Research on Classification Method of Film Damage Image Based on Improved ResNet50

Peiqiang Chen School of Computer Science and Engineering Xi'an Technological University Xi'an, China E-mail: 1529532803@qq.com

Abstract—In view of the challenges posed by traditional methods in accurately identifying the damage types of film images, this paper proposes an improved approach by leveraging transfer learning based on ResNet50, incorporating the channel attention mechanism from the Convolutional Block Attention Module (CBAM). Additionally, the AlphaDropout model is integrated with the SeLU activation function to enhance stability and performance, and the resultant model is named CBAM-ResNet50. The film dataset employed in this study comprises four types of damage: cracks, dewetting, particles, and scratches. Extensive training and testing of the CBAM-ResNet50 model on this dataset demonstrate significant improvements in classification accuracy. Compared to the original ResNet50 model, the proposed method achieves a remarkable 25.57% improvement in accuracy, reaching 90.58%. This work highlights the potential of combining attention mechanisms and advanced activation strategies to address complex image classification tasks. Furthermore, the approach paves the way for practical applications in quality inspection and automated defect detection in industrial processes.

Keywords-ResNet50; Convolutional Neural Network; Transfer Learning; CBAM Attention Mechanism

I. INTRODUCTION

Modern optical instruments use a large number of coated optical components. Under the action of strong laser, the damage of optical components is almost inevitable, and once the optical components are damaged, it can disrupt the normal operation of the entire system. Therefore, enhancing the laser resistance of optical films is of great importance [1]. Accurately judging whether the film is damaged or not is to accurately measure the LIDT (Laser Induced Damage Threshold) [2], LIDT is a measure Shuping Xu School of Computer Science and Engineering Xi'an Technological University Xi'an, China E-mail: 563937848@qq.com

of the laser-induced damage threshold, assessing the laser resistance of optical thin films. Only by accurately measuring LIDT, researchers can explore various approaches to enhance the laser resistance of optical thin films. Thereby reducing the use cost of the optical system.

With the advancement of computer science and technology in our country, Digital image processing technology has also been applied to the field of detecting whether the film is damaged. Compared with the traditional method of using microscope, it has higher accuracy and useability [3-4]. In our country, some scholars began to try to use the method of image classification to study whether the film image is damaged and the type of damage.

Chang Hao et al [5]. Aimed at the problem of film cracking caused by too high or too low temperature in the incubator during film processing. Combining with deep learning, a film rupture image detection technology is proposed, experimental results indicate that the proposed detection technology achieves an accuracy of 82.7%. which can meet the needs of industrial production. Wei Dong [6] proposed two unsupervised surface defect detection algorithms for thin films. The three metrics-precision, recall, and MAP (Mean Average Precision)—are evaluated comprehensively, and the test results are presented. The detection time of single image is 9.7 ms under GPU acceleration, and the detection speed reaches 103 FPS. Zhang Zhenhua et al. [7] introduced an optical film defect image recognition method utilizing an improved convolutional neural network. The experimental results demonstrate that the average classification accuracy achieved by this approach is 83.2%, and the training time is 964 s. Liu Peng [8] proposed to develop a set of thin film coating defect detection system to improve the quality of products. After testing, the host computer software functions well and can quickly and accurately process the collected film images. And the detection accuracy meets the required requirements.

In general, image-based methods have been widely used in the detection of thin film damage, However, there are still some problems that need further study. For example, how to improve the accuracy and reliability of detection, how to deal with the interference of complex environment and so on. Therefore, in view of the present situation, this paper presents a film damage image classification method based on an enhanced ResNet50, referred to as CBAM-ResNet50. The experimental results demonstrate 25% а improvement in classification accuracy with the proposed model. 57% compared with the original model, reaching 90. 58%. Moreover, CBAM-ResNet50 has fewer parameters, faster model training speed and shorter training time, which is more suitable for later embedded development. At the same time, it has good generalization ability and can be applied to other types of defect image detection.

II. FILM DAMAGE IMAGE CLASSIFICATION MODEL BASED ON IMPROVED CBAM-RESNET50

A. CBAM-ResNet50 Network Model

He Kaiming and others proposed the concept of residual network (ResNet) in 2015. Since its introduction, ResNet has garnered significant attention in the field of image processing and has delivered outstanding performance on numerous benchmarks. Especially in the ImageNet classification competition, ResNet stands out for its excellent performance. It quickly became one of the most advanced image classification models of its time [9].

