RESEARCH ARTICLE

Study on Improved VGGNet and SK Convolution Identification Model for Defect Classification of Single Molten Salt Battery

WENCHAO XU [1,2], **WENBO XU**[3], **SIXIANG ZHANG**[1], **YITAO HAN**[2], **YIN JIANG**[2], **AND HENGYI ZHAO**[2]

[1]College of Mechanical Engineering, Hebei University of Technology, Tianjin 300131, China
[2]College of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China
[3]College of Higher Vocational Education, Tianjin University of Technology and Education, Tianjin 300222, China

Corresponding authors: Wenchao Xu (xuwenchao@tjcu.edu.cn) and Sixiang Zhang (13502063552@163.com)

ABSTRACT Aiming at the lack of large public single molten salt battery data sets, to reduce the labour-consuming, to improve the insufficient learning ability of traditional diagnostic methods in the production of single molten salt battery, an image recognition model for molten salt battery defects based on transfer learning is proposed. First, some pre-processing operations and image enhancement on the single molten salt battery image are performed. Second, the backbone of recognition model is built based on VGG16 network, and the selective kernel (SK) convolution module is adopted after the bottleneck layer, convolution kernel with an appropriate size can be selected adaptively through input feature map; Third, the FC is taken the place of a GAP layer, a dropout layer, and other fine-tuning operations are added, a simplified model called V-VGGNet is got; Finally, the weight parameters obtained from the pre-training on the ImageNet data set are transferred to the single molten salt battery image recognition model V-VGGNet. For different network structures and different training strategies, comparative experiments of performance tests are conducted. The test data manifest that the accuracy rates of V-VGGNet network for three categories of defective images (Missing Negative Electrode, Broken Tab, and Missing Current Collector) and Assembly Normal images can reach 95.14%, 98.79%, 98.21%, and 99.41%, the average accuracy can achieve 97.91%, good performance improvement of the single molten salt battery is improved, it is about 3% higher compared to other well-knows networks, which verified the feasibility of V-VGGNet model and the effectiveness of the improvement.

INDEX TERMS Single molten salt battery, VGG network, SK convolution, deep learning.

I. INTRODUCTION

Amolten salt battery is a primary reserve battery, molten salt is used as its electrolyte, and it is activated by melting the molten salt. Thermal batteries have many advantages, including high specific energy and specific power, wide operating environment temperature, long storage time, rapid and reliable activation, compact structure, simple process, low

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

cost, and no maintenance [1]. As soon as it came out, it is favoured by the military and gradually developed into an ideal power source for advanced weapons such as anti-aircraft missiles, small nuclear devices, and shield ships. It has an extremely important position in the military field. To meet the strictly requirements of modern weapons, the LiMx/FeS2 molten salt battery (LiMx is a lithium alloy) developed in the 1970s not only overcame the long-term electrical noise but also greatly improved the specific energy [2], [3], [4]. High specific energy, high specific power, and high reliability

FIGURE 1. X-ray image.

are major technological breakthroughs in the development of thermal batteries, and they became the current development trend of thermal batteries.

The molten salt battery is generally welded and composed of several electric stacks, activation mechanisms, combined shells, combined covers, etc. The electric stack is composed of several single cells and heaters connected in series and parallel [5], [6]. The electric stack is placed in a combined shell, and it is connected with the terminal on the battery cover through the drain bar, a complete molten salt battery is formed. The X-ray image is shown in Figure 1.

The vital part is electric stack, which consists of a negative electrode (0.2 mm), an electrolyte (0.3 mm), a positive electrode (0.2 mm), a heating agent (0.5 mm), and a current collector (0.1 mm), and every electric stack is cyclically placed. A negative electrode, an electrolyte, a positive electrode, and a heating agent are composed of a single battery [7]. During the assembly process of thermal batteries, there may be some assembly errors, such as Missing Current Collector, Missing Negative Electrode, Broken Tab, etc. due to operational errors.

Traditional molten salt battery assembly defect detection methods are often manual detection; this detection method has the following shortcomings: (1) Subjective and non-standardized test results are influenced by subjective factors and objective factors; (2) Due to the inevitable mental exhaustion and fatigue of operators, test results such as missed detection and false detection are prone to occur; (3) It is difficult to manually detect small defects such as Broken Tabs; (4) During molten salt battery production and testing, improper operation may cause secondary damage to the single molten salt battery; (5) The traditional manual detection method is inefficient and time-consuming. Therefore, after the production of the molten salt battery is completed, the X-ray image of the molten salt battery needs to be tested to improve the efficiency and accuracy of manual detection.

In the past forty years, many defect detection algorithms are emerged, which consist of traditional image processing algorithms, machine learning algorithms based on manual features or shallow trajectory networks, and neural network-centric methods.

Traditional image processing algorithms mainly use the original image features for detection and segmentation [8], the features are presented on the surface of the defect, and the methods of structural, threshold, spectral, model-based,

etc. are included. Structural methods include edges [9], skeletons, template matching, and morphological operations [10]; Thresholding methods include iterative optimal thresholding [11], [12], Otsu thresholding algorithm, etc. [13], [14], [15]. Spectral methods typically include wavelets transform, etc. [16], [17], [18]. The methods based on handcrafted features or shallow neural networks usually involve the feature extraction stage and pattern recognition stage. The extracted features include the Local Binary Pattern (LBP) feature, GLCM, HOG, etc. [19], [20]. Most of the experimental results show that these detection algorithms have low detection accuracy in the application of defect detection.

Moreover, a great number of hyper parameters are needed to manually set in traditional image processing algorithms, and multiple thresholds are usually needed to manually set for defect features in actual scenes; the size of these thresholds is directly related to the background. When an algorithm is applied to a new task, hyper parameters are fine-tuned again, or even the algorithm needs to be redesigned.

In literature [21], Alasnanda conducted a morphology-based algorithm. Reasonable threshold processing is introduced to the canny operator for defect boundary extraction. To obtain continuous target boundary points, a mathematical morphological theory is used for corrosion operations, but the experimental results used in this method have much room for improvement. Felisberto in the literature [22] used empirical threshold to determine reasonable parameter values based on genetic algorithm, and two sets of different radiographic images are used to extract the weld. This method can extract the position of the weld in the radiographic image very well, but the effect of the extraction of the curved weld is not good.

In 2016, various common surface defects in the friction stir welding are classified into holes, grooves, cracks, keyholes, and flashes by Ranjan et al. [23], image pyramids and image reconstruction algorithms are used to identify defects based on the characteristics of defects, and machine learning methods are used to locate the defect area and analyse the degree of defect damage. So the efficient identification of holes, grooves, cracks, keyholes, and flash defects is realized. In 2018, traditional networks are used to classify and recognize ultrasonic signals of stainless steel weld defects by Florence et al. [24] at SSN College of Engineering in India, the classification of four types of defects including porosity, cracks, incomplete penetration and incomplete fusion is realized through the Back Propagation Network. Ultrasonic phased array technology is used by Murta [25] in the United States for defect analysis, and it is found that the classification of defects by this method relies heavily on experiences. The propagation in a two-dimensional medium is simulated, the k-nearest neighbour algorithm is used to link ultrasonic signal with modelling defect, and a good result is achieved.

Aiming at the problem of automatic defect detection, the study [10] used mathematical morphological filters to detect defects. Important texture features are got by the Gabor Wavelet Network (GWN), and an optimal morphological filter is constructed. The defect characteristics is described

well and the detection accuracy is high, but the visualization effect is poor, the interpretability is low, and it is not easy to understand. Jung et al. [26] used a lock-in amplifier-based EC instrument and a cup-core transceiver probe to detect crack defects. The edge line is generated by the percolation process, the panel crack detection and location are realized through the edge line evaluation. Kieselbach et al. [27] proposed a visual inspection system physically for detecting the car body paint surface defects, and an image retrieval system is built, which includes image SIFT features, textures, etc. The image retrieval system performs matching similarity on the image data for surface quality detection, and a better detection effect is realized. The multilayer perceptron method is used by Yahia in the study [28], the method is based on two types of texture features for defect detection, possible defects are segmented and extracted by an edge detector, and the extracted feature information is classified successfully.

The above methods can achieve the ultimate detection function to a certain extent, but such algorithms often combine shallow features such as shape, texture, or colour to identify a specific category, and simultaneous detection of multiple defects cannot be achieved. In practical applications, it is affected by objective factors such as environmental influences, lighting conditions, equipment motion blur, as well as the imaging quality recognition, effectiveness and reliability of target detection are not ideal, and such methods are not scalable.

Compared with the traditional methods in the above-mentioned literature, deep learning methods have unique potential and advantages in object recognition and scene classification, feature extraction is simple, classification is accurate. And deep learning methods have the ability to increase the accuracy as the number of samples increases.

As image processing technology has become an indispensable means in scientific research and technical application, especially the in-depth use of hardware such as graphics processing unit (GPU) with powerful computing capabilities, deep learning techniques is performed well in image recognition and target detection. Excellent network models such as YOLOV3, VGG16, GoogLeNet, and ResNet have appeared successively. The advantage of CNN is that it does not require complex operations such as manual image pre-processing and additional features design, and feature learning is performed by deep networks.

