Computing and Informatics, Vol. 43, 2024, 126–148, doi: 10.31577/cai_2024_1_126

# RESEARCH ON DISCRIMINATION METHOD OF CARBON DEPOSIT DEGREE OF AUTOMOBILE ENGINE BASED ON DEEP LEARNING

Qian Huang, Youxing Chen*, Zhaoba Wang, Xiang Chen

*School of Information and Communication Engineering*
*North University of China*
*Taiyuan, 030051, China*

*e-mail:* chenyouxing@nuc.edu.cn

Asoke Nandi, Maozhen Li

*Brunel University London*
*Uxbridge, UB8 3PH, UK*

Yong Jin

*School of Information and Communication Engineering*
*North University of China*
*Taiyuan, 030051, China*

**Abstract.** The detection of carbon deposit degree is of great significance to the maintenance of automobile engine. Due to issues with poor feature aggregation, inter-class similarity, and intra-class variance in carbon deposit data with a small number of samples, model-based discriminative approaches cannot be widely implemented in the market. In order to overcome this technical barrier, the article examines the impact of DCNNs (Deep Convolutional Neural Networks) level on the recognition effect of the degree of carbon deposit, introduces a dropout structure and data enhancement strategy to lower the risk of overfitting brought on by the small

* Corresponding author

dataset, and suggests a recognition method based on the kernel of dual-dimensional multiscale-multifrequency information features to enhance the differentiation characteristic. After experimental testing, the accuracy of this method is 86.9 %, the F1-score is 87.2 %, and the inference speed is 190 FPS, which can meet the practical requirements and provide basic support for the large-scale promotion of the model discrimination.

**Keywords:** Determination of carbon deposit degree, small datasets, fine-grained images, feature enhancement, deep learning

## 1 INTRODUCTION

Carbon deposit in automobile engines are oxides produced by incomplete combustion of fuels and lubricants under high temperature and high-pressure conditions, mainly distributed in several parts of the engine interior. Severe engine carbon deposit will affect the intake efficiency of the intake valve, the degree of atomization of the oil nozzle, and the ignition efficiency of the sparking plug, resulting in a decrease in engine power, incomplete gasoline combustion, and misfiring of the engine. It may even result in engine scrap if the carbon deposit is not cleaned for a long time. The degree of carbon accumulation plays an important role in the cleaning of the engine, therefore it is crucial to research a method for detecting the degree of carbon deposit in automobile engines.

Physical quantity-based and vision-based detection can be applied to assess the degree of carbon deposit in an automobile engine. Direct and indirect measurements are the two primary types of physical quantity-based detection. Direct measurement utilizes a sensor device to collect physical data, such as the thickness of the carbon layer [1] and optical signals on the carbon surface [2], to quantify the degree of carbon deposit. Indirect measurements combine physical quantities such as ignition failure rate [3], oil passage temperature [4], and intake manifold differential pressure [5] with corresponding thresholds to discriminate the degree of carbon deposit. Physical information discrimination can be interfered with in advance to prevent the occurrence of faults. However, the detection method based on physical quantities necessitates gathering driving state data and multiple measurements to ascertain the state of carbon deposit, which is difficult to promote in practical applications. Vision-based detection uses the image collected by the endoscope to discriminate the degree of carbon deposit, including manual and model discrimination. Yan et al. [6] artificially distinguish the degree of carbon deposit by imaging display, which is more intuitive and convenient than the physical detection method. Manual discrimination, however, is ineffective and is significantly influenced by the environmental factors. Xiao et al. [7] use the model discrimination method to differentiate the degree of carbon deposit. The process begins with the extraction of image features using machine learning, followed by a similarity comparison with the

pre-discriminative model trained after a large number of samples, finally, the discriminative result of the degree of carbon deposit was obtained. The model-based discrimination method uses a large amount of data to form a quantitative criterion, which is more scientific and credible than manual discrimination. Traditional machine learning techniques, however, require the combination of several feature extraction algorithms and various features in order to categorize the data, therefore their discriminative impact is subpar for the multifarious feature representations of carbon deposit images. Some sample images of carbon deposit in the combustion chamber of an automobile engine captured by an endoscope are shown in Figure 1. The boundary of determining the level of carbon deposit in the image has a wide range of fuzzy distribution, and is subject to interference from oil stains and bright spots. Highly similar subordinate classes result in small inter-class variance, while large intra-class variance due to interfering factors such as light, angle, and depth of field, posing a great challenge to the detection of the degree of carbon deposit.

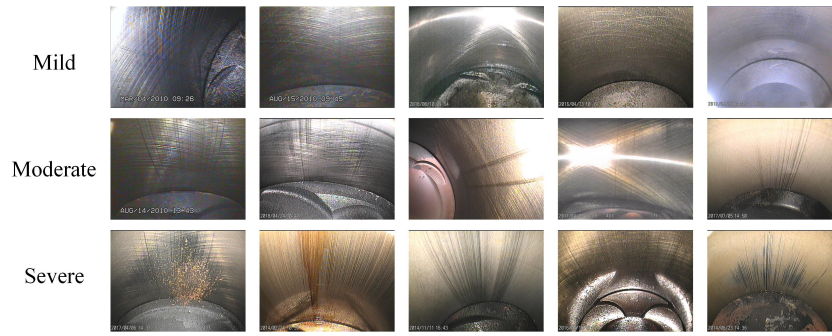



Figure 1. Sample images of carbon deposit. Categorized into three degrees: mild, moderate, and severe.

The deep learning-based method automatically learns high-dimensional abstractions from images through neural networks, which reduces the composition of feature engineering compared to traditional machine learning methods and performs well in the similar task of discriminating the degree of carbon deposit in automobile engines. Yao et al. [8] achieved excellent accuracy in classifying a small and unbalanced skin lesion dataset based on RegNetY network combined with regularization operations, a modified data enhancement strategy, a multi-weighted loss function, and an end-to-end learning strategy, and its discriminative results comparable to or even better than those of a dermatologist's diagnosis. Fan et al. [9] designed a deep residual network to obtain the dust concentration in the area of PV panels to address the problem of determining the degree of dust accumulation in PV panels. This method can effectively obtain the non-uniform two-dimensional surface feature point set, which provides theoretical support for the intelligent operation and maintenance of PV systems. Therefore, using deep learning to the job of discriminating between carbon degrees is advantageous to provide high-quality, highly discriminative fea-

tures, and thereby resolving the issue of challenging carbon degree discrimination and poor discrimination accuracy.

