

Received May 31, 2020, accepted June 22, 2020, date of publication June 29, 2020, date of current version July 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005450

# A New Image Recognition and Classification Method Combining Transfer Learning Algorithm and MobileNet Model for Welding Defects

HAIHONG PAN<sup>1</sup>, ZAIJUN PANG<sup>1</sup>, YAOWEI WANG<sup>1</sup>, YIJUE WANG<sup>2</sup>, AND LIN CHEN<sup>1</sup>

<sup>1</sup>School of Mechanical Engineering, Guangxi University, Nanning 530004, China

<sup>2</sup>Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269, USA

Corresponding author: Lin Chen (gdxcl@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51465005, in part by the Guangxi Science and Technology Major Project under Grant AA18118002, and in part by the Nanning Key Research and Development Project under Grant 20181018-1 and Grant 20181018-3.

**ABSTRACT** Welding quality directly affects the welding structure's service performance and life. Hence, the effective monitoring welding defects is essential to ensure the quality of the weld structure. Owing to the non-uniformity of the shape, position and size of welding defects, it is a complicated task to analyze and evaluate the acquired welding defects images manually. Fortunately, deep learning has been successfully applied to image analysis and target recognition. However, the use of deep learning to identify welding defects is time-consuming and less accurate due to the lack of adequate training data samples, which easily cause redundancy into the classifier. In this situation, we proposed a new transfer learning model based on MobileNet as a welding defect feature extractor. By using the ImageNet dataset (non-welding defect data) to pre-train a MobileNet model, migrate the MobileNet model to the welding defects classification field. This article suggested a new TL-MobileNet structure by adding a new Full Connection layer (FC-128) and a Softmax classifier into a traditional model called MobileNet. The entire training process of TL-MobileNet model has been successfully optimized by the DropBlock technology and Global average pooling (GAP) method. They can effectively accelerate the convergence rate and improve the classification network generalization. By testing the proposed TL-MobileNet on the welding defects dataset, it turned out our model prediction accuracy has arrived at 97.69%. The experimental results show that in several aspects, TL-MobileNet have better performance than other transfer learning models and traditional neural network methods.

**INDEX TERMS** Welding defects classification, feature extraction, deep learning, DropBlock, transfer learning, MobileNet.

## I. INTRODUCTION

As one of the main methods to connect workpieces, welding is an important part of the machine manufacturing line. Due to the influence of the environment and welding process, it is inevitable to produce various defects such as porosity, cracks and slag inclusion. Therefore, the study of the welding defects detection method has far-reaching significance for controlling product quality, improving service life and

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Tedesco<sup>1</sup>.

economic benefits. Usually, non-destructive testing (NDT) methods for welding defects mainly contain three categories: visual inspection [1], X-ray testing [2], [3] and ultrasonic testing [4], etc. In general, identification of defects in X-ray images is considered to the basic requirement for controlling welding quality in many industries. Based on X-ray evaluation theory, defects testing ways can be divided into manual the evaluation and computer-aided detection. Computer-aided detection technology relies on artificial intelligence technology to resolve disadvantages in manual evaluation process, such as time consumed and subjective

evaluation results. Thus, it has become a hot spot for more and more researchers and engineers.

Some image preprocessing techniques like image noise reduction [5], contrast enhancement [6], [7] and area of interest segmentation were used to segment the weld seam area in X-ray image to make the defect target area more prominent. In the defect classification stage, it is necessary to design a reliable classifier to distinguish different types of defects, presently many researchers have studied and discussed the development of different classification algorithms. Machine learning methods such as artificial neural network (ANN), support vector machine (SVM) and fuzzy system are the most widely used in the field of X-ray image defects recognition. The prime application of fuzzy theory in the field of welding defects detection was in the late 1990s [3], Liao [8] studied a fuzzy expert system method for classification of X-ray defect types, which has better classification accuracy than fuzzy k-nearest neighbor and multi-layer perceptron. Baniukiewicz [9] investigated a new type of compound classifier composed of fuzzy system and ANN. But there is a compromise between accuracy and interpretability in fuzzy defect detection. SVM and ANN are the most commonly used methods in defect detection. El Ouafi *et al.* [10] established a welding quality evaluation method of ANN by simulating welding parameters (welding time, current, voltage, thickness, etc.). Zapata *et al.* [11] modified the ANN to improve the detection accuracy of individual and overall defect characteristics. Yuan *et al.* [12] studied adaptive organization and adaptive feed-forward neural network to figure out the essential features of defects and effectively reduce identification errors. In order to obtain high accuracy and improve the efficiency of classification. Mu *et al.* [13] proposed an automatic classification algorithm combining principal component analysis (PCA) and SVM for selecting the optimal dataset. Inspired by this, Chen *et al.* [14] applied bees algorithm (BA) to extract defect features, used hierarchical multi-class SVM to obtain the accuracy up to 95%. Qi *et al.* [4] and Manasa and Nagarajah [15] provided an idea on how to optimize the feature redundancy process and improve classification efficiency and accuracy. Also, Extreme learning machine (ELM) is often used in image classification research because of its advantages in learning rate and generalization ability [16]. Su *et al.* [17] established an automatic defect identification system for solder joints by extracting texture features of welding defects. Han *et al.* [18] combined M-estimation with ELM and proposed a new ME-ELM algorithm, the algorithm can effectively improve the anti-interference and robustness of the model, and has high accuracy in the prediction of welding defects. Usually, these shallow machine learning methods are combined with the feature extraction process, which ultimately affects the machine learning prediction results. However, it is difficult to know which features should be extracted. Consequently, it is necessary to design efficient Deep Learning (DL) methods to realize automatic feature learning and welding defects prediction.

