A Deep Transfer Learning Model for Packaged Integrated Circuit Failure Detection by Terahertz Imaging

Yao Lu, Qi Mao, and Jingbo Liu

ABSTRACT Terahertz time-domain spectroscopy imaging system (THz-TDS) is becoming a promising tool for packaged integrated circuit (IC) failure detection due to its nonmetal penetrability and low radiation. However, two major obstacles are hindering the industrial application of the THz-TDS based IC detection method: 1) the low resolution of THz images may affect the detection accuracy; 2) the failure detection tasks are always carried out manually, which is inefficient and inaccurate. Thus, in this paper, we firstly enhanced the quality of IC THz images with a deconvolution algorithm and a mathematically simulated point spread function (PSF), and then we proposed a deep convolutional neural network (CNN) based failure detection framework to achieve end-to-end IC inspection automatically. Besides, we introduced transfer learning to overcome the limitation of the IC dataset size. The result demonstrated that our proposed method achieved excellent performance concerning both accuracy and efficiency.

INDEX TERMS THz-TDS imaging, Image enhancement, Failure detection, Convolutional neural networks, Transfer learning

I. INTRODUCTION

Detection of hidden failures in integrated circuit (IC) packages is vital as there is an increasing demand for quality and reliability in ICs. There are many causes of IC failures. One of the most common types is electrical overstress (EOS), which induces thermal shock, and therefore, generates hidden internal defects, such as cracks, delamination, and voids of packaging materials [1, 2].

Accordingly, a variety of failure analysis systems have been investigated to detect hidden defects in IC packages. Conventional failure detection methods include X-ray [3] and scanning acoustic microscopy (SAM) [4]. These approaches are practical but still have shortcomings [5]. To be more specific, X-ray transmission imaging may be harmful to both samples and operators due to its ionizing photon energy. Besides, X-ray cannot detect defects caused by thermal fatigue or EOS, as its energy does not change when passing through air in the defect locations [1, 2]. SAM, on the other hand, requires immersion of samples in a liquid environment, which may lead to the oxidation of these electronic devices [6].

Terahertz time-domain system (THz-TDS) is an emerging non-destructive inspection technology, which can penetrate nonmetallic materials and offer information about the composite damages [7]. Terahertz wave has no known safety issues to both samples and human bodies as its low-energy radiation (0.1–10 THz). In addition, it does not require samples exposed to liquid while testing, which enables easy and safe inspection. For these reasons, several studies have investigated the use of terahertz imaging for packaged IC failure detection. According to the distance from the aperture to the detected objects, the imaging systems used in these studies could be classified into two major categories: the near-field and the far-field THz imaging systems. The near-field imaging system could overcome the distortion effects of diffraction and provide better spatial resolution. For instance, the laser terahertz-emission microscope (LTEM) system achieves a resolution of 3 µm, and it has been successfully applied to inspect the cut interconnection lines in damaged integrated circuits [8, 9]. Besides LTEM, the Electro Optical Terahertz Pulse Reflectometry (EOTPR) was also developed

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515110703 and No.2020A1515110146).
to locate and clarify the fault of 3D packaged ICs [10, 11], and it has been used in semiconductor failure analysis laboratories in AMD, IBM Intel and TSMC. However, in near-field imaging systems, the objects should be placed at a subwavelength distance from the lens. Thus, objects thicker than approximately 100 µm cannot be imaged using this technique [12, 13]. Far-field THz imaging is a more feasible solution for some applications, for example, far-field THz imaging could be used for detecting failures of ICs caused by EOS and ESD [1]. And it can also measure the depth of the hidden voids in IC packages caused by overheating [2]. In addition, it can distinguish counterfeit ICs from authenticated ones by detecting unexpected materials of IC packages [13].

The first IC THz image was derived by Hu and Nuss in 1995 [21]. Years later, researchers have applied deep learning in automated detection of packaged ICs due to its unique strengths [3, 7]. The first IC THz image was derived by Hu and Nuss in 1995. Then, researchers have applied deep learning in automated detection of packaged ICs due to its unique strengths [3, 7]. The first IC THz image was derived by Hu and Nuss in 1995. Years later, researchers have applied deep learning in automated detection of packaged ICs due to its unique strengths. The first IC THz image was derived by Hu and Nuss in 1995.

