Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

A Lightweight Rust Detection Method of Power Equipment Based on DA-MobileNet

Qingzhu Shao[1], Min Xie[1], Wei Wang[1], Jun Zhang[1], Bo Gao[2*]
[1]State Grid Anhui Electric Power Co., Ltd., AnHui, China
[2]State Grid AnHui Electric Power Research Institute, AnHui, China
*Corresponding author

Abstract:

As a very important part of power system fault detection, corrosion detection of power equipment needs to be quickly and accurately identified. In view of the low efficiency of manual inspection of power equipment, a lightweight corrosion detection method based on DA-MobileNet is proposed in this paper. Firstly, the algorithm model is greatly compressed by MobileNet optimization method based on dense connection. Secondly, an up-sampling feature fusion strategy based on the dual attention model is proposed to compensate the loss of precision caused by the reduced model structure. Finally, based on the standard SSD and the deep separable convolution of MobileNet, an improved lightweight neural network model is constructed by combining dual attention mechanism. The experimental results show that compared with the standard SSD, the proposed algorithm can reduce the number of parameters by 63.6%, and improve the accuracy by 10.47% and average accuracy by 5.99% besides a speed increase of 46.7%, which can meet the requirements of rapid and accurate identification of rust detection of power equipment.

Keywords*: Target detection, Multi-scale fusion, Lightweight neural network, Attention mechanism, Convolutional neural network.*

I. INTRODUCTION

In China, increasing scale of power system, more complex structure and higher electric pressure bring a challenge to safe operation of the power system. As a very important part of power system operation, maintenance of power equipment plays a key role in safe operation of the entire power grid [1]. For some power equipment such as overhead transmission lines and box-type transformers, which are generally situated outdoors, metal parts exposed to wind and sun are very likely to be rusted. Besides, the terminal blocks placed inside the terminal box are also easy to be rusted in a damp, dusty, and confined space. Thus, safe operation of power equipment and even the entire power system is threatened. [2] At present, operation and maintenance of power equipment is mainly completed by inspectors. As the working intensity is high and the inspection results are susceptible to experience and sense of responsi-

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

bility of inspectors, an effective power equipment corrosion detection method is in need to support the safe and stable operation of the power grid.

With extensive use of UAVs and surveillance cameras, application of image recognition technology to replace manual inspection has become an effective solution for equipment inspection [3]. According to color characteristics of corrosion faults, Literature [4, 5] set up HSI space and RGB model to identify, segment and detect corrosion areas. With advancement of deep learning, target detection methods based on convolutional neural networks have surpassed traditional digital image processing methods from many aspects. In the aspect of rust fault detection, there are many researches on recognition methods based on deep learning. Literature [6] combined the HIS model and deep learning, and proposed a new idea for corrosion fault detection in power transmission lines. Nash W et al. [7] further segmented and extracted the corrosion scenes, although the results did not meet expectations. Zhou Ziqiang et al. [8] introduced transfer learning to solve the problem of small data samples, which also improved detection accuracy to some extent. However, because the current target detection algorithms rely on large convolutional neural network, many problems have arisen such as too large parameters and too slow detection speed, resulting in a failure to meet the real-time response requirements of power equipment fault detection.

Therefore, to solve the problem of low efficient manual inspection of power equipment, a light weight corrosion detection method based on DA-MobileNet is proposed in this paper. Firstly, the algorithm model in this paper aims at the current large-scale neural network must have strong computing and storage capabilities. It uses the MobileNet optimization method based on dense connections to greatly compress the model. Secondly, an up-sampling feature fusion strategy based on the dual attention model is proposed to compensate the loss of precision caused by the reduced model structure. Finally, based on the standard SSD [9] and the deep separable convolution of MobileNet, an improved lightweight neural network model is constructed by combining dual attention mechanism. Based on power equipment corrosion dataset, experimental verification is carried out against the algorithm proposed. The experimental results show that the proposed algorithm can still exceed the model with large parameters under the condition of low parameter amount and low inference time, and accurately identify the rust area.

II. RELATED WORK

In the safety monitoring of the power system, it’s necessary not only to identify corrosion image, but to position corrosion area, but also the rust area needs to be located. Currently, the mainstream target recognition algorithms that is applicable to the monitoring system in realism are generally one-stage, among which SSD and YOLOv3 are the most widely applied to directly identify and position corrosion targets.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

2.1 SSD Target Detection Algorithm

The target detection method based on deep learning has become the current mainstream algorithm since its first debut raised by R-CNN [10], and has had many variations such as two-stage algorithms including Fast R-CNN and Faster R-CNN [11] and one-stage algorithms including SSD and YOLO [12]. Among them, the one-stage algorithms are more suitable for industrial and robotic applications owing to its faster detection speed, because classification and regression problems are integrated so that final detection results can be obtained only after a single detection. Fig.1 gives the algorithm structure diagram of a standard SSD.