Considering the number of parameters and the training effect of each model, this paper chooses ResNet50 model to study. ResNet50 is composed of 50 convolutional layers, including residual

blocks, these are directly linked across layers to decrease the number of parameters. Enhance model performance. To boost the classification accuracy of ResNet50, the original ResNet50 architecture is modified. Based on the original model, transfer learning, channel attention mechanism in CBAM attention mechanism, AlphaDropout module and SeLU are used. Activation function to build an improved CBAM-ResNet50 network model. Transfer learning can accelerate the network's training process and help mitigate overfitting issues; CBAM attention mechanism has the ability to screen invalid information, and it can deal with the features extracted from the network model. Only those valuable features are transferred to the next layer, thus optimizing the information transfer efficiency the whole network; of The AlphaDropout module and the SeLU activation function maintain consistency in the mean and standard deviation between the input and output, preserving the normalization property. The network model of CBAM-ResNet50 is shown in Figure1.



Figure 1. CBAM-ResNet50 Network Model

As shown in the figure, the convolutional and pooling layer structures of the original ResNet50 remain unchanged. model The feature representation of the target image can be enhanced by adding the Channel Attention Mechanism (CAM) after the convolutional layers; An AlphaDropout module and a SeLU activation function are added after the average pooling layer, This ensures that the mean and standard deviation of both the input and output are kept consistent, and maintain the stability of their normalization properties; Because this paper only studies the classification of four types of film damage defects, The Softmax classification layer is thus changed into a Softmax classifier of 4 labels, Corresponding to the four defect types of cracks, dewetting,

particles and scratches in the film damage image dataset.

B. Construction of CBAM-ResNet50 network model

1) Transfer learning

When the network is relatively large, that is, the corresponding network parameters are relatively large, and the number of data sets is relatively small. It is not enough to train the whole network, so there will be overfitting, and the training results will be very bad. However, by using the transfer learning method, we can leverage the pre-trained model parameters to train our own smaller datasets. It can also train a better effect [10-11]. Therefore, the method of transfer learning is used in this study. The weights trained on the ImageNet dataset are used to train the film damage dataset. The ImageNet dataset contains over one million images across 1,000 categories, and the number of image categories is sufficient. Although it does not include the film damage image, the judgment of the film damage type is mainly based on its contour and other characteristics. Therefore, transfer learning can be carried out through the ImageNet dataset. The weight file for ResNet50 pre-training using ImageNet can be downloaded from the Tensorflow website.

2) CBAM Attention Mechanism

CBAM (Convolutional Block Attention Module) integrates both channel and spatial attention mechanisms to enhance model performance by emphasizing the importance of different regions in the input image. The channel attention module identifies the most relevant channels in the feature map, while the spatial attention module highlights the spatial locations that contain the most informative content. For the special type of film damage image, its characteristic is that the damage area is often presented in the form of tiling. This means that the distribution of lesions over the entire image space is relatively uniform, so in this case, Increasing the spatial attention mechanism will blur the recognition focus of the model, resulting in a decline in classification accuracy. Based on the above considerations, this study decided to use only the channel attention mechanism to process the film damage images. By increasing the weight of the

effective channel and suppressing the weight of the ineffective channel, the model can be designed to focus more on the features that are crucial for the classification task. By more efficiently extracting and utilizing the key information from the image, the model improves its feature representation capabilities, ultimately leading to higher classification accuracy [12-14]. The workflow of the CAM attention mechanism is illustrated in Figure 2.



Figure 2. CAM Attention Mechanism Diagram

During the feature extraction process in deep learning, global pooling technology plays a vital role. Firstly, two different pooling strategies, global maximum pooling and global average pooling, are used to aggregate the wide and high dimensional features of the feature vector f. The global maximum pooling operation focuses on the maximum position of each channel in the feature map, it can effectively capture the most prominent features in the image. Global average pooling is a comprehensive perception of each pixel in the feature map to obtain more comprehensive information. Next, the features obtained from global maximum pooling and global average pooling are concatenated. This step aims to fuse the complementary features extracted by both pooling methods, and a richer and more comprehensive feature representation is obtained. Then, the concatenated features are then fed into a multi-layer perceptron (MLP), which consists of two fully connected layers and a ReLU activation function. This MLP is used to further compress and abstract the features through the linear transformation of the fully connected layers and the nonlinear mapping the ReLU activation, provided by Highdimensional feature vectors can be compressed to lower dimensions while retaining information useful for classification tasks. Finally, the output of MLP is normalized by sigmoid function, and the normalized features are merged. A C \times 1 \times 1 dimensional channel feature vector is obtained

 $M_{c}(f) \in \mathbb{R}^{1 \times 1 \times C}$, $M_{c}(f) \in \mathbb{R}^{1 \times 1 \times C}$ which contains the global information of each wide-high channel feature map $f_{c} \in \mathbb{R}^{1 \times H \times W}$. The calculation process is illustrated in formula (1).

$$M_{c}(f) = \sigma(MLP(AvgPool(f)) + MLP(MaxPool(f)))$$

$$= \sigma \left(W_1(f_{max}^c) \right) + W_1 \left(W_0(f_{max}^c) \right) \quad (1)$$

In (1): W × H — width and height of the feature map; C — number of feature map channels; f(i, j)— feature map; MLP-Multilayer Perceptron.

