Visual inspection for transformer insulation defects by a patrol robot fish based on deep learning

Hongxin Ji1 | Xiwang Cui2 | Weiyan Ren1 | Liqing Liu3 | Wei Wang3

1 School of Aerospace Engineering, Tsinghua University, Beijing, China
2 School of Instrument Science and Opto-electronics Engineering, Beijing Information Science and Technology University, Beijing, China
3 State Grid Tianjin Electric Power Research Institute, Tianjin, China

Abstract
In order to effectively detect the internal insulation defects of large transformers, a miniature patrol robot fish is designed and used to observe the pressboard inside the transformer, which can visually inspect the insulation condition of the pressboard. In the process of visual inspection by a patrol robot fish, insulation defects observed in the transformer (such as discharge carbon marks, pressboard cracks, etc.) are small size, indistinctive color contrast and different shapes. The key purpose of the patrol robot fish is to identify the defect targets and defect types intelligently and quickly from the images taken by the camera on the fish. Considering that there are abundant insulation defects in transformers and a lack of samples for learning, a vision detection method based on deep learning network is proposed in this study. The proposed method integrates the variable autoencoder into the traditional Faster-RCNN target detection network and constructs an improved Faster-RCNN that enhances feature extraction. This method expands the small-scale training sample set and improves the generalisation performance of the model. In order to verify the effectiveness of the proposed method, the improved network is trained and tested, and the test results show that the improved network training model can accurately identify and mark the carbon marks on the pressboard surface.

1 INTRODUCTION

Although the transformer monitoring and diagnosis technology has made remarkable progress, detection of the internal insulation of a transformer is still an important and difficult task. Due to the closeness of the metal shell of a large transformer, it is difficult to effectively detect the internal condition of the large transformer. The current traditional method requires the transformer oil to be emptied so that technicians can enter the interior of the transformer through a manhole above the transformer for manual inspection [1–5]. When the manual detection is completed, a large amount of transformer oil needs to be injected back into the transformer. In addition, when the technician enters a large transformer for manual inspection, there are many key parts of the transformer that need to be carefully observed. Therefore, the technician needs to spend a considerable amount of time to detect the transformer in the narrow space, and the task is relatively difficult.

At present, detecting the internal insulation of a transformer faces four major problems: (1) Slow inspection: Most of the traditional internal inspection processes are carried out by people by entering the transformer or by hanging over. The inspection process includes oil drainage, oil filling and oil filtration. The whole process lasts at least 15 days, resulting in a long power outage and serious economic losses. (2) Inaccuracy: Due to the complex internal structure and limited space of the transformer, manual inspection makes it difficult to accurately determine the type, location and severity of internal defects, thus this inspection method is inaccurate and blind. (3) Risks: The space in the transformer is small and narrow, so that the technician runs the risk of colliding with the components inside the transformer, resulting in personnel injury and secondary damage to the equipment. (4) Uneconomical: Before determining the type and severity of the defects in the transformer, it will incur higher costs in the transportation, pumping, lifting, maintenance and other operations. The small patrol robot has the...
advantages of having a small size and flexible movement and can replace the technician to enter the transformer from the manhole. The transformer oil does not need to be drained completely and the operators need not enter the dangerous transformer. This method improves the detection safety and allows the transformer resume its operation in time, which is of great significance for intelligent, fast and accurate maintenance of the large transformer.

Therefore, some research institutions have carried out related research of the transformer patrol robot. In 2018, ABB launched its first oil-immersed transformer internal patrol robot named Txplore [6,7]. Txplore is rectangular in shape and 18×20×24 cm in size. Four propellers are arranged on the shell of the robot. Two propellers in the horizontal direction drive the forward movement and steering of the robot, and two propellers in the vertical direction control the upward and downward movement of the robot. Multiple cameras are installed on the Txplore. Shenyang University of Technology has developed a spherical transformer robot with a diameter of 19 cm, named SSTIR, with reference to the structure of underwater robots [8,9]. The SSTIR is equipped with two fuel injection thrusters in the vertical direction. Since the SSTIR has a slight positive buoyancy, the vertical thrusters provide only downward thrust. Two horizontal thrusters provide forward and steering thrust. A camera is installed in the middle of the robot [10]. At present, when the transformer inspection robot moves in the transformer oil, its vision camera constantly takes pictures of the internal environment, which are transmitted to the computer workstation through wireless signals. Then, the technician manually observes whether there are target defects on the pictures.