In 2019, a weld defect dataset is created by TWI Company Bacioiu [29], the recognition of two types of weld defects are realized by constructing a CNN with Fully Convolutional Networks (FCN), the classification accuracy of two types of weld defects reached 89% and 95% on this dataset. The research uses X-ray photos for classification and identification, and the accuracy of multi-category defect detection needs to be improved. Chen et al. [30] introduced a new end-to-end network model to detect surface defects. The improved model can effectively suppress the background and highlight defective areas. However, in the attention mechanism

of the improved model, only spatial attention is adopted, and the weight generated by channel attention is not combined.VGG16 and Xception models are used by Westphal and Seitz [31] combined with transfer learning methods, a new recognition model is constructed for LCD panel defects. This model can distinguish between defects and indicators of background regions. After many times training and good results are achieved, among which the CNN with stacked integration technology played an important role.

For the feature parameter extraction of welding defects, a defect feature database is built by Kasban [32], Mel Cestrum parameters and polynomial coefficients are used according to the size and shape of different defects. An artificially designed model is used for matching. This method abandons the traditional method of geometric features, and it is a relatively new feature extraction method.

Gibert et al. [33] introduced an innovative fastener detection algorithm combined with CNN and SVM, experimental results show that the anomaly detector's performance can be optimized and adjusted under a multi-task Bayesian framework. ResNet [34] network model is proposed by Kaiming H et al., which solved the problem of the difficulty of deep network training through Shortcut Connection fitting residual items. By increasing the network depth the gradient disappearance problem is alleviated, and the model error rate is greatly reduced. The weld defect data set established in the document [35] has complex types and different shapes. There are long-shaped streaks and cracks, as well as small-shaped point-shaped defects and rounds. Aiming at slag inclusion, porosity, lack of fusion, roundness, and other defects, the experimental results of the two models show that the ResNet network has the lowest error rate and better recognition.

The document [36] introduced a region proposal network (RPN). Some useful and high-quality region suggestions are used in Fast R-CNN for defection. By sharing their convolutional features, the two networks are further combined into one structure. Aiming at the low efficiency in manual inspection of automobile body welding appearance quality, the literature [37] proposed a multi-scale convolution block and designed an attention block to calibrate the solder joint feature map. Combined with multiple multi-scale blocks and a feature fusion strategy the backbone network of CNN can enhance computational efficiency. The classification data indicate that the detection accuracy of ACMNet can reach 95.2%. Literature [38] in view of the bearing performance fault detection and identification based on bearing sensor data, which overcame the shortcomings of massive training data, a recognition network combined Dense Convolutional block and attention mechanism is proposed, this method has the advantage of obtaining higher accuracy with fewer unknown learning parameters.

In response to the requirements for rapid and accurate state detection of rails, fasteners, sleepers, etc. on railway

tracks, this paper proposed an intelligent method [39] for detecting multi-target defects based on the BOVW model, and this method uses spatial pyramid decomposition and an improved YOLOv3 model. The detection accuracy rate can reach 96.26%. The complexity is reduced; the detection speed is further improved. Study [40] aiming at the cases of multiple types of defects in the actual situation, a two-step crack detection strategy is proposed, and the YOLOv4 network model is used to discard images without cracks. In the newly generated coarse area, the hybrid expanded convolutional block network is used, and the detection accuracy of cracks can reach 97.8%. Literature [41] proposed a method for identifying defects in wafer images which includes nine types of defects. The parameters learned from ImageNet are transferred to Dense Net, and classifier is redesigned. The actual experimental verification on the production line shows that the recognition efficiency is improved, thereby the feature learning ability is improved, and category imbalance in wafer defect recognition is also solved. In the study [42], aiming at the lack of a large number of marked wafer defect images, a GAN model combined with transfer learning is proposed. The learning features are introduced into the tag learning block, which effectively reduced the difference in feature distribution, the weight parameters in the offline wafer image are transferred to GAN network, the accuracy of wafer defect-recognition reached 97%. A loss function that can adapt to multiple scales is proposed in literature [43], which solved the problem of poor detection effect caused by sample imbalance. Reference [44] used void convolution instead of traditional convolution in the CSPDarknet-53 backbone network to improve the detection of defects at different scales. In study [45], lightweight network model MobileNetV2 is adopted to replace the original backbone feature extraction network of YOLOv4, and 3 × 3 convolution is replaced by depth separable convolution, which greatly reduced the parameter scale of the model and improved the detection speed of the model.

In this paper, the idea of transfer learning and improved SK module are draw into the V-VGGNet recognition model. Training a model often requires numerous samples, without enough training samples, the network can't fully extract features during the training process or the trained model has poor generalization ability. The main advantage of transfer learning over traditional neural networks is that it does not require plenty of training samples, to find the defective molten salt battery in the production process more quickly and accurately, a public molten salt battery dataset is established, and the defect recognition model is proposed and convergence ability is further improved.

The novelty contribution of this paper is highlighted:

(1) A diverse data set of molten salt battery is established by preprocessing and image enhancement, which exceeded most of the defect data sets. It is helpful to reduce the influence of external environmental factors such as exposure rate during image collection, and improve the learning ability of V-VGGNet with the increase of the number of data sets.

(2) The transfer learning method is introduced to build the basic model of V-VGGNet, through different transfer strategies a large number of weight parameters can be transferred to the model, and the representation with strong classification ability can be learned from a lot of weak features, which is beneficial to improve the recognition rate of the three categories of assembly defect images.

(3)The improved SK convolution module is introduced to the V-VGGNet, which is can adaptively select a convolution kernel with an appropriate size according to input feature map. The recognition accuracy and speed are improved by the molten salt battery assembly detection model V-VGGNet, which is nearly 3% higher than proposed by Zhao et al. [46]. Compared with five traditional detection methods, the V-VGGNet network model has a better recognition effect.

The structural components of this paper are as follows:

Section II represents the acquisition, pre-processing, and enhancement of molten salt battery defect images. Section III demonstrates the network architecture establishment and principle introduction of the molten salt battery identification model in detail. Section IV represents four groups of comparative experimental analysis, visualization of the feature map, analysis of recognition rates, and confusion matrix. Some concluding remarks are drawn in Section V.

II. EXPERIMENTAL DATA
A. ACQUISITION OF DEFECT IMAGE DATASET
In this paper, the defect samples of the dataset used in the V-VGGNet model are mainly derived from the laboratory and the inspection in the assembly production line.

The 3 categories of defect images and Assembly Normal images included in the dataset are all caused by the single molten salt battery assembly process. In order to make the samples as diverse and extensive as possible, defect samples with different damages in different environments are also made. The dataset includes 4 categories of image samples. The 3 categories of defects images include: Missing Negative Electrode, Broken Tab, and Missing Current Collector. Assembly Normal image and an X-ray machine for image acquisition are shown in Figure 2. The 3 categories of defect images of single molten salt battery assembly error are shown in Figure 3.

B. IMAGE PREPROCESSING AND ENHANCEMENT
Convolutional neural networks may have over-fitting problems due to insufficient data. Over-fitting will make the generalization ability worse. For preventing its occurrence, the method of data enhancement is adopted to enrich single molten salt battery dataset.

(1) Horizontal or vertical translation: Randomly translate the image of single molten salt battery by 10 pixels horizontally or vertically, and blank spaces are filled most nearly.

(2) Horizontal or vertical flip: randomly flip in the horizontal or vertical direction.
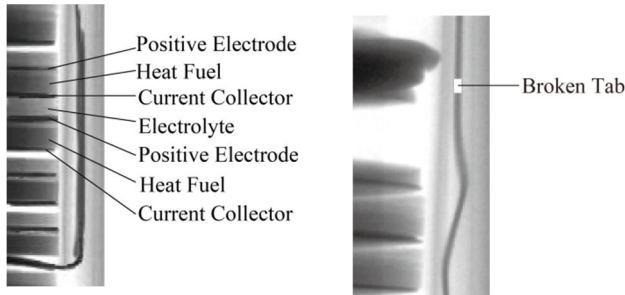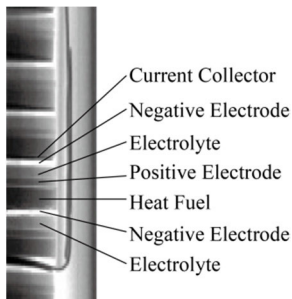
(a) X-ray machine.

(b) Single molten salt battery.

FIGURE 2. Assemble normal images and X-ray machine for image acquisition.

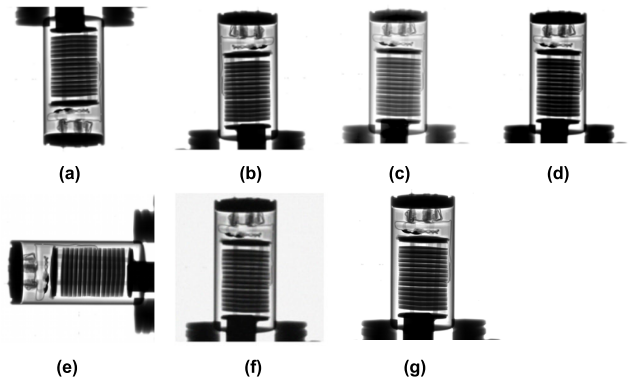(a) Missing Negative Electrode.   (b) Broken Tab.

(c) Missing Current Collector.

FIGURE 3. Three categories of assembly error defect images of a single molten salt battery.

(3) Zoom: The ratio is 0.85-1.15; blank spaces are filled most nearly.

(4) Rotation: Randomly rotate 0°-90°, blank spaces are filled most nearly.

(5) Artificial noise: adding Gaussian noise.

(a) vertical flip; (b) horizontal flip; (c) bright regulation; (d) contrast ratio; (e) rotate counterclockwise; (f) noise; (g) zoom;

FIGURE 4. Data enhancement.

(6) Adjust the brightness: increasing or decreasing the light according to the actual situation of the X-ray machine.

(7) Delete heavily polluted or useless images.