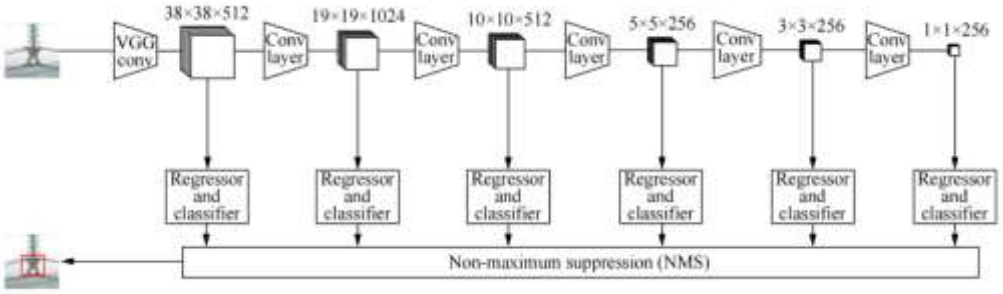

Fig 1: ssd algorithm structure

However, current target recognition algorithms are basically based on large-scale neural networks. For example, SSD uses VGG-16 as a feature extraction network, while YOLOv3's feature extraction network is DarkNet-53. Due to a large amount of parameters of these backbone networks, the detection speed is relatively slow, resulting in the impossibility to meet the requirements of real-time monitoring of the power system. Therefore, it is of great significance to study the application of target detection algorithms based on lightweight neural networks to detection of corrosion areas of power equipment.

2.2 Lightweight Network MobileNet

Based on stack design of depth-wise separable convolution, the network structure of MobileNet [13] has the basic idea that inter-channel correlation and spatial correlation are completely separated to greatly reduce the amount of calculation and parameters. Different from traditional convolutional network structures, this network structure performs convolution of each channel of the feature map (for example, 3×3×1), and merges various feature map channels processed by 1×1 convolution operation to reduce the number of channels. Since a large number of 3×3 convolution kernels are used in the MobileNet network structure, the amount of calculation is greatly reduced, with small impact on the accu-

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

racy of the model. This not only guarantees the accuracy of the model, but speeds up the calculation to meet the real-time requirements of the model. The network structure parameters of MobileNet are listed in Table I.

TABLE I. Parameter model of the basic network mobilenet

Convolutional layer name/step	size	Depth	Convolutional layer name/step	size	Depth
Standard convolution layer1/s2	3×3	32	Separable convolution layer7/s1	3×3	512
Separable convolution layer1/s1	3×3	64	Separable convolution layer8/s1	3×3	512
Separable convolution layer2/s2	3×3	128	Separable convolution layer9/s1	3×3	512
Separable convolution layer2/s1	3×3	128	Separable convolution layer10/s1	3×3	512
Separable convolution layer4/s2	3×3	256	Separable convolution layer11/s1	3×3	512
Separable convolution layer5/s1	3×3	256	Separable convolution layer12/s1	3×3	1024
Separable convolution layer6/s2	3×3	512	Separable convolution layer13/s1	3×3	1024

2.3 Attention Mechanism

The attention mechanism [14] is in essence derived from the visual mechanism in which humans only pay attention to certain objects according to their needs. In 2018, Hu J et al. proposed the SENet (Squeeze-and-Excitation Networks) [15], in which an attention mechanism among channels was introduced to calibrate the importance of different channels, and different weights were then put to improve efficient features and suppress ineffective characteristics. The algorithm structure of SE networks is shown as in Fig. 2.

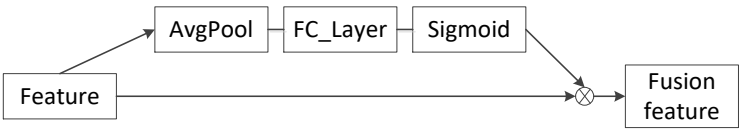

Fig 2: algorithm structure of se networks

In the SE networks, analysis and comparison depends on inter-channel interdependence and the spatial feature is the most important information among the image information. In order to reduce the amount of parameters and improve the efficiency of detection, 3×3 and 1×1 small convolution kernels are generally utilized to extract features. However, smaller convolution kernel means smaller receptive field, leading to the loss of global feature association. The network structure DANet proposed by Fu for semantic segmentation [16] adopts a position attention module (PAM) structure to establish an association model for pixels in the space. Then, the features are recalibrated in a weighted manner to complete the attention calibration in the spatial domain. The network structure of the PAM structure is shown in Fig. 3. In this paper, the weak attention mechanism based on the spatial domain and the channel domain

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

is cascaded and combined into a dual attention module to improve and strengthen the SSD-MobileNet network.