Few studies have been carried out on robots working in transformer oil. Referring to underwater vehicles, our team designed a new type of transformer patrol robot with fish fins. The robot’s fish body is a spherical sealing structure. Two micro-propellers are installed on both sides of the body to control the forward movement and turn of the fish. A visual camera is equipped on the upper part of the fish body. A swim bladder device is installed below the fish body and can accurately control the amount of oil absorbed or expelled by detecting the location of the piston, so as to accurately control the floating speed and sinking speed of the robot fish. There are four fins, two fins are mounted on the upper end of the fish body and the other two fins are mounted on the lower end of the fish body. The mechanical fins can control the deflection direction and the angle of the fish. Through the combination of the fins and the swim bladder device, a gliding movement of the robot fish can be realised even without power, which effectively reduces the power consumption and increases the inspection range of the robot fish. The camera of the patrol robot fish constantly takes pictures of the internal environment, which is transmitted to the computer workstation through wireless signals. The patrol robot fish has a relatively large patrol range, and only relying on the human eye to continuously observe the photos taken, it is easy to miss some small target defects and has huge workload. In response to this problem, this paper carried out the intelligent autonomous inspection based on the vision of the patrol robot fish and hopes that the defect targets can be identified automatically by artificial intelligence to improve work efficiency.

The detection precision of insulation defects depends on the target feature extraction from images taken by the patrol robot fish. Feature extraction and classification are the focus of traditional target detection methods. Therefore, many features extraction methods have been proposed in order to improve the deformation resistance and expression ability of features, and many forms of classifiers have been proposed in order to improve the speed and precision of classifiers [11]. Lowe [12,13] proposed the Scale Invariant Feature Transform (SIFT) by using the gradient information near key points of the image to describe the moving target. Dalal proposed the Histogram of Oriented Gradient (HOG) method, which was applied to solve the pedestrian detection problem of static images [14]. Felzenszwalb et al. proposed a De-formable part model (DPM) by combing the HOG and Support Vector Machine (SVM) [15]. These methods have become the classical target detection methods. Although the traditional target detection methods have achieved good development, high classification errors are generated due to the insufficient expression ability of extracted features to the target and poor feature separability. Meanwhile, the traditional target detection methods need to design features extraction manually, which are not suitable for different data. Compared with the features designed manually, the features acquired by deep learning are much richer and have stronger feature expression ability. Common deep learning models include restricted boltzmann machine (RBM) [16], autoencoder (AE) [17] and convolutional neural networks (CNN) [18]. With the development of deep learning, researchers have found that the precision of target detection using CNN can be greatly improved. This is because CNN extracts high-level features and improves the expression ability of features. In addition, the convolutional neural network integrates feature extraction, feature selection and feature classification into one model. Through the end-to-end training, CNN conducts overall functional optimisation and enhances features separability. Compared with traditional target detection methods, the detection method based on CNN has greatly improved precision and duration [19]. For this reason, this study adopts deep learning to visually detect the defects inside the transformer.

The transformer is completely enclosed by the metal shell, so that the image acquisition by the patrol robot fish in the transformer will be affected easily, by too strong or insufficient camera light supplementation, occlusion, low contrast, blurred motion and tilted angle of view, etc. These problems have led to poor sample quality (i.e. defective image samples) and the lack of “positive samples”. However, the current deep learning training usually requires a large number of samples. In order to solve the problem of sample imbalance in the actual system, a visual detection method of the patrol robot fish based on an improved Faster-RCNN was proposed in this study. In this method, the input layer of VGG16 deep network in the traditional target detection model is replaced by a VAE network, and then a self-encoding convolutional neural network with enhanced feature extraction is proposed. This method
expands the small-scale training sample set and improves the generalisation ability of the deep network model. The improved target network is trained and tested experimentally. The test results show that the improved network training model can accurately identify the carbon marks on the pressboard surface.