And finally, multiply that attention map with the weight of the channel corresponding $f_c \in \mathbb{R}^{1 \times H \times W}$ to the original feature map point by point, Generate the one-dimensional channel feature attention map containing the global feature information required by the spatial attention module f_1 , as shown in

by the spatial attention module J_1 , as shown in Equation (2).

$$f_1 = f \otimes M_c(f) \tag{2}$$

3) Alpha Dropout Module + SeLU Activation Function

The original ResNet50 network model has no Dropout module, and the activation function used is ReLU (Rectified Linear Unit).

To address the issue of overfitting, Dropout is applied when the model is large and the number of parameters is extensive. This technique achieves the regularization effect of the model in the training process by temporarily inactivating some neurons. To enhance the model's generalization, however, the distribution of activation values may change after each Dropout application, to solve this problem, researchers proposed AlphaDropout, which is an improvement of Dropout [15]. By using Alpha Dropout, the distribution stability of activation values can be better maintained, and the robustness of the model can be further enhanced. Even in large-scale models and complex tasks, it can effectively deal with the problem of overfitting. At the same time, a new activation function, SeLU (Scaled Exponential Linear Units) [16], is introduced. The combination of AlphaDropout and SeLU ensures that the mean and standard deviation of both the input and output remain consistent, thus

maintaining the stability of its normalized properties.

SeLU has a significant advantage over the activation function ReLU in that it does not have a deactivation region, as shown in Equation (3). The SeLU activation function has a saturation region at negative infinity, but this does not adversely affect the expressiveness of the model. In contrast, the SeLU activation function automatically normalizes the sample distribution to 0 mean and unit variance, the stability of the gradient in the training process is ensured, thereby effectively avoiding the problems of gradient explosion and disappearance. This feature of automatic normalization makes the SeLU activation function perform well when dealing with large-scale models and complex tasks.

$$SeLU = \lambda \begin{cases} x , x > 0 \\ \partial e^{x} - \partial, x \le 0 \end{cases}$$
(3)

Where ∂ and λ are hyperparameters λ of about 1.05 and ∂ about 1.67 and x are input quantities.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In This experiment, the self-made film damage image data set, ImageNet image data set and Northeastern University steel defect image data set are mainly used. There are four damage types in the film image data set, namely, cracks, dewetting, particles and scratches, which are used for classification and identification. The ImageNet image dataset is used in the transfer learning phase, the steel defect image dataset of Northeastern University is used to verify the generalization ability of CBAM-ResNet50 network model. Accuracy and recognition time are selected as the main evaluation indexes of the network model to facilitate the later embedded development of the model. The model memory footprint and the number of model parameters should also be considered. Combined with the confusion matrix,

precision, accuracy, recall, and F_1 score are calculated to comprehensively evaluate the network model's classification ability for thin film damage images.

A. Experimental environment

The experiments were performed on a Windows 10 operating system, featuring an Intel(R) Core(TM) i9-9900 CPU @ 3.60 GHz and 64GB of RAM. The operations were accelerated using an NVIDIA RTX 2080 Ti graphics card. The input images were normalized to a size of $224 \times 224 \times 3$, and the program was written in Python, the model was implemented using the TensorFlow deep learning framework. The initial learning rate was set to 0.0001, the dropout rate was 0.5, the batch size was 32, and the number of epochs was 100.

B. Dataset and Image Preprocessing

1) Film Damage Image Data Set

In this paper, all the original film damage images are numbered, classified and labeled (label 1 is crack, label 2 is dewetting, Label 3 is grain, label 4 is scratch), and finally all the images are imported into the computer in JPG format. The construction of four kinds of film damage image database is completed. All images in this database will be used as the input data set for the CBAM-ResNet50 convolutional neural network. The images in the training set, validation set, and test set are randomly selected by the computer.