In this study, we have introduced a CNN based packaged IC detection method. To strengthen its performance, we firstly enhanced the quality of IC THz images. To be specific, we simulated a point spread function (PSF) that could describe the degradation process of THz images accurately and then deconvoluted the THz images with the mathematically modeled PSF. Then, we introduced transfer learning to overcome the limitation of the IC dataset size. In the training process, the weights of the first few convolutional layers are pre-trained from ImageNet, the latter convolutional and FC layers are fine-tuned from the IC dataset. In this way, TL dramatically reduced the number of parameters needed to be trained and made it feasible to train a deeper network with a limited number of labeled data. In addition, to further reduce the computational complexity and memory requirement, we introduced Mobilenet V2 [29] as our base model since Mobilenet V2 is a lightweight CNN that needs comparatively less computing resource while maintaining good accuracy. Comparative experiments with VGG-16 [30] and Resnet-50 [31] demonstrated that our proposed method achieved excellent performance concerning both accuracy and efficiency. The main contributions of this paper can be summarized as follows:

1) We applied a THz image enhancement method based on the deconvolution algorithm and a mathematically simulated PSF. The method significantly enhanced the resolution of IC THz images.

2) We proposed a CNN based automated IC failure detection method and introduced TL to overcome the limitation of the IC dataset size, yield a better inspection accuracy.

3) We show that our proposed framework leads to a state-of-the-art result for packaged IC failure detection.

The rest of the paper is organized as follows. Section 2 discusses the related works. Section 3 presents the methodologies of the proposed approach, including Terahertz imaging and image enhancement, convolutional neural networks, and transfer learning. Section 4 presents the case studies. The conclusion and future researches are presented in Section 5.

II. RELATED WORK

A. TERAHERTZ IMAGING FOR IC DETECTION

Terahertz spectrum has been applied as a non-destructive testing method of packaged ICs due to its unique strengths. The first IC THz image was derived by Hu and Nuss in 1995. They have provided a feasible way for IC packages inspection, as THz radiation can transmit through plastic housing and reveal metal interconnects [7]. Years later, Keenan et al. presented an approach for detecting failures in ICs and PCBs caused by EOS and ESD. The damaged regions hidden in IC packages have been successfully identified through THz imaging [1]. Besides EOS and ESD, THz systems have also been investigated to detect other types of damages. For instance, Park et al. utilized the
reflection mode of the THz-TDS imaging system, successfully measured the thickness and depth of the voids in IC packages caused by overheating [2]. THz-TDS has also been applied to authenticate ICs. Ahi detected unexpected materials such as blacktopping layers and sanded and contaminated components by THz spectrum, in order to distinguish counterfeit ICs from authenticated ones [13]. The systems used in above-mentioned researches are all far-field THz imaging systems. In order to achieve micron-level detection, near-field THz imaging was applied for IC failure detection. For instance, Kiwa et al. proposed the laser terahertz-emission microscope (LTEM) system, which has clearly observed the cut interconnection lines in damaged chips [8]. Soon after, they redesigned their optical setup and improved the spatial resolution of the system to below 3 µm [9]. Yamashita et al. has also successfully utilized a LTEM to distinguish damaged circuits with disconnected wires from the normal ones [32]. Besides LTEM, the Electro Optical Terahertz Pulse Reflectometry (EOTPR) was also developed to locate and clarify the faults of 3D packaged ICs [10, 11]. It has been used in semiconductor failure analysis laboratories in AMD, IBM Intel, and TSMC.

B. TERAHERTZ IMAGE ENHANCEMENT

In general terms, there are two methods to enhance the quality of THz images. The first approach is multi-image super-resolution (MISR) methods. For instance, Xu et al. [33] have proved the effectiveness of projection onto convex sets (POCS) and iterative back projection (IBP) algorithms on THz image enhancement. However, it is time-consuming to obtain multiple images with the THz raster-scanning imaging system. Thus, the second method, single image super-resolution (SISR) is a more accessible method for practical applications. Some optical imaging algorithms were first investigated to improve the quality of THz images, such as, Liang et al. [34] used histogram equalization, wavelet denoising, and canny algorithms to enhance the contrast of the image, denoise and get the edge of the image, respectively. But these methods did not consider the blur problem caused by the convolution operation of the object and PSF. To resolve this issue, the Lucy–Richardson deconvolution algorithm was applied for THz image restoration. According to the PSF of the imaging system is known or not, the THz image enhancement researches can be classified into two categories: non-blind and blind methods. The blind methods only utilize the prior information of the blurred image itself, and then use regularization term (such as Wong et al. applied the total variation [35] and Krishnan et al. utilized the normalized sparsity measurement [36]) to estimate the PSF and improve its spatial resolution subsequently. Apart from blurred images, the non-blind methods also apply measured (Ning et al. [37]) or simulated (Ahi [38, 39]) PSF to restore the THz images.