2 STRUCTURE MODEL OF THE TRANSFORMER PATROL ROBOT FISH

The overall structure of the transformer patrol robot fish is shown in Figure 1, mainly including: robot fish body, robot swim bladder device, visual device, robot fish control system and robot fish fin device. The robot fish body is a spherical sealing structure. Two micro-propellers are installed on both sides of the body to control the forward movement and turn of the fish. A visual camera is equipped on the upper part of the fish body to inspect the insulation condition of the transformer. The control system is located in the middle of the robot fish body which mainly includes a motor drive and control module, a sensor acquisition module, a pose detection module, a wireless transmission module and a power supply module. A swim bladder is installed below the fish body and includes a piston oil cylinder, displacement sensors and a piston lift. Among them, the displacement sensor can accurately measure the amount of oil absorbed or expelled by detecting the location of the piston, so as to accurately control the floating speed and sinking speed of the robot fish. There are four fins, two fins are mounted on the upper end of the fish body and the other two fins are mounted on the lower end of the fish body. The mechanical fins can control the deflection direction and the angle of the fish. Through the combination of the fins and the swim bladder, a gliding movement of the robot fish can be realised even without power, which effectively reduces the power consumption and increases the inspection range of the robot fish.

The physical picture of the robot fish is shown in Figure 2. The robot fish is spherical with a diameter of 11 cm. There are a pair of fins at the upper and lower ends of the robot fish with a length (L) of 11 cm and a width (W) of 5 cm. The total width (TW) of the two fins is 18 cm and the total height (TH) of the upper and lower fins is 13 cm. The robot fish body is 3D printed with transparent resin, so the whole body appears transparent light blue. The installation position of the camera is 3D printed with colourless and transparent resin. Besides the robot fish, all other parts are 3D printed with highly ductile resin material.

Effective pixels of the vision camera installed in the patrol robot fish are 4 million. Horizontal field of view (HFOV) is 122°, vertical field of view (VFOV) is 93° and diagonal field of view (DFOV) is 150°. LED lights with adjustable brightness are used to fill in the light for the vision camera. Under conditions of strong supplementary light, the vision camera can take pictures in the dark. The pictures taken by the vision camera are then transmitted to the computer workstation via the 5.8G wireless digital image transfer module. The sealing performance and motion performance of the robot fish are tested in this study. The prototype has preliminarily realised the functions of rise up, sink down, plane 360-degree rotation, unpowered glide in transformer oil.

3 AN IMPROVED FAST-RCNN VISUAL TARGET DETECTION ALGORITHM

3.1 Typical Faster-RCNN

Faster-RCNN was proposed by Girshich et al. [20] in 2016 and has been widely used in the field of target detection at present. Faster-RCNN is an improved version based on the RCNN, SPP-Net and Fast-RCNN. Faster-RCNN consists of feature extraction network, regional recommendation network and classification regression network. The typical structure of Faster-RCNN is shown in Figure 3.

As shown in Figure 3, the steps of using Faster-RCNN for target detection are as follows: First, the whole image is input
into the input layer of the first CNN, and a feature map is obtained through convolution. Then, the feature map obtained by convolution is input into the Regional Proposal Network (RPN) to obtain feature information of the candidate box. The features of the candidate box are extracted and input into the classification regression network. Softmax classifier is used to identify calibrated categories. For the candidate boxes belonging to a certain feature, their position is further adjusted by regression.

3.2 Network structure and algorithm of the improved faster-RCNN

3.2.1 Feature extraction network based on VAE-CNN and VGG16

CNN is a kind of deep learning network with outstanding performance in image feature extraction. The structure of CNN consists of convolution, activation and pooling. When dealing with classification task, CNN establishes the mapping relationship between input data and specific feature space. After output image label through a fully connected network, the specific classification is realised [21]. VGG16 network is a CNN-based deep learning framework proposed by Synonym and Zisserman in 2015 [22]. The network has realised the training of 1.3 million \(224 \times 224 \times 3\) images belonging to 1000 categories, and the classification precision rate is 92.7\%. The parameters of the VGG16 model are open source, this makes VGG16 available for migration as a pre-training model in many classification models. The specific parameters of VGG16 can be seen in [22].

The CNN in VGG16 has a strong feature extraction capability, but there are two problems in the defect detection of an electric power system at present. One problem is that the dimension of the input image is much larger than \(224 \times 224 \times 3\) and the feature is more complex. The other problem is the lack of positive samples. Therefore, the feature extraction capability of VGG16 network will be limited, and the generalisation of the model is difficult to be guaranteed due to the lack of training samples. In recent years, the generation model has achieved remarkable results in small sample expansion [23]. Among them, variational auto-encoder CNN (VAE-CNN) can ensure the effect of feature extraction in small samples of complex data structures [24]. A self-encoding CNN detection model based on enhanced features is proposed by replacing part network in VGG16 with VAEs in this study, as shown in Figure 4. This detection model consists of two parts: variational encoder and CNN.