2) Steel Defect Image Data Set

The steel defect image dataset from Northeastern University is used to evaluate the generalization ability of the improved CBAM-ResNet50 network model. The images in the steel defect image data set of Northeastern University were collected by several teachers of Northeastern University. The data set includes seven defect types: Rolled-in scale, Patches, Crazing, Pitted, Surface, Inclusion, and Scratches, with 300 images for each defect type. In total, the dataset consists of 1800 grayscale images.

3) Data preprocessing

To remove irrelevant information and preserve the real, useful features in the image, some of the film damage images are corrupted with noise. To enhance the model's generalization ability, Preprocessing the acquired images is crucial. The main pretreatment steps are as follows: *a)* Image scaling, ResNet50 requires the size of the input image to be $224 \times 224 \times 3$. Therefore, the size of the acquired image is scaled to 224 * 224 * 3 pixels to adapt to the training of the model;

b) The main function of filtering and denoising is to suppress the noise of the image as much as possible on the premise of retaining the detailed information of the image [17];

c) Edge extraction, which uses edge detection algorithm to extract image features, can effectively remove redundant information in the image. It has certain advantages in improving the classification accuracy [18];

d) data enhancement, to enhance the generalization capability of the network model and prevent overfitting during training. data augmentation of the film damage image dataset is applied. In this paper, the original dataset is divided into three subsets: a training set with 49,760 images, a validation set with 6,220 images, and a test set with 6,220 images, following an 8:1:1 ratio. Only 80% of the training set undergoes data augmentation, while the validation and test sets do not undergo any data augmentation. In this paper, four kinds of image transformation techniques, including rotation, translation, shear and zoom, are used to enhance the images in the training set for seven times. There are 348320 images in the final training set [19];

C. Analysis of factors affecting model performance

1) Comparison of transfer learning effect

In order to avoid the problems of over-fitting and slow convergence, this experiment uses ILSVRC2012 sub-dataset in ImageNet for transfer learning. Figure3 shows the comparison of the relationship curve between the number of iterations of the validation set and the accuracy rate of the CBAM-ResNet50 network model with and without transfer learning. Figure 4 compares the relationship between the number of iterations and the loss value on the training set, with and without transfer learning.



Figure 3. Curve between the number of iterations and the accuracy with or without transfer learning



Figure 4. Curve between the number of iterations and the loss value with or without transfer learning

As shown in Figure 3, the ResNet50 + transfer learning model demonstrates higher accuracy and faster convergence compared to the original ResNet50 model on the validation set. Furthermore, the curve becomes more stable after reaching convergence. The ResNet50 + transfer learning model tends to be stable at 30 epochs, while the original ResNet50 model tends to be stable at 68 epochs. The accuracy of ResNet50 + transfer learning model is about 0.85, while the accuracy of ResNet50 model is only about 0.65. This shows that the training accuracy of the network is higher, the training speed is faster, and the network can be stabilized faster after using transfer learning.

As shown in Figure 4, the final loss value of the original ResNet50 network model is relatively high, approximately 0.75. The final loss value of ResNet50 + transfer learning model is 0.4 less than that of ResNet50 +, only 0.3, and the loss curve using transfer learning decreases faster. Achieve stability more quickly.

2) Effect comparison of CBAM attention mechanism

In the task of image classification, the internal feature transmission mechanism of the traditional convolutional neural network often lacks clear distinction. As a result, in the process of feature extraction and transmission, effective information is easily submerged in a large number of irrelevant information. This type of feature propagation not only impacts the model's ability to capture crucial information but may also lead to performance degradation in complex scenarios. To overcome this challenge, this paper incorporates the channel attention mechanism from the CBAM module into the traditional ResNet50 architecture. O as to realize the optimization and improvement of the feature transmission process.

The placement of the attention mechanism plays a crucial role in determining the model's recognition accuracy. Feature representations at different levels have different levels of abstraction semantic information. and therefore. and Embedding the attention mechanism in different locations of the network will produce different effects. To enhance the model's performance without altering its original structure, in this paper, we choose to add attention mechanism after the first and last convolution layer of ResNet50. This design can not only ensure that the attention mechanism can fully play its role, but it will not significantly affect the overall structure of the model. To verify the effectiveness of the attention mechanism's placement, four different addition schemes are designed in this paper, as shown in TABLE I.

TABLE I. FOUR CBAM ATTENTION MECHANISM ADDITION SCHEMES

| Programmer | Attention mechanism adding method | | |
|------------|--|--|--|
| Option 1 | Add an attention mechanism after the first convolution layer | | |
| Option 2 | Two attention mechanisms are added after the first and last convolution layer | | |
| Option 3 | Add 1 attention mechanism after the last convolutional layer | | |
| Option 4 | Do not add attention mechanism | | |

During the experiment, the same data set and consistent training parameters are used to ensure the fairness of the comparison results. The results of the experiment are displayed in Figure 5.



