Acta Mechanica Sinica, Volume 41, Issue 5, May 2025, Article number 424076
DOI: 10.1007/s10409-024-24076-x

A Cyclic Self-enhancement Technique for Complex Defect Profile Reconstruction based on Thermographic Evaluation

Haochen Liu [a*], Shuozhi Wang [b], Yifan Zhao [b*], Kailun Deng [b] and Zhenmao Chen [c]

[a] Research Center for Applied Mechanics, Key Laboratory of Ministry of Education for Electronic Equipment Structure Design, Shaanxi Key Laboratory of Space Extreme Detection, Xidian University, Xi'an, 710071, China

[b] School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, MK43 0AL, UK

[c] State Key Laboratory for Strength and Vibration of Mechanical Structures, School of Aerospace, Xi'an Jiaotong University, Xi'an, 710049, China

* Corresponding authors' *E-mail address:* liuhaochen@xidian.edu.cn ; yifan.zhao@cranfield.ac.uk

Abstract

Although machine Learning has demonstrated exceptional applicability in thermographic inspection, precise defect reconstruction is still challenging, especially for complex defect profiles with limited defect sample diversity. Thus, this paper proposes a self-enhancement defect reconstruction technique based on Cycle-consistent Generative Adversarial Network (Cycle-GAN) that accurately characterises complex defect profiles and generates reliable artificial thermal images for dataset augmentation, enhancing defect characterisation. By using a synthetic dataset from simulation and experiments, the network overcomes the limited samples problem by learning the diversity of complex defects from FEM modelling and obtaining the thermography uncertainty patterns from practical experiments. Then, an iterative strategy with a self-enhancement capability optimises the characterisation accuracy and data generation performance. The designed loss function structure with cycle consistency and identity loss constrains the GAN's transfer variation to guarantee augmented data quality and defect reconstruction accuracy simultaneously, while the self-enhancement results significantly improve accuracy in thermal images and defect profile reconstruction. The experimental results demonstrate the feasibility of the proposed method by attaining high accuracy with optimal loss norm for defect profile reconstruction with a recall score over 0.92. The scalability investigation of different materials and defect types is also discussed, highlighting its capability for diverse thermography quantification and automated inspection scenarios.

Keywords

Non-destructive Testing and Evaluation; Complex defect reconstruction; Generative Adversarial Network; Thermographic data augmentation; Self-enhancement;

Published by Springer. This is the Author Accepted Manuscript issued with: Creative Commons Attribution License (CC:BY 4.0).
The final published version (version of record) is available online at DOI:10.1007/s10409-024-24076-x. Please refer to any applicable publisher terms of use.

1. Introduction

Quantitative damage & degradation evaluation for key components, structures and materials are widely recognised as an indispensable cornerstone in industrial modernisation, maintenance science and life-cycle engineering management. The requirements in terms of Structural Health Monitoring (SHM), manufacturing quality control, damage inducement & propagation analysis raise an increased demand for inspection accuracy, resolution, and decision reliability. Non-Destructive Testing & Evaluation (NDT&E) provides multiple profound technologies for fault diagnosis and damage evaluation to understand the product life-cycle and support effective maintenance schedules [1–3]. Infrared Thermography (IRT) is widely developing due to its rapid, non-contact, intuitive, robust, non-invasive inspection, which has been deployed in multiple industries for integrity assessment and material property characterisation.

The thermal imagery or videos are the typical raw inspection outputs in IRT which contain not only the feature of defects during the heat transfer but also the material and structural thermal pattern in both spatial and time domains. Traditional IRT post-processing methods aim to enhance the defect contrast and improve the Signal-to-Noise Ratio (SNR) for higher detectability and measuring the location, size or depth of defects. Typical methods [4], including Thermographic Signal Reconstruction (TSR), Principal Component Thermography (PCT), and Pulse Phase Thermography (PPT), have been widely used to enhance the SNR of defective areas [5, 6]. However, accurate reconstruction of the defect profile is always time-consuming and needs numerous calibration samples as well as human pre-knowledge. It is especially difficult for small and complex irregular profiles, the boundary of which will be distorted and blurred in thermal images. Further, the deviation in the heating transfer and capturing process always result in uncertainties for decision-making in periodic inspection. Therefore, an automatic, accurate defect profile reconstruction method with less input from human pre-knowledge is highly demanded.