Since 2012, deep neural networks have expanded quickly because of improvements in computer power and the creation of enormous datasets. Researchers have developed and trained several traditional convolutional neural networks using the big dataset ImageNet, including VGGNet [10], GoogLeNet [11], ResNet [12], DenseNet [13], EfficientNet [14], and so on. The challenge of differentiating the degree of carbon deposit can be solved partly by selecting an appropriate backbone network for small datasets of carbon deposit images. In general, networks with more parameters and a complex structure can extract deeper features, but some few-shot tasks [8, 15, 16] demonstrate that modest networks outperform complex networks. We validate this argument through experimental comparisons in the carbon deposit dataset and select the best model suitable for discriminating the degree of carbon deposit in automobile engines. However, the model with strong universality is not targeted in the face of complex and variable carbon deposit images. By analyzing the distinguishing characteristics of the carbon deposit images in Figure 1, the cylinder wall surface scratches attached to the carbon deposit are highly distinguishable. Therefore, we enhanced the expression of features from both spatial and channel dimensions. The spatial dimension fuses the image's multi-scale information to consider the carbon deposit image's wide range of features and fine-grained characteristics. The channel dimension dynamically selects the feature map through the channel attention mechanism. It adaptively weights the features according to the importance of the input features, compensating for the problem of missing geometric features in the spatial dimension. Fewer carbon deposit data samples make the model prone to overfitting in the learning stage, which leads to poor generalization performance of the model. In order to increase data diversity and strengthen the generalization capabilities of the model, this work uses a variety of offline data enhancement techniques, including flipping and cropping and so on, as well as online data enhancement techniques, CutMix [17]. Literature [8] demonstrates that dropout is also a successful strategy for dealing with overfitting. To increase the model's robustness, randomly dropout high-dimensional features before the fully connected layer. What is more, the model's inference speed is just as crucial for the extension of the model discrimination approach as the model's generalization capabilities. As a result, the model design used in this research keeps its lightweight properties.

This paper uses the image of an automobile engine's combustion chamber as the research object. An improved DCNN-based model is proposed to automatically identify the degree of carbon deposit and the recommendations for the subsequent removal of engine carbon deposit are offered. The key contributions of this paper are as follows:

1. Deep learning-based strategy is used to determine the degree of carbon deposit in an automotive engine. We designed a model for carbon deposit detection and achieved excellent results.

2. Deep convolutional neural networks with various structures and their different variants are compared in this paper. The results demonstrate that the shallow network outperforms the deep network in the task of carbon deposit determination, however, the inference speed of the model is not entirely associated with its computational and parameter counts.
3. A lightweight Multi-Scale and Frequency Feature Extraction (MS & FFE) module is proposed to enhance the feature expression of carbon deposit images in this paper. The MS & FFE module is designed for images with a wide range of features and fine-grained characteristics to increase inter-class differences and decrease intra-class disparities.

## 2 METHODOLOGY

The technique design flowchart for this study is presented in Figure 2. The data is preprocessed first, and then the structure correction and parameter adjustment of DCNN are carried out according to the characteristics of carbon deposition. The test data apply to the adjusted model for degree discrimination, and finally, the confusion matrix generates to evaluate the model performance from multiple evaluation indexes.

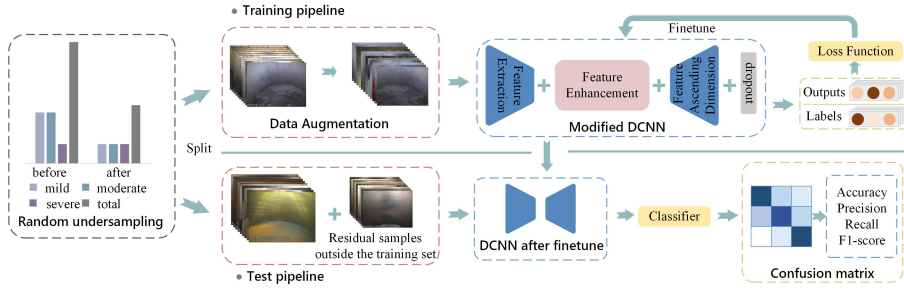



Figure 2. Flowchart of the method design. The upper part is the training, and the lower is the testing.

### 2.1 Dataset Introduction and Preprocessing

This paper mainly focuses on the degree of carbon deposit in the combustion chamber of an automobile engine, and a cooperative enterprise provides the carbon deposit data sets used. Table 1 introduces the category characterization of the carbon deposit dataset and the corresponding maintenance strategy.

Professionals label the dataset used for model training to ensure the reliability of the samples. Still, the uneven distribution of categories in the original dataset leads to bias in the model's training process, i.e., the prediction accuracy of the

| Label | Class | Training Set | Test Set | Characteristic Description | Maintenance Strategy |
|---|---|---|---|---|---|
| 0 | Mild | 8 800 | 900 | Cylinder wall surfaces are lightly covered with carbon deposit and have minor scratches. | Simple cleaning-based maintenance. The cleaning strategy mainly consists of caring for the external surfaces of the parts and cleaning steps on the inside and outside of each assembly and part, which is used to prevent corrosion of the pieces, reduce wear and tear of the elements and reduce fuel consumption. |
| 1 | Moderate | 8 800 | 900 | Cylinder wall surfaces are covered in an increased area with carbon deposit and had visible scratches. | Regular upkeep emphasizes cleaning and repair. Mainly include changing oil and filter components, replacing lubricants with the proper quality grade, and keeping the crankcase well-ventilated. |
| 2 | Severe | 8 800 | 110 | The cylinder wall surfaces adhesion layer is thick with carbon deposit, and scratches cover a wide area with almost no metal luster exposure. | Overhaul strategy based on clean repairs. The overhaul strategy includes replacing aged and worn parts and disassembling and cleaning steps. |

Table 1. Data set description. It includes the size of the data volume of the category training set and test set, the category characterization, and the corresponding maintenance strategy.

class types with a large amount of data is biased high. The prediction accuracy of the categories with a small amount of data is a little low. Random undersampling is performed on the original dataset to ensure a balanced number of classes. The data from the divided training set enlarge using a data enhancement approach to increase the model's generalization performance and robustness because too little raw data volume will result in the overfitting of the model. Table 2 lists the image transformations involved in this paper. a-f are the funda-

mental geometric transformations, which increase the number of images, the distribution of these images is consistent with the original images. To a certain extent, the geometric distortion of the images caused by Angle, perspective relationship, shooting and other reasons is eliminated. In contrast, g-r are the methods of processing pixels directly. The changes are more prominent than the geometric transformation. The geometrically modified extended image becomes increasingly complex through pixel processing, which increases the model's capacity for generalization. Overly intricate changes, however, might render the picture abstract and significantly alter its distribution. As a result, in this study, we use pixel processing modifications to pick no more than five images at a time at random.

