A New Cycle-consistent Adversarial Networks with Attention Mechanism for Surface Defect Classification with Small Samples

Long WEN, You Wang, and Xinyu LI, *Member, IEEE*

***Abstract*—Surface defect detection is the essential process to ensure the quality of products. Surface defect classification (SDC) based on deep learning (DL) has shown its great potential. However, the well-trained SDC model usually requires large training data, and the small intra-class differences between the defect and normal samples also degrades the performance of SDC model. To overcome these drawbacks, this research proposed a new Cycle-consistent Adversarial Networks with attention mechanism (AttenCGAN). Firstly, AttenCGAN is used for synthesizing defect samples to enlarge the samples volume. Secondly, the attention mechanism is adopted for the feature enhancement by finding the discriminative parts of the samples and enlarging the differences among the samples. AttenCGAN is tested on KolektorSDD and DAGM2007 datasets, and its accuracies are 98.53% and 99.57% with only a few samples. The experiment results show that AttenCGAN outperforms other published SDC methods based on DL and machine learning, which validates its potential.**

***Index Terms*—Cycle-consistent Adversarial Networks, Surface Defect Classification, Small Samples, Attention Mechanism**

I. INTRODUCTION

With the boosting of smart manufacturing, the cloud manufacturing becomes an effective manufacturing mode and it can transform the manufacturing resources and capabilities in the cloud to promote the efficiency of smart manufacturing [1]. The quality inspection of workpieces is vital in cloud manufacturing for the collaboration between the manufacturing for machine-to-machine collaboration [2], so the automatic defects detection for the surface of workpieces is necessary technique. Nowadays, machine vision-based surface defect classification (SDC) has been widely investigated in both industrial and academic fields [3] such as the complicated texture surface of the workpiece or product. It usually trains the deep learning (DL) models to automatically inspect the defects to avoid the unstable and unreliable by manual inspection and to promote its efficiency [4].

The researches on DL have achieved significant progresses on SDC [5]. These methods still suffer from several characteristics on the defect samples which can affect the defect classification performance heavily. As most DL methods follow the paradigm of supervised learning, the large training data is necessary in order to obtain the well-trained SDC model. However, the volume of the surface defect samples is small in most industrial environments, these characteristics of small samples on SDC would greatly limit the performance of the DL model [6]. The investigation on the SDC with small samples becomes essential to ensure the potential of the SDC models.

Several previous works have been done to handle with the small samples in SDC, such as data augment, transfer learning. But Generative Adversarial Network (GAN) provides an alternative way to handle the small samples in SDC. As GAN is familiar with synthesizing pseudo images, it can be regarded as the data augment technique to enlarge the volume of the SDC dataset. Several researchers used GAN to compensate for the volume of defect samples by generating new pseudo samples. Lu et al. [5] studied conditional GAN to classify the surface defect types on the thin-film transistor liquid crystal display, this method is trained with a limited training dataset and it also can provide the improved accuracy for the defects. Shao et al. [6] investigated the capsule GAN network and it adopted the superior competition mechanism into GAN. This method takes the GAN as the data augment technique on SDC and it can generate more realistic surface defect images.

On the other hand, even though the GAN based SDC has gained many pioneering works, it cannot handle with the situations when the differences between the normal and defect samples are minor while the differences between the defect samples are obvious. These indistinct intra-class and inter-class distance in the appearance of the defect samples makes it hard to be recognized by generator and discriminator network in GANs [7][8]. Several researchers attempted to adopt the attention mechanism to recognize the minor difference between different classes of samples. Liu et al. [7] studied the fabric

This work was supported in part by the Natural Science Foundation of China under Grant U21B2029 and 51805192, the Major Special Science and Technology Project of Hubei Province under Grant No. 2020AEA009 and sponsored by the State Key Laboratory of Digital Manufacturing Equipment and Technology (DMET) of Huazhong University of Science and Technology (HUST) under Grant No. DMETKF2020029. (Corresponding author: Xinyu LI)

Long WEN and You Wang are with School of Mechanical Engineering and Electronic Information, China University of Geosciences, No. 388 Lumo Road, Wuhan, 430074, China (e-mail: wenlong@cug.edu.cn, wangyou@cug.edu.cn).

Xinyu LI is with the State Key Laboratory of Digital Manufacturing Equipment & Technology, Huazhong University of Science and Technology, 1037#, Luoyu Road, Wuhan, 430074, China (lixinyu@mail.hust.edu.cn)

defect detection framework using GAN and Convolutional Neural Networks (CNN). In this method, the attention mechanism is adopted which can learn the discriminative features of defects and to better synthesize defect areas in the image further. Hu et al. [8] proposed the relative average GAN, which is driven by a dual attention mechanism aiming at generating the high-quality defect images on the metal workpiece. But when the number of defect samples is only a few, for example less than ten, it is still challenging to obtain the wonderful performance on SDC. So, it still needs further investigation on the SDC with small sample.

In this research, a new Cycle-consistent Adversarial Networks (CycleGAN) with attention mechanism is proposed for the SDC with the small samples, named AttenCGAN. Firstly, CycleGAN is adopted to synthesize many pseudo defect samples using small true defect samples to enlarge the defect dataset. Secondly, the attention mechanism is applied for the classification of the defect samples, which aims at finding the discriminative features of the defect samples and recognizing the discriminative parts on the defect samples to promote the proposed AttenCGAN. AttenCGAN is conducted on two famous datasets, and the results show that the proposed method has achieved a significant promotion by comparing with other DL and machine learning (ML) methods.

The rest of the study is as follows. Section II presents related work. Section III presents the proposed AttenCGAN. Section IV shows the case studies and the experimental results. Section V gives the conclusions.

II. RELATED WORK

The related work is described in this section, including GAN based SDC and the attention mechanism in SDC.

A. GAN based SDC

DL based SDC has achieved excellent performance in recent years [9]. Cheon et al. [10] studied the DL based SDC which can automatically classify different types of wafer surface damages. The method compares the features between the unseen class and the trained classes to effectively solve the shortage of samples. Zhu et al. [11] studied an intelligent recognition method based on DL to realize the nondestructive intelligent recognition of weld surface defects. This method combines CNN and random forest, and its accuracy is 98.75%. However, the defect samples are small in most real industrial environment, which could affect the performance of DL based SDC. Data augmentation (DA) and transfer learning (TL) techniques have been investigated to handle with the small samples in SDC. Wu et al. [12] investigated the feature extraction method based on TL to classify steel surface defects with the small defect samples, and the experimental results show that it can obtain better results. Liu et al. [13] incorporated the imbalanced sampler and TL to improve the imbalanced multi-label SDC, and it achieved 97% accuracy on the SDC on steel. Wan et al. [14] studied the improved TL based on VGG 19 network to handle with the few samples and imbalance dataset in the steel SDC case, and the results show that TL can promote the detection rate.