As a new field of machine learning, DL shows great potential in the field of defect detection, by continuously reducing the dimension in the process of feature learning to avoid the influence of feature extraction on the identification results, effectively improving the accuracy of defect detection. DL method has been applied to defect detection, including convolutional neural network (CNN) [19], deep belief network (DBN) [20] and sparse auto-encoder (SAE) [21]. Wang *et al.* [22] proposed a deep learn-based algorithm for X-ray image multi-defect type classification and automatic position recognition. Zhang *et al.* [21] studied SAE and particle swarm optimization (PSO) algorithm to realize real-time detection of welding defects. In addition, Hou *et al.* [19] adopted random oversampling, random under-sampling and synthetic minority over-sampling techniques to solve unbalanced sample defected dataset problem, and used deep convolutional neural network to identify porosity, cracks, slag inclusion and lack of penetration defects with an accuracy rate of 97.2%. Zhang *et al.* [23] achieved a high prediction accuracy on relatively small datasets of welding defects based on VGG-16 full convolution neural network. Nevertheless, in some areas the sample size is relatively small, which affects the prediction results. Thus, many researchers use transfer learning to overcome the problems of small samples, and use the deep CNN model trained on ImageNet as a feature extractor to migrate to the small dataset in another field and obtain good results [24]. It is worth noting that these small datasets are completely different from ImageNet. Zhang *et al.* [25] studied medical images by transfer learning methods and then obtained an identification accuracy of 97.041%. Ren *et al.* [26] researched the automatic surface detection of Decaf model based on deep transfer learning. Compared with other methods, the accuracy of Ren's method was improved by 0.66%-25.5% in classification task and 2.29%-9.86% in seven-minute defect detection. Yang *et al.* [27] used the mixed layer strategy to extract different scale features and obtained a high recognition accuracy in the small dataset military target recognition in the end. Since DL method achieves good effect in feature learning and avoids the influence on the prediction result, it has shown great potential in welding defects classification.

It is difficult to train the deep CNN model without a good training dataset, especially for the small sample size of welding defect labels. Thus, we proposed a new image recognition and classification method for welding defects, which combines the transfer learning algorithm and MobileNet model, namely TL-MobileNet model. This TL-MobileNet model has three advantages. (1) It can solve the problems of low prediction accuracy and time-consuming, which are induced by insufficient welding defects learning samples. Because this model combining transfer learning theory with trained MobileNet model form a welding defects feature extractor. (2) It has an enhanced feature extraction capability, since it added a new Fully Connected layer (FC-128) and a Softmax classifier after the MobileNet. The network layer structure gets deeper and the feature extraction level increases, the final

classification accuracy will be improved. (3) It can prevent the occurrence of over-fitting, and has a good generalization ability. Because the Global average pooling (GAP) and Drop-Block are integrated for the utilization of optimizing the entire training process in the TL-MobileNet model. This proposed TL-MobileNet will be tested on a welding defects dataset. And its good effect in welding defect recognition also will be proved by comparing with other methods (such as traditional MobileNet, Xception, VGG-16, VGG-19 and ResNet-50).

The rest of this paper is organized as follows: Section II, proposing the related welding defects classification model TL-MobileNet and DropBlock optimization algorithm; Section III, Experimental research on the classification of welding defects based on TL-MobileNet model; and Section IV, presenting the conclusion and future research work.

## II. ARCHITECTURE OF THE PROPOSED APPROACH

It is difficult to train a deep network structure with a small number of labeled samples in the welding defect recognition field, compared with the well-trained ImageNet dataset model with 14 million labeled images. Hence, the proposed TL-MobileNet model integrates Transfer learning & MobileNet for improving the classification accuracy of welding defects. the training process of TL-MobileNet is address, and the performance evaluation method of welding defect classification model is presented.

### A. TL-MOBILENET WELDING DEFECTS CLASSIFICATION MODEL

#### 1) TL-MOBILENET STRUCTURES

The weights and features of the MobileNet model are pre-trained in the source domain ImageNet [28] dataset (non-welded dataset), then they are transfer to the target domain for welding defects classification (Fig 1). The target domain does not use random initialization to start the data learning process from the beginning, and the model parameters are shared between the source domain and the target domain, so this method will help to improve the learning efficiency. In order to achieve the perfect classification effect of model training, the weights and features parameters of the training model of the migrated welding defect image dataset are fine-tuned.

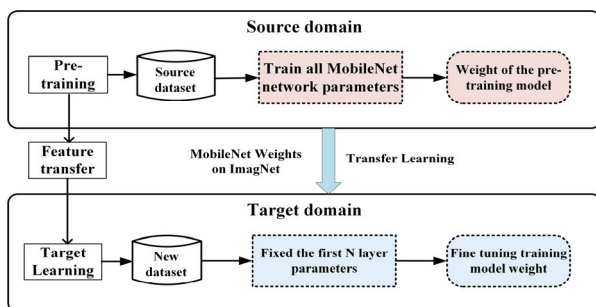


FIGURE 1. TL-MobileNet based on transfer learning model.

The structure of TL-MobileNet as shown in Fig 2, which includes three parts: 1) Data preprocessing, 2) Pre-trained MobileNet model initialization and 3) Defect classifier. The pre-training MobileNet model is composed of many convolutional layers, pooling layers and FC-1024. The number of neurons in the hidden layer of FC layer is 1024, which has 28 layers ( $1+2*13+1=28$ ) and is taken as the feature extraction layer for welding defects. The defect classifier has a Fully Connection layer FC-128 (new layer) and Softmax classifiers for improving the accuracy of welding defects classification. Thus, the TL-MobileNet has a depth of 29 layers. The Residual Connection Block (RCB) is the most important element for MobileNet. The RCB-1 and RCB-2 structures are used in the pre-trained MobileNet model (Fig 2b) to prevent gradient explosion. For welding defects classification, the multiple RCB-1 and RCB-2 blocks are superimposed after inputting the first convolution layer Conv3-32.

In Fig 2, the Conv3-128 indicates that the filter size in the convolutional layer is  $3 \times 3$  and its depth is 128. Conv1-128 indicates that the filter size in the convolutional layer is  $1 \times 1$ . FC-128 represents 128 neurons in the full connection layer. It is worth noting that the structure of Conv1-512 has 5 layers.

#### 2) RESIDUAL CONNECTION BLOCK

The RCB is based on the idea of shortcut connection to skip convolutional layers, which will help optimize the parameters of the training process and avoid the problem of gradient explosion in back propagation of errors.