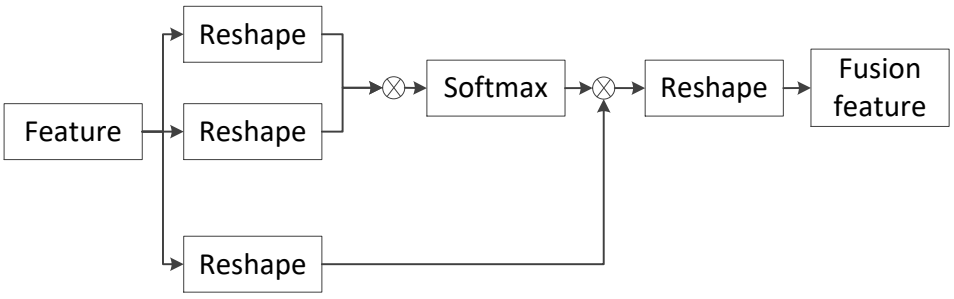

Fig 3: algorithm structure of pam module

III. LIGHTWEIGHT CORROSION DETECTION ALGORITHM BASED ON DA-MOBILENET

When image recognition is used to detect the corrosion areas of power equipment, it is necessary to consider the impact of different factors. Because the existing target detection methods are basically subject to large-scale neural networks which have huge parameters, they cannot recognize small objects accurately and thus it's impossible to meet the real-time monitoring requirements of power systems. The DA-MobileNet-based lightweight corrosion detection algorithm for power equipment proposed in this chapter can minimize the impact of external factors on detection of corrosion areas via self-learning, and can improve the detection effect by means of the following methods:

1. To train and learn from a large amount of data, without relying on prior knowledge and expert templates; the detection result surpasses the features manually designed;
2. To adopt the densely connected MobileNet model optimization method, according to which the number of parameters and the calculation amount are reduced by setting a smaller growth rate;
3. To design a dual attention model algorithm for small target detection. By adding samples, the algorithm has the ability to self-learn and update, further improving the accuracy.

Therefore, in this paper, a corrosion area detection algorithm for power equipment based on the lightweight neural network structure is proposed. The algorithm is verified to speed up detection, reduce amount of parameters, and be applicable to mobile devices.

3.1 Feature Extraction Optimization Method Based on Densely Connected MobileNet

Compared with the VGG-16 network, MobileNet, as a kind of lightweight network, uses a deeply separable convolution method to deepen the network and reduce the number of parameters and the calculated amount. Moreover, the classification accuracy is only 1% less in case of the ImageNet data set.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

However, in order to further decrease the number of parameters and calculations of the model, a dense structure is adopted, in which each layer will be densely connected with the output feature maps of all previous layers. Such connection mode can make full use of output feature maps of all previous layers, realizing recycling of features. Moreover, the number of output feature maps can be controlled by defining the hyperparameter growth rate. Besides, the dense structure can also solve the vanishing gradient problem to a certain extent. In this paper, the network layer in MobileNet is packed with dense blocks as the basic units, and the amount of parameters and calculations is decreased by defining a smaller growth rate. The concept of dense block defined in DenseNet is introduced to into the MobileNet structure model. The convolutional layer with the same input feature map size in the MobileNet model is regarded as a dense block, and connected with the previous layers so that the output feature maps of all the previous layers can be used as input. Such connection can alleviate the vanishing gradient problem to a certain extent. Meanwhile, since each layer is connected to all the previous layers, recycling of previous feature maps helps produce more feature maps with fewer convolution kernels. As shown in Fig. 4 that gives a four-layer densely connected block structure with a growth rate of 4, each layer uses the feature map outputted by the previous layer as its input. Because the deep convolution algorithm performs single-channel convolution of the feature map, the number of feature maps outputted by the deep convolutional layer in the middle is identical to the number of input feature maps of this layer, namely the sum of the feature maps outputted by all the previous layers.

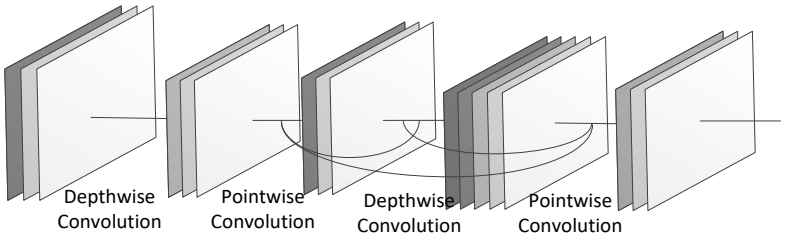

Fig 4: schematic diagram of mobilenet model based on dense connections

In the MobileNet model based on dense connections, the deep convolutional layer and the point convolutional layer in the deep separable convolutional layer of each layer realize convolution of the images by formula (1) (2) as shown below:

$$O_{dc}(y,x,j)=\sum_{u,v=1}^{s}K(u,v,j)\square I(y+u-1,x+v-1,j) \quad (1)$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

