may be due to the ROI pooling operation [40] in the two-stage detector to some extent can alleviate the impact of the change of object scale on the detection performance.

Furthermore, just as the experiment results with the dataset RBB-FO shown in Table 2, the scale matching method can still improve the detection performance of detector with oriented bounding box regression strategy, which is applied to the dataset with horizontal bounding box annotation though.

An obvious comparison is that the results on the RBB-FO dataset are generally worse than HBB-FO. We speculate that this may be caused by two reasons: the pretrained set (MS COCO) has a different labeled protocol, which is labeled with horizontal bounding boxes; and the task to regress oriented bounding boxes is more difficult since the output requires the tilt angle and the spatial extent of the objects which is more stringent with less background noise.

Considering the significant performance gain brought by Scale Histogram Matching, detection models will all be pretrained by the dataset processed by Scale Histogram Matching in the subsequent experiments.

2) EVALUATION OF ORIENTED BOUNDING-BOX REGRESSION

The commonly applied methods for object detection use the horizontal bounding boxes to specify the spatial extent of objects, such as Faster R-CNN [40], YOLO [43] and SSD [41]. Because the label protocols of the two datasets are different, it is not a simple matter to make a fair quantitative comparison between regression strategy with oriented bounding boxes and that with horizontal bounding boxes. The task of regression with rotated boxes for the dataset RBB-FO is more difficult, since it has more requirements such as the oriented angle and stricter spatial extent. Therefore, it is not appropriate to simply use the accuracy performance like mAP to compare the performance of the detectors.

Here, we show the advantage of using oriented bounding boxes to detect foreign objects qualitatively. We design two hybrid versions of our CNN model, denoted as *DFB-NN*$_h$ and *DFB-NN*$_r$, to regress horizontal bounding boxes and oriented bounding boxes respectively. Just as Figure 5 shows, both the methods using horizontal bounding boxes and oriented bounding boxes can detect foreign objects well when the targeted objects are upright and distributed scattered like forklift trucks in Figure 5. However, when the objects are tilted very severely and maybe densely distributed such as the tower cranes, our method using oriented bounding boxes can detect targeted objects better and locate them more precisely. Besides, the method using horizontal bounding boxes may not be able to distinguish crowded objects, which is a natural disadvantage of the post-processing non-maximum-suppression (NMS).

E. COMPARISONS WITH OTHER STATE-OF-THE-ART DETECTORS

To show the benefit of our proposed method, we carry out comparative experiments to verify the advanced full model proposed in this paper by comparing it with the state-of-art methods that were designed for traditional generic object detection with horizontal bounding boxes.

Table 3 presents the AP for each category, together with the mAP for the global performance of the detectors. The results show that our proposed oriented bounding box regression model achieves the best detection performance, outperforming all other upright object detection methods. We assume that this advancement is mainly benefited from the consideration of orientation of foreign objects in our proposed detector, by which our model is more robust to the appearance variance caused by rotation and suffered less from background noise around the targeted objects.