Some researchers adopted GAN as the DA method to handle with the small samples in SDC. GAN was firstly proposed by Goodfellow for image generation, and it is used to generate surface defect samples with only a few labelled defect samples and large numbers of unlabeled samples. Jain et al. [15] studied the GAN to create synthetic images and generate new defect samples from random noise, and it yielded better results using the synthetically augmented model. He et al. [16] used deep convolutional GAN and multiple training of residual network for pursuing the high confidence defect samples, and then they would be regarded as the training dataset to improve the performance of the SDC method. He et al. [17] studied a semi-supervised framework for surface defect classification on steel. It can synthesize enough labelled images using GAN, and it used multiple training on two classifiers to process labelled and unlabeled samples separately. Zhao et al. [18] investigated the surface defect detection model with only defection-free sample training by combining autoencoder and GAN. In this model, defect image samples are generated by GAN, and LBP is used to compare defection-free samples and defection-containing samples to find the surface defects on fabric. Lian et al. [19] studied GAN with CNN models to generate surface defect samples and expand the limited sample which contains the tiny defect on a large surface, the results show that the accuracy of GAN with CNN model has been improved. Sha et al. [20] added additional monitoring modules to CycleGAN, which can be used to synthesize specific types of defects, and the performance of synthetic data enhancement is 10% higher than traditional data enhancement methods alone. Niu et al. [21] studied the CycleGAN based surface defect segmentation method using the weakly supervised manner. It trained the model by the image-level annotations, and its results are superior to the supervised method on three industrial datasets. Tsai et al. [22] used two CycleGAN to synthesize the defect pixels, and then the U-Net network is trained for the defect segmentation.

As GAN has the good ability on synthesizing defect images to enlarge the defect samples, this research adopts the CycleGAN to generate more realistic pseudo defect samples in order to handle with the situation on SDC with only a few numbers of samples, for example less than 10, and even with only one defect samples.

B. Attention Mechanism in SDC

In some cases, the defects in the images are tiny and various, making the defect images and the normal images are same, this can be formulated as the fine-grained classification. The fine-grained SDC has been investigated in the SDC field. Liu et al. [23] investigated the lightweight detection method for the fine-grained surface defects on the aerospace seal O-rings, and the accuracy of this method has achieved 95.10% for five types of rings. Cui et al. [24] investigated the fine-grained detection method to detect the small defects. The proposed method has been tested on the public dataset and achieved significant results. Zhang et al. [25] studied the pavement defect segmentation method using the fine-grained method, and the network contains the spatial, contextual and boundary parts to extract the rich discriminative feature and promote the results. Huang et al. [26] applied a fine-grained bilinear CNN model to defect detection of railway infrastructure for the first time and proposed a new bilinear deep network space transformation and bilinear low-rank model. It is a good solution to the situation that the defects of railway power infrastructure damage or wear are difficult for children to detect. This method is superior to the machine learning method based on manual features and the classical deep neural network method. Tu et al. [27] proposed a real-time defect detection method for rail components based on case segmentation, which combined case segmentation, traditional image processing, and image classification, realized the accurate detection of fastener defects and rail defects. In addition, the geometric characteristics of fasteners are used to detect fastener defects under unbalanced conditions. Li et al. [28] proposed a new method based on two-stage learning. By integrating global context features and local defect features, the target detection network simultaneously performs defect detection and fine-grained classification of sewage pipelines in all defect areas. The technique is also the first to use learning techniques to classify and evaluate the level of defects in sewage pipes.

As attention mechanism can find the discriminative areas in the images by assigning a larger weight to the important areas and a smaller weight to the areas that are not needed, several studies applied the attention mechanism to realize the fine-grained SDC. Tang et al. [29] combined the spatial attention mechanism of deep CNN and dual-line pooling, which improved the representation ability of CNN and realized the full image classification of defective casting and defective-free casting. Tao et al. [30] proposed the DeepScratchNet weak scratch detection method, which can realize automatic detection of weak scratches by gathering multidimensional features to represent scratches. This method uses attentional feature fusion block to highlight the scratch feature and attenuate the noise, and it uses the dual attentional mechanism to fuse the high-level semantic feature and the low-level detail feature tightly. Hu et al. [31] studied the CNN model using image-level tag training only for the detection of tiny casting defects in complex industrial scenes, while the object-level attention is developed and the bilinear pooling is used to improve the model's ability to detect local casting defects. This strategy can well reduce the detection difficulty caused by small detection targets and a small number of tags. Su et al. [32] studied the complementary attention network, and this method solves the problems of similar defect features and complex background features in automatic defect detection of electroluminescent images.

As the defect samples in most SDC has shown the typical fine-grained characteristics, and it becomes more challenging when the defect samples are small, so the attention mechanism is applied as well as the CycleGAN to extract the discriminative feature of defect samples and promote the performance of AttenCGAN further.

III. The Proposed AttenCGAN for SDC

This research proposes a new SDC method with small defect samples, namely AttenCGAN. In AttenCGAN, the pseudo defect samples can be synthesized, and then they are used to train the surface defect convolutional neural network (SDCNN) with attention mechanism to promote the effectiveness. The architecture of the proposed AttenCGAN is shown in Fig 1.

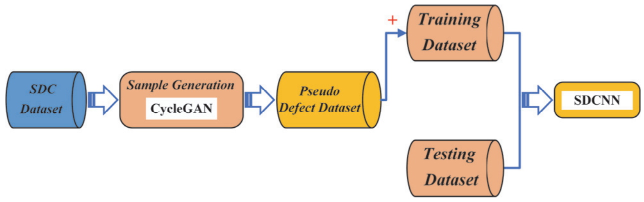

Fig 1 The architecture of AttenCGAN

A. Cycle-consistent Adversarial Network

GAN is a kind of unsupervised DL model, and the framework of GAN is to generate the needed model through supervision training through the adversarial process. The basic framework of GAN contains the Generator Network (*G*) and the Discriminator Network (*D*), which aims to produce relatively good output in the process of Adversarial Learning. *G* generates samples, its generating ability is improved as much as possible in training, and it can infinitely approximate the real image *x*. *D*'s discrimination ability can be improved as much as possible in training and it can identify accurately the false image generated by noise *Z*. Therefore, the loss of traditional GAN refers to (1). Where $\min_G \max_D$ means that *V(D, G)* is maximized from the perspective of Discriminator *D*, and then *V(D, G)* is minimized from the perspective of Generator *G*.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data(x)}}[\log D(x)] + E_{z\sim P_{zdata(Z)}}[\log(1-D(G(z)))] \quad (1)$$

This research studied a Cycle-consistent Adversarial Network (CycleGAN) model for SDC, which is a new variant of GAN. It skillfully symmetries two traditional GAN into two pairs networks to achieve unsupervised images translation and complete inter-domain image translation without paired images. Assume that GAN is the one-way generation, and its direction is X→Y. Different from GAN, CycleGAN not only needs to learn the image features of X and Y domains, but also needs to learn the cyclic translation in the X→Y→X domain and Y→X→Y domain. It contains two generators (denoted as *G* and *F*), and two discriminators (denoted as D_X and D_Y) where *G*: X→Y, *F*: Y→X.

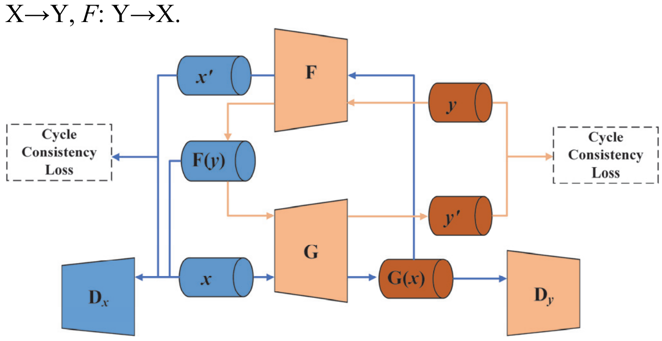

Fig 2 The architecture of Cycle-consistent Adversarial Network

Cyclic translation is the key process in CycleGAN. As GAN is an unsupervised process, the adversarial trainings on mapping *G*: X→Y and *F*: Y→X has no constraints, making the network has a huge capacity and there is no guarantee on that the learnt mapping can map each input X to its desired Y. To reduce the possible space of the adversarial training, the cycle-consistent is introduced. As shown in Fig 2, each image *x* in the input dataset X should be able to return to its origin version using the circular translation. Namely, the forward and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168432, IEEE Transactions on Industrial Informatics

backward circulation is consistent, i.e. $x \to G(x) \to F(G(x)) \approx x'$ and $y \to F(y) \to G(F(y)) \approx y'$.