The data enhancement operation is used to expand the sample of the single molten salt battery. The goal of these transformations is to generate more samples for creating a larger dataset and expand the amount of training data associated with the learning object. This not only increases the amount of data but also improves the quality, and learn features better and more, and play a certain role in alleviating model over-fitting. At the same time, learning other irrelevant features is avoided, and faster model convergence is promoted. The data enhancement is shown in Figure 4. 2149 original single molten salt battery images are expanded to 9085 images, which are proportionally separated into training, validation, and test, they account for 60%, 20%, and 20% respectively. The distribution of Assembly Normal (Nor) images and three categories of defect images (Missing Negative Electrode, M-N-E; Broken Tab, B-T; Missing Current Collector, M-C-C) in the single molten salt battery data set are shown in Figure 5. After a series of operations such as cropping, removing redundancy, and data enhancement, as the input of the network, the single molten salt battery images are resized to 224 × 224 and batch normalized.

III. IDENTIFICATION MODEL FOR DEFECT IMAGE OF SINGLE MOLTEN SALT BATTERY

A. STRUCTURE ANALYSIS OF VGG16

VGG16 is regarded as an outstanding image recognition model. As a classic model, VGG16 has the advantages of concise structure and easy implementation and still has high research value.

In V-VGGNet model, the classic VGG16 is the pre-training backbone, and combined with feature transfer some improvements are made on the basis of the VGG16 network model. A simplified description of the VGG16 network structure is shown in Figure 6.

The VGG16 has 16 layers, consisting of 13 CON layers, 5 down-sampling layers, and 3 FC layers. The Relu function
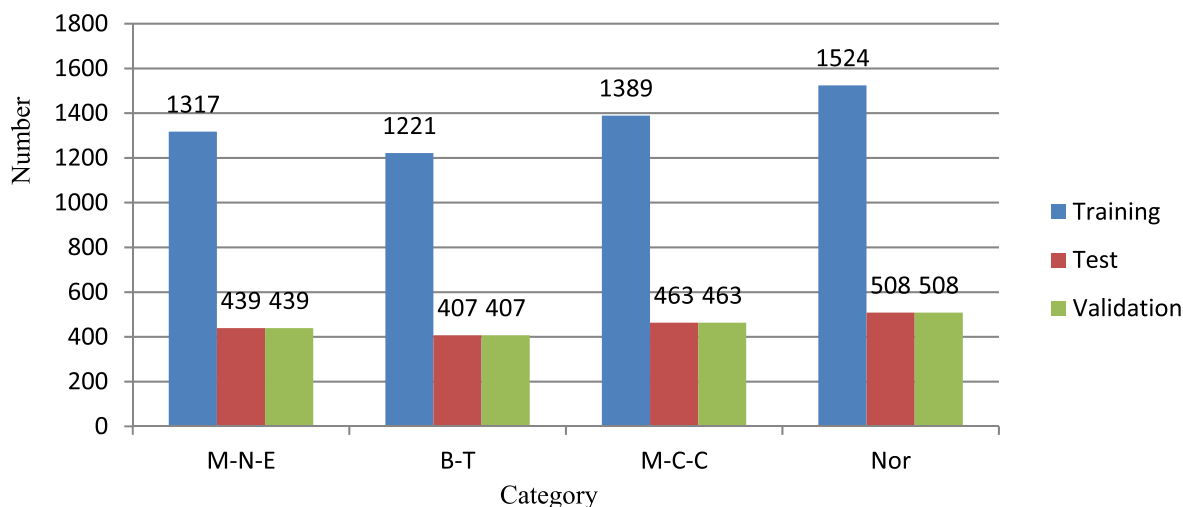
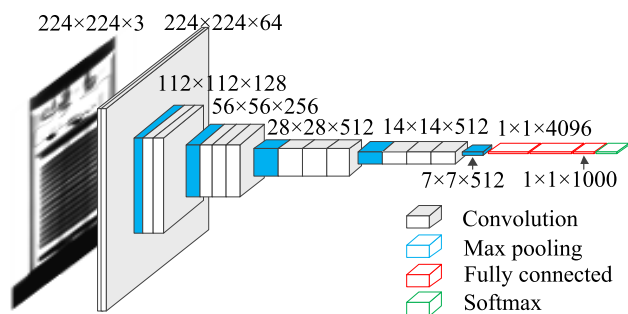**FIGURE 5.** Distribution of single molten salt battery dataset.

FIGURE 6. The VGG16 network.

is after the convolution layer, and the sampling strategy of the down-sampling layer is Max Pooling. The numbers of neurons in three FC layers are 4096, 4096, and 1000. 1000 represents the number of categories in the image data set. The VGG16 has 138,357,544 parameters. The network before the FC can be divided into 5 blocks. The number of convolutional layers and pooling layers included in bolck1, bolck2, bolck3, bolck4, and bolck5 are 2, 2, 3, 3, 3, and 1, 1, 1, 1, 1, respectively, and Block5 is also known as the bottleneck layer.

B. ESTABLISHMENT OF IDENTIFICATION MODEL FOR SINGLE MOLTEN SALT BATTERY

The learned knowledge is recognized and processed by tansfer learning through FC layer to complete the learning task under the new data set. The flow chart of migration learning is shown in Figure 7. In the VGG16 network model, on the one hand, the size of the receptive field cannot be adjusted adaptively; On the other hand, the complex parameter optimization operation of FC layer will lead to overfitting and increase training time.

The proposed network is improved based on the VGG16 model. First, an SK convolution module is connected to the

bottleneck layer. The network can automatically choose a convolution kernel of suitable size according to input features, the level of multi-scale network and the ability to identify small defects in single molten salt battery are improved; Secondly, the GAP is used to replace the fully connected layer of VGG16 model, and fine-tuning operations such as the dropout layer are added, so a simplified deep neural network model V-VGGNet is obtained; Finally, the weight parameters obtained from the pre-training learning on ImageNet are transferred to single molten salt battery image recognition model V-VGGNet, and network parameters are initialized and retrained. Replacing the FC layer with GAP is useful to speed up network convergence, reduce model parameters, and prevent overfitting caused by the optimization of considerable parameters in FC layer. The improved V-VGGNet is shown in Figure 8.

C. ADAPTIVE SELECTION OF CONVOLUTION KERNEL IN SK CONVOLUTION MODULE

Mostly the same size convolution kernel is set on each feature layer of traditional CNN. For example, multiple convolution kernels are set in GoogLeNet, but the convolution kernel cannot be adjusted adaptively according to the size of the input features, which affects the efficiency of feature extraction.

In V-VGGNet network, the SK convolution module from the SKNET network [47] is added after the bottleneck layer of V-VGGNet, which allows the V-VGGNet network to adaptively select a convolution kernel with an appropriate size through input features, thereby the efficiency of feature extraction is improved. The SK convolution module consists of three parts: separation, fuse, and selection.

1) SEPARATION

In the separation operation, multiple convolution kernels and input feature vector X perform convolution operations to form multiple branches, then perform convolution operations of
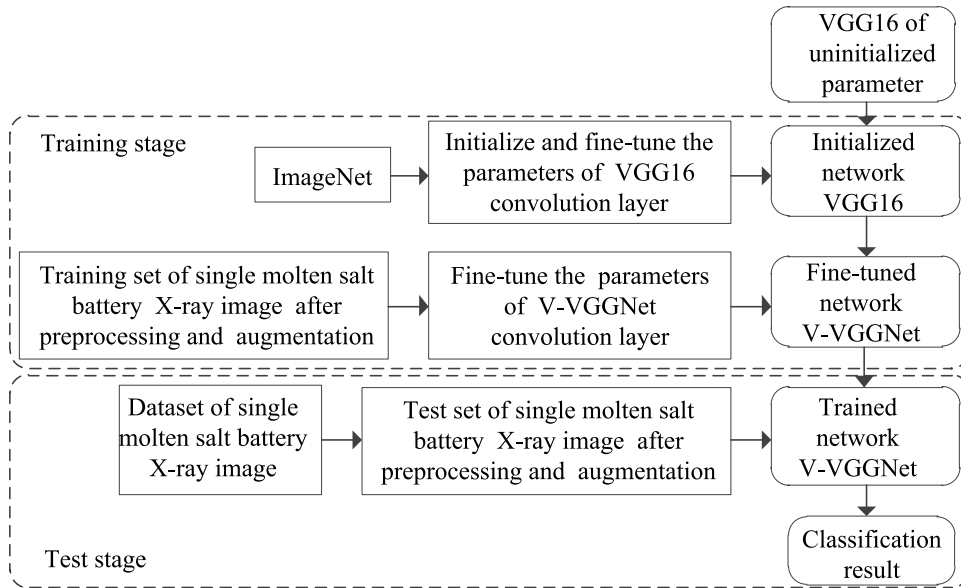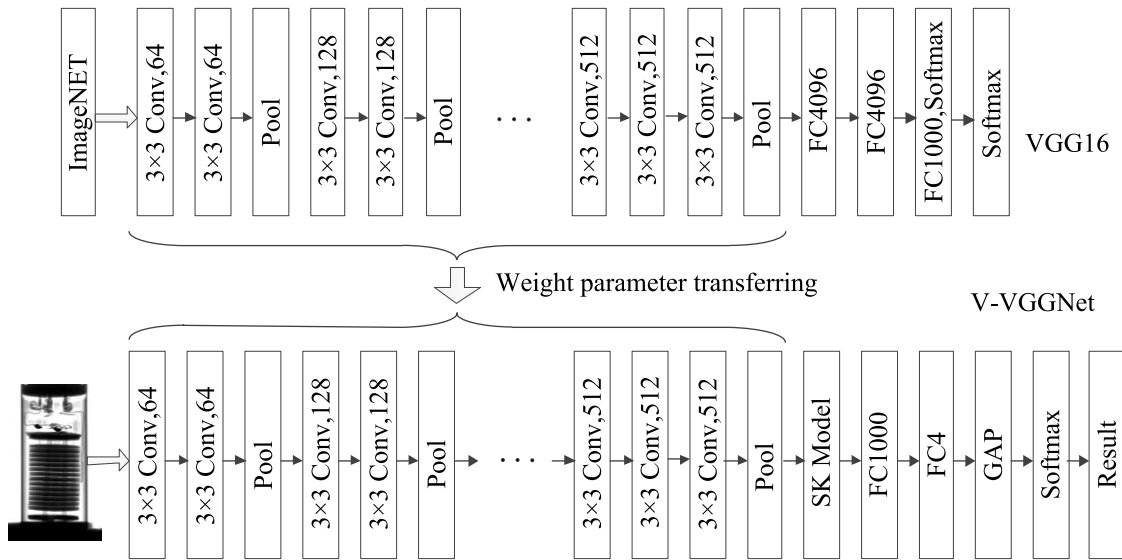
FIGURE 7. Transfer learning flow chart.