C. CONVOLUTIONAL NEURAL NETWORKS AND TRANSFER LEARNING

Convolutional neural networks (CNNs) were first proposed by LeCun et al. [40] in 1989. However, they did not draw much attention until AlexNet won the ImageNet competition in 2012 [41]. Since then, CNN has been successfully applied in many aspects due to its outstanding feature extraction abilities, such as image classification, face recognition, object detection, etc. [42]. Besides optical images, it has also been used to deal with terahertz images. Zhang et al. [43] adopted a faster region-based convolutional neural network (Faster R-CNN) to detect concealed weapons carried on personnel and achieved an encouraging result. Yang et al. [44] improved the effectiveness and efficiency of the concealed objects’ detection by taking two measures, firstly, they employed a deeper CNN model (VGG-16) for THz images classification. Secondly, they utilized the spatio-temporal information via a sparse and low-rank decomposition model. Motivated by these works, this paper proposed a CNN based automated IC failure detection method by using THz images.

Generally, the volume of labeled samples is relatively small, and it is almost impossible to train a deep CNN model without a large amount of labeled data. A practical solution is pretrained deep CNNs on ImageNet and then fine-tuned its parameters on the target dataset. CNNs pretrained on ImageNet can also perform well on a small dataset in another different domain [23]. It has been proved in many fields. Donahue et al. [26] evaluated the transferability of the deep convolutional activation features extracted from a large labeled dataset, concluded that DeCAF has strong representational and generalization ability to perform visual tasks. Tajbakhsh et al. [27] compared the performance of pre-trained CNNs with fully trained CNNs on four distinct medical imaging applications, demonstrated that fine-tuned CNNs could yield a better result for medical imaging applications. Cetic et al. [28] conducted plenty of CNN fine-tuning experiments for five different art-related classification tasks and achieved state-of-the-art classification results. Shao et al. [24] developed a deep learning framework for fault diagnosis. They used a pre-trained VGG-16 network and fine-tuned the last two convolution blocks and fully connected layers with three different testing datasets. Comparative analysis showed that their proposed approach achieves state-of-the-art results.

Though TL greatly improved the performance of CNNs, some conventional networks with complicated structures and numerous parameters, such as VGG and ResNet, still require a great amount of computing power and time, which in turn limited their application on mobile or embedded devices [45]. To solve this issue, Baheti et al. [46] modified the architecture of VGG-16 by replacing its FC layers with convolution layers. Experiment results showed that this modification not only reduced the number of the parameters from 140 million to 15 million, but also enhanced its performance. Pan et al. [47] proposed a welding defect
III. METHODS

A. TERAHERTZ IMAGING AND IMAGE ENHANCEMENT

THz time-domain spectroscopy system (THz-TDS) is a powerful tool for detecting packaged object defects. When obtaining a time-domain THz image, the sample is raster-scanned point by point in a 2-D plane, and the transmitted or reflected THz beam is recorded simultaneously. In this way, we can get a 3-D data cube with the third dimension being time [49]. Then a fast Fourier transform (FFT) is always performed on the time dimension to get frequency-domain THz images, which can present information about the absorption, reflection, or scattering losses and usually give better imaging results. The frequency-domain THz images can also be viewed as a 3-D data cube, with the third dimension being frequency. After FFT transformation, a high pass filter (HPF) is usually applied to filter out the low-frequency spectrum, since the distortion effects of the diffraction are higher for the lower wavelengths. Then a low pass filter (LPF) is performed to suppress the noise. As higher-frequency spectrum usually has a lower signal to noise ratio (SNR).

The scanning process could be expressed as follows:

\[ i(x, y, z) = o(x, y, z) \otimes PSF(x, y, z) + n \]  

where \( i(x, y, z) \) is the THz image, \( o(x, y, z) \) is the object function, the \( x-y \) plane is the raster-scanned coordinate (focal plane), the \( z \)-axis is the traveling path of the THz beam and \( z = 0 \) is the focal plane, PSF is the point spread function, \( \otimes \) denotes the convolution operation, which is a mathematically modeling of the raster-scanning operation, and \( n \) is the additive system noise.