CycleGAN reduces the search space by adding the cyclic consistency loss. This is the regularization of the generation models *G* and *F*, and can lead the generator to synthesize more real and compact defect samples. When the defect samples are small, the training for models *G* and *F* becomes more difficult, this regularization can reduce the search space and provide more reliable translation for the small samples. As shown in Fig 3, the structure of CycleGAN is like the cooperation of two basic GAN. The forward generation of CycleGAN can be viewed as the extension of GAN and it is X→Y→X, meaning that the generated Y can be inversely transformed into X through the reverse generator. In the forward step, the network is composed of generator *G*, *F* and discriminator D_Y, where D_Y is used to judge whether the generated sample belongs to the Y. In the reverse step, the network consists of generators *F*, *G* and discriminator D_X.

Before CycleGAN training, automatic data augment is used to expand the dataset. Several typical data augment methods including center_crop, color_jitter, random_grayscale, random_rotate and their random combinations are used as well.

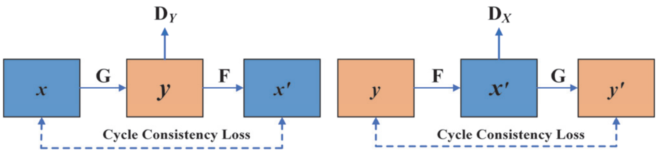

Fig 3 The Cycle Consistency Loss

1) Adversarial Loss

CycleGAN applies the least square loss and designs a cycle consistency loss to match the bidirectional generation process. As shown in Fig 2, *G* strives to make the generated image *G(x)* which resemble the images in the Y domain as much as possible. D_Y tries to distinguish $G(x)$ from the real data y as much as possible. The data distribution obeys for the mapping $x \sim p_{data(x)}$, $x_i \in \boldsymbol{X}$ and $y \sim p_{data(y)}$, $y_i \in \boldsymbol{Y}$. The loss function of function G and D_Y refers to (2). In the same way, for *F* and D_X, the loss function refers to (3).

$$L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x))] \quad (2)$$

$$L_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y \sim p_{data}(y)}[\log(1 - D_X(F(y))] \quad (3)$$

The cross-entropy loss uses its entropy value to reflect the difference between the data distributions on real data *p* and the generated data *q*. The smaller the cross-entropy loss is, the closer the two distributions are and the better the generation effect is. Cross entropy can effectively classify the true and false samples, but it cannot help the positive samples far away from the decision boundary to continue optimization, resulting when the generation network in gradient update will encounter diffusion. The model adopts the least square loss. This loss penalizes the false samples that deviate from the decision boundary, thus providing direction for gradient descent. The standard calculation formula of least squares (LS) refers to (4). In the Cycle-consistent Adversarial Network, the values of *a*, *b*, *c* are 1, 0 and 1 respectively.

$$\begin{aligned} \min V_{LSGAN}(D) &= \frac{1}{2} E_{x \sim P_{data(x)}}[(D(x)-b)^2] + \frac{1}{2} E_{z \sim P_{zdata(z)}}[(D(D(z))-a)^2] \\ \min V_{LSGAN}(G) &= \frac{1}{2} E_{z \sim P_{zdata(z)}}[(D(D(z))-c)^2] \end{aligned} \quad (4)$$

2) Consistency Loss

Circular structure cannot guarantee that input is constantly approaching the target field, as both generator and discriminator network need to be trained independently. If *G* does not learn anything related to *X* domain at all, but directly generates images from *Y* domain and outputs them to *D* for discrimination, it is enough to deceive the discriminator. However, such learning is invalid. To complement the design of the two-way network, consistency loss is developed. As shown in Fig 3, it is assumed that the input image *x* is eventually restored to *x'* after translation by generators *G* and *F*. To ensure that the input can maintain consistency as much as possible after two networks, the L1 loss is used to constrain them. For the reverse generation, it is important to ensure that the output y' prime approximates the input y. The two consistencies above can be expressed as *x→G(x)→F(G(x))≈x* and *y→F(y)→G(F(y))≈y*. Therefore, the cycle consistency loss function refers to (5).

$$L_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\| F(G(x)) - x \|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\| G(F(y)) - y \|_2] \quad (5)$$

3) Cycle Consistency Full Loss

The CycleGAN full loss refers to (6), where λ controls the relative importance of the two targets.

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (6)$$

As GAN, the final optimization function of CycleGAN is still to solve the maximum and minimum antagonism problem.

$$G^* F^* = \arg \min_{G,F} \max_{D_X, D_Y} L(G, F, D_x, D_Y) \quad (7)$$

In the experiments, λ is set to be 10. The learning rate is 0.0002, the number of iterations is 200. The learning rate is maintained in the first 100 epochs, and then it linearly decays to zero in the next 100 epochs. The batch size is set to be 1.

4) Evaluation Metric

In this research, the FID distance is used to calculate the distance between the generated sample and the real sample in the feature space. If the mean and variance of two Gaussian distributions are the same, the two distributions are judged to be the same. Therefore, a lower FID represents a closer image distribution, resulting in higher image quality and richer patterns. The FID refers to (8), where *Tr* stands for the sum of diagonal elements of the matrix, μ is the mean value and covariance, *x* is the real sample, and *y* is the generated sample.

$$FID = \left\| \mu_x - \mu_y \right\|^2 + Tr\left(\Sigma_x + \Sigma_y - 2\left(\Sigma_x \Sigma_y \right)^{\frac{1}{2}} \right) \quad (8)$$

B. Surface Defect CNN Backbone

Surface Defect CNN (SDCNN) is a new variant of classification network in AttenCGAN, which is attached to the attention mechanism. Fig 4 shows the components of SDCNN, including Auto Data Augmentation, Feature Extraction, Bilinear Attention Pool (BAP) Fusion, and Attention-guided Data Augmentation.

1) Auto Data Augmentation

Auto Data Augmentation is portable and does not require fine-tuning weights for pre-training of additional data. In this study, the optimal strategy obtained by pre-training on the ImageNet dataset is directly migrated to the surface defect dataset used in this study, and the surface defect image is automatically enhanced.

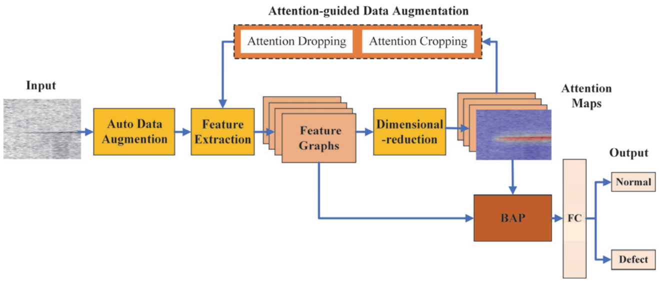

Fig 4 The structure of SDCNN in AttenCGAN

2) Feature Extraction

The feature extraction part uses a pre-trained CNN model to extract the image features. The CNN model used in this research is Inception V3. Suppose that the feature map obtained by the CNN model is $Fea \in R^{C\times H\times W}$, where *C*, *H* and *W* denote the channel number, width and height of the image feature. The attention is obtained by the 1*1 convolutional layer on *F*, and the attention map $Att \in R^{M\times H\times W}$ can be formulated as $Att = conv(Fea)$.