FIGURE 8. The improved network model structure V-VGGNet.

sizes 3×3 and 5×5 on the vector X, and two feature vectors $U' \in R^{H \times W \times C} U'' \in R^{H \times W \times C}$ are obtained.

The schematic diagram of the SK convolution module is shown in Figure 9.

2) FUSION

A gating mechanism is designed to distribute the information flow into the next convolutional layer. In the fusion process, the U' and U'' are obtained by the previous convolution operation, they are fused as shown in equation (1).

$$U = U' + U'' \tag{1}$$

For the fused U, the global information is compressed through GAP. As shown in Figure 9, s_c is obtained after GAP of U_c, s_c represents the c-th dimension feature of s, U_c represents the c-th dimension feature of U. As shown in equation (2).

$$s_c = f_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c(i,j) \tag{2}$$

After global average pooling, features are processed by the FC layer, and which are further compressed, the extraction efficiency is improved.
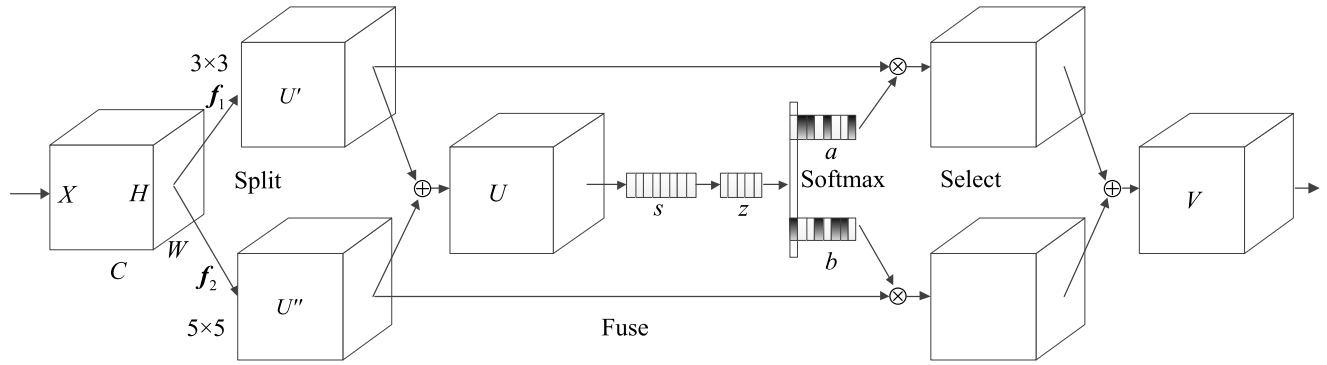
FIGURE 9. Schematic diagram of SK convolution model.

Compressed features are expressed as Equation (3).

$$z = \delta(\mathrm{B}(w_s)) \tag{3}$$

In equation (3): δ represents the function ReLU; B represents batch normalization; $z \in \mathbf{R}^{d \times 1}$, $w_s \in \mathbf{R}^{d \times c}$, the size of the parameter $\mathbf{d}$ is determined by setting the attenuation ratio $\mathbf{r}$, and its expression is expressed as Equation (4).

$$\mathbf{d} = \max(\mathrm{C}/\mathrm{r}, \mathrm{L}) \tag{4}$$

In equation (4): the size of L is set to 32.

(3) Choose. After the softmax operation, the soft attention vectors of $\mathbf{a}$ and $\mathbf{b}$ is obtained through the attention mechanism.

$$\mathbf{a}_c = \frac{\mathbf{e}^{A_c z}}{\mathbf{e}^{A_c z} + \mathbf{e}^{B_c z}} \quad \mathbf{b}_c = \frac{\mathbf{e}^{B_c z}}{\mathbf{e}^{A_c z} + \mathbf{e}^{B_c z}} \tag{5}$$

In equation (5): $A, B \in \mathbf{R}^{C \times d}$, A_c represents the c-th row of A, $\mathbf{a}_c$ represents the c-th element of $\mathbf{a}$; The meanings of B_c and $\mathbf{b}_c$ are consistent with the meanings of A and $\mathbf{a}_c$ respectively.

Through the weighted summation of the corresponding weight matrix and the convolution kernel, the output feature V is obtained.

$$V = [V_1, V_2, \cdots, V_c] \tag{6}$$

In equation (6): $V_c \in \mathbf{R}^{H \times W}$, $V_c = \mathbf{a}_c U' + \mathbf{b}_c U''$, $\mathbf{a}_c + \mathbf{b}_c = 1$.

D. AVERAGE POOLING LAYER

After feature extraction, numerous parameters are generated from the FC layer. The optimization of these parameters will generate lots of computation, prolong the convergence time of the model, and lead to overfitting. By adopting global average pooling, not only the slow model convergence problem caused by the optimization of excessive parameters is solved, but also the anti-overfitting ability of the network can be enhanced. The one-dimensional feature vector is calculated by averaging pixel values, and parameters optimization is not performed in this process, so overfitting is prevented.

Adding a global average pooling layer after the SK module in V-VGGNet can directly achieve dimensionality reduction

and massive reduction of network parameters (Actually, a FC layer has the largest proportion of parameters in CNN), classification performance of the V-VGGNet network is guaranteed, and training speed is also accelerated.

E. FULLY CONNECTED LAYER

The last part of a complete CNN often connects several FC layers, because the FC layers can map extracted features to the label space of the sample by the network. FC1 and FC2 are set before the GAP layer. Due to limited experimental samples, Dropout layer is added after FC2 for avoiding the occurrence of over-fitting, and the Dropout rate is 0.5. Meanwhile, function Relu is put on to process the problem of the gradient dispersion after each FC layer. Finally, the Softmax classifier containing 1000 neurons in the original VGG16 model is replaced by a Softmax classifier containing 4 neurons. Based on the improvement of transfer learning and the VGG16 network structure, the V-VGGNet image recognition model for assembly errors of single molten salt battery is obtained. Then, the VGG16 is pre-trained on ImageNet, learned parameters are migrated to V-VGGNet network model.

IV. EXPERIMENT AND DISCUSSION
A. EXPERIMENTAL ENVIRONMENT

All experiments are carried out 64bit Microsoft Win10 OS, Intel Core i7-9800T CPU @ 3.20 GHz, 32GB RAM, 3T hard drive, and 8G NVIDIA GeForce GTX 1080Ti GPU. Python version is 3.7. The deep learning framework is Tensorflow.

B. TRAINING STAGE

In the training process, a part of training network parameters are carried out by the SGD algorithm; the Momentum term is introduced to suppress the oscillation of SGD.

The iterative formulas are equations (7) (8).

$$\mathrm{W} = \mathrm{W} - \alpha \mathrm{dW} \tag{7}$$

$$\mathbf{b} = \mathbf{b} - \alpha \mathrm{d}\mathbf{b} \tag{8}$$

The Momentum is imported:

$$\mathbf{v}_{\mathrm{dW}} = \beta \mathbf{v}_{\mathrm{dW}} + (1 - \beta)\mathrm{dW}$$

$$v_{db} = \beta v_{db} + (1 - \beta)db$$
$$W = W - \alpha v_{dW}, \quad b = b - \alpha v \tag{9}$$

After the introduction of Momentum, the last updated value is taken into account on each update of the parameters. The learning rate is 0.001 and adjusted by exponential decay algorithm, the Momentum is 0.9, the parameter value of weight decay is 1×10^{-4}, and each batch uses 32 images for training.

The expression of the learning rate update in the exponential decay algorithm is equation (10).

$$lr = lr_0 \cdot (dr)^{(\lfloor \frac{gs}{ds} \rfloor)} \tag{10}$$

As shown in equation (10), lr represents the attenuated learning rate, dr represents the attenuation coefficient, ds represents the attenuation step size, and || represents rounding down. In the training process, the cost function adopts the cross-entropy loss function; the loss is calculated by the Softmax function. The L2 regularization term is added to the loss function, and the final expression is shown in equation (9).

$$J_\theta = -\sum_x P(x, \theta) \ln q(x, \theta) + \lambda \|\theta\|^2 \tag{11}$$

As shown in equation (11), θ represents the weight, x represents the batch of training samples, λ represents the regularization term coefficient, P represents the expected category probability, and q represents the predicted category probability.