Previous studies have already revealed that the key aspect of THz image restoration is the construction of the PSF. In this study, we applied the mathematical modeling of the PSF proposed by Ahi [39], which has been proved technically feasible in many studies [13, 38, 50], and could be illustrated in Fig. 1 and expressed as:

\[ PSF(x, y, z, f) = \sum_{i} \left( \frac{2}{\pi \omega(z, f)} \right)^{2} \exp \left( -2 \left( \frac{x^2 + y^2}{\omega(z, f)^2} \right) \right) \]  

where \( x, y \) is the position from the center of the beam, \( f \) is the frequency of the THz beam, \( \omega(z, f) \) is the spot radius of the beam at distance \( z \) from the focal plane and could be calculated by:

\[ \omega(z, f) = \omega(0, f) \left( 1 + \left( \frac{\lambda z}{2 \pi \omega(0, f)} \right)^2 \right)^{1/2} \]  

and \( \omega(0, f) \) is the spot radius at the focal plane:

\[ \omega(0, f) = \frac{0.610}{\sqrt{2 \ln 2}} \cdot \frac{c}{f \cdot NA} \]

where \( c \) is the speed of light, and NA is the numerical aperture of the imaging system.

FIGURE 1. Visualization of the blur kernels: (a) Modeled PSF profile and (b) Modeled PSF from above. [1 pixel = 0.05mm].

B. CONVOLUTIONAL NEURAL NETWORKS

CNNs are designed to simulate the way that biological neural networks handle images. Unlike traditional artificial neural networks (ANNs), CNNs can extract hierarchical features automatically, which are beneficial to image processing tasks.

A CNN generally consists of three elements: convolution layers, pooling layers, and fully connected (FC) layers. The convolution layers, pooling layers are integrated to form feature extractors, and the FC layers are used as image classifiers. Fig. 2 illustrates a typical CNN architecture.
convolution blocks, FC layers and SoftMax functions are followed to perform classification. This process can be represented as follows:

$$Y = \text{soft max} \left( W_f X + b_f \right)$$  \hspace{1cm} (8)

where $Y$ is the predicted labels; $w_f$ and $b_f$ represent the weights and biases between FC layer and output layer. $X$ is the node values of the last FC layer.

C. TRANSFER LEARNING AND FINE-TUNING

Transfer learning is implemented by duplicating the parameters of the pre-trained model to the target model [23]. Typically, the weights of a CNN are initialized randomly and then updated iteratively based on the loss function. However, a deep CNN usually has millions of weights and biases to be trained, which, in consequence, requires large numbers of labeled training data. Otherwise, the network may be overfitting or not converging.

Transfer learning is a promising method that can assist in the training of CNNs. Earlier convolutional layers are able to learn low-level features, such as curves and edges, which are applicable to most image classification tasks, whereas later convolutional layers are able to learn high-level features, which are more specific to different applications [23]. Thus, the weights of earlier convolutional layers which represent low-level feature extractors can be transferred. Only the weights of later convolutional and FC layers need to be learned from the new training dataset. The operation of updating the weights of last few convolutional and FC layers of a CNN is called fine-tuning. Fine-tuning a CNN generally begins with copying the weights of all the layers from a pre-trained model to the network we would like to train. Thus, the feature extraction and classification ability are transferred from the pre-trained model to the new one. Then, the last or last few FC layers are replaced by other FC layers with the same number of neurons as classes in the new application. After the weights of the new FC layer are initialized, the network can be fine-tuned, starting from only tuning the last FC layers to tuning all the layers in the CNN. The number of layers need to be fine-tuned is different according to the distance between the source data and the target data. For a similar dataset, only FC layers need to be fine-tuned, whereas, for a significantly different dataset, early convolutional blocks need to be fine-tuned as well. Compared with training from scratch, fine-tuning a CNN is more accurate and less time consuming as it remarkably reduces the number of parameters that need to be trained.

IV. EXPERIMENTS AND RESULTS

To test the performance of the proposed detection method for packaged ICs and verify its effectiveness. Firstly, we obtained and enhanced the THz images of the tested ICs. Then, we constructed the THz IC detection dataset. Finally, we input these images into the proposed models for failure

![Figure 2: Typical convolutional neural network architecture.](image-url)
detection. The procedure of TL based IC detection is given in Fig. 3.

The THz time-domain spectroscopy system used in this study is T-Ray 5000 produced by Advanced Phonotix, Inc. The setup of this system in transmission mode is shown in Fig. 4 (a), and Fig. 4 (b) is a photograph of the actual experimental setup. The spectral bandwidth of our TDS system ranges from 0.01 THz to 5 THz theoretically. However, the real range is from 0.01 THz to 1.0 THz, since the SNR of the signal whose frequency is higher than 1.0 THz is significantly low. The numerical aperture (NA) of this system is 0.125, which was determined by the diameter of the lens (38 mm) and focal length (150 mm).