3) BAP Fusion

The attention map and feature map are fused using BAP fusion, and then the fusion matrix $P \in R^{M\times N}$ fed back to the fully-connected layer (FC) for classification. The BAP operator fuses the feature map *Fea* and the attention map *Att* to strengthen the features of the images.

4) Attention-guided Data Augmentation

As attention map is obtained by the attention mechanism to indicate the important areas in the image, it can be used to guide the data augmentation. The attention guided data augment contains the Cropping and Dropping Augmentation. These two augmentations use the attention map to find the discriminative areas of the images and to use this discriminative information to train SDCNN. The attention guided data augment can be seen in Fig 5.

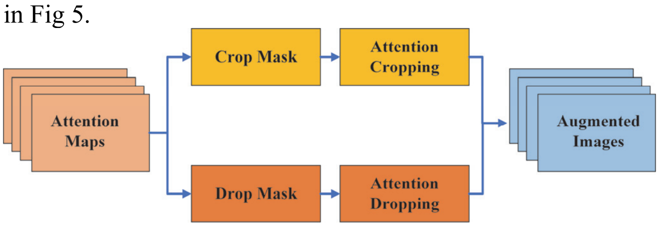

Fig 5 Attention-guided Data Augmentation

The cropping augmentation selects the discriminative areas for the augmentation. Firstly, a random channel of the attention map is chosen $k\in[1,M]$. Then A_k is normalized as shown in equation (9). A^* is used to guide the augmentation.

$$A^* = \frac{A_k - \min(A_k)}{\max(A_k) - \min(A_k)} \tag{9}$$

The cropping augmentation uses the clipping mask C_k to indicate the discriminative area of the images. When the element A^* is greater than the threshold θ_c, the clipping mask element is set to 1 and the rest is set to 0, as shown in equation (10). The default θ_c is set to be 0.5. The area which the element in C_k is 1 is the right discriminative area. So, these areas are clipped for the image and then they are resized as the data augment samples. These samples can force the model to pay more attention to these discriminative areas and promote the efficiency of SDCNN further.

$$C_k(i,j) = \begin{cases} 1, & if\ A^*(i,j) > \theta_c \\ 0, & otherwise \end{cases} \tag{10}$$

The dropping augmentation is familiar with the cropping augmentation, but it drops these discriminative areas in the images to enhance the effects of other information in the images on the final classification. The dropping mask D_k can be obtained as equation (11) using the dropping threshold θ_d and A^*. In this way, the small differences between classes can be recognized.

$$D_k(i,j) = \begin{cases} 0, & if\ A^*(i,j) > \theta_d \\ 1, & otherwise \end{cases} \tag{11}$$

IV. Case Studies and Experimental Results

In this section, the proposed AttenCGAN is tested on two famous surface defect detection datasets. The experimental results are shown to validate the potential of AttenCGAN.

A. Datasets Introduction

1) KolektorSDD Dataset

The KolektorSDD dataset provided and annotated by Kolektor Group has a total of 399 images of defective electrical commutators, and of which 50 images contain visible defects. It can be found in the link https://www.vicos.si/resources/kolektorsdd/. Several defect images in KolektorSDD are shown in Fig 6.

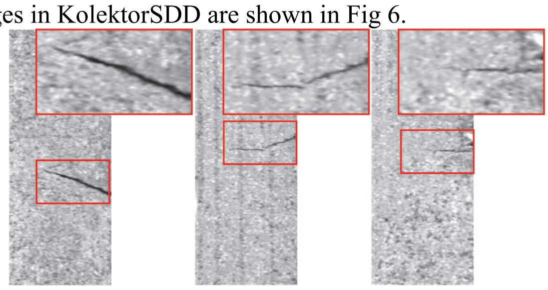

Fig 6 Defect samples in KolektorSDD dataset

2) DAGM 2007 Dataset

The DAGM 2007 dataset is the well-known surface defect detection benchmark database. It contains 10 types of surface images with artificial defects. Each type of dataset has 1150 images, including 1000 defect-free images and 150 defective images. The dataset can be found in the link https://hci.iwr.uni-heidelberg.de/content/weakly-supervised-learning-industrial-optical-inspection. Several defect images in DAGM2007 are shown in Fig 7. In this research, the Class 1 to Class 6 are used to validate the performance of AttenCGAN.

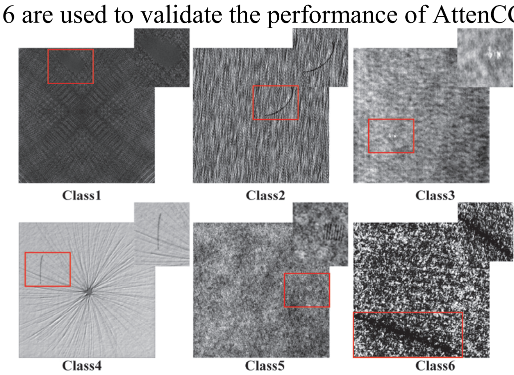

Fig 7 Defect samples in DAGM2007

B. Image Generation based on AttenCGAN

In this research, the defect images are randomly selected from the dataset and then fed into AttenCGAN to synthesize 200 pseudo defect samples, and then these samples are regarded as the training dataset to train the SDCNN. The number of the selected defect samples are 1, 2, 4, 6 and 8 on the KolektorSDD dataset, while the values are 1, 2, 4, 6, 8, 10, 20, 30 and 40 on the DAGM 2007 dataset. The synthetic pseudo samples by AttenCGAN are presented in Fig 8.

In Fig 8, there are three images for each row. The first images are the non-defect image from the dataset, which are used as the input X for AttenCGAN. The second images are the synthesized defect images and it is the $G(X)$ in AttenCGAN. The third images are the reconstructed images using the cyclic translation in AttenCGAN, and it is $F(G(X))$.

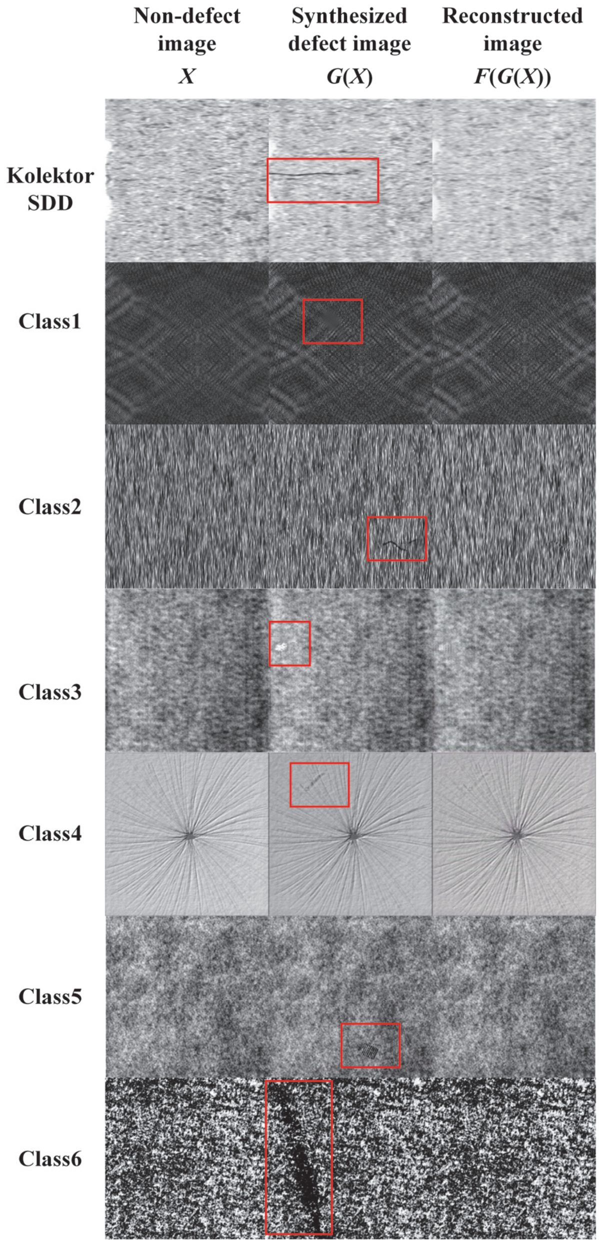

Fig 8 Synthesized defect image of KolektorSDD and DAGM 2007

From Fig 8, it can be seen that AttenCGAN can learn the mapping between defect image and non-defect image and it can generate the defect images well. What's more, the synthesized defect image can be reconstructed to the real non-defect image in high quality. According to the cyclic consistency, AttenCGAN can reconstruct its synthetic defect image back to the real image. The term FID is used to evaluate the quality of synthesized image, and it is presented on TABLE I.