In the training process, when the loss value becomes stable, lr is adjusted down again, until the optimal recognition model is obtained when the minimum value is reached, the final lr is 0.0005. The learning rate is initially set a little larger and then gradually reduced, to avoid the model failing to converge due to too large learning rate.

In the SK convolution module, the final set of SK convolution is determined by three important hyper parameters. n represents the number of paths that determine the choice of different convolution kernels aggregate, L represents the number of groups that control the cardinality of each path, and r represents the reduction ratio. The typical setting of the SK convolution module is expressed as SK [n, L, r], and its value is set to SK [n = 2, L = 32, r = 16].

C. DESIGN OF EXPERIMENTAL COMPARISON

To compare the impact of the effectiveness of transfer learning on classification accuracy, the impact of different networks on classification accuracy of defect recognition, the impact of improved the SK convolution module and migration learning on the time complexity of the V-VGGNet network structure, four groups of comparative experiments are designed in this paper.

The first group of comparative experiments: To test the effect of transfer learning, the parameters learned by V-VGGNet are discarded in this group of experiments, and the following two networks are compared experimentally

to verify the effect of defect recognition: V-VGGNet with parameter transferring and V-VGGNet without parameter transferring.

The second group experiments: To measure defect recognition rate on V-VGGNet network with parameter transferring, this experiment uses the following six network structures for experimental comparison: V-VGGNet, VGG16, Res-Net50, InceptionV3, MobileNetV2, and Xception.

The third group experiments: To verify the effect of improved SK convolution on defect recognition rate, this group of experiments uses the following two structures for experimental comparison: (1) The V-VGGNet network structure with the improved SK convolution module and weight parameter transferring. (2) VGG16 network with weight parameter transferring.

The fourth group experiments: To test the influence of network structure with weight parameter transferring and SK convolution module on time complexity, this group of experiments uses the following four network structures for experimental comparison: (1) The V-VGGNet network model proposed in this paper with SK convolution module and weight parameter transferring. (2) VGG16 network model with weight parameter transferring. (3) VGG16 network model without weight parameter transferring. (4) V-VGGNet network model without weight parameter transferring.

D. ANALYSIS OF RECOGNITION RATE

Precision, Recall, and the harmonic mean (F1-score) of Precision and Recall are used as one of the evaluating indicators. They are used to evaluate the performance of experimental networks (V-VGGNet, VGG16, Res-Net50, InceptionV3, MobileNetV2, and Xception). The Precision reflects the accuracy of the model. The Recall reflects the sensitivity of the model. The F1-score value range is between 0 and 1, which is used to measure the overall performance.

First, make the following settings, the TP, FP, and FN calculation methods of three parameters are shown in equations (12-14).

$$Pecision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1 - score = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{14}$$

Table 1 is the evaluation indicators of six network models under the test set. 4 categories of single molten salt battery images include Missing Negative Electrode (M-N-E), Broken Tab (B-T), Missing Current Collector (M-C-C), and Assembly Normal (Nor) image. The recognition accuracy of 4 categories of images is 95.14% (Missing Negative Electrode), 98.79% (Broken Tab), 98.21% (Missing Current Collector), and 99.41% (Assembly Normal) respectively, the overall recognition rate reached 97.91%, the V-VGGNet network's recognition accuracy is the most effective of six network models, which is nearly 3% higher than molten salt battery

TABLE 1. Accuracy, recall, and F1-score of six models.

Model	Precision(%)				Recall(%)				F1-score(%)			
	Nor	M-C-C	B-T	M-N-E	Nor	M-C-C	B-T	M-N-E	Nor	M-C-C	B-T	M-N-E
V-VGGNet	99.4	98.2	98.8	95.1	99.6	94.6	99.2	98.1	99.5	98.6	98.6	96.6
ResNet50	99.6	96.0	98.3	94.0	99.2	94.2	98.5	96.1	99.4	95.4	98.4	95.0
InceptionV3	98.8	95.1	97.8	93.5	99.0	92.4	98.7	95.2	98.9	93.8	98.3	94.3
VGG16	98.2	94.5	97.5	90.4	98.4	89.6	98.0	94.8	98.1	92.0	97.8	92.5
Xception	97.2	94.0	96.4	88.8	98.0	87.5	97.3	93.6	97.6	90.6	96.8	91.1
MobileNetV2	97.5	93.5	96.1	87.8	98.0	86.4	97.3	93.2	97.7	89.8	96.7	90.4

defect recognition method proposed by Zhao et al. From the experimental data, the recognition rate of the V-VGGNet network is higher.

E. COMPARISION OF FOUR GROUPS OF EXPERIMENTAL RESULTS WITH DIFFERENT NETWORK MODELS

The first group of experimental analysis:

As shown in Figure 10, the accuracy of the V-VGGNet network with weight parameter transferring (P-T) reached 97.91%. The accuracy rate of the V-VGGNet network model on the test set without weight parameter transferring reached 83%. Figure 10 shows the loss curve of V-VGGNet after 40 iterations, in which the loss value of the V-VGGNet network model with weight parameter transferring is about 0.15. It shows that the model has learned the defect features on the ImageNet data set, and these features can help to identify the defect images of single molten salt battery assembly errors.
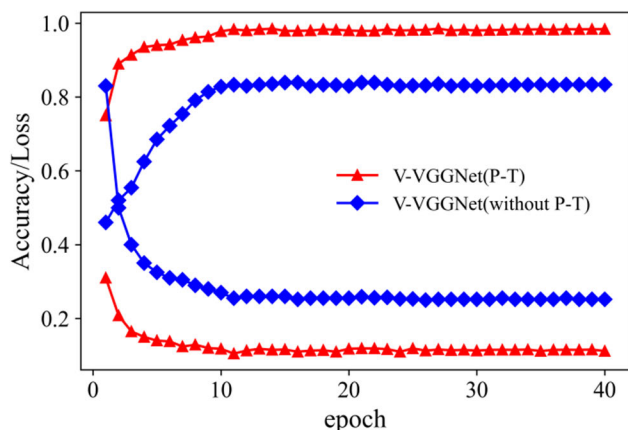

FIGURE 10. The comparative experiment of model transfer learning.

The second group of experimental analysis:

Figure 11 shows the experimental results about defect recognition rate of six different network structures. These six networks include V-VGGNet, VGG16, ResNet50, InceptionV3, MobileNetV2, and Xception. According to the accuracy curves of the training set and the validation set, the recognition effect of the V-VGGNet is the highest, the

TABLE 2. Experimental comparison of adding SK modules to six networks.

Model	Accuracy		FPS (Frame/s)
	Training set	Validation set	
V-VGGNet	97.91	97.50	44
VGG16+SK	96.67	96.31	51
ResNet50+SK	96.89	96.08	18
InceptionV3+SK	95.64	95.12	13
Xception+SK	94.73	94.17	14
MobileNetV2+SK	94.26	93.65	6

MobileNetV2 is the lowest, and the V-VGGNet is close to entirely convergence on the 15th epoch.

From Figure 11 (a), the recognition accuracy of V-VGGNet is nearly 1% higher than that of ResNet50, and nearly 4.2% higher than that of the lowest MobileNetV2. Compared with Figure 11 (b) and (d) in contrast, V-VGGNet's loss is about 0.06, which is the smallest and the first to achieve a stable state. Average recognition rate of V-VGGNet network is 97.91%, InceptionV3 network is 96.40%, ResNet50 network is 97.06%, MobileNetV2 network is 93.73%, Xception network is 94.11%, and VGG16 network is 95.21%. In addition, the V-VGGNet network model starts to converge on the 4th epoch and completely converges on the 10th epoch, which is faster than other networks and has a better overall effect.

The third group of experimental analysis:

The experimental results in Figure 12 illustrate the effect of the improved SK convolution module on the defect recognition rate, and the comparative experimental results of the V-VGGNet network model and the VGG16 network model with weight parameter transferring. The data in Figure 12 shows the improved SK convolution module is increased by about 4.5% compared with the VGG16 network with weight parameter transferring.