Being grateful to the powerful feature mining of machine learning algorithms and increasing application in IRT data, been developed for aiding automatic inspection decision-making [7]. Neural networks (NNs), with their various variations techniques, are reported to be effective in damage and defect identification. Recurrent NNs (RNNs) like Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) [8] have been applied for applications for enhancement of the probability of detection, and classification of defect types utilising temporal thermal features within pixel-level. Defect features [9, 10] like thermal conductivity, temperature contour, and temporal derivative peaks are well-learned to discriminate the defect position, infill, depth and size. Additionally, for defect boundary characterisation, imagery-based Convolutional NNs (CNNs) also present effectiveness at spatial feature learning against noise for defect profile segmentation. Xie et al. [11] proposed an unsupervised learning Autoencoder model to improve the visibility of rear surface cracks in inductive thermography. Marani et al. [12] used CNN to detect defects with multiple sizes and depths. As another type of CNN-based method tending to explore a single thermal image, VGG, CNN variations and transfer learning models have been applied in cooling radiators, solar photovoltaic modules, CFRP manufacturing [13], solar energy key components [14], machine faults diagnosis [15, 16], and crack detection

in integrated structures and materials in electronic packaging components [17, 18]. In combination networks, transfer learning and deep learning are structured to improve defect detection in composites. Further, the comparison works [19] have been conducted to report the effectiveness of the synthetic dataset, including simulation and experimental data using YOLO, Fast-RCNN, and Mask-RCNN. It reported that the synthetic dataset is practical and helpful in scenarios without enough experimental data [20]. However, these researches mainly deal with regular shape defects like circles or squares rather than irregular complex shape defects, and they need adequate training data in advance to guarantee accuracy.

A key factor for the machine learning-based data analysis for inspection is the database, which is required to be both comprehensive and diversified to ensure the network can learn sufficient cases and reliable features of interest, and avoid overfitting simultaneously. An insufficient dataset may result in cognitive uncertainties and inaccurate characterisation. However, it is costly and sometimes impossible to acquire numerous defect samples from real-world NDT applications. To address this challenge, the synthetic data, fused with multiple types of data sources (simulations and experiments), can effectively provide complementation, which utilises the data diversity from simulations and conforms to realistic factors from experimental data. Generative Adversarial Networks (GANs) is proposed to serve as an effective synthetic data augmentation tool to generate new data with similar analytical features [21-23]. GANs are employed to enrich fault samples to detect the faults in bearing faults [24]. It is also utilised to generate minority samples as additional sources for unbalanced datasets to diagnose faults [25]. For IRT, the hypothesis is that GAN can learn the structural, material, and thermal transfer mechanics and produce the data of new defective samples [26, 27]. Similarly, the adversarial learning capability of GANs can also be utilised for defect inspection to resolve the noise influence in thermography [28, 29]. The GAN was structurally improved in the joint loss to enhance its capability for training stability, which results in a higher detection rate in multiple defect objectives [30, 32]. The Cycle-GANs network, proposed by [31], provides a dual-direction concept for image translation between graphical styles with double-trained GANs. The evolved networks have been reported for machine failure diagnosis [33], crack inspection by vision detectors [34], and fault component detection in rail tracks. It lays down a promising architecture for thermographic inspection for both data augmentation purposes and defects characterisation enhancement using the augmented database. However, there is still very limited research to investigate its applicability for IRT.

This paper proposes a self-enhancement thermography inspection network to upgrade the automatic complex defect reconstruction using a cyclic dual image generation strategy. The novelty of the proposed strategy lies in the following aspects:

1. It provides an automatic and accurate characterisation approach for internal complex defect profiles in thermographic NDT.
2. A FEM+GAN strategy combined with a synthetic database of simulation and experiment is introduced to overcome the challenge of limited defect samples for network training.

3. A Cycle-GAN-based strategy is proposed for thermal inspection data augmentation and defect reconstruction simultaneously.
4. It is equipped with an iterative self-enhancement capability to find out the optimal defect characterisation accuracy and thermographic data generation quality.

The rest of the paper is organised as follows. The overall methodology and the network architecture are described in Section II. Then, the dataset preparation is presented in Section III. Section IV reports results and performance evaluation. The network loss optimisation and scalability of the proposed method are discussed in Section V. Finally, the conclusion and future work are given in Section VI.