RCB consists of multiple convolutional layers, batch normalization [29] and Rectified linear unit [30] (ReLU) function. Two different structures RCB-1 and RCB-2 are shown in Fig 3. RCB-1 represents stride=1 and the input and output feature sizes are the same, so the input and output are directly added, and  $F(x)$  represents the non-linear function of the convolution path, then the output of RCB-1 can be expressed as equation (1). RCB-2 means that the stride=2 and the input and output feature sizes are different, then the output of RCB-2 can be expressed by equation (2).

$$y = F(x) + x \quad (1)$$

$$y = F(x) \quad (2)$$

### B. OTHER RECOMMENDATIONS

#### 1) DATA AND PRE-PROCESSING

For x-ray images, segment image blocks establish a welding defect dataset according to the size of the network input. Then we divide the dataset into training data and test data, and label different defects.

#### 2) PRE-TRAINED MOBILENET MODEL INITIALIZATION

The weights of the pre-trained model on the ImageNet dataset are re-saved to the MobileNet feature extractor to classify welding defects (Fig. 2). Then the depth-separate convolution applies several filters to the local area of the input

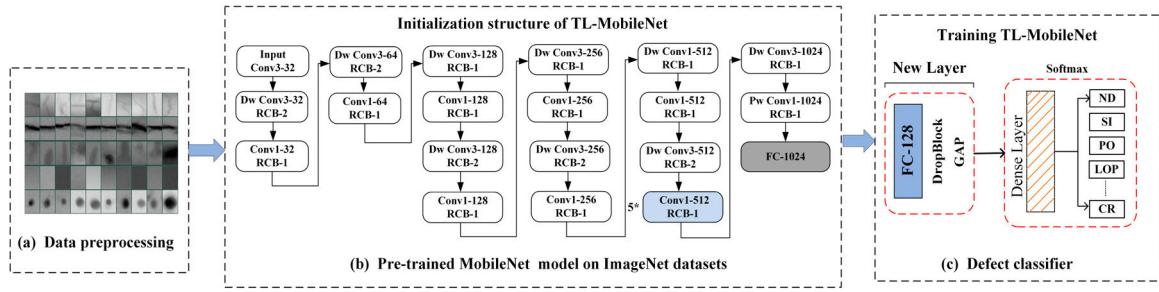


FIGURE 2. Technology Roadmap of the TL-MobileNet for welding defects classification.

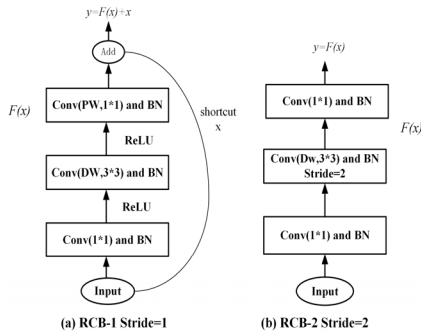


FIGURE 3. Residual connection block.

image to obtain the feature map of the welding defect image. When the input image block size is  $m*m$ , the given filter convolution process can be presented as in equation (3). Generally, the pooling layer is used to implement the downsampling operation behind the convolution layer for reducing the feature dimension and preventing the over-fitting. Thus, this model can extract a wider range of defect features. The maximum pooling process is shown in equation (4).

$$f_{i,s} = \text{soft max}(w^i x_s + b^i) \quad (3)$$

$$\text{pool}_s = \max(x_s) \quad (4)$$

where  $w^i$  represents the weight of the filter,  $x_s$  represents the input data,  $b^i$  is the bias of the filter  $i$ , and  $\sigma$  represents the activation function.

### 3) DEFECT CLASSIFIER

Defect features, obtained through pre-training, are input to the New Layer FC-128 and Softmax defect classifiers for training, and the final output is the probability of different classes of welding defects. During the training process, regularized DropBlock technology and GAP are used to reduce the amount of parameter calculation for the Fully Connected layer, prevent network over-fitting, and improve the accuracy of defect classification.

### C. PERFORMANCE EVALUATION USING CONFUSION MATRIX

A traditional confusion matrix method is used to evaluate the learning performance of TL-MobileNet. In the image classification, the confusion matrix is mainly used to compare the classification with the actual measurement value for describing the accuracy of model classification intuitively and accurately. the distribution of five kinds of welding defects can be directly identified by using the confusion matrix, and the evaluation indicators are shown in equation (5) and equation (6).

$$\text{precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{sensitivity} = \text{recall} = \frac{TP}{TP + FN} \quad (6)$$

where TP (True positive) is the real example, FN (False negative) is the false negative example, FP (False positive) is the false positive example, TN (True negative) is the true negative example, precision indicates accuracy, and recall indicates recall rate.

The  $F_{score}$  is used to evaluate the classification performance of TL-MobileNet on different types of defect datasets, which is shown in equation (7). If the value of  $F_{score}$  is closer to 1, then the performance of model classification for various defects is better.

$$F_{score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

## III. EXPERIMENTAL RESULTS

### A. WELDING DEFECTS DATASET

The dataset for subsequent experimental studies was from the public database (namely GDXray), which was provided by the BAM federal institute for materials research and testing in Berlin, Germany [2]. The ‘‘welding’’ defects in this database contain 88 defect images with different types and sizes. Based on prior knowledge cropping, defect samples were selected from the defect images by manual, and then the different defects were assigned class labels and added to the ‘Weld defect’ dataset as shown in Fig 4. There are total 6,208 defect

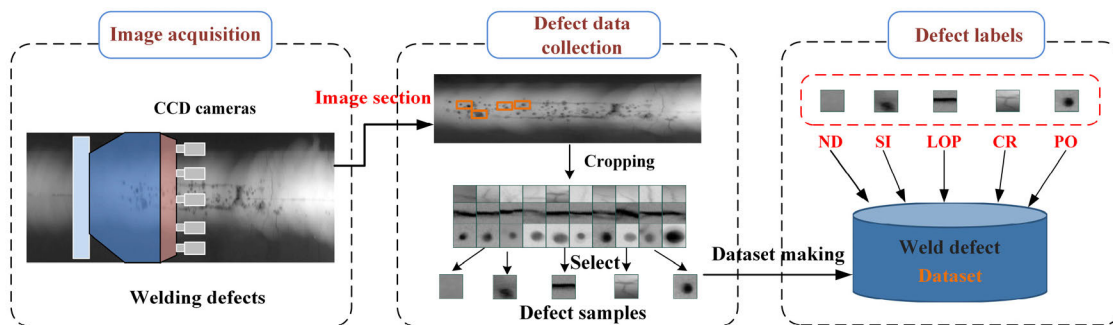


FIGURE 4. The establishment of the welding defects dataset.