Figure 5. Performance comparison of CBAM-ResNet50 model with different addition modes of attention mechanism

As can be seen from Figure 5, when an attention module is added to the ResNet50 network, its recognition performance is slightly improved. This indicates that the attention mechanism can greatly improve the model's feature extraction capabilities, allowing it to more effectively capture the essential features within the image. When two attention modules are integrated into the network, the recognition accuracy of the film damage images reaches its optimal performance. This outcome strongly validates the effectiveness of the channel attention mechanism in image recognition tasks. The model can mine the potential information in the image more deeply, and further improve its accuracy on the validation set. It is also important to note that while adding more attention modules may result in higher performance gains, but this will also increase the complexity and computational cost of the model. Therefore, in practical applications, it is necessary to balance the performance and complexity of the model according to the specific task and resource constraints. Select the most appropriate number of attention modules and where to add them.

3) Effect comparison of AlphaDropout module + SeLU activation function

In this paper, The Alpha Dropout module and SeLU activation function are innovatively incorporated to enhance the performance of film damage image classification. The combination of these two techniques not only helps to maintain the stability of the distribution of the feature map, it can also effectively prevent the occurrence of overfitting phenomenon, this approach also accelerates the training process and improves the model's convergence speed.

First of all, the AlphaDropout module is used in the training process. The output of the neurons in the network is set to zero randomly with a certain probability, thereby avoiding the excessive dependence of the model on the output of a specific neuron, the randomness introduced by this approach helps enhance the model's generalization ability. To avoid overfitting, where the model performs well on the training set but shows a significant drop in performance on the test set, the AlphaDropout module is employed. This module effectively improves the model's robustness. It can classification maintain stable performance. Secondly, the SeLU activation function is selected. which has the characteristic of self-normalization. The distribution of the output value can be automatically adjusted during the activation process to be close to the standard normal distribution. This property helps preserve the stability of the feature map distribution, The information loss or distortion caused by the action of the activation function in the layer-by-layer transmission process is prevented, and meanwhile, The SeLU activation function also has a faster convergence rate and can achieve better classification results in a shorter training period.

In order to verify the effect of AlphaDropout combined with SeLU on the classification of film damage images by ResNet50 network, Based On the ResNet50 network using transfer learning and attention adding CBAM mechanism. AlphaDropout + ReLU is tested respectively. AlphaDropout + SeLU and the original ResNet50. The recognition results of the three models under the same conditions are shown in TABLE II. Compared with using ReLU activation function alone, the combination of AlphaDropout and SeLU achieves the highest accuracy of 90.58%. Compared with the combination of Alpha Dropout and ReLU, its accuracy is increased by 0.13%, which is mainly attributed to the activation function of SeLU. It effectively solves the problem of neuron "inactivation", This, in turn, this enhances the expressive capability of the network model. The normalization feature of SeLU, coupled with AlphaDropout, ensures that the output data retains a mean of 0 and a standard deviation of 1. As a

result, the model's convergence speed is further accelerated.

 TABLE II.
 PERFORMANCE COMPARISON OF CBAM-RESNET50 MODEL UNDER DIFFERENT ACTIVATION FUNCTIONS

| Activation function | Classification accuracy (%) |
|---------------------|--------------------------------|
| AlphaDropout+SeLU | 90.58 |
| AlphaDropout+ReLU | 90.45 |
| ReLU | 89.16 |

D. Confusion Matrix of Film Damage Classification and Identification Results on CBAM-ResNet50

The confusion matrix, being an intuitive and effective tool, provides a clear view of the classification model's performance across each category. It can not only reflect the recognition accuracy of the model for each category, but also reveal the confusion of the model between different categories. It provides an important basis for model optimization. In this study, four different types of film damage were classified and identified, and the improved CBAM-ResNet50 model was used to model. To provide a more intuitive illustration of the model's classification performance, this paper presents the corresponding confusion matrix, as illustrated in Figure 6.



Figure 6. Confusion Matrix of Film Damage Classification Identification Results on CBAM-ResNet50

In this study, the performance indicators of CBAM-ResNet50 model on the task of film damage identification are calculated by using the confusion matrix. As shown in TABLE III.