2. Methodology

2.1. Overall methodology

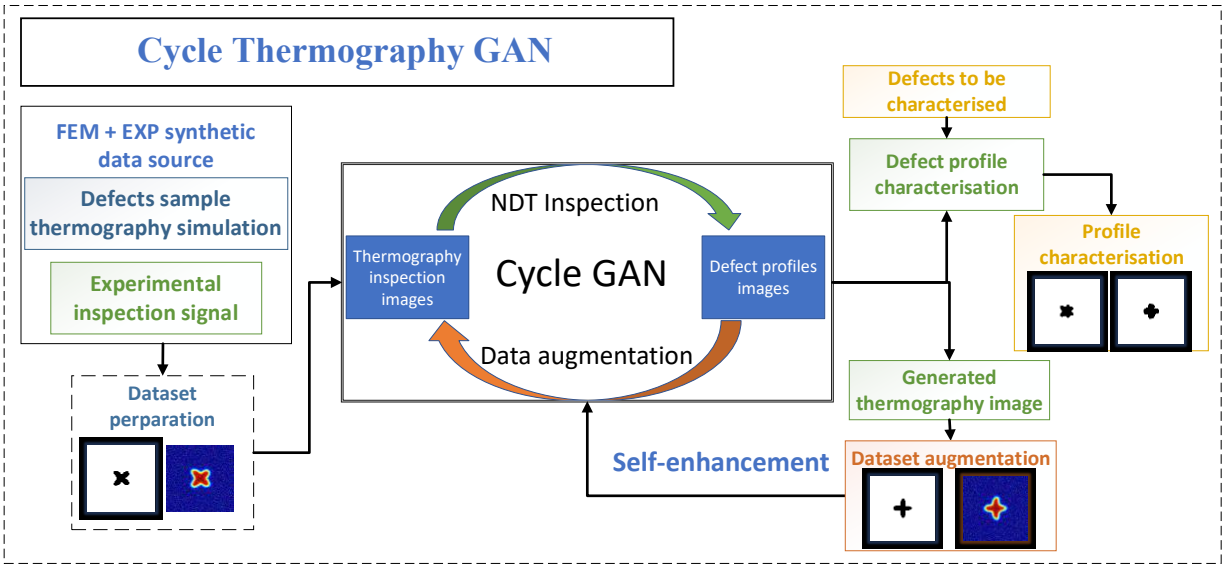


Figure 1. The proposed overall methodology

Figure 1 presents the principle and overall methodology of the proposed technique. In the first step, a synthetic dataset is established with both FEM simulation and experimental inspection results of the thermographic inspection for initial training of the cycle-GAN network. The experimental data provides realistic inspection uncertainty and thermal noise, while the simulation data provides diversity and comprehensiveness of the defect profiles, which is usually difficult to achieve in experiments. Then, the dataset of defect profiles and their corresponding thermal inspection image is fed to the network to learn the thermal transfer pattern of defects during thermographic inspection. By utilising dual generative discriminating capabilities, the network can generate thermal images from defect profiles in the form of binary images and can also determine the defect profile from thermographic inspection data. The generated thermal image data is used to self-improve the network iteratively by augmenting the thermal image dataset. This effectively addresses the limitation of AI-based decision-making for an NDT inspection with a small database. Finally, thanks to the self-enhanced network, the inspection discriminator is able to reconstruct complex defect profiles accurately. Benefitting from the comprehensiveness of synthetic data and generative self-improvement

architecture, this bilateral AI-based thermography technique can provide powerful applicability for complex defect reconstruction. It also can be used as an effective data generator for the application of other machine learning methods.

2.2. Network Architecture

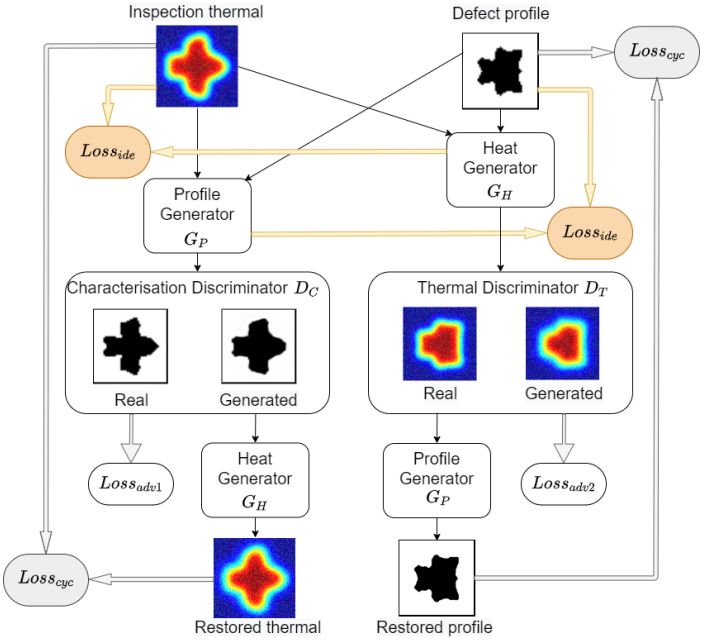


Figure 2. Network architecture

Generative adversarial models have shown effective applicability in image denoising, image translation, linguist interpretation and semantic segmentation. A traditional GAN-based network consists of a generator and discriminator denoted by G and D, respectively. The generator G focuses on producing samples to data space from a prior noise distribution, while the discriminator aims to judge the real/fake probability of a data space sample. Two operators (generator and discriminator) compete to win in the minimax campaign while respectively improving themselves from opposite judgement. The loss function of a GANs-type model is defined as Eq. (1)