In TABLE I, it shows that with the increasing number of defect images (N) used for AttenCGAN, the FID value is lower, showing that the closer between the synthesized defect image and the real defect image. On the KolektorSDD dataset, even the number of defect images is as small as 2, the FID value can be kept at a relatively low level, which is below 200. When there is only one defect image used for training, the FID value increases largely to 315. On all the six classes in DAGM2007, the FID value increases too along with the decreasing on the number of the defect images. The FID values are higher than the KolektorSDD dataset with the same number of defect images. However, most of FID values are also near 200 when the number of defect images is 2 except on Class 4. When only one defect image is fed to AttenCGAN, FID is also increasing a lot. These results show that AttenCGAN can obtain good results with small defect samples as small as 2 defect samples, but with fewer the defect samples, the performance of AttenCGAN can be degraded too.

TABLE I
THE FID VALUES OF DIFFERENT NUMBER OF DEFECT IMAGES IN ATTENCGAN

N	Kolektor SDD	DAGM2007					
		Class1	Class2	Class3	Class4	Class5	Class6
1	314	395	396	413	451	459	397
2	196	201	205	227	301	311	201
4	131	197	195	221	299	305	197
6	128	189	188	198	203	216	189
8	122	168	169	175	195	197	168
10	-	137	139	127	146	200	127
20	-	123	118	121	119	198	119
30	-	100	97	90	96	196	99
40	-	90	92	89	110	187	89

C. Discriminative Part Recognition by Attention Mechanism

In AttenCGAN, the discriminative parts of defect images are copied by the attention mechanism. The recognitions on the discriminative parts are shown in Fig 9 and Fig 10.

In Fig 9 and Fig 10, the left image is the raw image and the right image is its attention map. From the figures, it can be seen that the attention mechanism can find out the important parts of the image on both datasets, which can help AttenCGAN to promote its performance. In this research, the attention mechanism is used to deal with the situation that the difference between normal and defect samples is small, which can be validated by Fig 9 and Fig 10.

With the guidance of attention, AttenCGAN can not only learn the salient features from the defect samples, but also can pay more attention to the discriminative parts in the defect image while ignoring the other irrelevant features. This can help the AttenCGAN to recognize the small differences between the samples.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168432, IEEE Transactions on Industrial Informatics

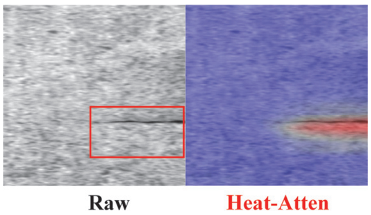

Fig 9 Visualization of defect images of KolektorSDD by attention mechanism

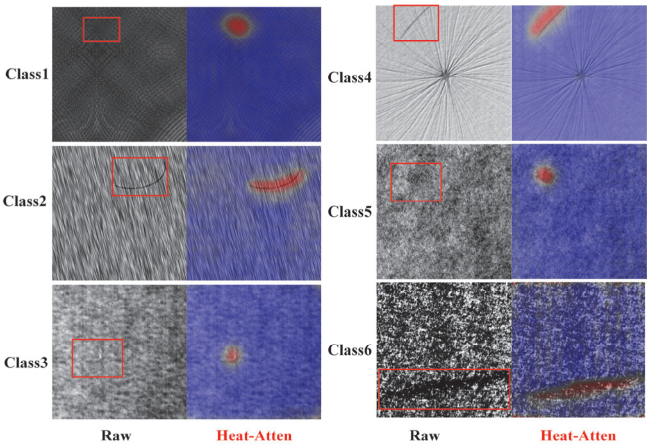

Fig 10 Visualization of defect images of DAGM 2007 by attention mechanism

D. Experimental Results

To validate the effectiveness on AttenCGAN with small samples in SDC, AttenCGAN is compared with the famous CNN models in the field. In the experiment, several defect samples are randomly selected from the dataset to train the MobileNet-V2, ResNet-50 and VGG16. On the KolektorSDD, the numbers of defect samples are 1, 2, 4, 6 and 8, while these are 1, 2, 4, 6, 8, 10, 20, 30 and 40 on the DAGM 2007 dataset. For the contrast, the SDCNN model without the synthesized defect images in AttenCGAN is used for the comparison.

The training hyper-parameters of AttenCGAN in all experiments in this study are as following. Epoch is 200, batch size is set to be 8 and learning rate is 0.001.

1) Experimental Results on KolektorSDD

The comparison results of AttenCGAN on KolektorSDD dataset is presented on TABLE II. It shows that the classification accuracies of MobileNet-V2, ResNet-50 and VGG16 are greatly affected by the sample size of surface defects. When the number of the defect samples changes from 8 to 1, the classification accuracies of them are decreased from near 95%~97% to only 62% to 66%. It also can be seen that these three methods rely on the number of the defect samples heavily on this dataset, since the results of them degrade significantly with the decreasing on the defect samples.

TABLE II
THE ACCURACIES OF THE COMPARED METHODS ON KOLEKTORSDD (%)

Methods	1	2	4	6	8
MobileNet-V2	62.00	78.41	81.41	93.06	95.06
ResNet-50	64.45	79.41	83.55	94.12	96.06
VGG16	65.77	79.58	84.42	96.06	97.06
SDCNN	86.44	94.79	96.05	96.52	97.06
AttenCGAN	**87.72**	**97.06**	**98.46**	**98.51**	**98.53**

Although AttenCGAN is also affected by the samples number, it has shown the good performances and the performances of AttenCGAN are higher than MobileNet-V2, ResNet-50 and VGG16 on all experiments with the 1 to 8 defect samples. When the number of defect images decreases from 8 to 1, the classification accuracy degrades from 98.53% to 87.72%. However, even though there are only two defect samples, the classification accuracy of AttenCGAN are as high as 97.06%, while these values of MobileNet-V2, ResNet-50 and VGG16 are 78.41%, 79.41% and 79.58%. These results show that the AttenCGAN can promote a lot on the small defect samples situations and it need fewer defect images than the baseline methods.

TABLE III
COMPARISONS WITH PUBLISHED METHODS ON KOLEKTORSDD

Methods	Average Accuracy (%)
Li [33]	99.16
LFCSDD [34]	98.74
Xu [35]	98.00
AttenCGAN	**98.53**

When there is only one defect image, the accuracy of AttenCGAN degrades a lot, but it is still superior to the baseline methods. The accuracy of AttenCGAN with 1 defect sample is 87.72% while these values of MobileNet-V2, ResNet-50 and VGG16 are 62.00%, 64.45% and 65.77%. This show that AttenCGAN is potential on with the extreme small defect samples. On the other hand, the AttenCGAN is also compared with the SDCNN which is trained without the synthesized defect images. The results show that AttenCGAN can be promoted on all situations with different numbers of defect samples.