TABLE 1. Welding defects dataset example.

Label	Type	Example	Sample
0	PO		1248
1	LOP		1496
2	CR		1104
3	SI		1168
4	ND		1192

TABLE 2. Experiment setting variables parameters.

Variables	Variable definitions	Values	
		TL-MobileNet	other Transfer Learning
Batch_size	Number of samples processed for each batch of training	64	64
Basic learning rate	The learning rate of the initial training setting	0.001	0.001~0.0001
Learning rate decay	After every 5 training, the learning rate decay coefficient	0.5	0.5
Training epoch	Total training iterations	50	50
Adam	Optimizer for iteratively updating parameters based on training data	$\beta_1=0.9, \beta_2=0.999$	$\beta_1=0.9, \beta_2=0.999$
DropBlock	A regularization technique in the convolution process	0.8	--
Learning strategy	Step: Learning rate change strategy	Step	Step

Note: '--' means no parameters; the  $\beta_1$  and  $\beta_2$  are the exponential decay rate.

samples, which contain five types of defects: non-defective (ND), lack of penetration (LOP), porosity (PO), slag inclusion (SI) and crack (CR) as exhibited in Table 1. The training/test ratio of the experimental data is set as 8: 2. That is, 80% of the experimental training data is randomly selected from the defect database, and the remaining 20% is used as the test dataset. It should be noted that the samples in all the training dataset are completely different from the samples in the test dataset.

**B. EXPERIMENTAL IMPLEMENTS**

The experimental running environment is Ubuntu18.04 with GTX 1080Ti GPU, and the programming environment is

implemented by python3.5. Where in Keras uses TensorFlow as the backend. All models can be found in Keras application website: <https://keras.io/applications/>.

Here, the feasibility of applying transfer learning to welding defects classification is demonstrated based on the Fig.2 (b). The performances of TL-MobileNet is compared with different transfer learning models (MobileNet, Xception, VGG-16, VGG-19, ResNet-50). During the training process, the variables / adjusted parameters were used as shown in Table 2. We use the “step” learning strategy. The basic learning rate of other transfer learning models is set to 0.001~0.0001 and decreases every 5 epochs with the factor of learning rate decaying 0.5. DropBlock is set to 0.8 in the

**TABLE 3.** Comparison results of TL-MobileNet model and transfer learning models in  $32 \times 32(\%)$ .

Run NO	TL-MobileNet	Transfer Learning Models				
		MobileNet	Xception	ResNet-50	VGG-16	VGG-19
1	<b>96.70</b>	96.62	97.50	62.08	92.03	92.91
2	<b>96.05</b>	96.62	97.26	76.09	95.33	93.40
3	<b>97.02</b>	97.10	96.20	89.05	95.25	93.48
4	<b>97.18</b>	94.12	96.46	62.88	96.46	95.33
5	<b>96.22</b>	95.25	96.50	75.28	96.78	91.55
6	<b>97.42</b>	97.10	96.42	74.88	95.89	95.81
7	<b>96.70</b>	95.65	97.18	75.28	96.94	94.93
8	<b>96.86</b>	94.77	97.34	85.83	96.86	95.89
9	<b>96.86</b>	96.62	96.94	66.99	95.57	93.16
10	<b>97.75</b>	95.49	96.22	73.19	94.85	95.94
Mean	<b>96.88</b>	95.94	96.80	74.16	95.59	94.18
Std	<b>0.48</b>	0.98	0.47	8.31	1.38	1.40

**TABLE 4.** Detection results of TL-MobileNet model and transfer learning models ( $32 \times 32 \times 3$ ).

Models	Mean Accuracy (%)	Running time(s)	Model size (MB)
TL-MobileNet	96.88	182.46	12.5
MobileNet	95.94	175.29	12.5
Xception	96.80	277.53	79.9
ResNet-50	74.16	335.35	90.4
VGG-16	95.59	186.99	58.9
VGG-19	94.18	193.96	76.4

TL-MobileNet model, whereas other transfer learning models do not have such optimized technique in the convolution process. Unless otherwise stated, all models are trained using Adaptive Moment Estimation (Adam).

### C. RESULTS AND DISCUSSION

The following section analyzed effects of different defects pictures sizes on the model prediction accuracy, the recognition efficiency and the model size, and discussed the feasibility of applying transfer learning to welding defects classification. In this research, the performance of different transfer learning models in extracting defect features and defect classification was proved.

#### 1) EVALUATION THE EFFICIENCY OF WELDING DEFECTS IDENTIFICATION( $M=32$ )

The prediction accuracy of different models at an image size  $m$  of  $32 \times 32$  is shown in Table 3. All models ran 10 times. The mean accuracy and the minimum accuracy of TL-MobileNet are 96.88% and 96.05%, respectively. The best accuracy of TL-MobileNet is 97.75%, which is better than the accuracy of other methods. The mean prediction accuracy of MobileNet, Xception, ResNet-50, VGG-16 and VGG-19 models are 95.94%, 96.80%, 74.16%, 95.59% and 94.18%, respectively. These indicate that TL-MobileNet model had good classification performances and was obviously better than other models.

The training parameters of experiment were set as Table 2. And the results of the mean accuracy, running time and model size of the entire classification process for different models are presented in Table 4. The prediction accuracy of TL-MobileNet increases by 0.94% and 2.7% than

MobileNet and VGG-19, respectively. The TL-MobileNet achieved 96.88% accuracy with running 182.46s. The accuracy of TL-MobileNet is 0.08% higher than that of Xception, but its model size is only 12.5MB and the spent time is about 2/3 of Xception. Because the TL-MobileNet employed the deep separable convolution to compress and accelerate in the training process, which can greatly reduce the number of model parameters. The TL-MobileNet model has less calculation time and model size compare with other transfer learning models, but it can acquire a higher prediction accuracy than other models. This indicates the TL-MobileNet has a potential in welding defects detection.