| Number | Operation Name | Description | Operation Probability or Range |
|---|---|---|---|
| a | Flip | Perform horizontal/vertical flipping of with random probability. | 0.5/0.2 |
| b | Crop | Crop the image by a random portion of its height/width. | $(0, 0.1)$ |
| c | Scale | Randomly scales the height/width of the image to different degrees. | $(0.8, 1.2)$ |
| d | Translate | Percent Randomly make translations of different proportions to the x- and y-axes of the image. | $(-0.2, 0.2)$ |
| e | Rotate | Rotate the image by a random angle. | $(-45, 45)$ |
| f | Shear | Clip the image to a random degree. | $(-16, 16)$ |
| g | Superpixels | Take N random superpixel values for the image and replace the original image with a random probability. | $(20, 200)$, $(0, 1.0)$ |
| h | Blur | Blur the image randomly using different kernel size blurring operations (Gaussian blur, mean blur, and median blur selected randomly). | $(0, 3.0), (2, 7)$, $(3, 11)$ |
| i | Sharpen | The image sharpens using a random degree of sharpening, with an arbitrary range of sharpening. | $(0.75, 1.5)$, $(0, 1.0)$ |
| j | Emboss | The image enhances using a randomized degree of relief, a randomized range of enhancement of the image. | $(0, 2.0)$, $(0, 1.0)$ |
| k | Edge Detect | Mark image edges in a black and white image to superimpose on the original image with random probability. | $(0, 0.7)$ |
| l | Additive Gaussian Noise | Adds Gaussian noise to the image, randomly sampled by channel and pixel. | $(0, 0.05 * 255)$ |
| m | Dropout | Randomly removes some pixels, i.e., sets some pixels to black. | $(0.01, 0.1)$ |
| n | Invert | Inverts the channels of each image with a set probability. | 0.05 |
| o | Dithering | Performs operations on the pixel values of each image channel with a probability of 0.5, including randomly adding pixel values, changing the brightness of the image, changing the contrast. | $(-10, 10)$, $(0.5, 1.5)$, $(0.5, 2.0)$ |
| p | Grayscale | Convert the image to grayscale, superimposed on the original image with random probability. | $(0, 1.0)$ |
| q | Elastic Transformation | Randomly shifting image pixels under a distortion field with a fixed intensity of 0.25. | $(0.5, 3.5)$ |
| r | Piecewise Affine | Distort images with random intensity. | $(0.01, 0.05)$ |

Table 2. List of all image transformations applied in the data enhancement strategy

In order to test the generalization and practicality of the model, the test set is supplemented with the remaining samples after random undersampling. The

test set is sampled independently and identically from the true distribution of the samples, ensuring that the test set is mutually exclusive with the training set and has not been used in the training set. The flow of data processing is shown in Figure 2 and the distribution of data volume after data processing is given in Table 1.

### 2.2 DCNN Models

Multiple network structures and models with different capacity sizes were comparatively analyzed, as shown in Table 3. Among these, VGG [10] extracts feature through stacked convolutional layers, which have a more straightforward structure and uses a smaller convolutional kernel to reduce the model complexity. ResNet [12] uses residual structure to address the issue of network deterioration, improving accuracy while significantly reducing computation. DenseNet [13] takes shortcut connectivity to the extreme and enhances feature reuse to improve model performance metrics. EfficientNet [14] uses neural architecture search principles to design the underlying network architecture, and extracts features by mobile inverted bottleneck convolutional blocks to make the model lighter and more efficient. Convolutional kernels of various sizes are employed by the Inception structure introduced in GoogLeNet [11] to extract multi-scale features and augment feature information. The pre-trained weights of the model on the large dataset ImageNet is loaded to DCNNs by the transfer learning method. The number of output classes of the modified model matched the discriminant categories of carbon deposition degree. The performance of several DCNN models is contrasted to determine which model is best appropriate for carbon degree discrimination in automobile engines. And make model selection recommendations for similar projects using the carbon deposit degree discriminating challenge. ResNet-18 will be used for the remaining work in this paper because the experimental results in Table 5 demonstrate that it achieves the highest accuracy in the carbon deposit test set and has the excellent inference speed performance, both of which meet the design requirements of the carbon deposit degree discriminative model.

As shown in Figure 3, the training accuracy of the models is ideal in the comparative experiments. However, there is a wide discrepancy in the test set. The model is overfitting due to excessive redundant parameters, which reduces its capacity to generalize to the test set. In order to avoid overfitting, a three-layer structure is created in this research and placed in front of ResNet-18's fully connected layer, as illustrated in Figure 4. In this paper, two lightweight convolution layers are added in front of a single dropout layer. Specifically, where Depthwise Separable Convolution is used to reduce the number of parameters in the feature transfer process, Group Convolution maps the input features to a higher dimensional feature space. The feature distribution disperses by boosting the feature dimensions to discard more redundant information. The setting of the Dropout parameter $p$ will be discussed in the experimental section.

| Model | Flops [B] | Params [M] | Model | Flops [B] | Params [M] |
|---|---|---|---|---|---|
| VGG-11 | 7.6 | 132.9 | DenseNet-161 | 7.8 | 28.9 |
| VGG-13 | 11.3 | 133.0 | EfficientNet-B0 | 0.39 | 5.3 |
| VGG-16 | 15.5 | 138.3 | EfficientNet-B1 | 0.7 | 7.8 |
| VGG-19 | 19.6 | 143.7 | EfficientNet-B2 | 1.0 | 9.2 |
| ResNet-18 | 1.8 | 11.7 | EfficientNet-B3 | 1.8 | 12 |
| ResNet-34 | 3.6 | 21.8 | EfficientNet-B4 | 4.2 | 19 |
| ResNet-50 | 3.8 | 25.6 | EfficientNet-B5 | 9.9 | 30 |
| ResNet-101 | 7.6 | 44.6 | EfficientNet-B6 | 19 | 43 |
| ResNet-152 | 11.3 | 60.2 | EfficientNet-B7 | 37 | 66 |
| DenseNet-121 | 2.9 | 8.0 | GoogLeNet | 1.6 | 6.2 |
| DenseNet-169 | 3.4 | 14.2 | Inception-v3 | 5.0 | 22.3 |
| DenseNet-201 | 4.3 | 20.0 | Inception-v4 | 6.2 | 41.3 |

Table 3. Compare the list of models, containing five structures

### 2.3 Feature Enhancement Module

The MS & FFE module is used to solve the poor aggregation of carbon image features and the fine-grained characteristics of minor inter-class differences and significant intra-class differences, as shown in Figure 5.