Several published methods in literatures are selected for the comparison, and the results are presented on TABLE III. Compared with the accuracy of 99.16% and 98.74% obtained by Li [33] and LFCSDD [34], AttenCGAN can obtain the competitive accuracy with small samples, and AttenCGAN is better than Xu [35], which is 98.00%. These results can validate the performance of AttenCGAN further.

2) Experimental Results on DAGM 2007

The comparison results on DAGM 2007 are shown in TABLE IV. From the result, it can be inferred that small samples have shown the great effects on the performance of all methods. Almost all methods are stable when the number of the defect sample are higher than 10. But then they degrade then the number of the defect sample are less than 6. Compared with MobileNet-V2, ResNet-50 and VGG16, AttenCGAN can still achieve near 90% when the number of the defect samples are 6. But on the extreme cases where there are only 1 defect image, the AttenCGAN has achieved near 76% on Class 1 to 6, where these values of MobileNet-V2, ResNet-50 and VGG16 are near 60%. These results can provide the solid support that AttenCGAN can be promoted a lot on the situation of small samples on DAGM 2007 dataset. But AttenCGAN can achieve better results with the extreme small samples than the baseline methods.

On all experiments with different numbers of defect samples, the performance of AttenCGAN are all higher than MobileNet-V2, ResNet-50, VGG16 and SDCNN. With 40 defect samples, the accuracies of AttenCGAN are 100% on Class 1, Class 2, Class 3 and Class 6, and 99.57% and 97.83% on Class 4 and Class 5.

TABLE IV
THE ACCURACY OF DIFFERENT CLASSIFICATION METHODS ON DAGM2007 (%)

Data No.	Methods	1	2	4	6	8	10	20	30	40
Class1	MobileNet-V2	59.84	62.31	64.19	70.34	74.66	83.52	84.39	85.86	87.78
	ResNet-50	60.03	64.17	65.50	73.55	78.51	84.78	85.03	85.83	86.65
	VGG16	61.24	65.39	66.21	74.38	79.43	86.57	86.93	87.96	88.89
	SDCNN	74.57	77.63	84.31	86.68	93.58	98.09	99.11	99.67	99.79
	AttenCGAN	**76.35**	**79.52**	**86.79**	**88.97**	**95.37**	**98.80**	**99.42**	**99.95**	**100**
Class2	MobileNet-V2	59.89	62.39	64.63	70.39	74.76	94.05	95.66	97.23	98.63
	ResNet-50	61.07	64.43	66.01	73.60	78.62	94.78	95.71	96.44	98.50
	VGG16	61.27	65.37	66.07	74.42	80.03	95.38	96.72	97.88	98.99
	SDCNN	74.41	77.35	83.97	86.72	93.29	96.79	97.33	97.67	97.83
	AttenCGAN	**76.31**	**79.43**	**86.82**	**89.07**	**95.35**	**98.77**	**99.89**	**99.97**	**100**
Class3	MobileNet-V2	58.78	61.44	63.72	69.52	73.98	86.09	87.45	90.71	95.13
	ResNet-50	60.23	63.55	65.67	72.76	77.94	86.09	88.09	90.56	96.09
	VGG16	60.46	63.68	66.01	73.67	80.00	86.09	88.09	89.87	97.29
	SDCNN	73.91	76.90	83.79	86.72	92.41	97.13	98.50	99.79	99.96
	AttenCGAN	**75.89**	**78.73**	**85.93**	**88.98**	**94.35**	**97.79**	**98.87**	**99.94**	**100**
Class4	MobileNet-V2	58.56	61.21	63.54	69.18	73.58	85.65	87.70	90.62	93.43
	ResNet-50	60.03	63.05	65.23	72.06	77.60	85.65	86.34	89.09	94.06
	VGG16	60.34	63.46	66.03	73.35	80.01	85.65	90.04	92.55	95.22
	SDCNN	72.98	76.77	83.21	86.61	92.44	98.13	98.56	99.37	99.53
	AttenCGAN	**75.04**	**78.23**	**85.13**	**88.35**	**94.30**	**98.78**	**98.89**	**99.48**	**99.57**
Class5	MobileNet-V2	59.78	61.66	64.02	69.68	74.11	87.83	89. 38	90.17	92.61
	ResNet-50	60.53	63.63	65.71	73.41	78.16	87.83	90.83	95.22	96.09
	VGG16	60.66	63.78	66.33	74.15	80.97	87.83	89.87	91.26	96.39
	SDCNN	73.08	76.85	83.35	86.76	92.57	96.42	96.91	97.45	97.81
	AttenCGAN	**75.29**	**78.47**	**85.30**	**88.85**	**93.79**	**96.90**	**97.08**	**97.63**	**97.83**
Class6	MobileNet-V2	60.03	62.35	64.49	70.46	74.91	86.52	89.87	90.67	93.32
	ResNet-50	61.14	64.23	65.59	73.67	78.88	86.52	91.30	93.04	94.39
	VGG16	61.89	65.50	66.44	74.78	79.76	86.82	90.87	93.41	95.79
	SDCNN	74.59	77.68	84.55	86.71	93.62	98.03	99.25	99.53	99.81
	AttenCGAN	**76.40**	**79.53**	**86.77**	**89.02**	**95.55**	**98.98**	**99.88**	**99.92**	**100**

The results of AttenCGAN are compared with several famous SDC methods in literature, and the results are shown in TABLE V. From the results, it can be seen that the accuracy obtained by AttenCGAN is 99.57% (The average accuracy on Class 1 to Class 6), and it outperforms VGG16[37], Tree2vector[36], LSDDN-v2[38] and Solar cell CNN[36], which has the accuracy of 88.00%, 96.92%, 97.23% and 99.26%. AttenCGAN is competitive with LFCSDD[34] which obtained 99.72%, and AttenCGAN is slightly inferior to LFCSDD with the small defect samples.

TABLE V
COMPARISONS WITH THE PUBLISHED METHODS ON DAGM2007 (%)

Methods	Average Accuracy
VGG 16[37]	88.00
Tree2vector[36]	96.92
LSDDN-v2[38]	97.23
Solar cell CNN[36]	99.26
LFCSDD[34]	99.72
AttenCGAN	**99.57**

E. Discussion

In this section, Firstly, AttenCGAN and other GANs methods are compared on the two datasets. Secondly, the classification accuracy of TL and other fine-grained SDC methods on these two datasets is compared respectively. Finally, the advantages of the proposed SDC method are summarized.

1) Comparisons with GANs

AttenCGAN is compared with GAN and StarGAN [39] to compare the FID values. 8 defect images are randomly selected from each defect category in KolektorSDD and DAGM2007 datasets for the experiment.

TABLE VI
COMPARISONS FID VALUE WITH GANS

Methods	Kolektor SDD	DAGM2007					
		Class1	Class2	Class3	Class4	Class5	Class6
GAN	232	263	269	285	301	300	271
StarGAN	175	211	215	217	228	231	223
AttenCGAN	**122**	**168**	**169**	**175**	**195**	**197**	**168**

It can be seen from Table VI that the training of GANs is significantly affected by the training data. Among them, the FID value obtained by GAN is the largest in both datasets, that is to say, the quality of the synthesized pseudo-defect image is the worst among these three methods. StarGAN is one of improved GANs, and the synthesized image is greatly improved compared with the original GAN, and its FID value is near 70 lower at most than that of GAN. The minimum FID value obtained by AtenCGAN is 122 on the KolektorSDD dataset, while these are 232 and 175 by GAN and StarGAN respectively, which are 90 and 53 higher than AttenCGAN. The proposed AttenCGAN can have the best FID values on all the cases, showing that it can perform well under the small defects in the industrial environment.