#### 2) THE INFLUENCE OF IMAGE SIZE $m$ ON PREDICTION ACCURACY

The influence of different input size  $m$  on the prediction accuracy of the model was researched. The value of  $m$  was set to  $32 \times 32$ ,  $64 \times 64$ ,  $96 \times 96$ ,  $128 \times 128$ , respectively, which was more suitable for defect identification of different sizes. The statistical parameters of running 10 times for each model, such as the maximum (Max), minimum (Min), mean accuracy and standard deviation (Std), were shown in Table 5. It is obvious that with the increase of the input defect image size, the accuracy of the model prediction will be improved to a certain extent, and the mean accuracy and standard deviation of TL-MobileNet are all better than those of VGG-16, VGG-19 and ResNet-50.

When the value of  $m$  is  $96 \times 96$ , TL-MobileNet achieves its best prediction result, and the best prediction accuracy is 98.95% by Xception model. Moreover. The prediction accuracy of TL-MobileNet was 3.96%, 0.84% and 4.01% higher than that of VGG-16, VGG-19 and ResNet-50

**TABLE 5.** Detection results of TL-MobileNet model and transfer learning models under different input sizes (%).

Size		TL-MobileNet	MobileNet	Xception	VGG-16	VGG-19	ResNet-50
$m=32 \times 32$	Max	<b>97.75</b>	97.10	97.50	96.94	95.94	89.05
	Min	<b>96.05</b>	94.12	96.20	92.03	91.55	62.08
	Mean	<b>96.88</b>	95.94	96.80	95.59	94.18	74.16
	Std	<b>0.48</b>	0.98	0.47	1.38	1.40	8.31
$m=64 \times 64$	Max	<b>98.39</b>	98.31	98.15	95.73	94.77	95.89
	Min	<b>96.94</b>	91.06	96.78	92.35	90.66	84.46
	Mean	<b>97.64</b>	96.59	97.39	94.40	92.80	90.32
	Std	<b>0.47</b>	2.03	0.46	1.06	1.30	3.81
$m=96 \times 96$	Max	<b>98.63</b>	98.15	<b>98.95</b>	95.97	97.18	96.30
	Min	<b>97.26</b>	97.23	96.05	90.58	96.05	91.38
	Mean	<b>97.69</b>	97.68	97.52	93.73	96.85	93.68
	Std	<b>0.45</b>	0.46	0.90	1.99	0.52	1.62
$m=128 \times 128$	Max	<b>99.03</b>	98.23	98.47	95.65	97.50	96.62
	Min	<b>96.05</b>	93.64	93.96	94.69	95.65	87.04
	Mean	<b>97.44</b>	97.23	97.33	94.94	96.68	93.18
	Std	<b>0.94</b>	1.38	1.28	0.79	0.55	2.87

models, respectively. The prediction accuracy of Xception is similar to that of TL-MobileNet. When  $m$  is  $128 \times 128$ , the maximum of prediction accuracy of TL-MobileNet is increase, but the mean accuracy decrease 0.25 than  $96 \times 96$ . However, the standard deviation of TL-MobileNet and Xception all increase. The results in Table 5 indicate that the image size can affect the defects identification and prediction accuracy. With the increase of image size, the accuracy of TL-MobileNet becomes significantly higher than that of MobileNet, which demonstrates the proposed TL-MobileNet is effective in welding defects identification.

Combined with the analysis of the experimental results in Table 5, when the welding defects image size is increased from  $32 \times 32$  pixels to  $96 \times 96$  pixels, the prediction accuracy of various transfer learning models is improved. This is because when the size of the picture is increased, the details extracted from the image that can describe the target feature are enlarged, accordingly the constructed TL-MobileNet model can obtain and learn more welding defects features from the enlarged picture, and improve the accuracy of defect prediction. Nevertheless, when the welding defect picture continues to increase from  $96 \times 96$  pixels to  $128 \times 128$  pixels, TL-MobileNet model's prediction accuracy rate drops. This is precise because the image resolution is inversely proportional to the picture size (resolution = pixel / size), when the pixel size is fixed, the larger the size will reduce the resolution of the picture, cause the picture to be blurry and some defects to be overlapped, and also affect the prediction accuracy of the model. In other words, it is not the larger the size of welding defect image is, the higher the prediction accuracy of small defects is. When the size of the welding defect image exceeds a size (for our test the size is  $96 \times 96$ ), the accuracy of welding defect prediction begins to decrease.

Confusion matrix is the most basic and intuitive method to measure the accuracy of classification model. According to equation (5) ~ equation (7), the confusion matrix of the TL-MobileNet model is calculated, which represents the recognition results of welding defects for different

size images. In Fig 5, the row of confusion matrix represents the actual weld defect type and the column is the predicted defect type. When the size of the input image is  $32 \times 32$ , the accuracy of the PO prediction is 97.30% and the probability of being misclassified as SI is 2.7% (Fig.5a). When the input image size is  $96 \times 96$ , the best results of prediction accuracy achieved for different welding defects (Fig.5c): the prediction accuracy of PO, LOP, CR, and ND is 99.61%, 99.65%, 100 % and 100%, respectively. However, when the input image size is  $128 \times 128$ , the probability of SI being misclassified as PO is as high as 8.29%, and the features of PO and SI feature details begin to overlap which affect the model recognition. It is obvious that the LOP, CR and ND are easily identified, but PO and SI are very difficult identified and easily lead to classification errors for any size.