$$\min_{G}\max_{D}\mathcal{L}_{GAN}(G,D_Y,X,Y)=\mathbb{E}_{y\sim p_{data}(y)}[logD_Y(y)]+\mathbb{E}_{x\sim p_{data}(x)}\left[log(1-D_Y(G(x))\right] \quad (1)$$

where X and Y denote the random noise space and target sample space, respectively. The generator G attempts to generate samples $G(x)$ that target the domain Y, while the task of the discriminator D_Y is to distinguish between translated samples $G(x)$ and the ground truth y. For inspection applications, most studied GANs are normally utilised for single detection purposes, such as defect localisation or segmentation from thermal images or thermal image augmentation from defect profiles.

A cycle thermography GAN design is proposed in this paper for thermography data augmentation and complex defect reconstruction simultaneously based on the cycle-GANs principle. This network forms a double-channel and nose to tail structure to implement double generating of thermography thermal/defect profile images, while simultaneously dually discriminating their probability of generated validation. Based on

the cycle-GAN principle, double adversarial training models can learn to map the outputs of two generators G and F loopily that yield the same distribution as the target domains Y and X, respectively.

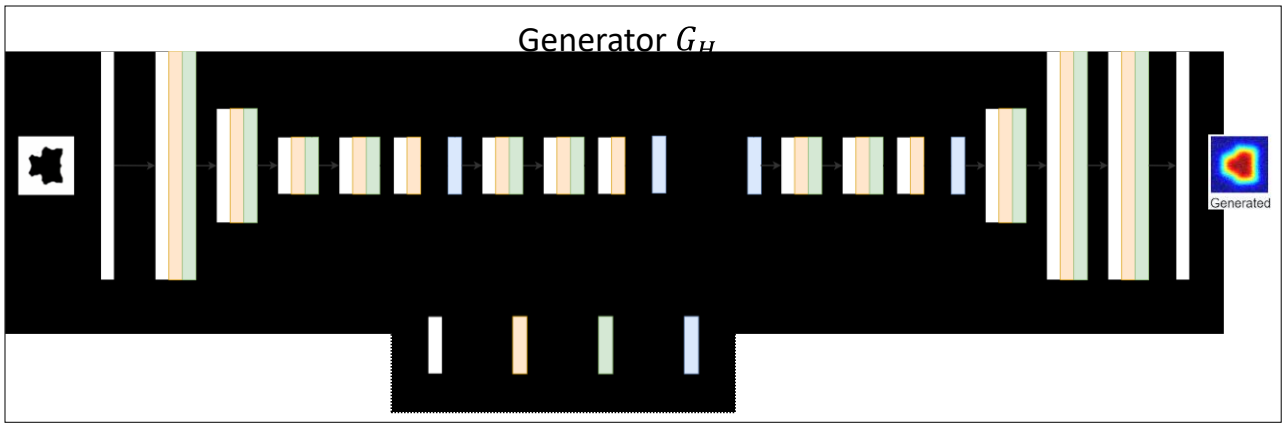


Figure 3. Structure of the heat generator in half cycle GAN

The proposed network architecture is shown in Figure 2. The first GAN includes a profile generator G_P and a characterisation discriminator D_C. The second GAN includes a heat generator G_H and a thermal discriminator D_T. In the first half cycle, a thermographic inspection image is fed into the profile generator G_P to generate (detect) its defect profile and try to trick the characterisation discriminator D_C. Then the generated profile image is translated back to the thermal image by G_H. Similarly, in the second half cycle, a defect profile binary image is fed into the thermal generator G_H to generate (estimate) its thermographic inspection image and try to trick the infrared discriminator D_T. Then the generated thermal image is translated back to the profile binary image by G_P. The overall loss function can be expressed by Eq. (2).

$$\mathcal{L}_{overall} = \mathcal{L}_{adv1} + \mathcal{L}_{adv2} + \lambda\mathcal{L}_{cyc} + \beta\mathcal{L}_{ide} \tag{2}$$

where $\mathcal{L}_{adv1}$, $\mathcal{L}_{adv2}$ denote the two GANs loss functions respectively. We believe that the cycle consistency loss $\mathcal{L}_{cyc}$ and the identity loss $\mathcal{L}_{ide}$ is important for this network. The former controls the overall cycle generative transferability, which means that the results after the cycle should be close to the actual original input. $\mathcal{L}_{ide}$ is the essential identity loss reducing the degree of variation if the data class of the input is similar to the desired output, which, to a certain extent, constrains the transferability of each generator. In this thermographic inspection, the accuracy of generative data has the top priority, and the translation between thermal and binary images is not as complex as other image style transfers. Therefore, $\mathcal{L}_{ide}$ is important in this self-enhancement method. The loss type selection is discussed in Section V. $\mathcal{L}_{cyc}$ and $\mathcal{L}_{ide}$ can be expressed by Eqs. (3) and (4).