2) *Comparisons with TL and Fine-Grained SDC Methods*

The comparison of the proposed AttenCGAN and TL and fine-grained SDC methods are presented to show the performance of AttenCGAN under the situation of small samples situation and the characteristic of small inter-class gap between normal samples and defective samples. Firstly, the TL models which are pre-trained on the ImageNet dataset are conducted for the defect classification. The pre-trained Mobilenet-V2, ResNet-50, and VGG16 are used. Secondly, the fine-grained classification methods selected for comparison are

PMG [40] and NTS-Net [41]. The number of defect images randomly selected from KolektorSDD and DAGM2007 datasets is 8 for each defect case.

TABLE VII
COMPARISONS WITH TL AND FINE-GRAINED SDC METHODS (%)

Methods	Kolektor SDD	DAGM2007					
		Class1	Class2	Class3	Class4	Class5	Class6
MobileNet-V2(TL)	96.12	80.60	80.53	79.97	79.66	80.47	81.00
ResNet-50 (TL)	96.89	81.26	81.07	80.99	79.89	81.09	81.55
VGG16 (TL)	97.71	82.13	83.15	83.02	83.00	83.52	82.69
PGM	93.62	88.17	92.13	88.41	86.53	88.44	87.62
NTS-Net	90.31	86.45	89.18	86.39	85.52	86.51	86.08
AttenCGAN	**98.53**	**95.37**	**95.35**	**94.35**	**94.30**	**93.79**	**95.55**

Table VII shows that AttenCGAN has higher classification accuracy than TL and fine-grained SDC methods. Combined with Table II and Table VII, it can be seen that the TL models have been significantly improved. Whereas, the classification accuracy of AttenCGAN is also better than the TL models. On KolektorSDD dataset, the classification accuracy of AttenCGAN can promote near 2% higher than that of TL models. On DAGM2007 dataset, the classification accuracy of three TL models is all below than 90%, and VGG16(TL) achieved the highest accuracy of 83.00% on Class 5, while AttenCGAN is much higher than the selected TL models. On the aspect on the fine-grained method, AttenCGAN can promote near 6-8% on the KolektorSDD dataset. On DAGM2007 dataset, the highest accuracies of PGM and NTS-Net is 92.13% and 89.18% on Class 2, while AttenCGAN can still promote 3.42% and 6.37%. These results show the performance of AttenCGAN.

3) The Discuss of AttenCGAN

In this research, a new AttenCGAN is proposed for the SDC with small samples. It has been conducted on two datasets, and the results validate that it has achieved a good promotion.

Firstly, the AttenCGAN is applied to synthesize many pseudo defect samples using only small defect samples. The results show that it can work well compared with other famous CNN models. AttenCGAN has better anti-degradation ability along with the decreasing of the defect samples. AttenCGAN can still work well when the defect samples are 2 on the KolektorSDD dataset and 8 on DAGM 2007 dataset. Even though there is only one defect sample, AttenCGAN also can promote a lot.

Secondly, AttenCGAN uses the SDCNN to classify the defect samples with attention mechanism. With the help of attention, the discriminative parts of the defect samples can be found, and the attention-guided data augment can also promote the performance of AttenCGAN in all the experiments with a different number of defect samples.

V. CONCLUSIONS AND FUTURE RESEARCH

This research studies a new CNN based SDC method with attention mechanism (AttenCGAN) under the small sample situation. The main contributions of this research are presented as followed. Firstly, a new CycleGAN based defect image synthesis method is investigated and it can generate high-quality defective images to avoid performance degradation by insufficient training samples. Secondly, the attentional mechanism is applied in AttenCGAN. With the help of attentional mechanism, AttenCGAN can pay more attention to the discriminative areas in defect images and recognize the small differences between the samples. AttenCGAN is validated on the KolektorSDD and DAGM2007. The results show that AttenCGAN can copy with both the situations of fewer defect samples and small intra-class differences in SDC, and it shows the better anti-degradation ability along with the decreasing of the defect samples. AttenCGAN is also compared with other published SDC methods, which validates its potential in the field of SDC.

The limitations of AttenCGAN include the following aspects. Firstly, AttenCGAN can perform well with a few samples, but it is still cannot work on the situations when no defect samples are provided. Secondly, the AttenCGAN uses the attention guided data augmentation to enhance the discriminative feature, but how to use more powerful attentions to promote the discriminative feature still need to be investigated. Based on these limitations, the future research can be conducted in the following ways. Firstly, the zero-shot surface defect detections can be investigated using AttenCGAN to enhance its usability on SDC. Secondly, a more powerful attention mechanism can be investigated, and then it can be used to find the discriminative area and promote AttenCGAN further.

REFERENCES

[1] P. Helo, Y. Hao, R. Toshev, and V. Boldosova, “Cloud manufacturing ecosystem analysis and design,” Robotics and Computer-Integrated Manufacturing, vol. 67, pp. 102050, 2021.

[2] C. Yang, Y. Wang, S. Lan, L. Wang, W. Shen, and G. Q. Huang, “Cloud-edge-device collaboration mechanisms of deep learning models for smart robots in mass personalization,” Robotics and Computer-Integrated Manufacturing, vol. 77, pp. 102351, 2022.

[3] J. Lian, W. Jia, M. Zareapoor, Y. Zheng, R. Luo, D. K. Jain, and N. Kumar, “Deep-Learning-Based Small Surface Defect Detection via an Exaggerated Local Variation-Based Generative Adversarial Network,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1343-1351, 2020.

[4] D. Zhang, K. Song, Q. Wang, Y. He, X. Wen, and Y. Yan, “Two Deep Learning Networks for Rail Surface Defect Inspection of Limited Samples with Line-Level Label,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6731-6741, 2021.

[5] H.P. Lu and C.T. Su, “CNNs Combined with a Conditional GAN for Mura Defect Classification in TFT-LCDs,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 34, no. 1, pp. 25-33, 2021.

[6] G. Shao, M. Huang, F. Gao, T. Liu and L. Li, “DuCaGAN: Unified Dual Capsule Generative Adversarial Network for Unsupervised Image-to-Image Translation,” *IEEE Access*, vol. 8, pp. 154691-154707, 2020.

[7] J. Liu, B. G. Zhang and L. Li, “Defect detection of fabrics With Generative Adversarial Network Based flaws modeling,” 2020 Chinese Automation Congress (CAC), Nov. 2020, pp. 3334-3338.

[8] J. Hu, P. Yan, Y. Su, D. Wu and H. Zhou, “A Method for Classification of Surface Defect on Metal Workpieces Based on Twin Attention Mechanism Generative Adversarial Network,” *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13430-13441, 2021.

[9] L. Wen, X. Li and L. Gao, “A New Reinforcement Learning Based Learning Rate Scheduler for Convolutional Neural Network in Fault Classification,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 12, pp. 12890-12900, 2021.

[10] S. Chevon, H. Lee, C. O. Kim and S. H. Lee, “Convolutional Neural Network for Wafer Surface Defect Classification and the Detection of Unknown Defect Class,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 2, pp. 163-170, 2019.

[11] H. Zhu, W. Ge, and Z. Liu, “Deep Learning-Based Classification of Weld Surface Defects,” *Applied Sciences*, vol. 9, no. 16, 2019.