In order to verify the prediction accuracy of the proposed TL-MobileNet model, the TL-MobileNet model was compared with other models for welding defect detection. The other models include: back propagation (BP) [4], K-nearest neighbors (KNN) [4], Extreme Learning Machine [18], Histogram of Oriented Gradients (HOG) [19], Convolutional neural networks (CNN) [31], artificial neural network (ANN) [11], support vector machines with principal component analysis (PCA-SVM) [13] and Extreme learning machine [17]). Table 6 represents the comparisons between the accuracy of the proposed method and that of other researchers in the prediction of welding defects (the input image size  $96 \times 96$ ). The TL-MobileNet and ELM models have stronger robust than the KNN, ANN and BP models. There are large differences (14.69%) between the ANN and CNN models. It is worth mentioning that the prediction accuracy of the Single-ELM model is 95.45%, higher than that of other traditional neural network methods. However, the prediction accuracy of TL-MobileNet is 97.69%, ranking the second highest in the queue, and close to that of Ensemble-ELM with a value difference of only 0.24%.

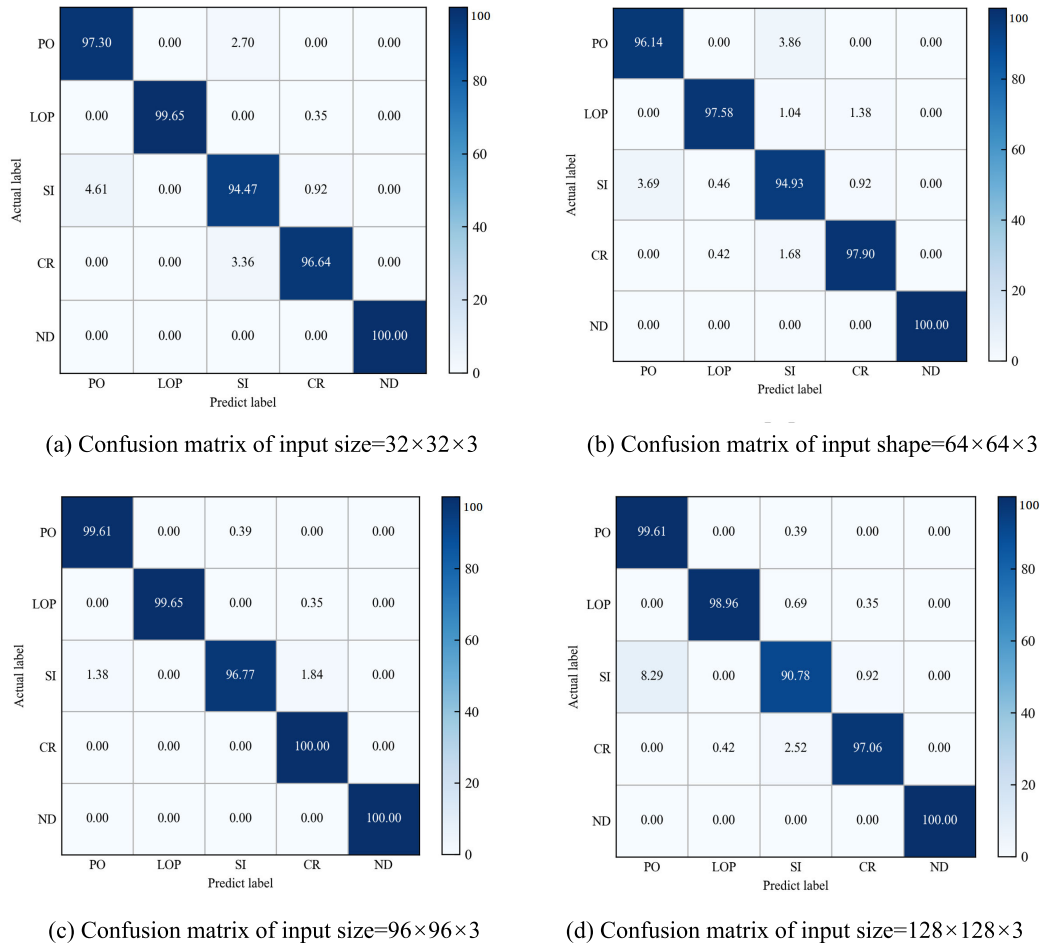


FIGURE 5. Confusion matrix of TL-MobileNet results.

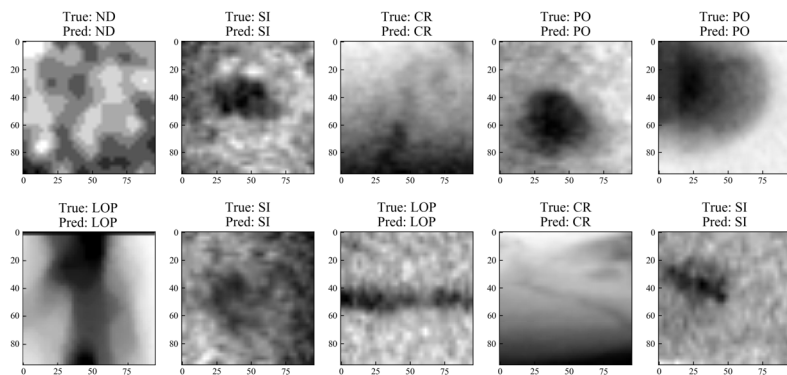


FIGURE 6. Visualization results of welding defects dataset for input size=96 × 96 × 3.

The visualization prediction results of a welding defects by TL-MobileNet are shown in Fig 6. Obviously, in this test identification process, the TL-MobileNet model has better recognition for each welding defect type, and prediction type is the same as the actual type without misjudgment. The features of PO and LOP are significantly distinguishable.

However, sometimes the defect images for SI and CR are similar and they are difficult to distinguish (Fig 6), which can lead to misjudgments (Fig 5).

Table 3~Table 6 indicate the proposed TL-MobileNet model can obtain good results for different types of welding defects, even in small sample datasets. Because the



TABLE 6. Mean accuracy of different methods.