$$\mathcal{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}\left[\| F\big(G(x)\big) - x \|_{\text{norm}}\right] + \mathbb{E}_{y\sim p_{data}(y)}\left[\| G\big(F(y)\big) - y \|_{\text{norm}}\right] \tag{3}$$

$$\mathcal{L}_{ide}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}\left[\| G(y) - y \|_{\text{norm}}\right] + \mathbb{E}_{y\sim p_{data}(y)}\left[\| F(x) - x \|_{\text{norm}}\right] \tag{4}$$

In the detailed structure, Figure 3 and Figure 4 illustrates the two key network components of heat generator G_H and thermal discriminator D_T in the half cyclc-GAN (right part of Figure 2), respectively. The profile

generator $G_P$and characterisation discriminator D_C can be recognised as an inverse replica of G_H and D_T. The heat generator is composed of image convolution, normalisation, activation layer and concatenate blocks to translate out the thermal pattern of the defect sample. The thermal discriminator is composed of convolution, normalisation and a Leaky Rectified Linear Unit to determine the quality of the generation. In Figure 3 and Figure 4 the white blocks represent the 2D convolution layer, and the orange ones represent the normalisation layer. The activation in Figure 3 is ReLU (described as green blocks). The leakyReLU is applied for the discriminator in Figure4. Each generator is composed of encoder for input feature extraction, transformer (contains 9 residual blocks) for style translating, and decoder (contains 2 deconvolution blocks) for restoring back to original dimension.

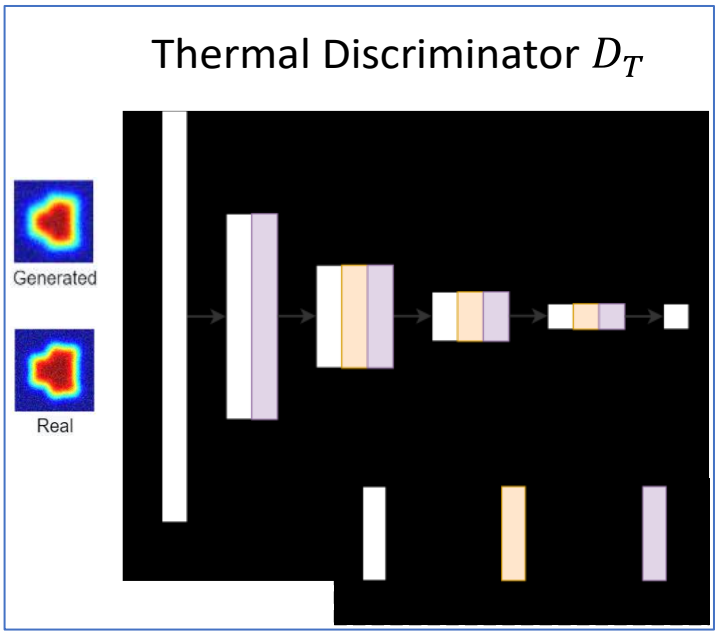


Figure. 4. Structure of the thermal discriminator in half cycle GAN

2.3. Network self-enhancement

Self-enhancement is one of the top features of using generative results for data augmentation. The CT-GAN mainly focuses on thermography image generation for dataset augmentation from random defect profiles. It improves the diversity of defect profile samples and highly accelerates the data augmentation speed compared to the FEM simulation. As shown in Figure 2, the augmented dataset is fed back to the network to iteratively enhance the learning capability.

Considering the risk of dataset contamination and degradation from random defect samples, this study proposes to use Peak Signal to Noise Ratio (PSNR) (shown as Eq. (5)) and Structural Similarity Index Metric (SSIN) (shown as Eq. (6)) as criteria to analyse the quality of generated augmenting data. The feedback augmenting samples are requested to be detected by the network to have a good enough (judged by PSNR and SSIM) result of the inspected (generated) profile. Those samples that have both scores over specific thresholds are qualified samples for feedback augmentation. Those samples that cannot be detected accurately by the current network, are not included. It guarantees a gradually self-improving behaviour rather than an

unquestioning dataset increment. The PSNR and SSIM can be expressed as follow

$$PSNR = 20 \times log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{5}$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j) - Y(i,j)]^2$$

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \tag{6}$$

where MSE is the mean squared error between two images. μ_X, μ_Y is the mean of X and Y, respectively. σ_X, σ_Y, σ_{XY} denote the variance of X and Y and covariance, respectively.