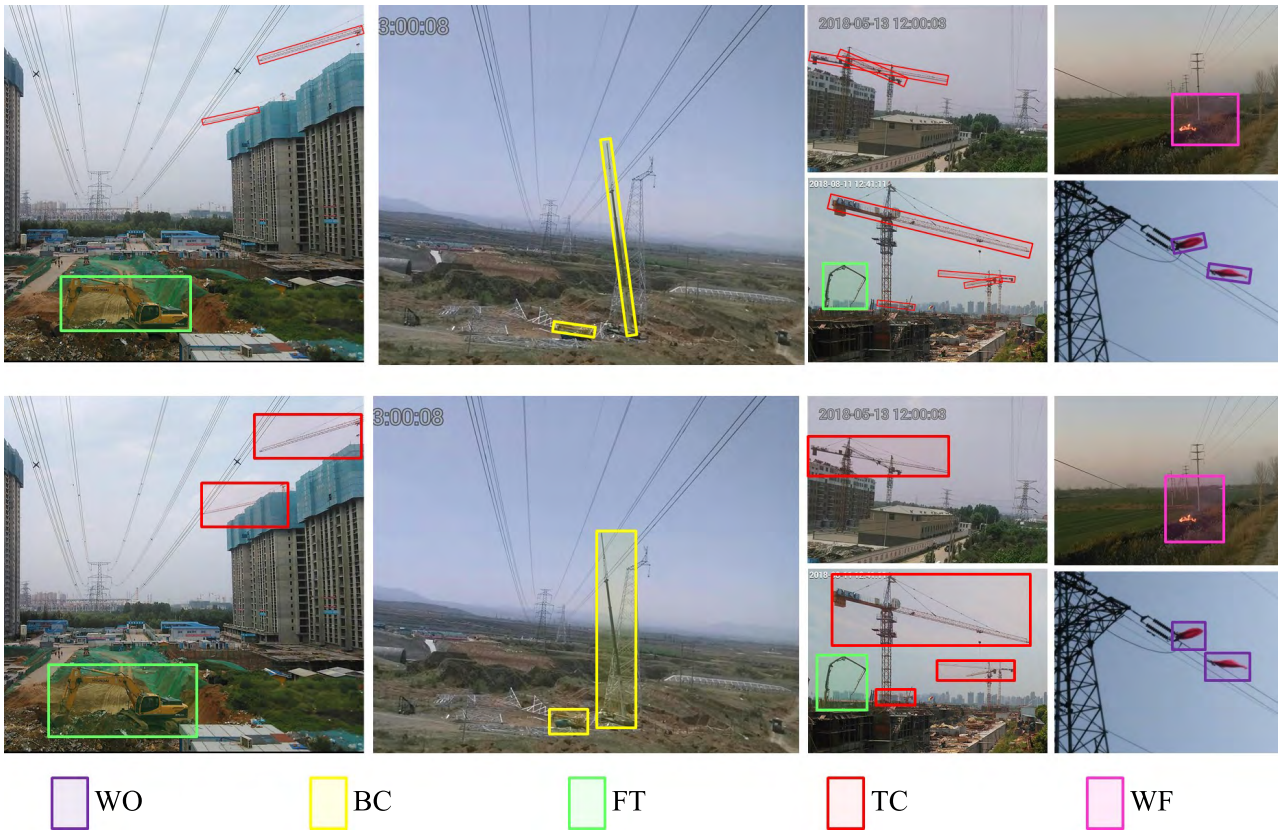


FIGURE 5. Comparison of detection results by our proposed *DFB-NN*$_r$ (images in the top row) and *DFB-NN*$_h$ (images in the bottom row). Different types of detected foreign objects are displayed in bounding boxes with different colors.

TABLE 3. Comparison results with other state-of-the-art detectors. Abbreviations explanation: MS represents whether the network uses multi-scale feature structure; STAGE: is the network one-stage (1) or two-stage (2) detector; BB: are horizontal (h) or rotated (r) bounding boxes used for regression.

Detector	MS	STAGE	BB	AP_{WO}	AP_{BC}	AP_{FT}	AP_{TC}	AP_{WF}	mAP	FPS
Faster R-cnn		2	h	87.0	88.3	88.7	89.1	84.8	87.4	7
FPN FRCN [44]	√	2	h	88.1	88.5	89.5	89.2	83.2	88.2	6
YOLO v3$_h$	√	1	h	80.1	77.6	87.7	86.8	72.1	84.3	67
YOLO v3$_r$	√	1	r	79.2	78.5	80.3	88.7	70.3	83.2	52
SSD$_h$		1	h	73.3	78.7	80.3	79.5	71.0	77.2	33
SSD$_r$		1	r	70.9	77.7	79.9.3	78.2	71.3	74.8	19
Retinanet$_h$	√	1	h	82.3	80.2	87.9	88.2	80.1	85.3	45
Retinanet$_r$	√	1	r	82.5	80.9	85.7	87.9	81.8	85.5	31
DFB-NN$_h$	√	1	h	86.5	87.7	88.2	88.3	85.2	87.5	53
DFB-NN$_r$ w/o		1	r	82.1	82.3	89.2	89.3	83.2	85.7	61
DFB-NN$_r$	√	1	r	87.4	83.4	90.2	90.7	87.2	88.1	45

Table 3 also lists the time consumption of each model in terms of frames per second (FPS). Since the two-stage detector includes an extra network for region proposals, these detectors (such as faster RCNN) are much less efficient than one-stage detectors, which indicates that these networks are unsuitable for scenarios with high real-time requirements such as online monitoring of transmission.