$$O_{pc}(y,x,j)=\sum_{i=1}^{m}K(i,j)\square I(y,x,j) \tag{2}$$

Where, $O_{dc}(y,x,j)$ represents the value of point (y,x) in the j th feature map; $K(u,v,j)$ represents the value of point (u,v) in the j th convolution kernel. $u=1,2\cdots s$, $v=1,2\cdots s$, where s refers to the size of the convolution kernel; $I(y+u-1,x+v-1,j)$ represents the value of point $(y+u-1,x+v-1)$ on the j th input channel; $O_{pc}(y,x,j)$ represents the value of the point (y,x) in the j th feature map; $K(i,j)$ represents the value of channel in the th convolution kernel; $I(y,x,j)$ represents the value of the point (y,x) on the i th input channel; $i=1,2\cdots m$; m represents the number of deep convolutional kernels in the previous layer.

The concept of dense block in DenseNet is introduced to the MobileNet model structure based on dense connections, so that the amount of parameters and calculation is less than that of the ordinary MobileNet models by defining a smaller hyperparameter growth rate. In the ordinary MobileNet models, dimensionality of the feature map should be reduced by a deep convolution with a step size of 2 for every two deep separable convolution layers. Since the input feature maps should have the same size in the same dense block, there are only two deep separable convolutional layers in one dense block. The dense connection-based MobileNet model proposed in this paper regards the deep separable convolutions as two separate layers, and four convolutional layers with the same input feature map size as a dense block, greatly reducing the amount of parameters and calculations of the MobileNet network.

3.2 Upsampling Strategy Based on Dual Attention Model Mechanism

In this chapter, the spatial attention module and the channel attention module are combined to build a cascaded dual attention model. Then, an attention feature for each position in each channel is constructed after simultaneous calibration of the spatial domain and the channel domain. A cascaded model of spatial attention and channel attention stiffens the detection effect, and the feature fusion is adopted to finally propose an upsampling strategy based on the dual attention model mechanism.

The cascaded dual attention model is composed of a spatial attention module and a channel attention module, between which the former first flattens in channels the original feature map in C×H×W size to C×N and then transposes it into N×C. After multiplication of these two feature matrices, a N×N feature calibration matrix is obtained, in which each position represents the relationship among pixels in the original feature. At this point, the feature calibration matrix is normalized and calibrated using the two-dimensional softmax function to get a feature mask matrix (FMM), in which the value of each position is equal to the proportion of amount of information provided by each pixel in the original feature

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

maps. After matrix multiplication of the proportion with the flattened original feature map C×N, features of the original feature maps can be re-calibrated. At last, the original feature map is added back to the same residual structure. See the followings for its expression:

$$E_c = \alpha \sum_{i=1}^{N} (s_{ij} D_i) + A_j \tag{3}$$

Where, E is the calibrated feature map, D is the feature map before transpose, A is the original feature map added back to the residual structure, and s_{ij} the weight value of the (i, j) th position, which is obtained using the softmax function:

$$s_{ij} = \frac{\exp(B_i C_j)}{\sum_{i=1}^{N} \exp(B_i C_j)} \tag{4}$$

Where, B is the flattened N×C feature map, and C is the flattened C×N feature map.

The Squeeze operation of channel attention is then performed. The feature maps are processed by Average Pooling and compressed into a 1×1 feature vector which is a complex of all the information in the feature map and can be used as the main basis for judging the importance of features. The specific step of Squeeze operation is given as follows, where E is the feature map, $E(i, j)$ the pixel in the feature map, and H and W the sizes of the feature map:

$$z_c = F_{sq}(E_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} E_c(i, j) \tag{5}$$

The Excitation operation is subsequently performed. A fully connected layer is set up and densely connected with the above-mentioned feature vector in order to form a small learnable and trainable network that can be used as a basis to judge the importance of feature vectors and provide a back propagation path. Then, the sigmoid function normalizes all channel information to 0 - 1, and explicitly reflects the amount of information acquired by each channel to form a mask vector. The specific step of Excitation operation is given as follows, where W is the adjustable parameter and δ is the activation function:

$$S = F_{ex}(z, W) = \sigma(w_2 \delta(W_1 z)) \tag{6}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{7}$$

In the final Reweight operation, the mask of each channel is used as the weight to multiply each pixel of the feature map, so that the proportion of channel information is weighted to each feature map to realize the feature recalibration at the channel level:

$$x_c = F_{re}(E_c, S_c) = S_c \square E_c \tag{8}$$

The final feature map is the map that has been calibrated by space attention and channel attention. The complete structure of the cascaded dual attention model is shown in Fig. 5.