Packaged ICs, as shown in Fig. 5 (a) and 5 (f), were scanned in this study. When obtaining the THz time-domain image, we first fixed the IC on the XY 2D moving stage, and the THz image is constructed point by point by physically moving the sample on the stage. The raster-scanning step size was 0.25 mm, and the imaging range is 220 × 65 (the interval of the pixels is 0.25 mm). Then, we performed a FFT to get the frequency-domain THz images, which usually give better imaging results, as shown in Fig. 5 (c) and 5 (h). Afterwards, an HPF with 1.0 and LPF with 0.6 were applied respectively to suppress noises and minimize diffraction effects, as shown in Fig. 5 (d) and 5 (i). The wavelength of the derived THz beam is from 0.3 mm to 0.5 mm. Finally, after deconvolution with the PSF modeled by equation 2, the restored THz image is shown as Fig. 5 (e) and 5 (j).

A. DATA DESCRIPTION

1) TERAHERTZ IMAGING AND ENHANCEMENT RESULTS

---

FIGURE 3. Flowchart of packaged IC detection based on fine-tuning pre-trained model.

FIGURE 4. (a) Schematic of the T-Ray 5000 system and (b) image of the THz-TDS system operating in the transmission mode.

FIGURE 5. Two ICs, one qualified: (a) Optical image, (b) X-ray image, (c) Original THz image, (d) THz image after HPF and LPF (e) Restored THz image; another defective: (f) Optical image, (g) X-ray image, (h) Original THz image, (i) THz image after HPF and LPF (j) Restored THz image.
As can be seen, comparing the optical and X-ray images of the qualified and defective ICs, we cannot observe any defect. Though we can notice some difference between the original THz images of qualified and defective ICs, their resolution is too low to identify the key information, such as the actual shape of the die frame, bond-wires, and the defect region, which are shown more clearly in the restored images.

We also analyzed the waveform of the qualified and defective ICs to clarify their image difference, as shown in Fig. 6. Fig. 6 (a) illustrates that the traversed THz pulses from the defective IC show more time delays and attenuations, and Fig. 6 (b) shows that the detective IC have smaller amplitudes due to the energy attenuation of THz wave passing through the defective region. Thus, the defective region has lower brightness in the frequency-domain THz images.

![FIGURE 6. (a) THz time-domain pulses of the reference and IC samples. (b) THz frequency-domain pulse of the reference and IC samples.](image)

We have collected 94 packaged ICs, and half of these IC samples were subjected to 30V DC to simulate failure caused by EOS. We then divided all the collected samples into two classes: the samples subjected to 30V DC are defective ones, whereas the others are qualified ones. Afterwards, these 47 IC pairs (47 qualified ICs and 47 defective ones) are randomly divided into a training set with 33 sample pairs and a testing set with the remaining 14 pairs. Due to the difficulty and expense of collecting IC THz images, the training samples are insufficient, which may influence the generalization of the trained CNN model and lead to overfitting [51]. Thus, a maximum of 15% zoom, width shift, height shift and a rotation varied over an interval of 15° were employed to augment the training set. After augmentation, there were 660 pairs of training images. And they were further divided into 520 pairs of training images and 140 pairs of validation images.

**B. CONVOLUTIONAL NEURAL NETWORK TRAINING AND FINE-TUNING**

1) **CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE**

![FIGURE 7. Architecture of MobileNet V2 network.](image)

The CNN used in this study is MobileNet V2 [29]. MobileNet V2 is a lightweight CNN proposed by the Google team. The key improvement in MobileNet is that they replaced regular convolutional layers with depthwise separable convolutional blocks [45]. Each separable convolutional block consists of a 3x3 depthwise convolutional layer to filter the input, and then followed by a 1x1 pointwise convolutional layer to combine these filtered values. The separable convolutional block has far fewer parameters, and thus, it is much faster than the conventional convolutional layer with nearly the same accuracy. Besides separable convolution, MobileNet V2 also adopts two other approaches to improve its performance further. The first improvement is the application of the expansion layer, which enables the model to extract features in higher-dimensional spaces to avoid information loss. The second improvement is the residual connection which could help the flow of gradients through the network. The basic unit of MobileNet V2 is the bottleneck residual block, which mainly consists of three convolutional layers: a 1x1 expansion convolutional
layer, a depthwise convolutional layer, and a 1×1 pointwise convolutional layer. The architecture of MobileNet V2 is shown in Fig. 7. It is composed of 17 bottleneck residual blocks in a row, followed by a regular 1×1 convolution, a global average pooling layer, and a classification layer with class number k.

2) TRANSFER LEARNING SCENARIOS AND TRAINING SETTINGS

Generally, two approaches can be applied to achieve TL. In the first approach, the pre-trained CNN is used as a feature extractor, that is, during the training process, all the convolutional layers of the pre-trained model are frozen, only the weights of the FC layers are updated with the new dataset. The second approach which could further increase the performance of the model is to fine-tune a few selected convolutional layers of the pre-trained CNN alongside the FC layers [52].