The spatial dimension enhances the control of features with different distribution ranges by fusing the multi-scale characteristics of the image, which effectively improves the correlation of the location between the parts. Multi-scale feature extraction can be accomplished by parallelizing several Dilation Convolutions with varied dilation rates since a Dilation Convolution may produce a broader receptive field and handle the wide range of feature distributions in the carbon deposit picture. However, Dilation Convolution-based models are huge and ineffective [18, 19, 20, 21], which results in ineffective model prediction. Therefore, inspired by the feature extraction module in ESPNetv2 [22], we firstly map the input features from the high-dimensional space to the low-dimensional space by Group Convolution, and secondly, we use Depthwise Separable Dilated Convolution (DSDC) instead of the Dilated Convolution with high computational overhead in this paper. Multi-scale feature fusion is realized by Hierarchical Feature Fusion (HFF), and it concatenates in the channel dimension to recover the feature dimension space. The role of HFF is to solve the gridding effect of dilation convolution. The node output function $X_n$ of HFF defines as Equation (1).

$$X_n(x) = N\left(D_n(x), R_n(x)\right) \tag{1}$$

$x$ is the output of the group convolution in front of the DSDC, $n$ is the sequence number of parallel DSDC, $N$ is the aggregation node, $D_n(x)$ is the output of the $n^{\text{th}}$

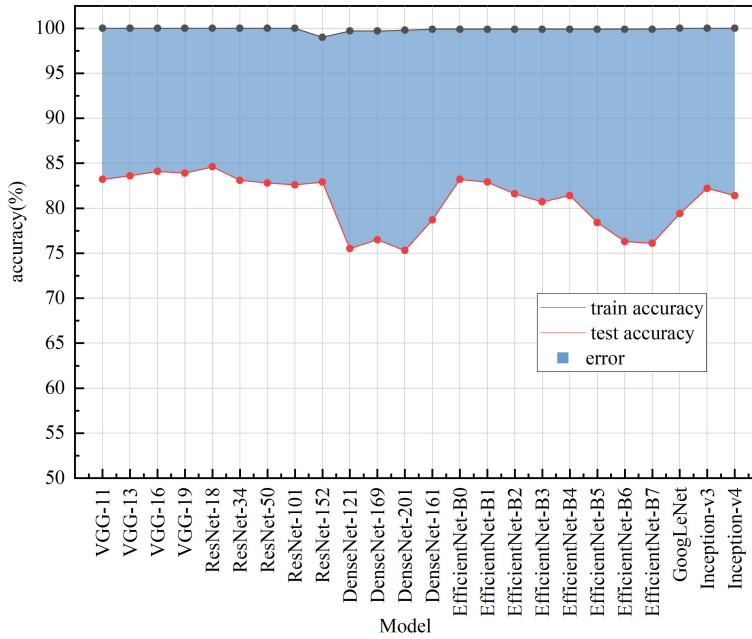



Figure 3. Numerical comparison plot of training accuracy and test accuracy, there is a large error gap

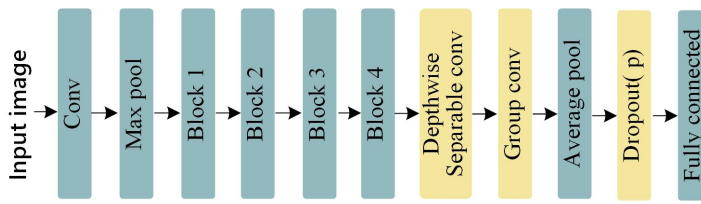



Figure 4. Modified ResNet-18 architecture

DSDC, and $R_n(x)$ is defined as in Equation (2).

$$R_n(x) = \begin{cases} 0, & n = 1, \\ D_{n-1}(x) + R_{n-1}(x), & \text{otherwise.} \end{cases} \tag{2}$$

The feature output of the spatial dimension is denoted as:

$$F_{\text{spatial}}(x) = concat\left(X_1, X_2, X_3, X_4\right). \tag{3}$$

The channel dimension utilizes the channel attention mechanism to focus on important information and compensate for the loss of feature structure due to the

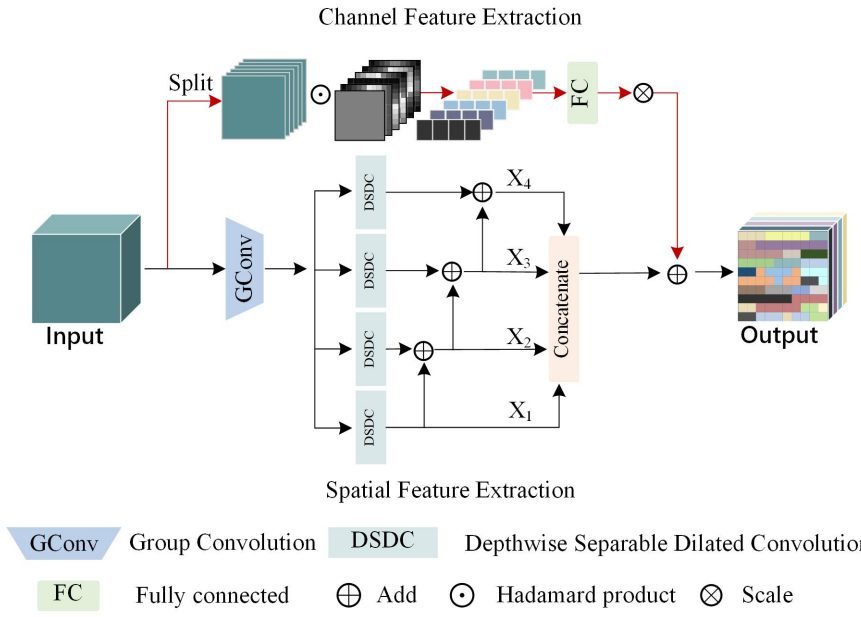



Figure 5. MS & FFE module incorporates multi-scale features of spatial and multi-frequency features of channels

spatial feature extraction module. In this paper, the Global Average Pooling (GAP) in SENet [23] is extended to multiple frequencies to realize the weighting of feature channels [24]. Firstly, the weights at different frequencies are obtained by doing Discrete Cosine Transform (DCT) on the input feature mapping. Then $m$ frequency components with the highest performance are selected. Then the input features are divided into $m$ equal parts, and the corresponding DCT weights are assigned to each feature quantity to do Hadamard Product to obtain the frequency features. The frequency characteristics can be expressed as Equation (4).

$$\begin{aligned} F^i &= 2DDCT^{u_i,v_i}\left(Y^i\right) \\ &= \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} Y^i_{:,h,w} B^{h,w}_{u_i,v_i} \end{aligned} \tag{4}$$