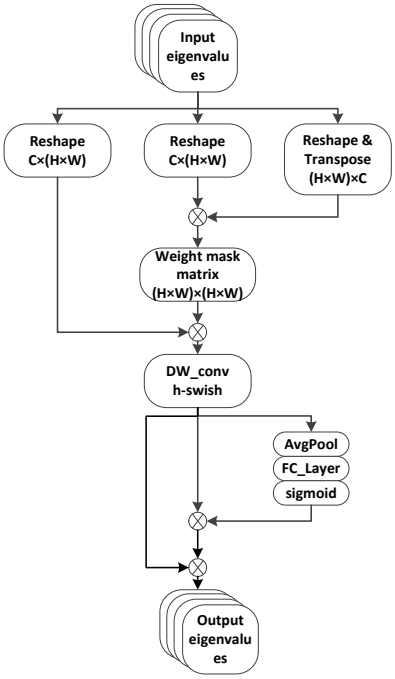

Fig 5: cascade dual attention model

In order to improve the detection effect of small targets, the standard SSD target detection algorithm utilizes multi-scale features for simultaneous detection. However, due to lack of feature fusion, small targets cannot be well detected. Furthermore, although the use of deep separable convolution can greatly reduce the number of parameters, the detection accuracy is reduced to a certain extent due to loss of many adjustable parameters. Therefore, this paper adopts the concept of multi-scale fusion, and carries out up-sampling fusion of the multi-scale features detected separately by the standard SSD. As for fusion

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

strategy, although the elementwise add method used by FPN [17] has higher requirements on feature similarity of single-channel feature maps, content matching of the up-sampled high-level feature maps may not be realized. The concat method pays more attention to the feature information in different channels, and utilizes an attention model to perform correlation calibration of the fused features so that more valuable feature information can be selected. However, as the concat method will double the number of channels, a convolutional layer is added to cut the number of channels in half when fusing the features. The structure of the up-sampling strategy based on the attention mechanism is shown in Fig. 6.

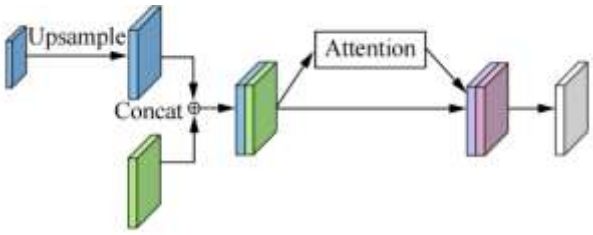

Fig 6: upsampling strategy based on attention mechanism

3.3 Lightweight corrosion target detection algorithm based on DA-MobileNet

In this paper, lightweight operation is performed by combining MobileNet's deep separable convolution and the standard SSD. First, the feature extraction network VGG-16 of SSD is replaced by a MobileNet feature extraction network based on dense connection. As for application of multi-scale features, the up-sampling strategy based on the dual attention model proposed in this paper adopts the up-sampling FPN structure, and combines the concat method and the attention mechanism for feature fusion, which compensates loss of accuracy caused by lightweight feature extraction. Finally, a DA-MobileNet-based lightweight power equipment corrosion target detection algorithm is finally raised. The complete algorithm structure is shown in Fig. 7.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

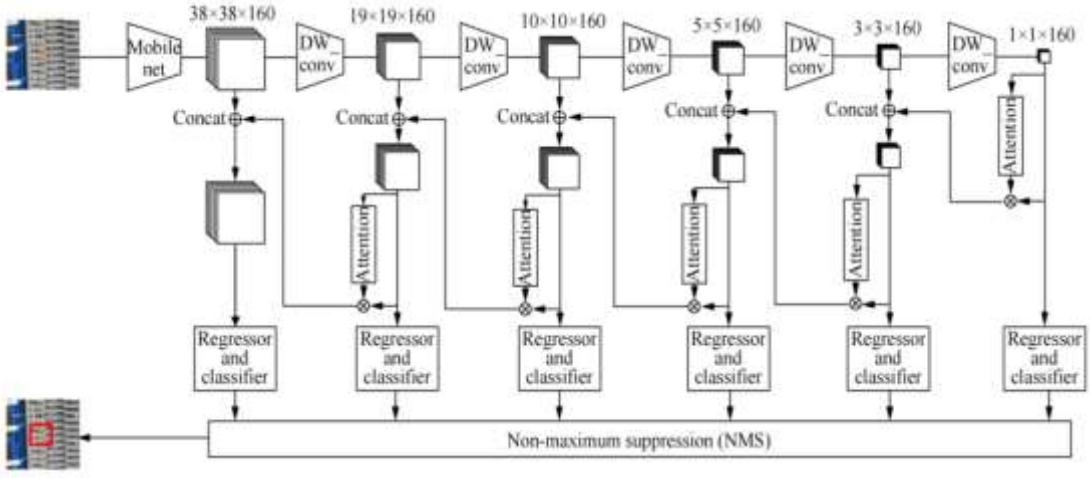

Fig 7: algorithm structure of lightweight power equipment corrosion target detection based on da-mobilenet

On the basis of VGG-16, SSD replaces the last three fully connected layers with a convolutional layer for feature extraction, and uses the densely connected MobileNet for further feature extraction. In order to ensure the effectiveness of feature extraction, a block structure is used for convolution operation. In other words, multiple deep separable convolution layers are used as a whole to optimize the extracted features using the multi-layer structure.