In this study, we trained Mobilenet V2 with the second method, which achieved better results in many studies [24, 27, 28, 52, 53], especially when the target dataset (the collected dataset) is different from the source dataset (ImageNet [54]). The number of convolutional layers needs to fine-tune differs with regard to the target dataset size and the distance between source data and target data [27]. Therefore, we designed three different fine-tuning schemes to find the best solution, one fine-tuning (1FT), 2FT and 3FT.

1) 1FT: only the weights of the last bottleneck residual block and the followed layers are updated while the rest of the layers are frozen.

2) 2FT: the weights of the last two bottleneck residual blocks and the followed layers are trainable.

3) 3FT: the weights of the last three bottleneck residual blocks and the followed layers are trainable.

For comparison, we also fine-tuned VGG-16 and Resnet-50 with the same scenario, and then we trained all the three CNN models from scratch with the same training data.

We tested our proposed fine-tuning scenarios on the constructed THz IC image dataset. Since the original THz IC images are gray-scale images with the size of 220 × 65, during the training process, we firstly resized the THz images to 224 × 224, and we then duplicated them into three channels, as the input size of the three fine-tuned models is all 224 × 224 × 3. Then, we trained them with the SGD optimizer, the learning rate is 0.001, and the batch size is set to be 32. To prevent overfitting, the training was stopped when the validation accuracy did not show any improvement over ten epochs. All experiments are carried out on the GPU with GeForce GTX 1080Ti, by using Keras library [55] for training and fine-tuning the CNN models.

C. RESULTS

The classification accuracy for different fine-tuning schemes is given in Table I, with the best indexes shown in bold. We can see that CNNs trained from scratch are outperformed by all the fine-tuning scenarios when we only have a limited number of training data. For the fine-tuning scenarios, 2FT achieved the best performance for all the three CNN models, thus it is the optimal fine-tuning scenario.

| Training Scenarios | CNN Models |
|---------------------|------------|
|                     | VGG-16     | Resnet-50 | MobileNet V2 |
| 1FT                 | 96.4       | 89.3      | 96.4         |
| 2FT                 | 100        | 92.9      | 100          |
| 3FT                 | 100        | 82.1      | 100          |
| From Scratch        | 85.7       | 75.0      | 89.3         |

Table II shows the classification accuracy, training time and the number of parameters when the number of fine-tuned blocks is set to 2. As seen, though both VGG-16 and MobileNet V2 achieved a 100 percent accuracy rate of classification, combine with the training time and parameter numbers, we can conclude that the lightweight MobileNet V2 gain the best performance in the least time.

| CNN Models   | Accuracy (%) | Time (s) | Parameters (million) |
|--------------|--------------|----------|----------------------|
| VGG-16       | 100          | 308.5    | 138                  |
| Resnet-50    | 92.9         | 195.9    | 23                   |
| MobileNet V2 | 100          | 176.3    | 2.5                  |

V. CONCLUSIONS

In this paper, we aim to present a practical solution for terahertz-based IC failure detection, we firstly enhanced the quality of THz images to improve detection accuracy, and then we proposed a CNN based failure detection framework to achieve end-to-end IC inspection. Besides, we investigated whether TL can promote the performance of the IC detection networks. Comparison results between fine-tuned and fully-trained CNNs showed that knowledge transfer from natural images to IC THz images is possible, even though their distance is relatively large. Thus, TL provides a feasible solution for CNN training when we could not collect large quantities of labeled training data. The comparison results between 3 different fine-tuning scenarios showed that fine-tuning the last two bottleneck residual blocks (or convolutional blocks) achieved the best performance. The assumption is that though last convolutional layers are able to learn high-level features, which is helpful to the specific application, fine-tuning too many layers may lead to overfitting due to complex model structure and limited training data. Moreover, we can conclude the lightweight MobileNet V2 is an ideal CNN model for IC failure detection, since compared to VGG-16 and ResNet-50, it gains the best performance in the least time.
However, our proposed packaged IC detection method may have some limitations. Firstly, due to the difficulty and expense of collecting IC THz images, we could not collect as many training samples as possible. Though we adopted a data augmentation algorithm, the limited training data size may still influence the testing accuracy, especially for the trained from scratch scenario. Secondly, our CNN could only detect defects caused by EOS. We would collect ICs of more different failure types to train our model, so that our model could recognize more failure types. Finally, we only classified the tested ICs into qualified and defective ones. We would train our proposed networks to determine the exact location of the cracks and voids hidden in the IC packages in future works.