$Y^i$ is the characteristic component, $i \in \{0, 1, \ldots, m-1\}$, $u_i, v_i$ are the indices of the 2D frequency components corresponding to $Y^i$, $H, W$ are the height and width of the input features, $h \in \{0, 1, \ldots, H-1\}$, $w \in \{0, 1, \ldots, W-1\}$. The function $B^{h,w}_{u_i,v_i}$ of 2D DCT is:

$$B^{h,w}_{u_i,v_i} = \cos\left(\frac{\pi u_i}{H}\left(h+\frac{1}{2}\right)\right)\cos\left(\frac{\pi v_i}{W}\left(w+\frac{1}{2}\right)\right). \tag{5}$$

The multi-frequency features are concatenated into 1-dimensional vectors by fully connected, and the 1-dimensional vectors are transformed to the input feature size by scale transformation, as in Equation (6).

$$F_{\text{channel}} = \text{scale}\left(FC\left(F^0, F^1, \ldots, F^{m-1}\right)\right). \tag{6}$$

Then the output of the feature enhancement module is:

$$F_{\text{output}} = F_{\text{spatial}} + F_{\text{channel}}. \tag{7}$$

The structural design of the feature enhancement module is inspired by the residual structure, which solves the network degradation problem by amplifying the minor differences through the identity function. Our primary purpose is to highlight the essential features to reduce the redundant parts, so the channel feature extraction module is used as a shortcut to connect the fusion features to the spatial feature extraction module, as shown by the red line in Figure 5, so that the network can maximize the acquisition of the feature differences when passing through the MS & FFE module. The module's design allows feature resampling at different spatial levels of the CNN network, enriching the kernel and selectively emphasizing crucial information without changing the size of the feature map while ensuring a lightweight network.

### 2.4 Model Structure

Table 4 displays the precise structure of the network of the degree discrimination model for carbon deposit. ResNet-18 consists of one convolutional layer, four convolutional blocks, and one fully connected layer. Two double-layer convolutional structures make up each convolutional block. ResNet-18 outperforms other DCNN models in terms of performance, but it lacks the specificity needed to differentiate between different degrees of carbon deposit. Therefore, in this paper, the MS & FFE feature enhancement module and the three-layer structure to reduce the risk of overfitting are designed for carbon deposit degree discrimination's characteristics and application requirements. The MS & FFE modules are embedded at the end of each block, and the arrangement of the three-layer structure is shown in Figure 4. The structure list contains information on the network design characteristics, such as the size of the convolution kernel, the stride, and the number of channels.

## 3 EXPERIMENTS

### 3.1 Experimental Parameter Settings

The experiments in this paper are conducted under the PyCharm integrated compilation environment. The GPU is NVIDIA GeForce GTX 1080 Ti∗3, the experiments

| Layer Name | Output Size | Structure |
|---|---|---|
| Conv | $112 \times 112$ | $7 \times 7$, stride 2, 64 |
| Block1 | $56 \times 56$ | $3 \times 3$ max pool, stride 2, 64 |
| | | $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$ |
| | | $[MS \,\&\, FFE, 64] \times 1$ |
| Block2 | $28 \times 28$ | $\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$ |
| | | $[MS \,\&\, FFE, 128] \times 1$ |
| Block3 | $14 \times 14$ | $\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$ |
| | | $[MS \,\&\, FFE, 256] \times 1$ |
| Block4 | $7 \times 7$ | $\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$ |
| | | $[MS \,\&\, FFE, 512] \times 1$ |
| DConv | $7 \times 7$ | $3 \times 3$, stride 1, 512 |
| GConv | $7 \times 7$ | $1 \times 1$, stride 1, 1 024 |
| | $1 \times 1$ | Average pool<br>Dropout<br>Fully connected<br>Softmax |

Table 4. Description of the detailed network structure of the carbon deposit degree discrimination model

use parallel training, the training batch size is 32, the number of training epochs is 100, the optimizer uses the SGD optimizer, the initial learning rate is 0.001, the learning rate is adjusted using the cosine decay strategy to decay to 0.0001, and the loss function uses the cross-entropy loss function. Due to the dataset's small original data volume, which is insufficient to support the complex neuron population of deep neural networks, training data expand for a large number of new instances, but this increases the similarity between the expanded data and the original figure's characteristics, which increases the risk of overfitting. In order to make the training data more complicated and improve the model's capacity for learning, the CutMix [17] operation is introduced to the training in this study. This operation crops and mixes any two images, $x \in \mathcal{R}^{W \times H \times C}$ represents the input image and $y$ is the image label, and incorporates any two input images $(x_A, y_A)$, $(x_B, y_B)$ to generate a new training sample $(\tilde{x}, \tilde{y})$. The mixing operation is defined as in Equations (8) and (9):

$$\tilde{x} = M \odot x_A + (1 - M) \odot x_B, \tag{8}$$

$$\tilde{y} = \lambda y_A \oplus (1 - \lambda) y_B, \tag{9}$$

where $M \in \{0, 1\}^{W \times H}$ is a binary mask used to extract and fill the image, $\oplus$ denotes the combination of two labels, $\lambda$ conforms to the beta distribution $\lambda \sim \text{Beta}(\alpha, \beta)$,

and in the experiments $\alpha$, $\beta$ is set to 1, i.e., $\lambda$ obeys a uniform distribution of $(0, 1)$. The mixing probability of each processed batch is set to 0.1.

### 3.2 Evaluation Metrics

The experiments use Accuracy, F1-score, and Inference Speed as evaluation metrics. Among them, F1-score utilizes the precision rate and recall rate to evaluate the model's classification performance comprehensively. Inference Speed is a measure of the prediction speed of the model, and the unit of FPS indicates the number of pictures that can be predicted per second. The computational expression of the evaluation index is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{10}$$

$$F1\text{-}score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}, \tag{11}$$

where,

$$Precision = \frac{TP}{TP + FP}, \tag{12}$$

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

The F1-score is generated by adding a weighting factor to each category since the number of test sets is not balanced, and the weighting factor is based on the percentage of data volume allocation. The score for each metric is then derived using the following equation:

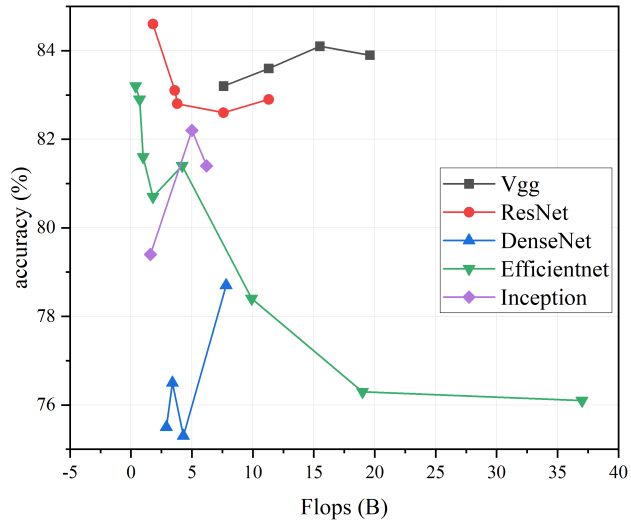
$$S = \frac{N_{c_0}}{N_{\text{total}}} S_0 + \frac{N_{c_1}}{N_{\text{total}}} S_1 + \frac{N_{c_2}}{N_{\text{total}}} S_2, \tag{14}$$

where $N_{c_i} (i = 0, 1, 2)$ is the number of categories, $N_{\text{total}}$ is the total number of test sets, and $S_i$ $(i = 0, 1, 2)$ is the corresponding category F1-score.