The most important part of the variational encoder is the generation of data samples. For each input image \(x \in \mathbb{X}\), its specific distribution \(p(x \mid z)\) can be obtained by encoding the
image. The recognition model \( g(z \mid x) \) is adopted to approximate the true posterior probability \( p(z \mid x) \). KL divergence is adopted to measure the similarity degree of the two distributions:

\[
KL(q \| p) = \int q(t) \log \frac{q(t)}{p(t)} dt = E_q[\log q] - E_q[\log p]
\]

After obtaining the distribution of real data according to the decoder, the data \( Z \) generated by the generator can recover as much original information as possible by decoding. The loss function of this process is expressed as:

\[
L_{AE} = MSE + KL,
\]

where \( MSE \) is the mean square error of identifying model and real data.

### 3.2.2 Region proposal network

RPN based on the feature extraction network generates the recommendation region by classifying and coordinate regressing the image background and detection target. Therefore, the input of RPN is the feature map obtained by convolution, and the output of RPN is the generated rectangular recommendation box.

RPN is fully connected to the feature map of the last convolutional layer through a sliding window. Each sliding window generates a corresponding low-dimensional vector. The rectangular recommendation box generated by RPN through window sliding is also called anchor box. The recommended region is obtained by modifying any anchor frame with four parameters, and the modified formula is expressed as:

\[
\begin{align*}
    x &= wx_f + x_o \\
    y &= wy_f + y_o \\
    w &= w_o \exp(t_w) \\
    h &= h_o \exp(t_h)
\end{align*}
\]

where, \( x, y, w, h \) represent the central horizontal coordinates, vertical coordinate, width and height of the predicted recommended area variables, respectively; \( l_x, t_x, l_y, t_y \) represent the four correction parameters.

At the beginning of the training, RPN determines the recommended window conforming to the calibration from a large number of windows by setting the threshold. The loss function of the network includes the classification part and the regression part, expressed as:

\[
L_{RPN} = \frac{1}{N_{cl}} \sum_i L_{cls}(p_i, p_i^* \| t_i, t_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)
\]

where, \( p_i \) represents the probability that the \( i \)th predicted anchor frame; \( p_i^* \) represents the actual label of the \( i \)th anchor frame; \( p_i^* L_{reg} \) represents whether regression correction is made to the anchor frame without foreground object. \( p_i^*=0 \) or \( p_i^*=1 \) is logical indicator; \( t_i \) represents the correction parameter of the predicted recommendation box; \( t_i^* \) represents the correction parameter of the anchor frame relative to the actual recommendation box; \( L_{cl} \) represents the logarithmic loss of dichotomies; \( L_{reg} \) represents regression loss of coordinate correction parameter.

\[
\begin{align*}
    L_{cls}(p_i, p_i^*) &= -\log \left[p_i p_i^* + (1 - p_i) (1 - p_i^*)\right] \\
    L_{reg}(t_i, t_i^*) &= \sum_{\alpha \in \{x, y, w, h\}} \text{smooth} b_{1,1}(t_i - t_i^*)
\end{align*}
\]

where, \( \text{smooth} b_{1,1} \) is smooth regression function which is used to filter the predicted correction parameters, as calculated by:

\[
\text{smooth} b_{1,1}(x) = \begin{cases} 
0.5x^2 & \text{if } \left| x \right| \leq 1 \\
\left| x \right| - 0.5 & \text{otherwise}
\end{cases}
\]

### 3.2.3 Classified regression network

Classification regression network is the key network for the final target detection. The classification regression network takes the feature graph output by the feature network and the recommended area box output by RPN as inputs and unifies the size of all features in the channel through the RoI pooling layer and generates the feature graph of the fixed area. Finally, two full connection layers are used to predict the defect category confidence of fixed recommended area frame and the regression of recommended area frame. In this network, two connected layers in a series are used to combine the feature maps of the candidate regions with a fixed size, then the feature vectors representing the candidate regions are obtained. The feature vector was input into the defect classification layer and the coordinate refinement layer respectively, and the score value and correction parameter of the candidate area belonging to each defect category are obtained. The Softmax layer is used to normalise the score value, and then the corresponding confidence degree is obtained. The coordinate information of the candidate area is obtained by using the correction parameter.