Our *DFB-NN* is a one-stage detector essentially, which is not separating detection and proposal, making the overall pipeline single-stage. One-stage detectors sacrifice part of detection performance in pursuit of speed. Therefore, all the one-stage detectors in Table 3 perform worse than the two-stage detectors, even though they can run in real-time way. According to the characteristics of the task-specified dataset and detection task, our proposed *DFB-NN* network can achieve quite competitive detection results on the basis of real-time running. In terms of horizontal bounding boxes regression task, our *DFB-NN*$_h$ network can reach 87.5% mAP, only 0.7% mAP lower than FPN FRCN. Besides, the *DFB-NN*$_h$ runs over 50 FPS, much faster than all the other two-stage methods. While the faster RCNN can not exceed 10 FPS under the same experimental conditions.

By modifying the *DFB-NN*$_h$ to *DFB-NN*$_r$, the network can still achieve satisfying results. The model's mAP reaches 88.1% and even is a little higher than *DFB-NN*$_h$, which demonstrates that our approach can detect foreign objects in

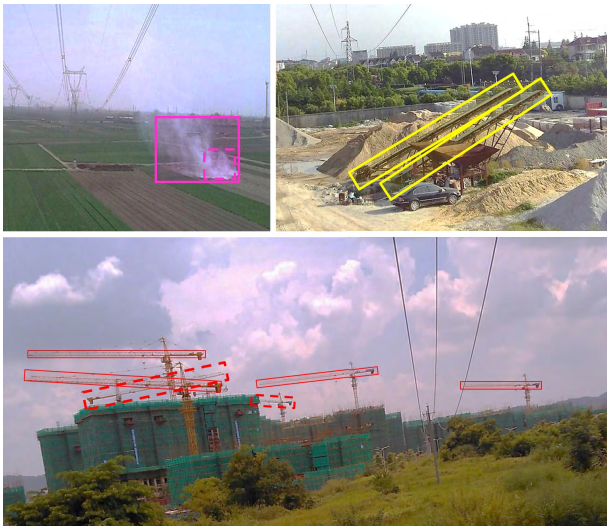


FIGURE 6. **Typical failed detections. Top left: the detection of fire smoke (pink dotted box) only covers part of the ground truth region (pink solid box); Top right: two inclined conveyor belts are mistakenly detected as on-board cranes; Bottom: two tower cranes (red dotted boxes) are missed in the detections, while red solid boxes represent the true positive detections.**

the transmission line environments precisely and is robust to the variations in object scale, oriented angle, and cluttered background. It should be noted that, by comparing APs for each category, the gain of detection accuracy mainly comes from the accuracy improvement of the on-board cranes and tower cranes categories, which are most likely to occur with tilt and dense phenomena. Experiment results show that objects from these two categories that are more likely to tilt are more appropriate for prediction using oriented bounding boxes, whose accuracy of these categories is even more than 90%.

F. ANALYSIS OF FAILED CASES

To figure out the gap between the detection results of our model and the ground truth annotations, we show some typical failed detection cases in Figure 6. There are still a small amount of failed cases: missed detections, false detections, or detection regions which are not very reliable. We argue that there is a large subjective willingness in the labeling procedure for foreign object regions such as fire and smoke. Besides, the cases of missed detection or false detection can be greatly optimized by increasing the training samples or adopting some few-shot learning tricks.

V. CONCLUSION

Foreign object detection is a fundamental step towards automatic inspection and maintenance of power transmission lines. Inspired by the recent success of deep learning techniques, we propose *DFB-NN* to detect invading foreign objects for inspecting power transmission lines. The *DFB-NN*, which is based on oriented bounding box regression strategy, can predict the oriented angle of tilted objects and can more accurately predict the spatial location in close proximity. Besides, we also construct a monitoring image dataset which contains five types of invading foreign objects. Considering the difference of scale distribution between additional pretrained datasets and task-specified datasets like HBB-FO and RBB-FO, we propose an easy but effective approach, Scale Histogram Matching, to boost the performance gained from pretraining. Experiments show that our proposed model has a competitive performance, especially for the oriented object detection.