To evaluate the performance of the proposed network, two indexes are employed to quantitatively analyse the results of generative thermal data and defect characterisation. The performance evaluation for defect profile characterisation uses the Recall score which is a useful index for measuring image segmentation accuracy. The Recall score is calculated by measuring the overlap between the predicted profile and ground truth. For the generated thermal images, the Fréchet inception distance (FID) is utilised to assess the performance. FID is a metric used to assess the quality of images created by generative models. It compares the distribution of generated images with the distribution of real images, which analyses the generated data quality (e.g., the similarity of generated images to real ones) in statistical multi-dimensional Gaussian distributions. It can be written as

$$FID = \|\mu_X - \mu_Y\|^2 - Tr(\sum X + \sum Y - 2\sum X \sum Y) \tag{7}$$

where X and Y are the real and generated images respectively assumed to be two multivariate normal distributions. μ_X and μ_Y are the magnitudes of vector X and Y. Tr denotes the trace of the matrix, $\sum X$ and $\sum Y$ denote the covariance matrices of X and Y, respectively.

3. Dataset preparation and pre-processing

The dataset for this research is the synthetic dataset combined FEM simulation and experimental results of a Pulsed Thermography (PT) inspection for flat-bottom hole defects in metal materials.

3.1. Defect modelling by FEM simulation

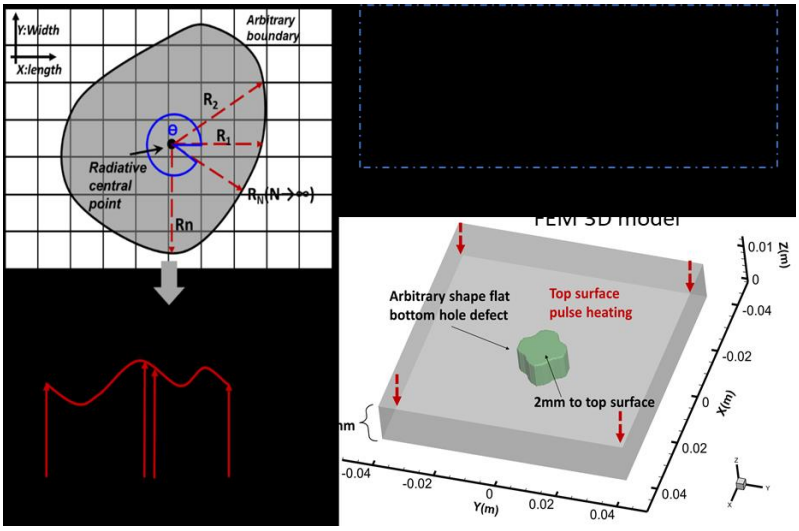


Figure 5. Defect modelling for complex profile bottom hole.

The governing equation of the PT heat transfer can be written as Eq. (8), and its FEM governing form is shown in Eq. (9).

$$\kappa\{\nabla\}^T\{\nabla\}T - \rho c\dot{T} = -Q|_\Gamma \tag{8}$$

$$[K]\{T(t)\} + [C]\left\{\frac{\partial T(t)}{\partial t}\right\} = \{Q(t)\} \tag{9}$$

where T and $\{\nabla\}$ denote the temperature and gradient vector respectively, and $-Q|_\Gamma$ is the excitation heat flux applied on the surface. κ, ρ and c denote the heat conductivity, density, and specific heat of the material, respectively.

To generate defects with a complex boundary, the polar coordinate defect profile parameterisation is employed. By using a circumferential radiative defect parameterisation and Fourier series fitting defect profile fitting strategy [35], planar arbitrary defect profiles can be represented and modified by changing the centre point location and several Fourier series coefficients. The scheme of defect samples modelling is shown on the left side of Figure 5. The FEM modelling defects are shown in the bottom-right corner of Figure 5. This modelling strategy can quickly generate a diversity of complex defect profiles with a small number of defect parameters, which supports the proposed FEM+GAN scheme with numerous various defect samples.