A series of experiments are also carried out compared with other five networks that added SK modules respectively. From Table 2, the proposed method is outperformed other
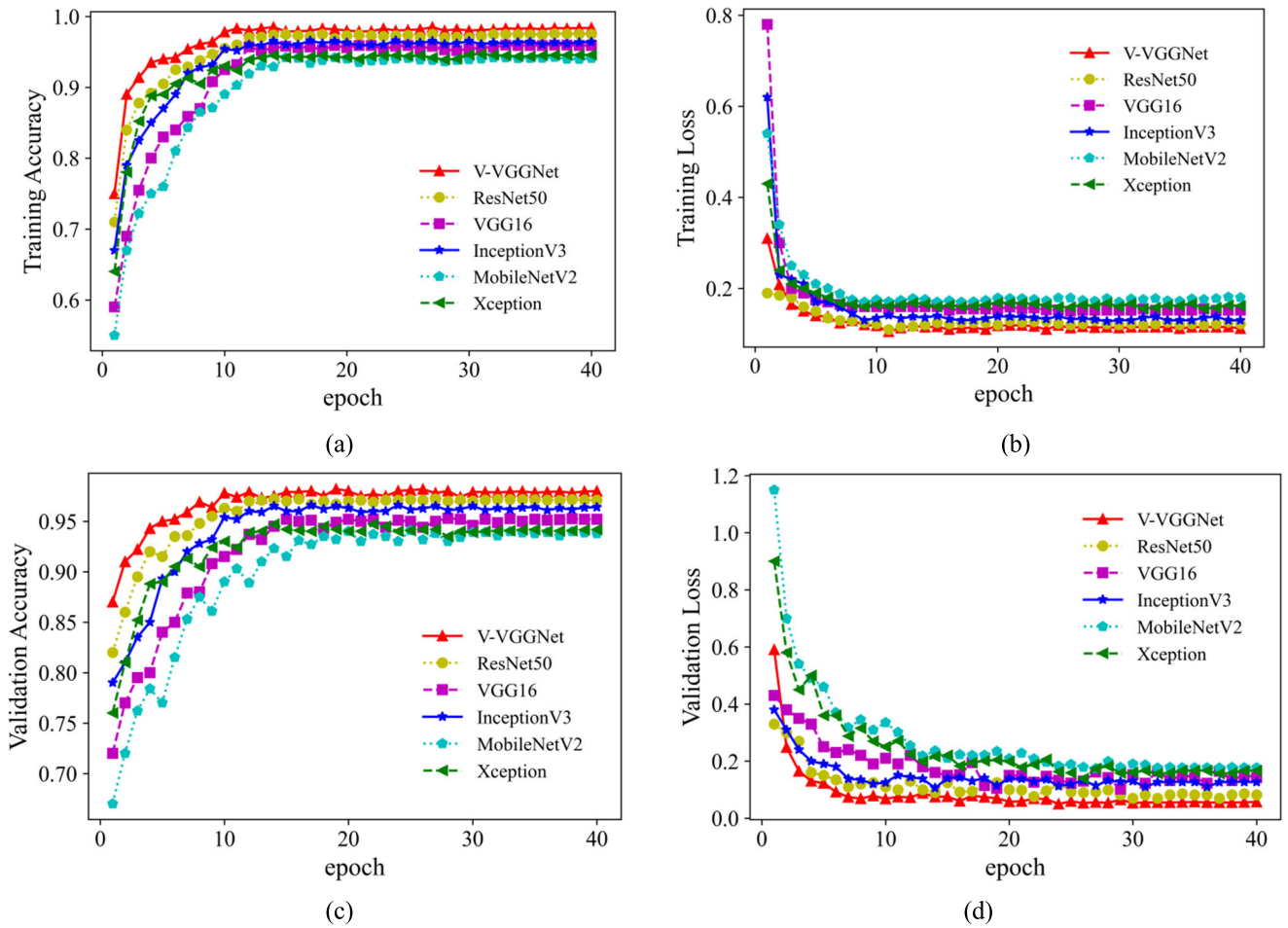
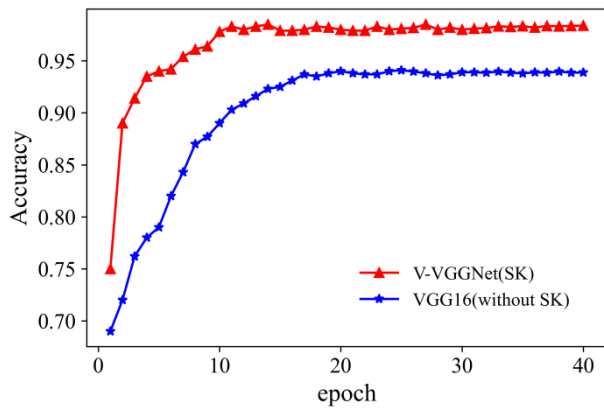**FIGURE 11.** The comparative experiment of six network models.

FIGURE 12. The comparative experiment of SK convolution module.
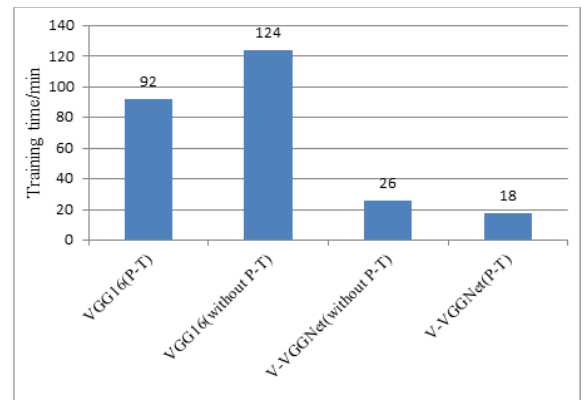

FIGURE 13. The comparative experiment of time complexity of network model.

advanced methods. Although the frame rate of the V-VGGNet is not the highest, the main purpose of the network is to improve the recognition accuracy, it is worth slow speed.

The fourth group of experimental analysis:

Figure 13 shows the time complexity results of four different networks for 40 epochs. These four networks include

(1) The V-VGGNet network model with SK convolution module and weight parameter transferring (P-T) proposed in this paper. (2) VGG16 network model with weight parameter transferring. (3) VGG16 network model without weight parameter transferring. (4)V-VGGNet network model without weight parameter transferring.
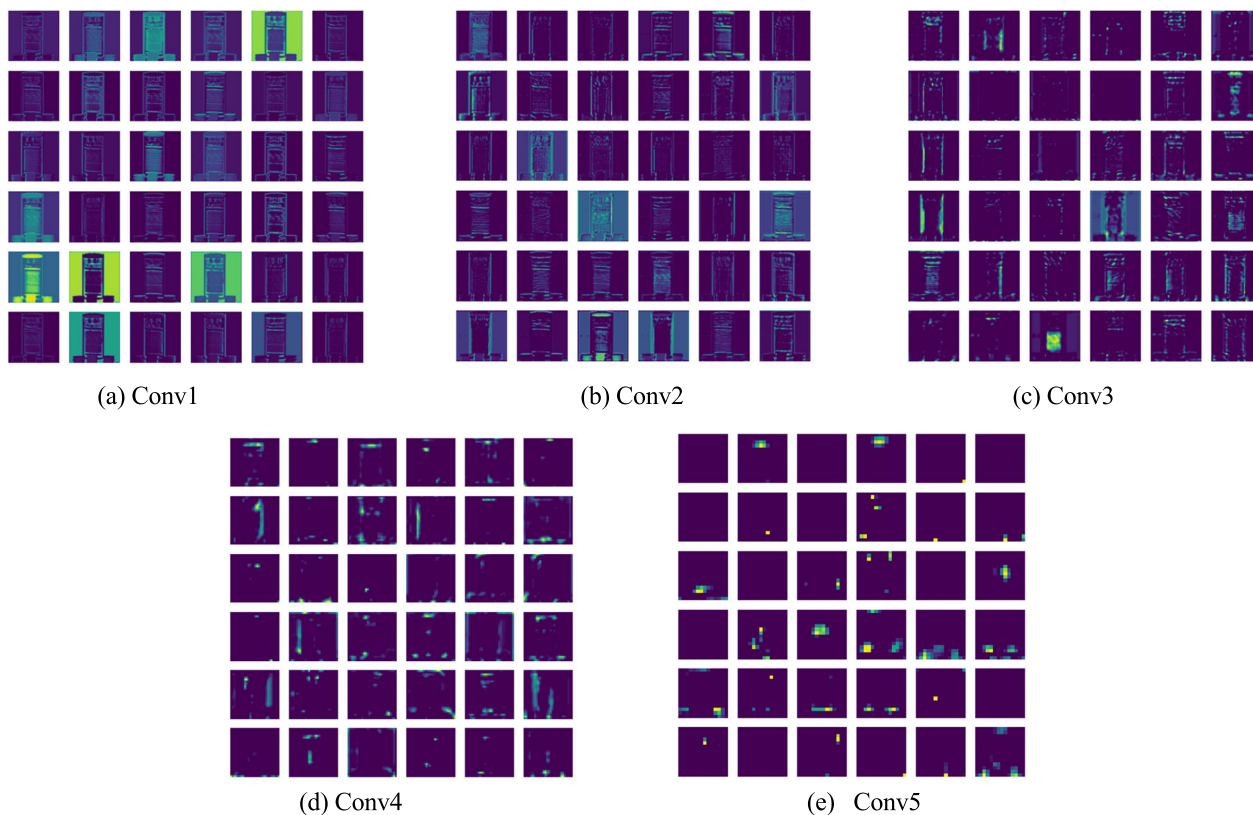
(a) Conv1   (b) Conv2   (c) Conv3

(d) Conv4   (e) Conv5

FIGURE 14. Feature maps visualization of molten salt battery image.

(a)V-VGGNet   (b) InceptionV3   (c) RESNET50

(d) MobileNetV2   (e) Xception   (f) VGG16
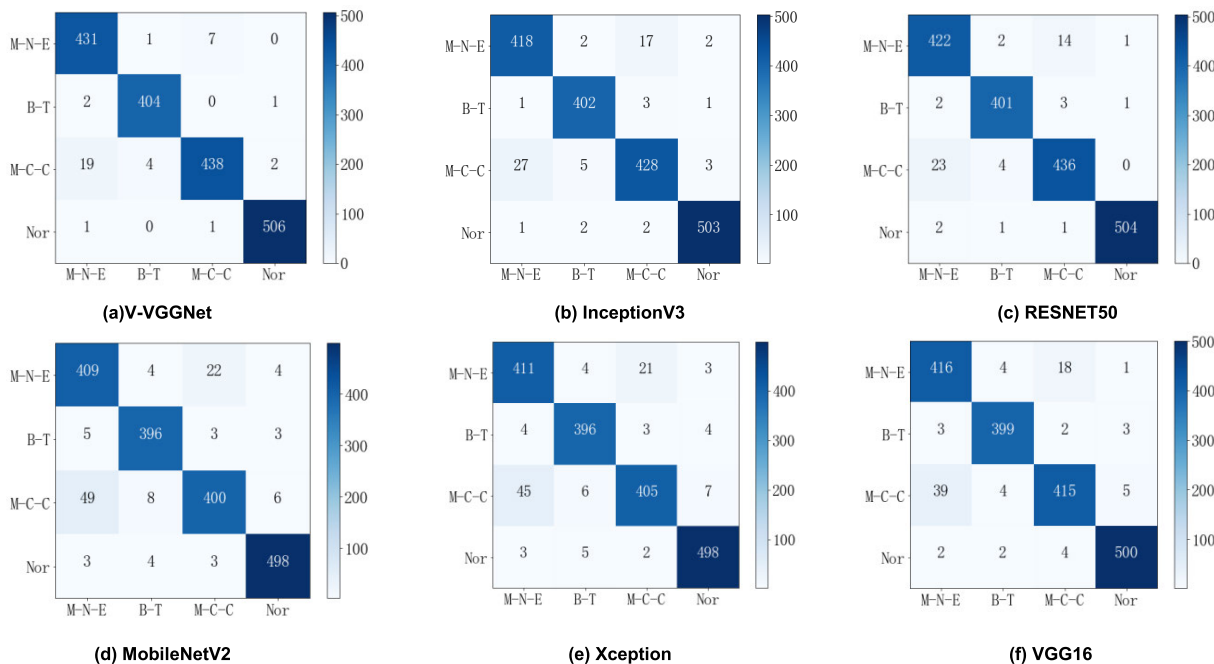
FIGURE 15. Comparison of confusion matrix of six network models.