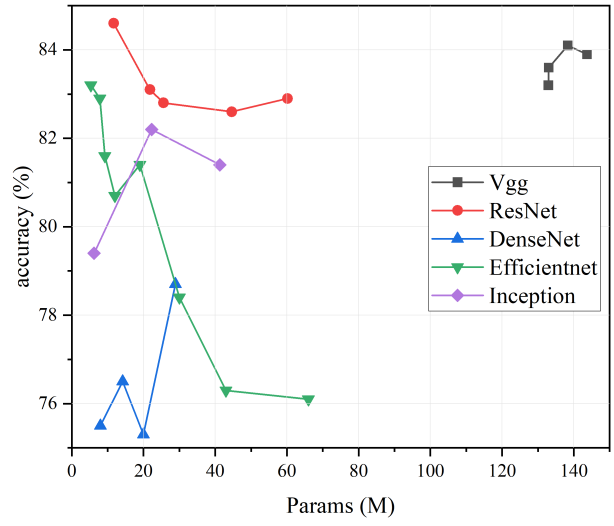
### 3.3 Experimental Results and Analysis

#### 3.3.1 Performance of Different DCNN Models

DCNNs with different structures and capacities are compared and analyzed in the carbon deposit dataset, and evaluate the performance of the models from the three indexes of Accuracy, F1-score, and FPS in combination with the practical requirements. The experimental results are shown in Table 5 and Figure 6 compares the accuracy of different models by combining the amount of computation and the number of parameters.




a) Flops vs. accuracy




b) Params vs. accuracy

Figure 6. Accuracy of models with different structures and capacity sizes on the carbon deposit dataset

| Model | Accuracy [%] | F1-score [%] | FPS | Model | Accuracy [%] | F1-score [%] | FPS |
|---|---|---|---|---|---|---|---|
| VGG-11 | 83.2 | 83.6 | **254** | DenseNet-161 | 78.7 | 76.9 | 56 |
| VGG-13 | 83.6 | 84.1 | 249 | EfficientNet-B0 | 83.2 | 83.5 | 117 |
| VGG-16 | 84.1 | 84.4 | 233 | EfficientNet-B1 | 82.9 | 83.1 | 93 |
| VGG-19 | 83.9 | 84.1 | 229 | EfficientNet-B2 | 81.6 | 82.2 | 90 |
| ResNet-18 | **84.6** | **84.8** | 234 | EfficientNet-B3 | 80.7 | 81.6 | 70 |
| ResNet-34 | 83.1 | 83.5 | 164 | EfficientNet-B4 | 81.4 | 81.7 | 61 |
| ResNet-50 | 82.8 | 83.2 | 131 | EfficientNet-B5 | 78.4 | 78.6 | 52 |
| ResNet-101 | 82.6 | 83.3 | 78 | EfficientNet-B6 | 76.3 | 76.6 | 45 |
| ResNet-152 | 82.9 | 83.1 | 64 | EfficientNet-B7 | 76.1 | 76.2 | 37 |
| DenseNet-121 | 75.5 | 76.3 | 72 | GoogLeNet | 79.4 | 79.3 | 157 |
| DenseNet-169 | 76.5 | 77.2 | 56 | Inception-v3 | 82.2 | 82.7 | 101 |
| DenseNet-201 | 75.3 | 75.9 | 43 | Inception-v4 | 81.4 | 82 | 54 |

Table 5. The experimental results of DCNNs were compared in terms of three evaluation metrics

According to the experimental findings, ResNet-18 has the most excellent Accuracy and F1-score in the carbon deposit dataset, with scores of 84.6 % and 84.8 %, respectively. However, the classification accuracy declines with the addition of more network layers. The classification accuracy of EfficientNet exhibits a similar pattern of declining accuracy with increasing model complexity. The remaining networks' accuracy trends are growing and subsequently falling. This result suggests that models with a larger and more complex number of parameters in the same structure have poor classification results for our task, which has a small amount of data and few category discriminative flags. In contrast, simple or moderately complex models have better classification performance. In Table 5, which compares the classification performance of various structural models, VGG and ResNet perform noticeably better than DenseNet in terms of classification accuracy and inference speed, indicating that deeper and more sophisticated models do not serve better when it comes to the task of discriminating the degree of carbon deposit. Models with shorter layers and simpler structures, on the other hand, are more suited for the task of differentiating the level of carbon deposit.

When comparing the inference speed of DCNNs, the experimental results in Table 5 demonstrate that the inference speed of similar structures essentially changes proportionally to the model complexity, i.e., the model's inference speed gradually declines with an increase in the model computation and the number of parameters. However, this conclusion is not applicable to models with different structures, e.g., VGG-11 has 24 times more parameters than EfficientNet-B0, but the inference speed is two times faster than EfficientNet-B0. This result suggests that the inference speed of a model is not entirely related to the computational volume and the number of parameters of the model but is also closely related to the composition of the model building blocks.