Method	Mean Accuracy	
Traditional Methods	BP [4]	89.00%
	KNN [4]	93.00%
	HOG [19]	81.60%
	CNN [31]	94.69%
	ANN [11]	80.00%
PCA-SVM [13]	90.75%	
ELM Method	Single-ELM [17]	95.45%
	Ensemble-ELM [17]	97.93%
<b>Proposed Method</b>	<b>TL-MobileNet</b>	<b>97.69%</b>

TL-MobileNet model combined the transfer learning, FC-128 (new layer) and Softmax classifiers, it has the advantages in feature learning. Furthermore, TL-MobileNet is better than MobileNet, Xception, VGG-16, VGG-19, ResNet-50 and traditional methods (such as BP, KNN, HOG, CNN, ANN, PCA-SVM, ELM) in terms of average prediction accuracy and standard deviation, which further shows the potential of this model in welding defects detection.

#### IV. CONCLUSION AND FUTURE RESEARCHES

This paper proposes a TL-MobileNet for welding defects detection. The experiments of welding defects classification using the 'Weld' dataset verify that the TL-MobileNet can accurately identify specific defects on a limited number of training samples. A number of experiments have been carried out for various size images using different transfer learning models. It proves that the proposed TL-MobileNet method has better recognition accuracy with smaller model size and less calculation time. The results demonstrate that the feature extracted by TL-MobileNet is significantly better than the traditional method in the classification task. The proposed TL-MobileNet will be applied in the actual industry to improve the accuracy of the defect identification for welding products.

There are two limitations for TL-MobileNet: (1) building datasets of the welding defect usually requires the prior knowledge to manually select the defect image, which is easy to be misclassified into other types of defects in the selection process. (2) this study did not involve welding defects in on-line detection. In the future, we will collect more welding defect samples from diversified production and working environments to expand the dataset. Furthermore, to improve prediction accuracy and detection efficiency the TL-MobileNet will be optimized and it will be implemented on-line through deploying on the mobile terminal.

#### ACKNOWLEDGMENT

(Haihong Pan and Zaijun Pang contributed equally to this work and should be regarded as co-first authors.)

#### REFERENCES

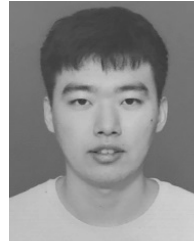
- [1] M. Shah Hairol Nizam, A. Rashid Mohd Zamzuri, S. Marizan, and S. Ahmad Zaki, "Vision based identification and classification of weld defects in welding environments: A review," *Indian J. Sci. Technol.*, vol. 9, no. 20, pp. 1–15, May 2016, doi: 10.17485/ijst/2016/v9i20/82779.
- [2] D. Mery, V. Rizzo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "GDxRay: The database of X-ray images for nondestructive testing," *J. Nondestruct. Eval.*, vol. 34, no. 4, pp. 34–42, Dec. 2015, doi: 10.1007/s10921-015-0315-7.
- [3] T. W. Liao and J. Ni, "An automated radiographic NDT system for weld inspection: Part I—weld extraction," *NDT E Int.*, vol. 29, no. 3, pp. 157–162, 1996, doi: 10.1016/0963-8695(96)00009-6.
- [4] A. L. Qi, J. F. Wang, F. Wang, U. Idachaba, and G. Akanmu, "Welding defect classification of ultrasonic detection based on PCA and KNN," *Appl. Mech. Mater.*, vols. 380–384, pp. 902–906, Apr. 2013, doi: 10.4028/www.scientific.net/AMM.380-384.902.
- [5] J. Kumar, R. S. Anand, and S. P. Srivastava, "Flaws classification using ANN for radiographic weld images," in *Proc. Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2014, pp. 145–150, doi: 10.1109/SPIN.2014.6776938.
- [6] K. Ali, M. Awan, A. Jalil, and F. Mustansar, "Localization and classification of welding defects using genetic algorithm based optimal feature set," in *Proc. Int. Conf. Inf. Commun. Technol. (ICICT)*, Dec. 2015, pp. 1–6, doi: 10.1109/ICICT.2015.7469485.
- [7] G. D'Angelo and S. Rampone, "Feature extraction and soft computing methods for aerospace structure defect classification," *Measurement*, vol. 85, pp. 192–209, May 2016, doi: 10.1016/j.measurement.2016.02.027.
- [8] T. W. Liao, "Classification of welding flaw types with fuzzy expert systems," *Expert Syst. Appl.*, vol. 25, no. 1, pp. 101–111, 2003, doi: 10.1016/S0957-4174(03)00010-1.
- [9] P. Baniukiewicz, "Automated defect recognition and identification in digital radiography," *J. Nondestruct. Eval.*, vol. 33, pp. 327–334, Dec. 2014, doi: 10.1007/s10921-013-0216-6.
- [10] A. El Ouafi, R. Bélanger, and J.-F. Méthot, "An on-line ANN-based approach for quality estimation in resistance spot welding," *Adv. Mater. Res.*, vol. 112, pp. 141–148, May 2010, doi: 10.4028/www.scientific.net/AMR.112.141.
- [11] J. Zapata, R. Vilar, and R. Ruiz, "Automatic inspection system of welding radiographic images based on ANN under a regularisation process," *J. Nondestruct. Eval.*, vol. 31, no. 1, pp. 34–45, Mar. 2012, doi: 10.1007/s10921-011-0118-4.
- [12] P. X. Yuan, C. C. Zhang, and Y. Yuan, "Research on welding line defect recognition of the in-service pipeline using X-ray detecting," *Appl. Mech. Mater.*, vols. 195–196, pp. 5–12, Aug. 2012, doi: 10.4028/www.scientific.net/AMM.195-196.5.
- [13] W. Mu, J. Gao, H. Jiang, Z. Wang, F. Chen, and C. Dang, "Automatic classification approach to weld defects based on PCA and SVM," *Insight-Non-Destructive Test. Condition Monitor.*, vol. 55, no. 10, pp. 535–539, Oct. 2013, doi: 10.1784/insi.2012.55.10.535.
- [14] Y. Chen, H.-W. Ma, and G.-M. Zhang, "A support vector machine approach for classification of welding defects from ultrasonic signals," *Nondestruct. Test. Eval.*, vol. 29, no. 3, pp. 243–254, Jul. 2014, doi: 10.1080/10589759.2014.914210.
- [15] M. Kesharaju and R. Nagarajah, "Feature selection for neural network based defect classification of ceramic components using high frequency ultrasound," *Ultrasonics*, vol. 62, pp. 271–277, Sep. 2015, doi: 10.1016/j.ultras.2015.05.027.
- [16] M. Ahmad, S. Shabbir, D. Oliva, M. Mazzara, and S. Distefano, "Spatial-prior generalized fuzziness extreme learning machine autoencoder-based active learning for hyperspectral image classification," *Optik*, vol. 206, Mar. 2020, Art. no. 163712, doi: 10.1016/j.ijleo.2019.163712.
- [17] L. Su, L. Wang, K. Li, J. Wu, G. Liao, T. Shi, and T. Lin, "Automated X-ray recognition of solder bump defects based on ensemble-ELM," *Sci. China Technol. Sci.*, vol. 62, no. 9, pp. 1512–1519, Sep. 2019, doi: 10.1007/s11431-018-9324-3.
- [18] Y. Han, Z. Li, and J. Ye, "Improving Stability of Welding Model with ME-ELM," in *Transactions on Intelligent Welding Manufacturing*. Springer, 2018, pp. 61–77. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-981-10-7043-3\\_4](https://link.springer.com/chapter/10.1007/978-981-10-7043-3_4), doi: 10.1007/978-981-10-7043-3\_4.
- [19] W. Hou, Y. Wei, Y. Jin, and C. Zhu, "Deep features based on a DCNN model for classifying imbalanced weld flaw types," *Measurement*, vol. 131, pp. 482–489, Jan. 2019, doi: 10.1016/j.measurement.2018.09.011.