Figure 13 shows the V-VGGNet's training time is 18 minutes, which is 74 minutes faster than the training of the VGG16 network with weight parameter transferring, the recognition speed is the fastest, and the effect is significant.

F. ANALYSIS OF MODEL FEATURE VISUALIZATION

The feature map can be regarded as the feature space of input image. Visualizing the feature map is conducive to understand internal feature representation of CNN, and it is a means to

probe the neural network black box, corresponding feature maps can be extracted and output. Through the CNN to visualize the feature map of single molten salt battery, it can be observed intuitively how the CNN transforms the image data of the single molten salt battery, so it is useful for better understanding the working principle of CNN and adjusting V-VGGNet parameters.

The V-VGGNet convolutional layer is shown in Figure 14. Conv1 and Conv2 extract the shallow layer features of single molten salt battery, the features are actually some edges and colours information. From Figure 14(c) Conv3, visual content is become less and less, more detailed contour and texture of single molten salt battery is reserved. Finally, the output feature maps of Conv4 and Conv5 become very abstract and sparse.

In the process of feature maps extraction, register_forward_pre_hook () function is used to obtain the feature graph. The parameter is a function name that needed to be implemented by itself. The parameters module, input and output are the module name, a tensor tuple input and a tensor output. Then torchvision.utils.make_grid () and torchvision.utils.save_image () are used to save the tensor containing the feature map as an image file for visualization and other operations.

There are two functions about feature map visualization: One is to improve the structure of the training network, the other is to delete redundant nodes to achieve model compression. The dead feature map location of fixed convolution layer is the same under different input data. These dead feature maps cannot provide effective information, and due to their fixed positions, corresponding convolution kernels can be removed from the network to play a role of model compression.

G. ANALYSIS OF CONFUSION MATRIX

The visualization confusion matrix of six models including V-VGGNet, Xception and so on model on validation set are illustrated in Figure 15. Accuracy of classification can be calculated, recognition results can be seen intuitively in image classification. It is composed of a matrix of n rows and n columns. The value on the diagonal represents the number of images correctly recognized by V-VGGNet. The value on off-diagonal line represents the number of images that the network recognizes incorrectly.

From Figure 15, each model has the ability to recognize three kinds of assembly error defects and Assembly Normal (Nor) images of single molten salt battery. The V-VGGNet model has the highest recognition rate for Broken Tab (B-T), and has a high misrecognition rate for the Missing Negative Electrode (M-N-E) and the Missing Current Collector(M-C-C), mainly because of these two situations: Missing Negative Electrode is misidentified as Missing Current Collector, and Missing Current Collector is misidentified as Missing Negative Electrode. Among them, the MobileNetV2 network has the highest false recognition

rate for Missing Current Collector and Missing Negative Electrode.

It is also illustrated V-VGGNet network model has the best performance, which is able to correctly identify 506 of the 508 Assembly Normal images of single molten salt battery images, and which 1piece of image is misidentified as Missing Current Collector and 1 as Missing Negative Electrode. A convenient method for the recognition of Missing Current Collector, Broken Tab, and Missing Negative Electrode assembly error defect images of single molten salt battery is constructed.

The reasons for the recognition error may be an error in the image acquisition, or it may be that the texture difference between the Missing Current Collector and the Missing Negative Electrode of single molten salt battery is not obvious, which makes the network more difficult to recognize the images. It can be seen that the wrongly identified images mostly occur between images with similar features, and the error is lower, the stability of the V-VGGNet network is further explained.

V. CONCLUSION

During inspection process of molten salt battery production line, there are problems such as time-consuming and labour-intensive manual defect detection, the insufficient ability of traditional diagnosis methods, and low classification accuracy. The deep learning method can improve the recognition rate with the increase in the number of molten salt battery samples, and the main advantage over traditional structure is that it does not require massive training data; it can fundamentally solve the current lack of large-scale public single molten salt battery datasets.

In this paper, the V-VGGNet recognition network of single molten salt battery defects is constructed; the innovations are as follows:

(1)Transfer learning and CNN are introduced in the V-VGGNet defect-recognition network, the representation with strong classification ability can be learned from a lot of weak features, which is beneficial to improve the recognition rate of the three categories of assembly defect images, the three defects of molten salt battery assembly error include Missing Negative Electrode, Broken Tab and Missing Current Collector.

(2) A diverse data set of molten salt battery is built by preprocessing and image enhancement, which is helpful to reduce the interference of environmental factors, and the learning ability of V-VGGNet is improved with the increase of the number of data sets.The VGG16 network structure is redesigned, the SK convolution module is used, it is achieved the purpose of reducing training parameters and shortening training time.

(3) For different network structures and different training strategies, performance tests and comparative experiments of different classification methods are carried out. The recognition accuracy of the V-VGGNet network is improved; it can reach 97.91%, which is nearly 3% higher than proposed

by Zhao et al. It provides a good solution for the manual detection of defects of single molten salt battery.

In the next step of research work, to further improve the recognition rate and wide applicability of the model, the image pre-processing will be further optimized, and it is necessary to construct novelty structure and improvement of activation functions and classifiers based on previous work [48]. At the same time, this model can also be applied to other types of molten salt batteries for preliminary defect detection.

ABBREVIATIONS

CNN	Convolutional Neural Network.
SIFT	Scale-Invariant Feature Transform.
FC	Fully Connected Layer.
GAP	Global Average Pooling.
VGG	Visual Geometry Group.
SK	Selective Kernel.
GPU	Graphics Processing Unit.
YOLO	You Only Look Once.
GAN	Generative Adversarial Network.
P-T	parameter transferring.
M-N-E	Missing Negative Electrode.
B-T	Broken Tab.
M-C-C	Missing Current Collector.
Nor	Assembly Normal.

REFERENCES

[1] B. Liu, Y. Huang, H. Cao, A. Song, Y. Lin, M. Wang, and X. Li, "A high-performance and environment-friendly gel polymer electrolyte for lithium ion battery based on composited lignin membrane," *J. Solid State Electrochemistry*, vol. 22, no. 3, pp. 807–816, Mar. 2018.

[2] S. Narayanan, X. Li, S. Yang, H. Kim, A. Umans, I. S. McKay, and E. N. Wang, "Thermal battery for portable climate control," *Appl. Energy*, vol. 149, pp. 104–116, Jul. 2015.

[3] P. Butler, C. Wagner, R. Guidotti, and I. Francis, "Long-life, multi-tap thermal battery development," *J. Power Sources*, vol. 136, no. 2, pp. 240–245, Oct. 2004.

[4] Y. Han, "Thermodynamic performance evaluation on a molten hydroxide direct carbon fuel cell with asymmetric anode and cathode," *Int. J. Electrochemical Sci.*, vol. 15, pp. 8849–8872, Sep. 2020.

[5] M. S. Ding, C. Krieger, and J. A. Swank, "Developing nanofoil-heated thin-film thermal battery," U.S. Army Res. Lab., ARL-TR-6664, 2013.

[6] Y. Niu, Z. Wu, J. Du, and W. Duan, "A new thermal battery for powering borehole equipment: The discharge performance of Li–Mg–B alloy/LiNO$_3$–KNO$_3$/MnO$_2$ cells at high temperatures," *J. Power Sources*, vol. 245, pp. 537–542, Jan. 2014.

[7] H. Ye, C. Strock, T. D. Xiao, P. R. Strutt, D. E. Reisner, R. A. Guidotti, and F. W. Reinhardt, "Novel design and fabrication of thermal battery cathodes using thermal spray," *MRS Proc.*, vol. 548, p. 701, Feb. 1998.

[8] R. Ren, T. Hung, and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE Trans. Cybern.*, vol. 99, no. 2, pp. 1–12, Mar. 2017.

[9] J. A. Tsanakas, D. Chrysostomou, P. N. Botsaris, and A. Gasteratos, "Fault diagnosis of photovoltaic modules through image processing and Canny edge detection on field thermographic measurements," *Int. J. Sustain. Energy*, vol. 34, no. 6, pp. 351–372, Jul. 2015.

[10] K. L. Mak, P. Peng, and K. F. C. Yiu, "Fabric defect detection using morphological filters," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1585–1592, Sep. 2009.