### 3.3.2 Discussion of Dropout Parameter $p$

We analyze the parameter $p$'s value for the dropout layer in Table 6. We set the dropout ratio of 0.1-0.8 to train the model accordingly. The experimental findings demonstrate that the model test accuracy increases to a peak and progressively decreases. The model test accuracy rises to a maximum of 85.0 % when the dropout ratio is set to 0.2, an improvement of 0.4 % over the accuracy before it was implemented.

| $p$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
|---|---|---|---|---|---|---|---|---|
| Accuracy [%] | 84.8 % | **85.0 %** | 84.3 % | 83.5 % | 83.1 % | 82.9 % | 82.4 % | 81.4 % |

Table 6. Discuss the MS & FFE of the value of the dropout layer parameter $p$ on the accuracy

### 3.3.3 Channel Attention Mechanisms in the MS & FFE Module

In this section, we discuss the impact of different channel attention mechanisms on the MS & FFE module and consider the problem of channel representation from a frequency domain perspective, comparing SENet [23], which uses GAP to obtain feature information from the zero frequency, and CBAM [25] and SRM [26], which add global maximum pooling and global standard deviation pooling to SENet. The experimental comparison results are shown in Table 7. No operations are added to prevent overfitting in the experiment, and only feature resampling modules are added after each block. The results indicate that adding a multi-frequency channel attention mechanism helps improve accuracy most while having the fastest inference speed.

| Module | Accuracy [%] | F1-score [%] | FPS |
|---|---|---|---|
| Multi-Scale and SENet | 84.8 % | 85.3 % | 198 |
| Multi-Scale and CBAM | 85.3 % | 85.5 % | 183 |
| Multi-Scale and SRM | 85.0 % | 85.2 % | 178 |
| MS & FFE | **85.7 %** | **85.7 %** | **201** |

Table 7. Performance effects of different channel attention mechanisms on MS & FFE module

### 3.3.4 Ablation Experiments

We conducted the corresponding ablation experiments, as indicated in Table 8, to verify the efficacy of the approach and module design employed in this research. All models use the augmented dataset with offline data. Model A is a separate addition to the online data enhancement CutMix. After cutting and mixing, the samples are more complex, thus improving the generalization ability of the model. Its accuracy

and F1-score compared to ResNet-18 have a slight increase. Model B is a single-layer dropout structure, and model C is a three-layer dropout structure designed in this paper. The experimental results show that dropout can reduce the impact of overfitting and effectively improve the classification accuracy, and at the same time, verifies the effectiveness of this paper's method of upgrading the features dimension and then dropping them. The D and E models provide feature enhancement from a single dimension, the former via dilation convolution-extraction of multi-scale spatial features and the latter by acquiring channel information features at various frequencies. According to the experimental findings, adding either the spatial feature improvement module or the channel attention module alone enhances model performance, therefore, combining the two feature modules further boosts model accuracy. The final method in this paper achieves an accuracy of 86.9 % and an F1-score of 87.2 % in the task of discriminating the degree of carbon deposit in an automobile engine, which is an improvement of 2.3 % and 2.4 %, respectively, compared to the original model ResNet-18.

In the model inference speed results, the inference speed is unchanged because the network structure of model A has not been changed. In contrast, model B speeds up inference by eliminating certain unnecessary parameters. The inference speed of the other models is lowered to varying degrees because they are based on the original model structure supplement. The ultimate inference speed of this article is 190 FPS, which is slower than ResNet-18 but still quicker than the inference speed of most of the models in Table 5. The DCNN model developed in this study can successfully perform the carbon deposit degree discriminating task in terms of accuracy and speed of inference.

| Model | CutMix | Dropout | | Feature Enhancement Module | | | Accuracy [%] | F1-score [%] | FPS |
|---|---|---|---|---|---|---|---|---|---|
| | | one layer | three layers | Multi-Scale module | Multi-Frequency module | MS & FFE module | | | |
| A | ✓ | | | | | | 84.8 % | 85.1 % | 234 |
| B | ✓ | ✓ | | | | | 85.2 % | 85.2 % | **245** |
| C | ✓ | | ✓ | | | | 85.5 % | 85.6 % | 228 |
| D | ✓ | | ✓ | ✓ | | | 86.0 % | 86.5 % | 218 |
| E | ✓ | | ✓ | | ✓ | | 85.8 % | 86.0 % | 221 |
| Proposed Method | ✓ | | ✓ | | | ✓ | **86.9 %** | **87.2 %** | 190 |

Table 8. Results of the ablation experiments

## 4 CONCLUSION

This study addresses the difficulties of complicated operation and low reliability of conventional carbon deposit degree detecting systems. The DCNN model suitable for the discrimination task of carbon deposition degree is explored by compar-

ing models with various structures and different capacity sizes. The experimental findings demonstrate that the shallow network performs better in the task of this research. Aiming at the issue of overfitting easily caused by the small data volume of the carbon deposit dataset, various data enhancement methods and dropout structures are used to reduce the risk of overfitting. The dropout structure raises the dimension of image features and discards some of them, effectively improving the model's classification accuracy. The MS & FFE feature resampling module is designed to enrich the transition features for the problems of poor aggregation of discriminative features and low inter-class separability in the carbon deposit image. The MS & FFE module acquires the spatial multi-scale features and the channel multi-frequency features, which effectively increases the inter-class differences, reduces the intra-class differences and improves the separability of the carbon deposit image among the classes. In order to match the real test effect, a large number of unprocessed images are used to test the performance of the model. The experimental findings indicate that the paper's structure has a faster inference speed while maintaining recognition accuracy, making it more suitable for use in real-world engineering.

**Acknowledgement**

This work was supported by the Shanxi Province Key Research and Development Program Projects (No. 202302020101008) and the Research Project Supported by Shanxi Scholarship Council of China (No. 2022-145) and the Graduate Science and Technology Project Supported by North University of China (No. 2022180506).

## REFERENCES

[1] Yang, W.—Liu, Y.—Hong, A.—Li, G.—Liu, H.—Liang, B.—Jiang, R.—Li, Y.—Li, S.—Wang, R.—An, C.: Engine Cylinder Cover, Engine, Vehicle and Sediment Treatment Method. 2022, https://worldwide.espacenet.com/CN113818970B (CN Patent No. CN113818970B).

[2] Prabhakar, B.—Conti, R.: Non-Contact Process for Engine Deposit Layer Measurement. 2023, https://worldwide.espacenet.com/US11650043B2 (US Patent No. US11650043B2).

[3] Glugla, C. P.—Huberts, G. J.—Morrow, N. W.—Qu, Q.: Spark Plug Fouling Detection for Ignition System. 2018, https://worldwide.espacenet.com/CN104728020B (CN Patent No. CN104728020B).

[4] Roe, A. P.—Thompson, S. A.—Petrou, A. T.—Campagna, M. J.: System for Determining Piston Damage Based on Carbon Deposit Growth. 2018, https://worldwide.espacenet.com/US9957887B2 (US Patent No. US9957887B2).

[5] Mu, D.—Lyu, X.—Song, G.—Hu, Y.—Liguo, W.: Method and Device for Detecting Carbon Deposit of Air Injection Valve and Electronic Equip-

ment. 2019, https://worldwide.espacenet.com/CN110005524A (CN Patent No. CN110005524A).

[6] Yan, J.—Shao, C.—Wang, L.—Huang, L.: Device for Detecting Thickness of Carbon Deposit in Fuel Engine. 2019, https://worldwide.espacenet.com/CN209131594U (CN Patent No. CN209131594U).