Softmax classifier is selected as the output function of the network to predict the defect category of the zone box [25]. The Softmax layer follows the full connection layer and receives the output of full connection layer, then calculates the prediction probability distribution of the region as follows:

\[
p(u | m) = \frac{e^{m_i}}{\sum_k e^{m_k}},
\]

where, \( p(u \mid m) \) represents the probability of category \( u \) under the condition of softmax layer input feature vector \( m \); \( m_i \) represents the \( i \)th input data of the eigenvector \( m \); \( N \) represents the total number of defect categories. For all values of \( i \), \( \arg \max p(u \mid m) \) is the predicted defect area of the softmax layer output.
TABLE 1  Experimental server configuration parameters

| Name             | Parameter                  |
|------------------|----------------------------|
| CPU              | Intel Xeon Gold-5217 (x2)  |
| GPU              | NVIDIA Tesla V100-16G *2   |
| Memory capacity  | 128G                       |
| Software         | cuda 10.1 + cudnn 7.6.4    |
| Compiled language| Python 3.6                  |

The final loss function includes classification loss and regression loss, expressed as:

$$L_{\text{cls}} = \begin{cases} L_{\text{cls}}^\prime (p, u) + \lambda L_{\text{loc}} (v^\prime, v^*) \\ L_{\text{cls}}^\prime (p, u) \end{cases},$$  \hspace{1cm} (8)

where, $L_{\text{cls}}^\prime$ is forecast defect category loss, expressed as:

$$L_{\text{cls}}^\prime (p, u) = - \log p (u | m),$$  \hspace{1cm} (9)

where, $p$ is confidence coefficient, $u$ represents the actual category tag corresponding to the input recommendation area.

$L_{\text{loc}}$ is the regression loss of the position correction parameter in the predicted area, expressed as:

$$L_{\text{loc}} (v^\prime, v^*) = \sum_{i \in \{x, y, w, h\}} \text{smoothl}_{1,1} (v^\prime, v^*_i),$$  \hspace{1cm} (10)

where, $v^\prime$ represents the final recommended area box correction parameter for the prediction; $v^*_i$ represents the correction parameter of the recommended area relative to the actual calibrated area.

4  | MODEL TRAINING PROCESS

The transformer defect detection can be realised through the proposed improved algorithm framework of faster CNN. The implementation process is described in detail as follows:

1. Training sample set construction

Transformer defect images are obtained through a high voltage experiment. The location and category of the defect are manually marked by a calibration software. The image matrix of the marked image is used as variable $X$ and the XML file formed
by calibration is used as variable \( Y \). \( X \) and \( Y \) form the training sample. The cross validation method was adopted to divide the training samples into training validation sets and test sets, and then input them into the model for training.

2. Determine the network structure and training algorithm

The training process of the model adopts the method of combining alternate training and joint training which is the same as the original network. The training process is divided into four steps. First, the regional recommendation network and feature extraction network are trained separately, in which the model parameters of VGG16 are directly imported by the pre-training model. The training samples are sampled using Mini-batch, each of which contained \( N \) anchor frames randomly extracted from an image. BP algorithm and STOCHASTIC gradient descent algorithm are used to optimise the loss functions of all anchors and carry out end-to-end training.

Second, the output of the regional recommendation network in step (1) is taken as the input of the detection network to conduct pre-training classification regression network. After the training, the unique parameters of shared convolutional layer and classified regression network are updated.

Third, retrain the area recommended network, fix network shared part of the parameters and only update its unique parameters.

Fourth, according to the retraining results, the entire firm-RCNN is fine-tuned, the parameters of shared parts of the network are fixed and only the parameters of unique parts of the firm-RCNN are updated.

3. Hyperparameter setting

The training network hyperparameter setting mainly includes input and output layers of each network, convolution kernel size, dropout inactivation rate, regular term coefficient and learning rate. Among them, the dimensions of input and output layers are determined by the number of variables. After many tests, the network modelling effect is good when the convolution kernel is set as 5*5, the pooling kernel size is set as 2*2, the sliding window of the anchor frame is set as 3*3 and the learning rate adopts variable learning settings ranging from 1E-4 to 0.1.

5 | ANALYSIS OF THE TRANSFORMER DEFECT DETECTION EFFECT

5.1 | Sample acquisition and calibration

In order to verify the detection effect of the proposed improved Faster-RCNN in transformer defect recognition, 100 transformer defect images are obtained through a high voltage experiment, which is used to construct the initial small sample defect recognition library. Some defect examples are shown in Figure 5.