 TABLE III.
 FILM DAMAGE CLASSIFICATION AND IDENTIFICATION RESULTS OF CBAM-RESNET50 MODEL

| Damage category | Accuracy (%) | Accuracy (%) | Recall (%) | F_1 Fraction |
|--------------------|-----------------|-----------------|------------|-------------------|
| Crack | 95.02 | 95.99 | 96.69 | 96.34 |
| Dewetting | | 96.21 | 96.78 | 96.49 |
| Particles | | 93.98 | 91.90 | 92.93 |
| Scratches | | 88.01 | 85.98 | 86.98 |

As shown in TABLE III, all metrics, except for the precision, recall, and score of scratches, are above 91%.

As shown in Figure 6, the total number of misrecognitions is relatively low. which indicates that the improved CBAM-ResNet50 model performs well. And can be applied to the identification of the film damage image.

E. Ablation Experiment on CBAM-ResNet50

Ablation experiments are conducted on the custom-built film damage image test set. Ablation experiments were carried out on all possible cases of CBAM-ResNet50 model. On the basis of the original model, transfer learning, CBAM attention mechanism. AlphaDropout + SeLU and AlphaDropout + ReLU are added in turnon the basis of the original model and transfer learning, four schemes of CBAM attention mechanism are added in turn; Building upon the original model, transfer learning and CBAM attention mechanism AlphaDropout scheme 2, +SeLU and AlphaDropout + ReLU are added in turn. TABLE IV displays the experimental results.

As shown in TABLE IV, the model accuracy improves when adding transfer learning, the four mechanism schemes, CBAM attention AlphaDropout + SeLU, and AlphaDropout + ReLU, sequentially based on the original model. Increased by 20.04%, 20.06%, 20.12%, 20.09%, 20.08%, 20.17% and 20.13% respectively; On the basis of the original model and transfer learning, the model's accuracy is also improved when the CBAM attention mechanism is added in turn. Improved by 0.65%, 4.12%, 1.41% and 1% respectively, and the classification accuracy of the model with CBAM attention mechanism scheme 2 was the highest; Based on the original model, transfer learning and CBAM attention mechanism scheme 2.

AlphaDropout + SeLU and Alpha Dropout are added in turn. When + ReLU, the model's accuracy is increased by 1.42% and 1.37%, respectively. Transfer learning, when applied to the original model, results in the highest improvement in accuracy. CBAM attention mechanism scheme 2 and AlphaDropout + SeLU are added. Finally, the classification accuracy of the film damage image reaches 90.58%, which basically reaches the estimated value of the film damage classification accuracy in this paper.

| Original model | Transfer learning | CBAM Protocol 1 | CBAM Scheme 2 | CBAM Scheme 3 | CBAM Protocol 4 | AlphaDrop out +SeLU | AlphaDrop out+ReLU | Accuracy (%) |
|-------------------|----------------------|--------------------|------------------|------------------|--------------------|------------------------|-----------------------|-----------------|
| \checkmark | | | | | | | | 65 |
| \checkmark | \checkmark | | | | | | | 85.04 |
| \checkmark | | \checkmark | | | | | | 85.06 |
| \checkmark | | | \checkmark | | | | | 85.12 |
| \checkmark | | | | \checkmark | | | | 85.09 |
| \checkmark | | | | | \checkmark | | | 85.08 |
| \checkmark | | | | | | | | 85.17 |
| \checkmark | | | | | | | \checkmark | 85.13 |
| \checkmark | \checkmark | \checkmark | | | | | | 85.69 |
| \checkmark | \checkmark | | \checkmark | | | | | 89.16 |
| \checkmark | \checkmark | | | \checkmark | | | | 86.45 |
| \checkmark | \checkmark | | | | \checkmark | | | 85.04 |
| \checkmark | \checkmark | | \checkmark | | | \checkmark | | 90.58 |
| \checkmark | \checkmark | | \checkmark | | | | \checkmark | 90.45 |

TABLE IV. ABLATION EXPERIMENTS ON A TEST SET OF SELF-MADE FILM DAMAGE IMAGES

F. Comparison between the algorithm in this paper and the existing algorithm

To assess the effectiveness of the algorithm, detailed experimental comparison and analysis are carried out. Several classic convolutional neural network models, such as AlexNet, GoogLeNet, VGG series and ResNet series, are selected. Testing on the same image classification task, the accuracy, training time and the number of model parameters on the test set are the three key indicators. The experimental results are detailed in TABLE V.