In this section, a total of 6 feature maps with the size of respectively 38×38, 19×19, 10×10, 5×5, 3×3 and 1×1 are used for border prediction and target classification, among which the shallow feature maps with larger size can be used to detect small objects and the deep feature maps with smaller size can be used to detect conspicuous objects. 38×38, 19×19, 10×10 and 5×5 feature maps each uses six preselected boxes with different sizes and aspect ratios, while 3×3 and 1×1 feature maps each uses four preselected boxes with different sizes and aspect ratios. Thus, there are 11,620 preselected boxes in total. Afterwards, hard sample mining is performed to remove the pre-selected boxes that can be easily distinguished, and select positive and negative samples (the ratio is 1:3) as training samples. In this way, the phenomenon that the entire samples tend to be negative because negative samples are far more than positive samples can be eliminated, which ensures the training to run on the rails with high recall rate and high accuracy. In the algorithm mentioned in this section, two sub-networks are used to perform target classification and border regression against the extracted features. The final output of the classification sub-network is the probability value of each category, while the regression sub-network finally outputs the coordinate value of each prediction. The cost function continues to use Loss of the SSD, the confidence coefficient of each category and the offset relative to the default box coordinates. Assume that $x_{ij}^{p} = \{1, 0\}$ means whether the i th default box matches the j th real box containing category p. If

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

$x_{ij}^{p}=1$, the ith default box matches the jth real box containing category p; otherwise, it is 0. The basis for matching is whether the IoU (Intersection-over-union, IoU) of the default box and the real box exceeds the threshold (0.5). Thus, the overall objective loss function L is calculated as follows:

$$L(x,c,l,g)=\frac{1}{N}(L_{conf}(x,c)+\beta L_{loc}(x,l,g)) \tag{9}$$

Where: N is the number of default boxes matched; L_{conf} the confidence loss; L_{loc} the location loss; β the weight coefficient of the location loss; x the input data; c the confidence; l and g represent the prediction box and the real box, respectively; Among them, the Softmax loss between the confidence of the background class and the confidence of the target class is used as L_{conf}, and the calculation method is as follows:

$$L_{conf}(x,c)=-\sum_{i\in Pos}^{N}x_{ij}^{p}\ln\hat{c}_{i}^{p}-\sum_{i\in Neg}\ln\hat{c}_{i}^{0} \tag{10}$$

$$\hat{c}_{i}^{p}=\frac{\exp c_{i}^{p}}{\sum_{p}\exp c_{i}^{p}} \tag{11}$$

Where: c_{i}^{0} represents the confidence of the background class, which corresponds to the neative (Neg) default box that does not contain the target object; c_{i}^{p} represents the confidence of the target class, which corresponds to the positive (Pos) default box that contains the target objects under category P. L_{loc} refers to the SmoothL1 loss (L_{s}) between the prediction box and the real box, and the calculation formula is as follows:

$$L_{loc}(x,l,g)=\sum_{i\in Pos}^{N}\sum_{m\in\{cx,cy,w,h\}}x_{ij}^{p}L_{s}(l_{i}^{m}-\hat{g}_{j}^{m}) \tag{12}$$

$$\hat{g}_{j}^{cx}=\frac{g_{j}^{cx}-d_{i}^{cx}}{d_{i}^{w}},\hat{g}_{j}^{cy}=\frac{g_{j}^{cy}-d_{i}^{cy}}{d_{i}^{h}} \tag{13}$$

$$\hat{g}_{j}^{w}=\ln\frac{g_{j}^{w}}{d_{j}^{w}},\hat{g}_{j}^{h}=\ln\frac{g_{j}^{h}}{d_{j}^{h}} \tag{14}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

Where: $\hat{g}$ represents the relative offset between the real box g and the default box d; the parameters m of the box are composed of 4 parameters describing the coordinates, where (cx, cy) is the center of the box, (w, h) is the width and height of the box. The specific calculation method of L_s is as follows:

$$L_s(X) = \begin{cases} 0.5X^2 & if\ |X| < 1 \\ |X| - 0.5 & otherwise \end{cases} \tag{15}$$

Where, $X = l_i^m - \hat{g}_j^m$ represents the coordinate offset between the predicted box and the real box.

IV. EXPERIMENT AND ANALYSIS

4.1 Experimental Process

Since use of target detection technology for fault detection of power equipment is still in the development stage and there has currently been no public data set for power equipment corrosion target detection, a RustDetection data set is built based on the existing power equipment corrosion images, which are composed of various corrosion failure pictures collected on site or in the network. After data argumentation is applied, 600 corrosion pictures are finally used as the training set, and 200 as the test set. Furthermore, the corrosion areas are labelled by LabelImg, and processed in the format of the VOC2012 data set. Fig. 8 renders the labeling effect.