The calibration of training samples is an important part in the process of constructing training samples. The samples obtained by the high voltage experiment are original defect graphs. The defects in the graphs need to be marked with Label Img soft-
ware to obtain the real labels of defects in the training sample.

The label format of defect graph is $Y = \{\text{‘carbon mark'}, \text{‘carbon mark’…, ‘Carbon mark’}\}$. In practice, defects only need to be detected and there is no need to classify defect types. Therefore, label format of $Y$ has only one category.

The labelled samples are divided into training verification samples and test samples according to 80% and 20% ratios, and then the cross-validation method is used for training. Suppose $D$ is the sample set which is composed of group $k$ defect pictures and their annotations:

$$D = \{\{X_1; Y_1\}, \{X_2; Y_2\}, \ldots, \{X_k; Y_k\}\}. \quad (11)$$

$D$ is divided into $k$ mutex subsets of similar size, which satisfies the following formula:

$$D = D_1 \cup D_2 \cup \ldots \cup D_k, \quad D_i \cap D_j = \emptyset \quad (i \neq j). \quad (12)$$

Take $k$-1 subset as training set and the remaining subset as test set:

$$\begin{align*}
D_{\text{test}} &= \text{rand}(D_1, D_2, \ldots, D_k) \quad D_{\text{test}} \subset D \\
D_{\text{train}} &= D_{\text{test}} \quad D_{\text{train}} \subset D
\end{align*} \quad (13)$$

Thus, $k$ group training/test data sets can be obtained and $k$ times of training and testing can be conducted.

### 5.2 Model evaluation index

The final task of the model is to detect the defects and mark them in the form of an anchor frame. The precision of labelled recommended area box is evaluated by IoU (Intersection over Union) index which is defined by:

$$\text{IoU} = \frac{(A \cap B)}{(A \cup B)}, \quad (14)$$

where, $A$ and $B$ represent the area of the anchor frame and the real marking frame, respectively.

For each graph, calculate the prediction box of each positive sample. According to the prediction box, the real labeled IoU value and the set IoU threshold value, the correct detection value TP number and the wrong detection value FP number of each category in each graph are obtained. Thus, the precision $P_r$ of each class can be calculated as below:

$$P_r = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (15)$$
Similarly, the number of missed targets FN and the recall rate can be obtained.

\[ R_c = \frac{TP}{TP + FN}. \] (16)

The performance of the training model can be verified by defining the model precision evaluation index. However, the transformer defect detection problem belongs to multi-classification problem and the image has more than one label. Therefore, the above evaluation criteria of single label classification are no longer applicable. mAP (Mean Average Precision) is a commonly used indicator for target detection in the world, which is composed of precision index, recall rate index and retrieval ranking. Its calculation formula is as follows:

\[ mAP = \frac{1}{N_c} \int_0^1 P_c(R_c) \, dR. \] (17)

In this paper, mAP is also used as the precision evaluation index of the model. The specific calculation process of mAP refers to the literature [26].

**TABLE 3** Comparison of various indicators and training time under different training epoch

| Training iteration | Precision index(mAP %) | Training time (s) |
|--------------------|------------------------|-------------------|
| 10                 | 33.46                  | 3                 |
| 20                 | 69.20                  | 7                 |
| 60                 | 66.74                  | 16                |
| 100                | 53.19                  | 27                |
5.3 | Training model training process and results

Pytorch 1.2.0 framework released by Facebook is adopted to build the learning network. The hardware configuration used in the experiment is shown in Table 1.

According to the procedure described in Section 2, each training sample is first divided into training verification set and test set. Next, the network for the training is set up. A VAE is added to the feature extraction network to realise the expansion of the small-scale training sample set. A total of 4800 sample sets are generated after the expansion. The results of a sample after variational self-encoder are shown in Figure 6.

It can be seen from Figure 6 that the original sample is expanded through variational encoder. By rotating the original image, the shape and defect depth of the image are expanded and the features are enriched. The network can simulate the low quality images caused by the environment or the shooting light when the robot fish moves in transformer oil. Thus, the generalisation of the model has been increased. Table 2 shows the change of precision index and training time of model test set during iteration process. Figure 7 shows the training and verification loss curves in the training process, respectively.