TABLE I
FEM SIMULATION SPECIFICATION

Item	Parameters of simulation models
Specimen dimension	Plate model: 80mm * 80mm * 6mm
Defect	Flat-bottom hole defects, depth: 2mm to the inspection surface.
Thermal properties	Carbon steel plate: ρ: 7.93 kg/m3; c: 500 J/(kg·K); κ: 45 W/(m·K);
Pulse heating	1 erg/G = 1 emu → 10^{-3} A·m^2 = 10^{-3} J/T
FEM modelling	320*320*12 hexahedral isotropic elements

The thermal inspection dataset includes synthetic simulation data of a plate structure with multiple flat-bottom hole defects. This work aims to investigate the inspection defect profile reconstruction capability. To minimise the mutual interference of multiple defects in one specimen, the defective specimen models are built individually in the same plate model, and the parameters of the exciting source are fixed. The model used is illustrated in Figure 5, where a complex profile flat-bottom hole with a 2 mm depth to the top surface is shown. The thermodynamic material parameters for the thermography of simulation models and the PT simulation specifications are listed in TABLE I. The top graph of Figure 6 shows the thermal images at 0.4s after the pulse which is at the optimal time range of the 1st derivative in the TSR method. The bottom graph illustrates the ground truth of the selected 18 defects, where relatively simple and complex profiles are included.

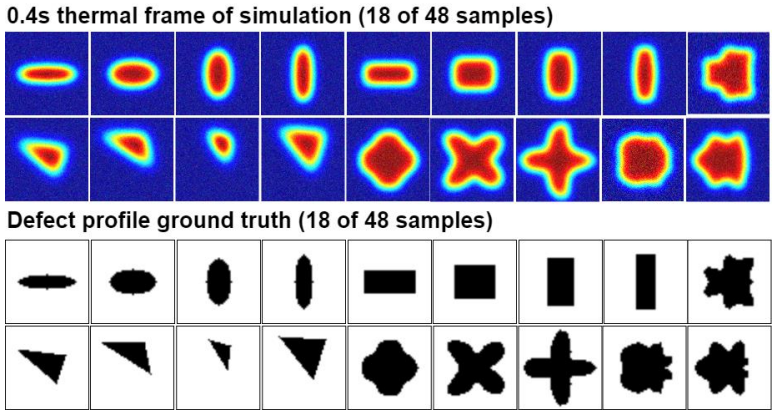
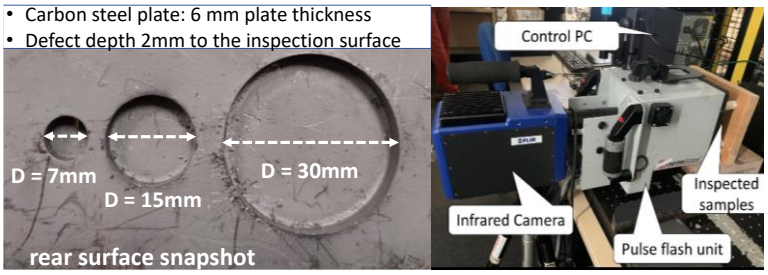


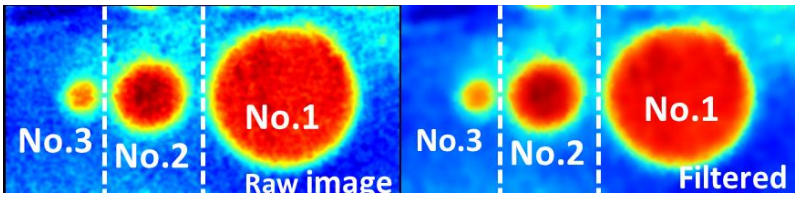
Figure 6. Simulation defects dataset

3.2. Experimental PT inspection

The experiments were conducted using the Thermoscope II pulsed active thermography system which comprises two capacitor banks powered by Xenon flash lamps, a synchronised control PC, and an infrared camera. A FLIR SC7000 series infrared camera system operating between 3-5.1μm and a spatial resolution of 640*512 pixels was used to perform the inspection. The sample was placed with the inspected surface perpendicular to the camera's field of view at a distance of 250mm. A 6mm thick carbon steel specimen with three flat bottom holes was inspected with PT. The sample was inspected from the black-painted surface opposite the drilled surface. The defects were at a depth of 2mm from the inspection surface shown in Figure 7(a). The diameters of the holes are 30mm, 16mm and 8mm, respectively. A sampling rate of 50 Hz was used. Considering the small portion of experimental data in the synthetic dataset, the experiment of the selected sample (containing 3 defects) was repeated 5 times to provide more realistic data variation for network learning. All 15 experimental data (5 inspection repetitions) are added to the dataset with a simulation-to-experiment ratio of 48:15. Figure 7(b).