REFERENCES

[1] E. Keenan, R. G. Wright, R. Mulligan, and L. V. Kirkland, “Terahertz and laser imaging for printed circuit board failure detection,” in Proceedings AUTOTESTCON 2004., 2004, pp. 563-569.

[2] S.-H. Park, J.-W. Jang, and H.-S. Kim, “Non-destructive evaluation of the hidden voids in integrated circuit packages using terahertz time-domain spectroscopy,” Journal of Micromechanics and Microengineering, vol. 25, p. 095007, 2015.

[3] T. D. Moore, D. Vanderstraeten, and P. M. Forssell, “Three-dimensional X-ray laminography as a tool for detection and characterization of BGA package defects,” IEEE Transactions on Components and Packaging Technologies, vol. 25, pp. 224-229, 2002.

[4] E. Martin, C. Larato, A. Clément, and M. Saint-Paul, “Detection of delaminations in sub-wavelength thick multi-layered packages from the local temporal coherence of ultrasonic signals,” Nd & E International, vol. 41, pp. 280-291, 2008.

[5] N. M. Burford and M. O. El-Shenawee, “Enhancement of terahertz imaging of packaged power electronic devices,” in 2015 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, 2015, pp. 1300-1301.

[6] G.-M. Zhang, D. M. Harvey, and D. R. Braden, “Microelectronic package characterisation using scanning acoustic microscopy,” NDT & E International, vol. 40, pp. 609-617, 2007.

[7] B. B. Hu and M. C. Nuss, “Imaging with terahertz waves,” Optics letters, vol. 20, pp. 1716-1718, 1995.

[8] T. Kiwa, M. Tonouchi, M. Yamashita, and K. Kawase, “Laser terahertz-emission microscope for inspecting electrical faults in integrated circuits,” Optics letters, vol. 28, pp. 2058-2060, 2003.

[9] M. Yamashita, K. Kawase, C. Otani, T. Kiwa, and M. Tonouchi, “Imaging of large-scale integrated circuits using laser terahertz emission microscopy,” Optics Express, vol. 13, pp. 115-120, 2005.

[10] Y. Cai, Z. Wang, R. Dias, and D. Goyal, “Electro Optical Terahertz Pulse Reflecrometry—an innovative fault isolation tool,” in 2010 Proceedings 60th Electronic Components and Technology Conference (ECTC), 2010, pp. 1309-1315.

[11] M. Nagel, A. Michalski, and H. Kurz, “Contact-free fault location and imaging with on-chip terahertz time-domain reflectometry,” Optics express, vol. 19, pp. 12509-12514, 2011.

[12] F. Blanchard, A. Doi, T. Tanaka, and K. Tanaka, “Real-time, subwavelength terahertz imaging,” Annual Review of Materials Research, vol. 43, pp. 237-259, 2013.

[13] K. Ahi, S. Shahbazmohamadi, and N. Asadizanjani, “Quality control and authentication of packaged integrated circuits using enhanced-spatial-resolution terahertz time-domain spectroscopy and imaging,” Optics and Lasers in Engineering, vol. 104, pp. 274-284, 2018.

[14] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, “A survey of deep learning-based network anomaly detection,” Cluster Computing, pp. 1-13, 2019.

[15] H. Zhao, H. Liu, W. Hu, and X. Yan, “Anomaly detection and fault analysis of wind turbine components based on deep learning network,” Renewable energy, vol. 127, pp. 825-834, 2018.

[16] Y. Bao, Z. Tang, H. Li, and Y. Zhang, “Computer vision and deep learning-based data anomaly detection method for structural health monitoring,” Structural Health Monitoring, vol. 18, pp. 401-421, 2019.

[17] Q. Yao, M. Zhuo, C. Lv, Y. Lu, X. Yan, S. Yan, et al., “Convolutional neural network model based on terahertz imaging for integrated circuit defect detections,” Optics Express, vol. 28, pp. 5000-5012, 2020.

[18] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, “Face recognition: A convolutional neural-network approach,” IEEE transactions on neural networks, vol. 8, pp. 98-113, 1997.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, pp. 2278-2324, 1998.

[20] J. Fan, W. Xu, Y. Wu, and Y. Gong, “Human tracking using convolutional neural networks,” IEEE Transactions on Neural Networks, vol. 21, pp. 1610-1623, 2010.

[21] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, and G. Penn, “Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition,” in 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP), 2012, pp. 4277-4280.

[22] R. Socher, Y. Bengio, and C. D. Manning, “Deep learning for NLP (without magic),” in Tutorial Abstracts of ACL 2012, ed, 2012, pp. 5-5.

[23] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” in Advances in neural information processing systems, 2014, pp. 3320-3328.