It can be seen from Figures 7 and 8 that the loss curves of regional recommendation network and classification regression network show a downward trend during the training. The total loss is also decreasing. The total loss continued to decline after 100 sessions of the training, indicating that the model does not reach the optimal state after 100 sessions of training. However, the fitting precision has been very high at this time and the mAP index has reached more than 95%, which is enough for transformer defect detection.

The model obtained after 100 times of the training is used to identify the test samples of the training set. Accurate identification results can be obtained, as shown in Figure 9.

Figure 9 shows that the model with 100 training sessions can accurately mark different shapes of carbon marks. Figure 9b shows that the carbon mark area can be accurately marked when the recognition precision is above 95%.

In order to verify the influence of the VAE proposed in this study on the whole target detection model, a small sample training set is trained by using the original fast-RCNN. Table 3 shows the model precision and training time under different training times.

Similarly, the model without VAE after 100 times of training is saved and used to identify the test set. The results are shown in Figure 10. It can be seen from Figure 10 that the precision of the model without VAE is very low and this model is difficult to accurately distinguish and label carbon marks of different categories. This indicates that it is difficult to effectively train the model without VAE with small sample size.

5.4 | Experimental results of target recognition of the transformer patrol robot

The visual inspection testing platform of transformer patrol robot fish is shown in Figure 11. The oil tank is made of plexiglass to observe the motion performance of the robot fish. The oil tank is 800 mm long, 450 mm wide and 500 mm high, full of transformer oil. An insulating board with carbon marks is arranged in the transformer oil tank, as shown in Figure 11. The patrol robot fish is put into transformer oil from the oil injection hole and starts its inspection. The visual camera in the robot fish continuously captures the images of the transformer’s internal environment. The images are then transmitted wirelessly to a computer workstation.

The camera on the transformer patrol robot fish continuously captures carbon marks on the insulating pressboard when the fish moves in the transformer oil. In order to verify the
FIGURE 11  Visual inspection testing platform of the transformer patrol robot fish

Identification effect of the patrol robot fish, carbon marks in different conditions are photographed, such as: (a) different distances between the camera and pressboard carbon marks; (b) different camera light intensity; (c) different degrees of camera shake caused by the rapid movement of the robot fish; (d) different number of target carbon marks on the same pressboard. The computer workstation can automatically and intelligently identify images with different conditions, and the recognition results are shown in Figures 12–15.

It can be seen from Figure 15 that there are three carbon marks on the insulating pressboard. The improved deep learning model based on Faster-RCNN accurately recognises the two carbon marks in the middle and on the right, while the carbon marks on the left are not detected. The main reason is that the size of this carbon mark is very small and the color of this carbon mark is close to the background of the image. In addition to this situation, the improved depth learning model based on Faster-RCNN designed in this study can accurately identify carbon marks in different conditions (different shooting distance, different light intensity and different camera shake degree) and has strong generalisation ability, which is suitable for transformer defect detection.

FIGURE 13  Recognition results in different light intensity at the distance of 14 cm

FIGURE 14  Recognition effect of carbon mark shot when the robot fish shakes quickly

6 | CONCLUSION

A visual defect detection method based on the transformer patrol robot fish and Faster-RCNN deep learning frame has been proposed in this study. In order to solve the problem that the transformer has fewer insulation defect samples, the proposed method embeds the VAE network into the traditional Faster-RCNN and constructs a self-encoding convolutional neural network based on enhanced features. The proposed
method has realised the offline modelling of small samples and can be applied in practice.

A visual inspection test platform has been built to verify the proposed method. The results show that the proposed method has high recognition precision not only for samples generated by the high voltage test, but also for images captured in the test platform. Therefore, this method has a high generalisation performance, high recognition precision and strong anti-noise capability in some environments. Compared with other visual detection methods, the proposed method does not need to obtain large-scale samples. Therefore, it saves the annotation time and model training time and is easy to update the model. Compared with the traditional oil drain detection, the proposed method has advantages of low cost and easy operation and can realise online identification and location.

In this article, the experimental condition of the robotic fish visual inspection method is new clean transformer oil. Since the patrol robot fish uses a camera to visually shoot the internal conditions of the transformer, the clarity of pictures depends on the cleanliness of the camera lens and the turbidity of the transformer oil. The patrol robot fish should be kept far away from sludges and deposits during the inspection and shooting. Since the colour and turbidity degree of transformer oil will change to different degrees with the different operating time of a large transformer, further research is needed on visual detection of transformer oil with different turbidity levels.