- [20] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 92–104, May 2016, doi: [10.1016/j.ymssp.2015.11.014](https://doi.org/10.1016/j.ymssp.2015.11.014).
- [21] Y. Zhang, X. Gao, D. You, and N. Zhang, "Data-driven detection of laser welding defects based on real-time spectrometer signals," *IEEE Sensors J.*, vol. 19, no. 20, pp. 9364–9373, Oct. 2019, doi: [10.1109/JSEN.2019.2927268](https://doi.org/10.1109/JSEN.2019.2927268).
- [22] Y. Wang, F. Shi, and X. Tong, "A welding defect identification approach in X-ray images based on deep convolutional neural networks," in *Proc. Int. Conf. Intell. Comput.*, 2019, pp. 53–64, doi: [10.1007/978-3-030-26766-7\\_6](https://doi.org/10.1007/978-3-030-26766-7_6).
- [23] X. Zhang, Z. Gong, and B. Liu, "Weld defect images classification with VGG16-Based neural network," *Proc. Digit. TV Wireless Multimedia Commun.*, 2017, pp. 215–223, doi: [10.1007/978-981-10-8108-8\\_20](https://doi.org/10.1007/978-981-10-8108-8_20).
- [24] Y. Jason, C. Jeff, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 3320–3328.
- [25] S. Zhang, F. Sun, N. Wang, C. Zhang, Q. Yu, M. Zhang, P. Babyn, and H. Zhong, "Computer-aided diagnosis (CAD) of pulmonary nodule of thoracic CT image using transfer learning," *J. Digit. Imag.*, vol. 32, no. 6, pp. 995–1007, Dec. 2019, doi: [10.1007/s10278-019-00204-4](https://doi.org/10.1007/s10278-019-00204-4).
- [26] R. Ren, T. Hung, and K. C. Tan, "A generic Deep-Learning-Based approach for automated surface inspection," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 929–940, Mar. 2018, doi: [10.1109/TCYB.2017.2668395](https://doi.org/10.1109/TCYB.2017.2668395).
- [27] Z. Yang, W. Yu, P. Liang, H. Guo, L. Xia, F. Zhang, Y. Ma, and J. Ma, "Deep transfer learning for military object recognition under small training set condition," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6469–6478, Oct. 2019, doi: [10.1007/s00521-018-3468-3](https://doi.org/10.1007/s00521-018-3468-3).
- [28] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. F. Li, "ImageNet: A large-scale hierarchical image database, series ImageNet," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2009, pp. 248–255. [Online]. Available: <https://ieeexplore.ieee.org/document/5206848>, doi: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- [29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [30] G. Hinton and V. Nair, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 1–8.
- [31] Zhu, Ge, and Liu, "Deep learning-based classification of weld surface defects," *Appl. Sci.*, vol. 9, no. 16, p. 3312, Aug. 2019, doi: [10.3390/app9163312](https://doi.org/10.3390/app9163312).



**ZAIJUN PANG** received the B.S. degree in mechanical manufacturing and automation from Wuhan Textile University, Wuhan, China, in 2018. He is currently pursuing the degree in mechatronic engineering with Guangxi University.

His research interests include machine learning and intelligent algorithm.



**YAOWEI WANG** received the B.S. degree in the IoT engineering from Hohai University, Nanjing, China, in 2016. He is currently pursuing the degree in mechanical and electronic engineering with Guangxi University.

His research interests include computer vision and deep learning.



**YIJUE WANG** received the M.Sc. degree in statistics from the University of Connecticut, Storrs, CT, USA, in 2017, where he is currently pursuing the Ph.D. degree in computer science.

Since 2015, he has been working with USA Insurance Company and Pharmacy Company on risk prediction problem with statistical analysis and machine learning algorithm. He has worked as a Research Assistant on different projects about closest pair algorithm and membership privacy

with deep learning model. His research interests include membership attack protection in machine learning model, closest pair algorithm, and battery management.



**HAIHONG PAN** received the M.E. degree in mechatronic engineering from Guangxi University, Nanning, China, in 1991, and the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007. He is currently a Professor and the Vice President of the College of Mechanical Engineering, Guangxi University. His research interests include numerical control, virtual reality, electromechanical detection and control technology use in rehabilitation robot and industrial robot manipulators, deep learning, computer vision, and intelligent manufacturing.



**LIN CHEN** received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2008. She is currently a Professor with the College of Mechanical Engineering, Guangxi University. Her research interests include machine learning, numerical control, virtual reality, electromechanical detection and control technology use in rehabilitation robot and industrial robot manipulators, and battery management system in electric vehicles.

...