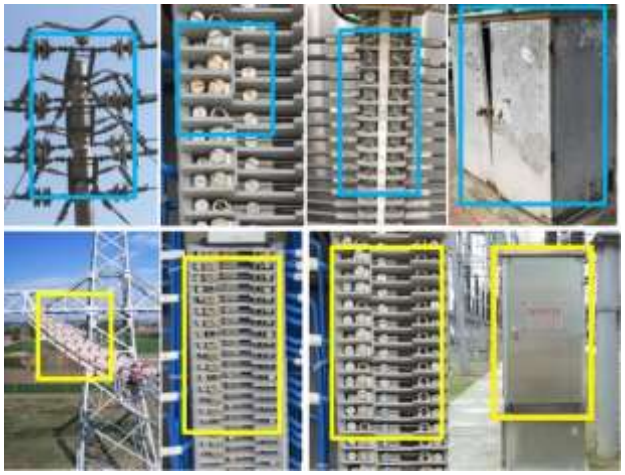

Fig 8: rust detection data set where the corrosion area is labelled in blue, and no-corrosion area is labelled in yellow

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

Because the sample size of the corrosion detection data set proposed in this paper is small and if the data set is directly trained, the network cannot converge quickly with poor detection results. Therefore, the VOC2012 common public data set containing a total of 17125 pictures (21 categories) is pre-trained, and the transfer learning method is then used to fine-tune the proposed RustDetection data set. Fig. 9 is a comparison chart for the loss caused after 100 rounds of training by random initialization and pre-training models. We can see from the picture that the use of transfer learning can effectively improve the detection accuracy and speed up the optimization efficiency of the model.

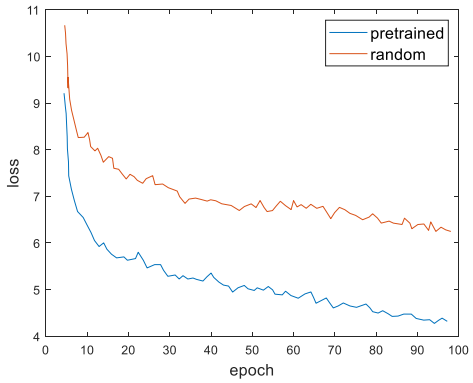

Fig 9: loss comparison between random initialization and pre-training models

At the training stage, the input images are scaled to 300×300×3 RGB images, before being normalized and trained. The training phase is completed on the NVIDIA GTX 1080Ti GPU. Transfer learning is applied. First, the established network model is trained against the VOC2012 data set for 300 rounds. After good detection effect is obtained, part of the multi-class sub-network is removed, and replaced by a two-class sub-network which is composed of 6 convolutional layers respectively designed for target prediction at 6 scales. The Kaiming initialization method is adopted to initialize the parameters. Compared with random initialization, this initialization method can effectively prevent the output value of the activation function from 0 to ensure that the training can proceed smoothly. Subsequently, the complete algorithm model is trained against the proposed RustDetection data set using a phased learning rate method. In other words, a large learning rate is applied at the initial stage of training and a small learning rate at the later stages so as to accelerate model convergence and speed up training. At the test and verification stages, the target images are input to the trained algorithm model, which finally outputs 11,620 candidate boxes. Each candidate box includes 2 classification values (corrosion, background) and 4 coordinate values (coordinates, length and width of the candidate box's central point). Then, all candidate boxes that are identified as backgrounds are filtered out, and the rest boxes are processed by non-maximum suppression. Finally the candidate frame with the largest IOU is selected as the target frame to complete

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

the corrosion detection.

4.2 Analysis of Experimental Results

The specific detection results are shown in Fig. 10, in which the corrosion area is marked in colors, and the upper left corner lies category tags of the area.

Fig 10: corrosion target detection results

In order to further verify the advantages of the proposed algorithm in model size, detection speed and accuracy, this paper compares the standard SSD model with VGG-16 and ResNet-50 as its backbone network and the proposed lightweight SSD model based on the attention upsampling strategy. The evaluation indexes include precision, recall and AP (Average Precision) value. See formula (16) and formula (17) for calculation of accuracy and recall:

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

Where TP represents the number of positive samples judged correctly, FP represents the number of positive samples judged incorrectly, and FN represents the number of negative samples judged incorrectly. To get AP value, the 11-Point method of VOC2007 is adopted, in which Recall is set as [0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0]. Find out the maximum precision value x, and then calculate the average precision according to formula (18)

$$AP_{11-point} = \frac{1}{11}\left(\sum_{x \in MaxPrecision} x \right) \tag{18}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

TABLE II give the detection results of different algorithm models under the RustDetection data set.