Firstly, the accuracy of the test set is used to evaluate the classification performance of each model on the same dataset. By comparing these accuracy results, the performance differences between models in image classification tasks can be clearly observed. Additionally, training time is another key metric for assessing the efficiency of the algorithms. This study records the time each model takes to complete training under identical hardware conditions. and makes a comparative analysis. This is helpful to understand the computational complexity of different models in the training process and their feasibility in practical applications; Finally, the number of model parameters is the key factor to measure the complexity and storage requirements of the model. This paper counts the number of parameters of each model. A comparative analysis is made, which is helpful to evaluate the differences of different models in terms of resource occupation. And their applicability in different application scenarios.

| Model | Test Set Accuracy (%) | Training time (H) | Number of parameters |
|---------------|-----------------------|-------------------|----------------------|
| AlexNet | 63.28 | 57 | 57.02×106 |
| GoogLeNet | 60.59 | 42 | 46.88×106 |
| VGG16 | 70.02 | 135 | 130.38×106 |
| VGG19 | 67.19 | 142 | 139.59×106 |
| ResNet18 | 64.32 | 21 | 21.80×106 |
| ResNet50 | 65.01 | 25 | 25.56×106 |
| ResNet101 | 64.57 | 41 | 44.55×106 |
| CBAM-ResNet50 | 90.58 | 23 | 23.48×106 |

TABLE V. FILM DAMAGE CLASSIFICATION PERFORMANCE OF DIFFERENT MODELS

As shown in TABLE V, CBAM-ResNet50 network model's accuracy proposed in this paper achieves 90.58% on the test set, ranking first. Moreover, CBAM-ResNet50 has fewer parameters and fast model training speed, and can be applied to the identification and classification of film damage types. It is also more convenient for later embedding and development. Figure 7 is a histogram of the data in TABLE V, so that the changes in the data look clearer and more intuitive.



Figure 7. Film Damage Classification Performance of Different Models

G. Verifying the Generalization Ability of the CBAM-ResNet50 Network Model

To validate the generalization capability of the CBAM-ResNet50 network model proposed in this paper, the steel defect image dataset from Northeastern University is used for experimental verification. The original dataset is split into three parts: a training set with 1440 images, a validation set with 180 images, and a test set with 180 images,

following an 8:1:1 ratio. Data augmentation is applied to 80% of the training set, while the validation and test sets remain unchanged. As a result, the final training set contains 10080 images.

1) Confusion Matrix of Steel Defect Classification Identification Results on CBAM-ResNet50

A confusion matrix serves as a visual tool to assess the performance of classification models, and it is one of the important indicators to evaluate the results of models. It is often used to evaluate classifier models. In this research, the steel defect image dataset from Northeastern University was employed to evaluate the generalization capability of the CBAM-ResNet50 network model. The enhanced CBAM-ResNet50 model was applied to classify and recognize the test set data, the corresponding confusion matrix is shown in Figure 8.



Figure 8. Confusion Matrix of Steel Defect Classification Recognition Results on CBAM-ResNet50

In this study, the performance indicators of CBAM-ResNet50 model in the task of steel defect image recognition are calculated by using confuse matrix. As shown in TABLE VI.

| Damage category | Accuracy (%) | Accuracy (%) | Recall (%) | F_1 Fraction |
|--------------------|-----------------|-----------------|------------|-------------------|
| Press-in of scale | 95.56 | 96.67 | 96.67 | 96.67 |
| Patch | | 93.33 | 93.33 | 93.33 |
| Cracking | | 96.67 | 96.67 | 96.67 |
| Pit | | 96.67 | 96.67 | 96.67 |
| Impurities | | 93.33 | 93.33 | 93.33 |
| scratches | | 96.67 | 96.67 | 96.67 |

TABLE VI. STEEL DEFECT CLASSIFICATION AND IDENTIFICATION RESULTS OF CBAM-RESNET50 MODEL

TABLE VI demonstrates that the precision, recall, and score for all six types of steel defects exceed 93%. Notably, for defects such as iron oxide scale pressing, cracking, pitting, and scratching, these metrics surpass 96.5%. The overall accuracy rate is 95.56%. $F_1 F_1$.

From Figure VI, it is evident that the overall number of misrecognitions is low. This result fully proves that the CBAM-ResNet50 network model proposed in this paper shows excellent performance in the task of defect recognition. It also reflects its good generalization ability, which makes the model in the face of different types and different forms of defects. Can keep higher recognition accuracy, thereby greatly enhancing the universality and the reliability in practical application.