ACKNOWLEDGEMENTS
This work is funded by the National Natural Science Foundation of China (No. 51907102) and the China Postdoctoral Science Foundation (No. 2018M640128).

REFERENCES
1. Meng, Q., et al.: Study on completely disassembly type of large capacity UHV AC power transformers. In: Proceedings of the CSEE’17, Barcelona, Spain, 2–4 April 2017
2. Guo, H., et al.: On-site drying technology of disassembly-transported UHV AC transformer. In: 2020 IEEE International Conference on High Voltage Engineering and Application, Beijing, China, 6–8 September 2020. https://doi.org/10.1109/ICHVE49031.2020.9290035
3. Guo, H., et al.: Key technology and engineering application of the on-site assembly UHV AC transformer. In: 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 21–23 November 2019. https://doi.org/10.1109/iSPEC48194.2019.8975151
4. Chen, Z.: Review of direction of arrival estimation algorithms for partial discharge localisation in transformers. IET Sci. Meas. Technol. 13(4), 529–535 (2019)
5. Yang, X., et al.: Comparison of PD detection methods for power transformers: their sensitivity and characteristics in time and frequency domain. IEEE Trans. Dielectr. Electr. Insul. 23(5), 2925–2932 (2016)
6. Shah, G.C.H.: ABB ability inspection for transformers-TXplore. ABB USCR Mechatronics Sensors Bloomfield, CT, USA (2018), https://www.hitachi-abb-powergrids.com/offering/product-and-system/transformers/transformer-service/advanced-services-for-transformers/asset-management-solutions/txplore-internal-transformer-inspection-robot
7. Shah, G.C.H., et al.: ABB TXplore robot redlines transformer inspection-transformer in renewables application-ABB special report transformers. ABB USA (2018). https://new.abb.com/news/detail/7870/abbs-txplore-robot-redlines-transformer-inspection
8. Choi, H.T., et al.: Development of an underwater robot, ODIN-III. In: Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), Las Vegas, NV, USA, 27–31 October, 2003
9. Yue, C., et al.: Mechatronic system and experiments of a spherical underwater robot: SUR-II. J. Intell. Rob. Syst. 80(2), 325–340 (2015)
10. Feng, Y., et al.: Hovering control of submersible transformer inspection robot based on ASMBM method. IEEE Access 8, 76287–76299 (2020). https://doi.org/10.1109/ACCESS.2020.2989378
11. Du, L., et al.: Target detection method based on convolutional neural network for SAR image. J. Electron. Inf. Technol. 38(12), 3018–3025 (2016)
12. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
14. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–25 June 2005
15. Felzenszwalb, PF., et al.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. 32(9), 1627–1645 (2010)
16. Hinton, G.E., et al.: A fast learning algorithm for deep belief nets. Neural Comput. 18(7), 1527–1554 (2006)
17. Hinton, G.E., Zemel, R.S.: Autoencoders minimum description length and helmholtz free energy. Adv. Neural Inf. Process. Syst. 6, 3–10 (1994)
18. Lecun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998)
19. Huang, X., et al.: A multi-feature fusion-based deep learning for insulator image identification and fault detection. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 1957–1960 (2020)
20. Ren, S.Q., He, K. et al.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149 (2017)
21. Krizhevsky, A., et al.: ImageNet classification with deep convolutional neural networks. Commu. ACM, 60(6), 84–90 (2012)
22. Simonyan, K., et al.: Learning local feature descriptors using convex optimisation. IEEE Trans. Pattern Anal. Mach. Intell. 36(80), 1573–1585 (2014)
23. Wang, Y., et al.: Generalizing from a few examples: A survey on few-shot learning. ACM Comput. Surv. 53(3), 1–33 (2020)
24. Kingma, D.P., Welling, M.: An introduction to variational autoencoders. now, (2019)
25. Chopra, P., Yadav, S.K.: Restricted Boltzmann machine and softmax regression for fault detection and classification. Complex Intell. Syst. 4(1), 67–77 (2018)
26. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, New York (2010)

How to cite this article: Ji H, Cui X, Ren W, Liu L, Wang W, et al.: Visual inspection for transformer insulation defects by a patrol robot fish based on deep learning. IET Sci Meas Technol. 2021;15:606–618. https://doi.org/10.1049/smt2.12062