[11] Z. Tong, S. Xie, H. Liu, W. Zhang, C. Pei, Y. Li, Z. Chen, T. Uchimoto, and T. Takagi, "An efficient electromagnetic and thermal modelling of eddy current pulsed thermography for quantitative evaluation of blade fatigue cracks in heavy-duty gas turbines," *Mech. Syst. Signal Process.*, vol. 142, Aug. 2020, Art. no. 106781.

[12] M. Malarvel, G. Sethumadhavan, P. C. R. Bhagi, S. Kar, and S. Thangavel, "An improved version of Otsu's method for segmentation of weld defects on X-radiography images," *Optik*, vol. 34, pp. 109–118, Aug. 2017.

[13] M. Win, A. R. Bushroa, M. A. Hassan, N. M. Hilman, and A. Ide-Ektessabi, "A contrast adjustment thresholding method for surface defect detection based on mesoscopy," *IEEE Trans. Ind. Informat.*, vol. 11, no. 3, pp. 642–649, Jun. 2015.

[14] T. Kalaiselvi and P. Nagaraja, "A rapid automatic brain tumor detection method for MRI images using modified minimum error thresholding technique," *Int. J. Imag. Syst. Technol.*, vol. 25, no. 1, pp. 77–85, Mar. 2015.

[15] L. Wang, Y. Zhao, Y. Zhou, and J. Hao, "Calculation of flexible printed circuit boards (FPC) global and local defect detection based on computer vision," *Circuit World*, vol. 42, no. 2, pp. 49–54, May 2016.

[16] X. Bai, Y. Fang, W. Lin, L. Wang, and B.-F. Ju, "Saliency-based defect detection in industrial images by using phase spectrum," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2135–2145, Nov. 2014.

[17] M. Vejdannik and A. Sadr, "Automatic microstructural characterization and classification using dual tree complex wavelet-based features and bees algorithm," *Neural Comput. Appl.*, vol. 28, pp. 1–13, Jan. 2017.

[18] Y. Li, Y. L. Li, H. Luo, M. Yu, G. Jiang and H. Cong, "Fabric defect detection algorithm using RDPSO-based optimal Gabor filter," *J. Textile Inst.*, vol. 126, no. 4, pp. 1–9, 2018.

[19] S. Susan and M. Sharma, "Automatic texture defect detection using Gaussian mixture entropy modeling," *Neurocomputing*, vol. 239, pp. 232–237, May 2017.

[20] Y.-G. Cen, R.-Z. Zhao, L.-H. Cen, L.-H. Cui, Z.-J. Miao, and Z. Wei, "Defect inspection for TFT-LCD images based on the low-rank matrix reconstruction," *Neurocomputing*, vol. 149, pp. 1206–1215, Feb. 2015.

[21] Alaknanda, R. S. Anand, and P. Kumar, "Flaw detection in radiographic weld images using morphological approach," *NDT E Int.*, vol. 39, no. 1, pp. 29–33, Jan. 2006.

[22] M. K. Felisberto, H. S. Lopes, T. M. Centeno, and L. V. R. de Arruda, "An object detection and recognition system for weld bead extraction from digital radiographs," *Comput. Vis. Image Understand.*, vol. 102, no. 3, pp. 238–249, Jun. 2006.

[23] R. Ranjan, A. R. Khan, C. Parikh, R. Jain, R. P. Mahto, S. Pal, S. K. Pal, and D. Chakravarty, "Classification and identification of surface defects in friction stir welding: An image processing approach," *J. Manuf. Processes*, vol. 22, pp. 237–253, Apr. 2016.

[24] S. E. Florence, R. V. Samsingh, and V. Babureddy, "Artificial intelligence-based defect classification for weld joints," *IOP Conf. Mater. Sci. Eng.*, vol. 402, no. 1, pp. 1–13, 2018.

[25] R. H. F. Murta, F. D. A. Vieira, V. O. Santos, and E. P. de Moura, "Welding defect classification from simulated ultrasonic signals," *J. Nondestruct. Eval.*, vol. 37, no. 3, pp. 1–10, Sep. 2018.

[26] H. K. Jung and G. Park, "Rapid and non-invasive surface crack detection for pressed-panel products based on online image processing," *Structural Health Monitor.*, vol. 18, nos. 5–6, pp. 1928–1942, Nov. 2019.

[27] K. K. Kieselbach, M. Nöthen, and H. Heuer, "Development of a visual inspection system and the corresponding algorithm for the detection and subsequent classification of paint defects on car bodies in the automotive industry," *J. Coatings Technol. Res.*, vol. 16, no. 4, pp. 1033–1042, Jul. 2019.

[28] N. B. Yahia, T. Belhadj, S. Brag, and A. Zghal, "Automatic detection of welding defects using radiography with a neural approach," *Proc. Eng.*, vol. 10, pp. 671–679, 2011.

[29] D. Bacioiu, G. Melton, M. Papaelias, and R. Shaw, "Automated defect classification of aluminium 5083 TIG welding using HDR camera and neural networks," *J. Manuf. Processes*, vol. 45, pp. 603–613, Sep. 2019.

[30] H. Chen, Q. Hu, B. Zhai, H. Chen, and K. Liu, "A robust weakly supervised learning of deep conv-nets for surface defect inspection," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11229–11244, Aug. 2020.

[31] M. Kim, M. Lee, M. An, and H. Lee, "Effective automatic defect classification process based on CNN with stacking ensemble model for TFT-LCD panel," *J. Intell. Manuf.*, vol. 31, no. 5, pp. 1165–1174, Jun. 2020.

[32] H. Kasban, O. Zahran, H. Arafa, M. El-Kordy, S. M. S. Elaraby, and F. E. Abd El-Samie, "Welding defect detection from radiography images with a cepstral approach," *NDT E Int.*, vol. 44, no. 2, pp. 226–231, Mar. 2011.

[33] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.

[34] K. He, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.

[36] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[37] M. Xiao, B. Yang, S. Wang, Z. Zhang, X. Tang, and L. Kang, "A feature fusion enhanced multiscale CNN with attention mechanism for spot-welding surface appearance recognition," *Comput. Ind.*, vol. 135, Feb. 2022, Art. no. 103583.

[38] S. Plakias and Y. S. Boutalis, "Fault detection and identification of rolling element bearings with attentive dense CNN," *Neurocomputing*, vol. 405, pp. 208–217, Sep. 2020.

[39] X. Wei, D. Wei, D. Suo, L. Jia, and Y. Li, "Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model," *IEEE Access*, vol. 8, pp. 61973–61988, 2020.

[40] W. Jiang, W. Jiang, L. Wu, S. Liu, and M. Liu, "CNN-based two-stage cell segmentation improves plant cell tracking," *Pattern Recognit. Lett.*, vol. 128, pp. 311–317, Dec. 2019.

[41] J. Yu, Z. Shen, and S. Wang, "Wafer map defect recognition based on deep transfer learning-based densely connected convolutional network and deep forest," *Eng. Appl. Artif. Intell.*, vol. 105, Oct. 2021, Art. no. 104387.

[42] J. Yu, Z. Shen, and X. Zheng, "Joint feature and label adversarial network for wafer map defect recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 1341–1353, Jul. 2021.

[43] L. Yj et al., "Real-time defect detection of hot rolling steel bar based on convolution neural network," *Chin. J. Sci. Instrum.*, vol. 42, no. 12, pp. 211–219, 2021.

[44] G. JQ et al., "Lithium battery defect detection method based on improved YOLOv4," *Electron. Meas. Technol.*, vol. 45, no. 15, pp. 144–150, 2022.

[45] D. F. Chen, K. Yan, and J. X. Xiong, "Detection method of workpiece surface defects based on improved YOLOv4," *J. Comput. Appl.*, vol. 42, no. 1, pp. 94–99, 2022.

[46] T. Zhao, "Research on defects detecting method in thermal battery based on local contrast enhancement algorithm," *Laser J.*, vol. 43, no. 1, pp. 24–28, 2022.

[47] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 510–519.

[48] W. Xu, S. Zhang, F. Bai, and T. Zhao, "Research on improved residual network classification method for defect recognition of thermal battery," *IEEE Access*, vol. 10, pp. 113234–113248, 2022.

SIXIANG ZHANG received the B.Sc. degree from Tianjin University, in 1982, the M.Sc. degree from Zhejiang University, in 1993, and the Ph.D. degree from Tianjin University, in 1996. He is currently a Professor and a Ph.D. Supervisor with the Hebei University of Technology. His main research interests include the theory and technology of mechanical geometric measurement and computer image processing technology.

YITAO HAN is currently pursuing the bachelor's degree with the Tianjin University of Commerce, China. His main research interests include pattern recognition and deep learning.
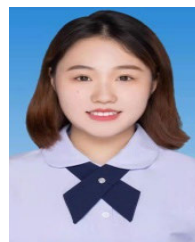

YIN JIANG is currently pursuing the bachelor's degree with the Tianjin University of Commerce, China. Her main research interests include pattern recognition and deep learning.

WENCHAO XU received the M.I.S. degree from the Hebei University of Technology, China, in 2009. He is currently an Associate Professor with the College of Information Engineering, Tianjin University of Commerce, China. His research interests include pattern recognition, intelligent control, and deep learning.

WENBO XU received the M.I.S. degree from the Tianjin University of Technology and Education, China, in 2017. He is currently a Lecturer with the College of Higher Vocational Education, Tianjin University of Technology and Education. His research interests include pattern recognition, intelligent control, and reducer design.

HENGYI ZHAO is currently pursuing the bachelor's degree with the Tianjin University of Commerce, China. Her main research interests include pattern recognition and deep learning.

• • •