[12] S. Q. Wu, S. Y. Zhao, Q. Q. Zhang, L. Chen, and C. R. Wu, "Steel Surface Defect Classification Based on Small Sample Learning," *Applied Sciences*, vol. 11, no. 23, 2021.

[13] Y. Liu, Y. Yuan, and J. Liu, "Deep learning model for imbalanced multi-label surface defect classification," *Measurement Science and Technology*, vol. 33, no. 3, pp. 035601, 2021.

[14] X. Wan, X. Y. Zhang, and L. L. Liu, "An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples an d Imbalanced Datasets," *Applied Sciences*, vol. 11, no. 6, 2021.

[15] S. Jain, G. Seth, A. Paruthi, U. Soni, G. Kumar, "Synthetic data augmentation for surface defect detection and classification using deep learning," *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-020-01710-x, 2020.

[16] Y. He, K. Song, H. Dong, and Y. Yan, "Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network," *Optics and Lasers in Engineering*, vol. 122, pp. 294-302, Nov. 2019.

[17] Y. He, K. Song, H. Dong, and Y. Yan, "Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network," *Optics and Lasers in Engineering*, vol. 122, pp. 294-302, Nov. 2019.

[18] Z. Zhao, B. Li, R. Dong, and P. Zhao, "A Surface Defect Detection Method Based on Positive Samples," in Proc. Proceedings of the PRICAI 2018: Trends in Artificial Intelligence, Nanjing, China, Aug. 2018, pp. 473-481.

[19] J. Lian, W. Jia, M. Zareapoor, Y. Zheng, R. Luo, D. K. Jain, and N. Kumar, "Deep-Learning-Based Small Surface Defect Detection via an Exaggerated Local Variation-Based Generative Adversarial Network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1343-1351, Feb. 2020.

[20] A. Jiangsha, L. Tian, L. Bai, and J. Zhang, "Data augmentation by a CycleGAN-based extra-supervised model for nondestructive testing," *Measurement Science and Technology,* vol. 33, no. 4, pp. 045017, Jan. 2022.

[21] S. Niu, B. Li, X. Wang, S. He, Y. Peng, "Defect attention template generation cycleGAN for weakly supervised surface defect segmentation," *Pattern Recognition*, vol. 123, pp. 108396, 2022.

[22] D. Tsai, S. Fan and Y. Chou, "Auto-Annotated Deep Segmentation for Surface Defect Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, 2021, Art no. 5011410.

[23] H. Liu, B. He, Y. He, X. Tao, "Lightweight detection algorithm for fine-grained surface defects of aerospace seal rings," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 6, pp. 11423-11440, 2021.

[24] L. Cui, X. Jiang, M. Xu, W. Li, P. Lv and B. Zhou, "SDDNet: A Fast and Accurate Network for Surface Defect Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-13, 2021, Art no. 2505713.

[25] Y. Zhang, J. Wu, Q. Li, X. Zhao and M. Tan, "Beyond Crack: Fine-Grained Pavement Defect Segmentation Using Three-Stream Neural Networks," *IEEE Transactions on Intelligent Transportation Systems*, doi: 10.1109/TITS.2021.3134374, 2021.

[26] H. Huang, J. Xu, J. Zhang, Q. Wu, and C. Kirsch, "Railway Infrastructure Defects Recognition using Fine-grained Deep Convolutional Neural Networks," in *Proc. 2018 Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, Australia, Dec. 2018, pp. 1-8. Z.

[27] Tu, S. Wu, G. Kang and J. Lin, "Real-Time Defect Detection of Track Components: Considering Class Imbalance and Subtle Difference Between Classes," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-12, 2021, Art no. 5017712.

[28] D. Li, Q. Xie, Z. Yu, Q. Wu, J. Zhou, and J. Wang, "Sewer pipe defect detection via deep learning with local and global feature fusion," *Automation in Construction,* vol. 129, pp. 103823, Sep. 2021.

[29] Z. Tang, E. Tian, Y. Wang, L. Wang and T. Yang, "Nondestructive Defect Detection in Castings by Using Spatial Attention Bilinear Convolutional Neural Network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 82-89, 2021.

[30] X. Tao, D. Zhang, W. Hou, W. Ma, and D. Xu, "Industrial Weak Scratches Inspection Based on Multifeature Fusion Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-14, 2021.

[31] C. Hu and Y. Wang, "An Efficient Convolutional Neural Network Model Based on Object-Level Attention Mechanism for Casting Defect Detection on Radiography Images," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 12, pp. 10922-10930, 2020.

[32] B. Su, H. Chen, P. Chen, G. Bian, K. Liu and W. Liu, "Deep Learning-Based Solar-Cell Manufacturing Defect Detection with Complementary Attention Network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4084-4095, 2021.

[33] Y. Li and J. Li, "An End-to-End Defect Detection Method for Mobile Phone Light Guide Plate via Multitask Learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-13, 2021.

[34] Y. Li, Y. Chen, Y. Gu, J. Ouyang, J. Wang, and N. Zeng, "A Lightweight Fully Convolutional Neural Network of High Accuracy Surface Defect Detection," Proc. Proceedings of the 29th International Conference on Artificial Neural Networks (ICANN 2020), Bratislava, Slovakia, Sep 2020, pp. 15-26.

[35] L. Xu, S. Lv, Y. Deng and X. Li, "A Weakly Supervised Surface Defect Detection Based on Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 42285-42296, 2020.

[36] H. Chen, Q. Hu, B. Zhai, H. Chen, and K. Liu, "A robust weakly supervised learning of deep Conv-Nets for surface defect inspection," *Neural Computing and Applications*, pp. 1-16, 2020.

[37] R. Li, "Metal surface defect detection based on few defect datasets," Proc. Proceedings of the 2019 5th international conference on green power, materials and manufacturing technology and applications (GPMMTA 2019), Dec. 2019, Vol. 2185, No. 1, pp. 020027.

[38] Y. Qin, Y. Xing, and J. Du, "LSDDN: A Lightweight End-to-End Network for Surface Defect Detection," Proc. ICIAI 2020: 2020 the 4th International Conference on Innovation in Artificial Intelligence, May. 2020, pp. 79-83.

[39] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789-8797.

[40] R. Du, D. Chang, A. K. Bhunia, J. Xie, Z. Ma, Y. Z. Song, and J. Guo, "Fine-grained visual classification via progressive multi-granularity training of jigsaw patches," in *Proc. European Conference on Computer Vision (ECCV)*, Springer, 2020, pp. 153-168.

[41] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, "Learning to navigate for fine-grained classification," in *Proc. Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, Sep. 2018, pp. 420-435.

Long WEN received his Ph.D. degree in industrial engineering from Huazhong University of Science and Technology, China, 2014.
He is a Professor in the School of Mechanical Engineering and Electronic Information, China University of Geosciences. He had published more than 30 refereed papers, and his research interests include deep learning, automatic machine learning, fault diagnosis and intelligent algorithm, etc.

You WANG received his bachelor's degree in Mechanical Design manufacture and Automation from China University of Geosciences, China, 2018.
He is a master in the School of Mechanical Engineering and Electronic Information, China University of Geosciences. He researches interests include deep learning, automatic machine learning, fault diagnosis and intelligent algorithm, etc.

Xinyu LI (M'19) received his Ph.D. degree in industrial engineering from Huazhong University of Science and Technology, China, 2009.
He is a Professor of the Department of Industrial & Manufacturing Systems Engineering, State Key Laboratory of Digital Manufacturing Equipment & Technology, School of Mechanical Science & Engineering, HUST. He had published more than 80 refereed papers. His research interests include intelligent algorithm, big data and machine learning etc.