[7] Xiao, J.—Ji, M.—Lan, L.—Huang, Y.: Engine Detection Method and System, Computer Equipment and Storage Medium. 2021, https://worldwide.espacenet.com/CN113284131A (CN Patent No. CN113284131A).

[8] Yao, P.—Shen, S.—Xu, M.—Liu, P.—Zhang, F.—Xing, J.—Shao, P.—Kaffenberger, B.—Xu, R. X.: Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification. IEEE Transactions on Medical Imaging, Vol. 41, 2022, No. 5, pp. 1242–1254, doi: 10.1109/TMI.2021.3136682.

[9] Fan, S.—Wang, Y.—Cao, S.—Zhao, B.—Sun, T.—Liu, P.: A Deep Residual Neural Network Identification Method for Uneven Dust Accumulation on Photovoltaic (PV) Panels. Energy, Vol. 239, 2022, Art. No. 122302, doi: 10.1016/j.energy.2021.122302.

[10] Simonyan, K.—Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, 2014, doi: 10.48550/arXiv.1409.1556.

[11] Szegedy, C.—Liu, W.—Jia, Y.—Sermanet, P.—Reed, S.—Anguelov, D.—Erhan, D.—Vanhoucke, V.—Rabinovich, A.: Going Deeper with Convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[12] He, K.—Zhang, X.—Ren, S.—Sun, J.: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[13] Huang, G.—Liu, Z.—Van Der Maaten, L.—Weinberger, K. Q.: Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[14] Tan, M.—Le, Q. V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Chaudhuri, K., Salakhutdinov, R. (Eds.): Proceedings of the 36th International Conference on Machine Learning (ICML 2019). Proceedings of Machine Learning Research (PMLR), Vol. 97, 2019, pp. 6105–6114, https://proceedings.mlr.press/v97/tan19a.html.

[15] Dong, H.—Song, K.—Wang, Q.—Yan, Y.—Jiang, P.: Deep Metric Learning-Based for Multi-Target Few-Shot Pavement Distress Classification. IEEE Transactions on Industrial Informatics, Vol. 18, 2021, No. 3, pp. 1801–1810, doi: 10.1109/TII.2021.3090036.

[16] Shen, S.—Xu, M.—Zhang, F.—Shao, P.—Liu, H.—Xu, L.—Zhang, C.—Liu, P.—Yao, P.—Xu, R. X.: A Low-Cost High-Performance Data Augmentation for Deep Learning-Based Skin Lesion Classification. BME Frontiers, Vol. 2022, 2022, Art. No. 9765307, doi: 10.34133/2022/9765307.

[17] Yun, S.—Han, D.—Chun, S.—Oh, S. J.—Yoo, Y.—Choe, J.: CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019,

pp. 6022–6031, doi: 10.1109/ICCV.2019.00612.

[18] Zhao, H.—Shi, J.—Qi, X.—Wang, X.—Jia, J.: Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[19] Chen, L. C.—Papandreou, G.—Kokkinos, I.—Murphy, K.—Yuille, A. L.: Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. CoRR, 2014, doi: 10.48550/arXiv.1412.7062.

[20] Yu, F.—Koltun, V.: Multi-Scale Context Aggregation by Dilated Convolutions. CoRR, 2015, doi: 10.48550/arXiv.1511.07122.

[21] Yu, F.—Koltun, V.—Funkhouser, T.: Dilated Residual Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 636–644, doi: 10.1109/CVPR.2017.75.

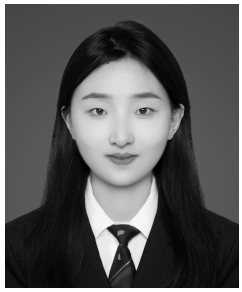
[22] Mehta, S.—Rastegari, M.—Shapiro, L.—Hajishirzi, H.: ESPNetv2: A Light-Weight, Power Efficient, and General Purpose Convolutional Neural Network. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9182–9192, doi: 10.1109/CVPR.2019.00941.

[23] Hu, J.—Shen, L.—Sun, G.: Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[24] Qin, Z.—Zhang, P.—Wu, F.—Li, X.: FcaNet: Frequency Channel Attention Networks. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 763–772, doi: 10.1109/ICCV48922.2021.00082.

[25] Woo, S.—Park, J.—Lee, J. Y.—Kweon, I. S.: CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.): Computer Vision – ECCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11211, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[26] Lee, H.—Kim, H. E.—Nam, H.: SRM: A Style-Based Recalibration Module for Convolutional Neural Networks. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1854–1862, doi: 10.1109/ICCV.2019.00194.



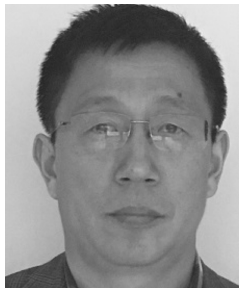
**Qian Huang** received her B.Eng. degree in electronic information engineering from the North University of China in 2022. She is currently working toward her Master's degree in information and communication engineering. Her research interests include deep learning, computer vision, and image processing.

**Youxing Chen** received his Ph.D. degree from the North University of China, in 2010. He is currently Professor with the School of Information and Communication Engineering, North University of China. His research interests include the areas of image processing, signal processing, and non-destructive testing.

**Zhaoba Wang** received his Ph.D. degree in instruments science and technology from the Nanjing University of Science and Technology in 2002. He is currently Professor with the School of Information and Communication Engineering, North University of China. His main research interests include information processing and reconstruction, computer vision, and image processing.

**Xiang Chen** received her Bachelor's degree from the North University of China in 2023. She is currently pursuing her Master's degree in information and communication engineering. Her research direction is digital signal processing and image processing.

**Asoke Nandi** received his Ph.D. degree in physics from the Cambridge University in 1978, and he is currently Professor in the Department of Computer Engineering at Brunel University London, UK, and a Fellow of the Royal Academy of Engineering. He has long been engaged in research on statistical signal processing, wireless communications, machine learning and biomedical signal processing, and has made great contributions to signal processing of genomes and brain signal processing applications.



**Maozhen Li** received his Ph.D. degree from the Institute of Software, Chinese Academy of Sciences, in 1997. He is currently Professor with the Department of Electronic and Computer Engineering, Brunel University London, UK. His main research interests include high performance computing, big data analytics, and intelligent systems with applications to smart grid, smart manufacturing, and smart cities. He has over 180 research publications in these areas, including four books. He has served over 30 IEEE conferences. He is also a Fellow of the British Computer Society and the IET. He is on the editorial board of a number of journals.

**Yong Jin** received his Ph.D. degree from the North University of China, in 2013. He is currently Professor with the School of Information and Communication Engineering, North University of China. His research interests are in the areas of image processing, online inspections, and big data analytics.