[24] S. Shao, S. McAleer, R. Yan, and P. Baldi, “Highly accurate machine fault diagnosis using deep transfer learning,” IEEE Transactions on Industrial Informatics, vol. 15, pp. 2446-2455, 2018.

[25] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on knowledge and data engineering, vol. 22, pp. 1345-1359, 2009.

[26] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, et al., “Decaf: A deep convolutional activation feature for generic visual recognition,” in International conference on machine learning, 2014, pp. 647-655.

[27] Tajbakhsh, Nima, Shin, Jae, Y., Gurudu, et al., “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” IEEE Transactions on Medical Imaging, 2016.

[28] E. Cetinic, T. Lipic, and S. Grbic. “Fine-tuning Convolutional Neural Networks for fine art classification,” Expert Systems with Applications, vol. 114, pp. 107-118, 2018.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottle-necks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510-4520.
[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[32] M. Yamashita, C. Otani, K. Kawase, K. Nikawa, and M. Tonouchi, "Noncontact inspection technique for electrical failures in semiconductor devices using a laser terahertz emission microscope," Applied Physics Letters, vol. 93, p. 041117, 2008.

[33] L.-M. Xu, W.-H. Fan, and J. Liu, "High-resolution reconstruction for terahertz imaging," Applied optics, vol. 53, pp. 7891-7897, 2014.

[34] Y. Liang and W. Fan, "Image enhancement techniques used for THz imaging," in International Symposium on Photoelectronic Detection and Imaging 2011: Terahertz Wave Technologies and Applications, 2011, p. 819515.

[35] T. M. Wong, M. Kahl, P. H. Bolivar, and A. Kolb, "Computational image enhancement for frequency modulated continuous wave (FMCW) THz image," Journal of Infrared, Millimeter, and Terahertz Waves, vol. 40, pp. 775-800, 2019.

[36] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in CVPR 2011, 2011, pp. 233-240.

[37] W. Ning, F. Qi, Z. Liu, Y. Wang, H. Wu, and J. Wang, "Resolution enhancement in terahertz imaging via deconvolution," IEEE Access, vol. 7, pp. 65116-65121, 2019.

[38] K. Ahi, "A method and system for enhancing the resolution of terahertz imaging," Measurement, vol. 138, pp. 614-619, 2019.

[39] K. Ahi, "Mathematical modeling of THz point spread function and simulation of THz imaging systems," IEEE Transactions on Terahertz Science and Technology, vol. 7, pp. 747-754, 2017.

[40] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, et al., "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, pp. 541-551, 1989.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, pp. 1097-1105, 2012.

[42] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Neural computation, vol. 29, pp. 2352-2449, 2017.

[43] J. Zhang, W. Xing, M. Xing, and G. Sun, "Terahertz image detection with the improved faster region-based convolutional neural network," Sensors, vol. 18, p. 2327, 2018.

[44] X. Yang, T. Wu, L. Zhang, D. Yang, N. Wang, B. Song, et al., "CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images," Signal Processing, vol. 160, pp. 202-214, 2019.

[45] A. G. Howard, M. Zhu, B. Chen, D. Kalenchenko, W. Wang, T. Weyand, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[46] B. Baheti, S. Gajre, and S. Talbar, "Detection of distracted driver using convolutional neural network," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 1032-1038.

[47] H. Pan, Z. Pang, Y. Wang, Y. Wang, and L. Chen, "A new image recognition and classification method combining transfer learning algorithm and mobilenet model for welding defects," IEEE Access, vol. 8, pp. 119951-119960, 2020.

[48] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," Sensors, vol. 21, p. 2852, 2021.

[49] H. Guerboukha, K. Nallappan, and M. Skorobogatyi, "Toward real-time terahertz imaging," Advances in Optics and Photonics, vol. 10, pp. 843-938, 2018.

[50] Z. Zhang, Y. Lu, C. Lv, Q. Mao, S. Wang, and S. Yan, "Restoration of integrated circuit terahertz image based on wavelet denoising technique and the point spread function model," Optics and Lasers in Engineering, vol. 138, p. 106413, 2021.

[51] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621, 2017.

[52] M. Akay, Y. Du, C. L. Sershen, M. Wu, T. Y. Chen, S. Assassi, et al., "Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model," IEEE Open Journal of Engineering in Medicine and Biology, vol. 2, pp. 104-110, 2021.

[53] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," Computer methods and programs in biomedicine, vol. 157, pp. 19-30, 2018.

[54] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255.

[55] A. Gulli and S. Pal, Deep learning with Keras: Packt Publishing Ltd, 2017.