However, the types of foreign objects specified in our collected dataset are still limited. In reality, there are other types of foreign objects that may cause dangers to transmission lines. Therefore, how to detect as many types of foreign objects as possible through incremental learning is one of the subsequent researches we will carry out in the future. In addition, our model still has a small number of detection errors. Continuously optimizing the accuracy of the detection model through online learning is also one of the significant directions for future work.

REFERENCES

[1] S. Peungsungwal, B. Pungsiri, K. Chamnongthai, and M. Okuda, "Autonomous robot for a power transmission line inspection," in *Proc. ISCAS . IEEE Int. Symp. Circuits Syst.*, vol. 3, May 2001, pp. 121–124.

[2] X. Qin, G. Wu, J. Lei, F. Fan, X. Ye, and Q. Mei, "A novel method of autonomous inspection for transmission line based on cable inspection robot LiDAR data," *Sensors*, vol. 18, no. 2, p. 596, Feb. 2018.

[3] 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.

[4] C. Sampedro, C. Martinez, A. Chauhan, and P. Campoy, "A supervised approach to electric tower detection and classification for power line inspection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 1970–1977.

[5] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 107–120, Jul. 2018.

[6] X. Hui, J. Bian, X. Zhao, and M. Tan, "Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection," *Int. J. Adv. Robotic Syst.*, vol. 15, no. 1, 2018, Art. no. 1729881417752821.

[7] X. Gong, Q. Yao, M. Wang, and Y. Lin, "A deep learning approach for oriented electrical equipment detection in thermal images," *IEEE Access*, vol. 6, pp. 41590–41597, 2018.

[8] Z. Li, Y. Liu, R. Hayward, J. Zhang, and J. Cai, "Knowledge-based power line detection for UAV surveillance and inspection systems," in *Proc. 23rd Int. Conf. Image Vis. Comput. New Zealand*, Nov. 2008, pp. 1–6.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

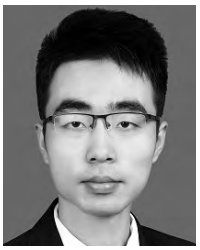
[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

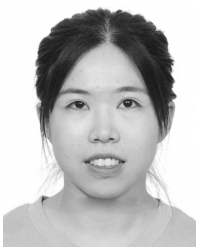
[13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 740–755.

[14] J. Zhu, Z. Yuan, and T. Liu, "Welding joints inspection via residual attention network," in *Proc. 16th Int. Conf. Mach. Vis. Appl. (MVA)*, May 2019, pp. 1–5.