Table II. Comparison of detection effects of different network models (%)

AlGORITHM MODEL	Recall	Precision	AP
SSD(VGGBASE)	78.04	86.49	65.36
SSD(ResNetbase)	75.61	93.94	70.92
SSD(MobileNetbase)	63.41	83.87	59.76
PROPOSED ALGORITHM	78,05	96,96	71.35

It can be seen from Table II that if only the lightweight MobileNet structure is used to process the SSD model, the detection effect will become worse due to parameter loss. According to the method proposed in this paper, the addition of upsampling and feature fusion modules effectively improves the detection effect, which is even superior to that of the original standard SSD algorithm. Table III shows the comparison results of models with different algorithm structures in the number of parameters, weight, and detection time in Intel Core i5-7200U CPU.

TABLE III. Comparison of different model sizes

Algorithm model	the number of parameters	Weight(MB)	detection time(s)
SSD(VGGbase)	23 745 908	90.58	1.84
SSD(ResNetbase)	25433 488	97.02	1.24
SSD(MobileNetbase)	4021 220	15.34	0.50
Proposed algorithm	8638 820	32.95	0.98

To sum up, the method proposed in this paper expands the up-sampling network structure and increases the number of parameters by 53.4%, compared to the lightweight MobileNet SSD model. However, compared to the standard SSD model with VGG-16 as the backbone network and huge number of parameters, this method decreases the number of parameters by 63.6%, and increase the speed by 46.7%, the accuracy by 10.47% and the average accuracy by 5.99%. Compared with the standard SSD with ResNet-50 as the backbone network, it can also improve the accuracy by 2.98% and the average accuracy by 0.43% when reducing the number of parameters by 66%.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

V. CONCLUSIONS

This paper proposes a lightweight power equipment corrosion target detection algorithm based on DA-MobileNet. In view of the huge number of parameters of the target detection model and the high requirement on computing power, a MobileNet feature extraction optimization method based on dense connections is proposed. An upsampling strategy based on the dual attention model is built to optimize the lightweight network structure, which makes up for the loss of precision caused by parameters reduction. The algorithm proposed in this paper has the detection accuracy of 96.96% and the average precision of 71.35% while greatly reducing the number of parameters. Besides, it only takes 980ms (only 240ms in case of GPU acceleration) to complete the detection, meeting the actual needs of power equipment safety monitoring. The future works include to transplant and load the network model into the terminal equipment and achieve real-time monitoring in the industrial scene.

ACKNOWLEDGE

This work is supported by the Science and Technology Project of State Grid Anhui Electric Power Co., Ltd. "Research and Application of Key Technologies for UHV DC Protection Fault Warning and Intelligent Decision Based on Panoramic State Intelligent Perception" (No. 52120019007Z)

REFERENCES

[1] N. Hatziargyriou et al, Definition and classification of power system stability revisited & extended, IEEE Trans. Power Syst., 2020.
[2] J. Zhao, et al., Power system dynamic state estimation: motivations, definitions, methodologies, and future work, IEEE Trans. Power Syst., vol. 34, no. 4, pp. 3188-3198, July 2019.
[3] Liao K W, Lee Y T. Detection of rust defects on steel bridge coatings via digital image recognition. Automation in construction, 2016, 71 (Nov.pt.2): 294-306.
[4] Z. Tian, G. Zhang, Y. Liao, R. Li and F. Huang, Corrosion Identification of Fittings Based on Computer Vision, 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 592-597, doi: 10.1109/AIAM48774.2019.00123.
[5] J. A. I. Diaz, M. I. Ligeralde, J. A. C. Jose and A. A. Bandala, Rust detection using image processing via Matlab, TENCON 2017-2017 IEEE Region 10 Conference, pp. 1327-1331, 2017.
[6] Rahman A, Wu Z Y, Kalfarisi R. Semantic, Deep Learning Integrated with RGB Feature-Based Rule Optimization for Facility Surface Corrosion Detection and Evaluation. Journal of Computing in Civil Engineering, 2021, 35(6): 04021018.
[7] Nash W, Drummond T, Birbilis N. Quantity beats quality for semantic segmentation of corrosion in images. arXiv preprint arXiv: 1807.03138, 2018.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 212-229
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

[8] L. Liu, E. Tan, Y. Zhen, X. J. Yin and Z. Q. Cai, "AI-facilitated coating corrosion assessment system for productivity enhancement," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018, pp. 606-610, doi: 10.1109/ICIEA.2018.8397787.

[9] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. European Conference on Computer Howard A, Sandler M, Chu G, et al. Searching for MobileNetv3. arXiv preprint arXiv: 1905.02244, 2019.

[10] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.

[11] Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91-99.

[12] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern.

[13] Howard A, Sandler M, Chu G, et al. Searching for MobileNetv3. arXiv preprint arXiv: 1905.02244, 2019.

[14] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention. Proceedings of the 27th International Conference on Neural Information Processing Systems.

[15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7132–7141.

[16] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.

[17] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2117–2125.