IV. CONCLUSIONS

Through image collection, image preprocessing and image enhancement, the data sets of four types of damage in film images are constructed. The data set is used to build the CBAM-ResNet50 network model. Transfer learning, CBAM attention mechanism, and the introduction of AlphaDropout module into the model with the use of SeLU activation function are used. The convolutional neural network shows improved classification accuracy on the experimental dataset. Our model outperforms traditional machine learning methods and other enhanced convolutional neural networkbased algorithms in terms of accuracy. The recognition rate is as high as 90. 58%, which proves that the CBAM-ResNet50 network model is suitable for the recognition and classification of film damage images. And the overall classification accuracy of CBAM-ResNet50 for the Northeastern University steel defect image data set reaches 95.56%, The results indicate that the CBAM-ResNet50 network model demonstrates strong generalization capabilities. It can also maintain a high recognition accuracy in the face of other types of defects.

REFERENCES

- [1] Li, X., Wang, Y., Zhang, L., et al. "Progress in the design and fabrication of advanced optical thin films for highpower laser applications." *Optics and Laser Technology*, 129, 106059, 2020.
- [2] Yang Haonan. Numerical simulation and film thickness prediction of optical thin film slit coating [D]. Xi'an University of Technology, 2023.
- [3] Li, Z., et al. "Photothermal photoacoustic detection for characterizing high-performance optical thin films." *Journal of Optical Society of America A*, 39(8), 1424-1431, 2021.
- [4] Jiang, X., Yang, L., et al. "Enhancement of laser-induced damage threshold in ZrO2 thin films using ion beam preirradiation." *Journal of Vacuum Science & Technology A*, 39(2), 021405, 2021.
- [5] CHANG Hao, YANG Zhenghao, WEI Yougong, et al. "Image detection technology of transverse tensile film fracture based on machine vision" [J]. *China Plastics*, 2021, 35 (12): 114-120.
- [6] Wei Dong. Research on surface defect detection algorithm of MEMS acoustic thin film based on microscopic machine vision [D]. Tianjin University, 2021.
- [7] ZHANG Zhenhua, LU Jingui, LI Haoran. "Optical film defect recognition based on improved convolutional neural network" [J]. Computer Applications and Software, 2021, 38 (04): 197-203.
- [8] Liu Peng. Design and Implementation of Defect Detection Software for Thin Film Coating Based on Digital Image Processing [D]. Heilongjiang University, 2019.
- [9] Zhai X, Meng Z, Wei J, et al. Intensity Measurement of Backscattered Light from Underwater Targets Using Femtosecond Pulse Laser[J]. Marine Geodesy,2019,42(3):316-326.
- [10] Wu, T., et al. "Study of laser damage identification of optical thin films by acoustic emission." *Applied Physics A*, 128(4), 227, 2021.
- [11] Su Junhong, Liang Haifeng, Xu Junqi. Discrimination of laser damage by light scattering method [J]. Journal of Vacuum Science and Technology, 2010, 30 (03): 325-328.
- [12] Andrea C, Ferm n M, Inmaculada P.Bunch transpiration is involved in the hastening of grape berry ripening under elevated temperature and low relative humidityconditions [J]. Plant Physiology and Biochemistry, 2024,206108258-.
- [13] Wu Xiaoye, Zhang Lichao, Shi Shi. Photothermal photoacoustic detection technology applied to characterization of high performance optical films [J]. Chinese Journal of Optics, 2014, 7 (05): 701-711.

- [14] Yang Lihong, Shi Wei, Su Junhong, et al. Study on improving damage threshold of ZrO2 thin films by laser pre-irradiation [J]. Journal of Xi' An University of Technology, 2012, 32 (02): 93-98.
- [15] ZHOU Zhanhong, LI Ji, HE Honglin, et al. Aeroengine residual life prediction method based on DBN-GRU [C]//Chinese Society of Aeronautics. Proceedings of the 6th China Aeronautical Science and Technology Conference. School of Aeronautical Manufacturing Engineering, Nanchang Aeronautical University, 2023:7.
- [16] Wei Dong. Research on surface defect detection algorithm of MEMS acoustic thin film based on

microscopic machine vision [D]. Tianjin University, 2021.

- [17] Shi Wei. Study on Laser Damage Identification of Optical Thin Films [D]. Xi' An University of Technology, 2012.
- [18] Suman K,P.D,Himanshu, et al. Exploring the impact of different annealing conditions on physical properties of CdSeTe thin films for solar cell absorberlayer applications [J]. Inorganic Chemistry Communications, 2024,159111779-.
- [19] Su Junhong, Lv Ning, Ge Jinman. Characteristics of plasma shock wave in laser film damage [J]. China Laser, 2016, 43 (12): 141-145.