[15] M. Liao, Z. Zhu, B. Shi, G.-S. Xia, and X. Bai, "Rotation-sensitive regression for oriented scene text detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5909–5918.
[16] J. Zhu, Z. Yuan, C. Zhang, W. Chi, Y. Ling, and S. Zhang, "Crowded human detection via an anchor-pair network," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1391–1399.
[17] W. Zhang, X. Liu, J. Yuan, L. Xu, H. Sun, J. Zhou, and X. Liu, "RCNN-based foreign object detection for securing power transmission lines (RCNN4SPTL)," *Procedia Comput. Sci.*, vol. 147, pp. 331–337, Jan. 2019.
[18] S. Zhu, C. Chen, and W. Sultani, "Video anomaly detection for smart surveillance," 2020, *arXiv:2004.00222*. [Online]. Available: http://arxiv.org/abs/2004.00222
[19] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 403–412.
[20] A. Noguchi and T. Harada, "Image generation from small datasets via batch statistics adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2750–2758.
[21] S. Li, Z. Zhang, B. Li, and C. Li, "Multiscale rotated bounding box-based deep learning method for detecting ship targets in remote sensing images," *Sensors*, vol. 18, no. 8, p. 2702, Aug. 2018.
[22] X. Yang, H. Sun, X. Sun, M. Yan, Z. Guo, and K. Fu, "Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network," *IEEE Access*, vol. 6, pp. 50839–50849, 2018.
[23] F. Zhang, Y. Fan, T. Cai, W. Liu, Z. Hu, N. Wang, and M. Wu, "OTL-classifier: Towards imaging processing for future unmanned overhead transmission line maintenance," *Electronics*, vol. 8, no. 11, p. 1270, Nov. 2019.
[24] A. Prasad, J. Belwin Edward, and K. Ravi, "A review on fault classification methodologies in power transmission systems: Part—I," *J. Electr. Syst. Inf. Technol.*, vol. 5, no. 1, pp. 48–60, May 2018.
[25] X. Liu, X. Miao, H. Jiang, and J. Chen, "Review of data analysis in vision inspection of power lines with an in-depth discussion of deep learning technology," 2020, *arXiv:2003.09802*. [Online]. Available: http://arxiv.org/abs/2003.09802
[26] O. Menéndez, M. Pérez, and F. Auat Cheein, "Visual-based positioning of aerial maintenance platforms on overhead transmission lines," *Appl. Sci.*, vol. 9, no. 1, p. 165, Jan. 2019.
[27] O. A. Menendez, M. Perez, and F. A. A. Cheein, "Vision based inspection of transmission lines using unmanned aerial vehicles," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Sep. 2016, pp. 412–417.
[28] G. Zhou, J. Yuan, I.-L. Yen, and F. Bastani, "Robust real-time UAV based power line detection and tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 744–748.
[29] J. Wang, J. Wang, J. Shao, and J. Li, "Image recognition of icing thickness on power transmission lines based on a least squares Hough transform," *Energies*, vol. 10, no. 4, p. 415, Mar. 2017.
[30] H. Zou and F. Huang, "A novel intelligent fault diagnosis method for electrical equipment using infrared thermography," *Infr. Phys. Technol.*, vol. 73, pp. 29–35, Nov. 2015.
[31] Z. Zhao, X. Fan, G. Xu, L. Zhang, Y. Qi, and K. Zhang, "Aggregating deep convolutional feature maps for insulator detection in infrared images," *IEEE Access*, vol. 5, pp. 21831–21839, 2017.
[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
[34] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2241–2248.
[35] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020.
[36] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," 2019, *arXiv:1905.05055*. [Online]. Available: http://arxiv.org/abs/1905.05055
[37] D. T. Nguyen, W. Li, and P. O. Ogunbona, "Human detection from images and videos: A survey," *Pattern Recognit.*, vol. 51, pp. 148–175, Mar. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320315003179
[38] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[39] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[41] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
[42] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[43] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767
[44] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
[45] L. Liu, Z. Pan, and B. Lei, "Learning a rotation invariant detector with rotatable bounding box," 2017, *arXiv:1711.09405*. [Online]. Available: http://arxiv.org/abs/1711.09405
[46] X. Yu, Y. Gong, N. Jiang, Q. Ye, and Z. Han, "Scale match for tiny person detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1257–1265.



JINGUO ZHU received the bachelor's degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2019, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include but are not limited to deep learning and computer vision.



YUE GUO received the B.S. degree from Northwest A&F University, in 2018. She is currently pursuing the master's degree with the Department of Electrical Engineering, Xi'an Jiaotong University. Her research fields are artificial intelligence technology and its application for condition monitoring and fault diagnosis of electrical apparatus.

FANDING YUE received the B.S. degree from Shandong University, in 2019. She is currently pursuing the master's degree with the Department of Electrical Engineering, Xi'an Jiaotong University. Her research fields have been involved in condition monitoring technique and fault diagnosis for electrical apparatus.

HUAN YUAN received the B.S. degree from Southwest Jiaotong University, in 2014, and the Ph.D. degree from the Department of Electrical Engineering, Xi'an Jiaotong University, in 2019. He is currently an Assistant Professor with Xi'an Jiaotong University. His current research interests include intelligent perception and system in complex electromagnetic environment, wireless power transfer, and artificial intelligence.

XIAOHUA WANG (Senior Member, IEEE) received the B.S. degree from Chang'an University, in 2000, and the Ph.D. degree from the Department of Electrical Engineering, Xi'an Jiaotong University, in 2006. He is currently a Professor with Xi'an Jiaotong University. His research fields have been involved in condition monitoring technique and fault diagnosis for electrical apparatus.

AIJUN YANG (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Xi'an Jiaotong University, in 2009 and 2014, respectively. He is currently an Associate Professor with Xi'an Jiaotong University. His field of interest is gas sensing based on nanomaterials.

MINGZHE RONG (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively. He is currently a Professor with Xi'an Jiaotong University. He focuses on the detection and diagnosis techniques for electrical equipment and on-line monitoring technique. He is an IET Fellow.