(a)



(b)

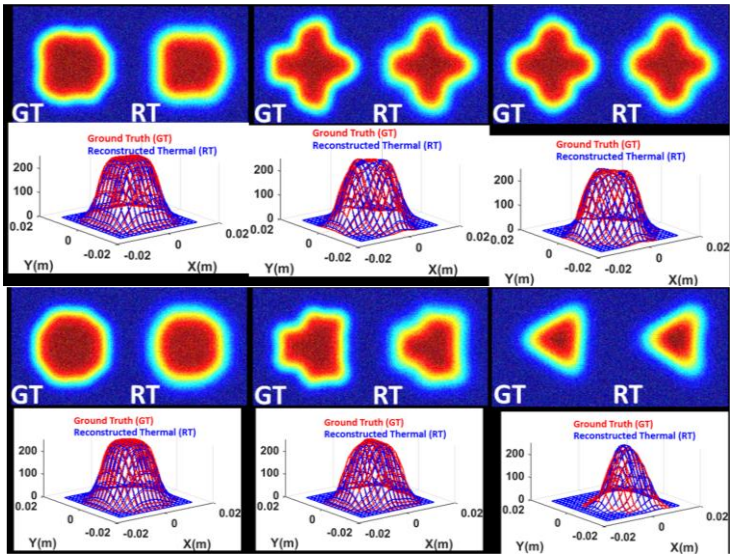
Figure 7. Experimental specimen and PT thermal images. (a) Defect specimen and PT experiment system; (b) Raw and pre-processed experimental data.

4. Results

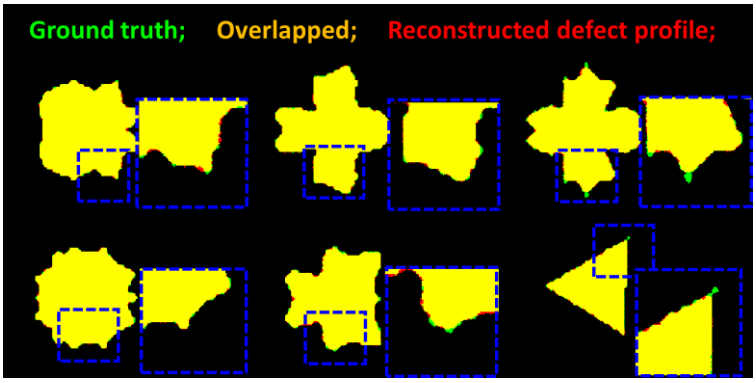
4.1. Self-enhancement capability

To validate the self-enhancement capability, the simulation dataset, described in Section III.A, was used for evaluating the characterisation and generation along 8 iterations of dataset augmentation. The dataset containing 48 defect models is randomly distributed into a ratio of 39:9 for training and testing. The FID score and Recall score were employed to quantify the quality of the thermal image generation and defect profile characterisation. Between two iterations of training, instead of unquestioning data increment, the 9 testing samples are first evaluated with PSNR and SSIM on inspected profile with the current network. Only the generated images (both thermal and profile images) with PSNR over 27dB and SSIM over 0.96 are added to the next round of training dataset. This screening process two thresholds ensure the augmented samples with satisfied standard criteria, which guarantees a high-quality iterative augmentation.

Figure 8 presents the comparison between the generated results (thermal and profile image) and the ground truth (GT). Figure 8(a) demonstrates that the network provides high-quality thermal images with high consistency to the ground truth. Figure 8(b) shows that the reconstructed defect profiles overlapped with the ground truth with high accuracy, where both shape and sharp corners and edges are well aligned.



(a)

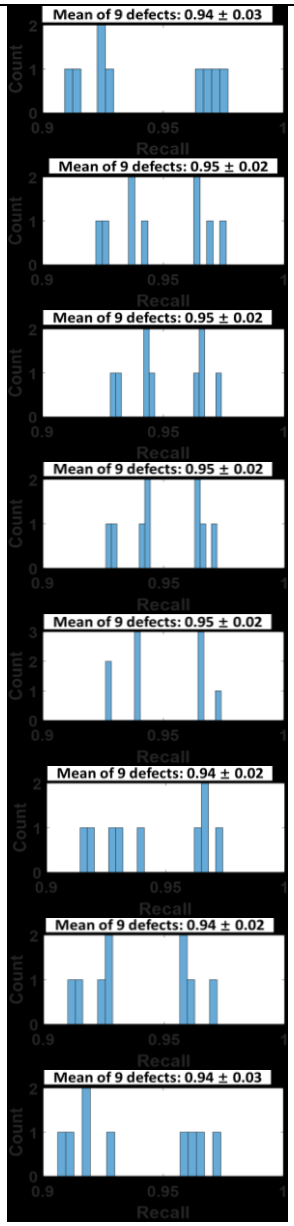


(b)

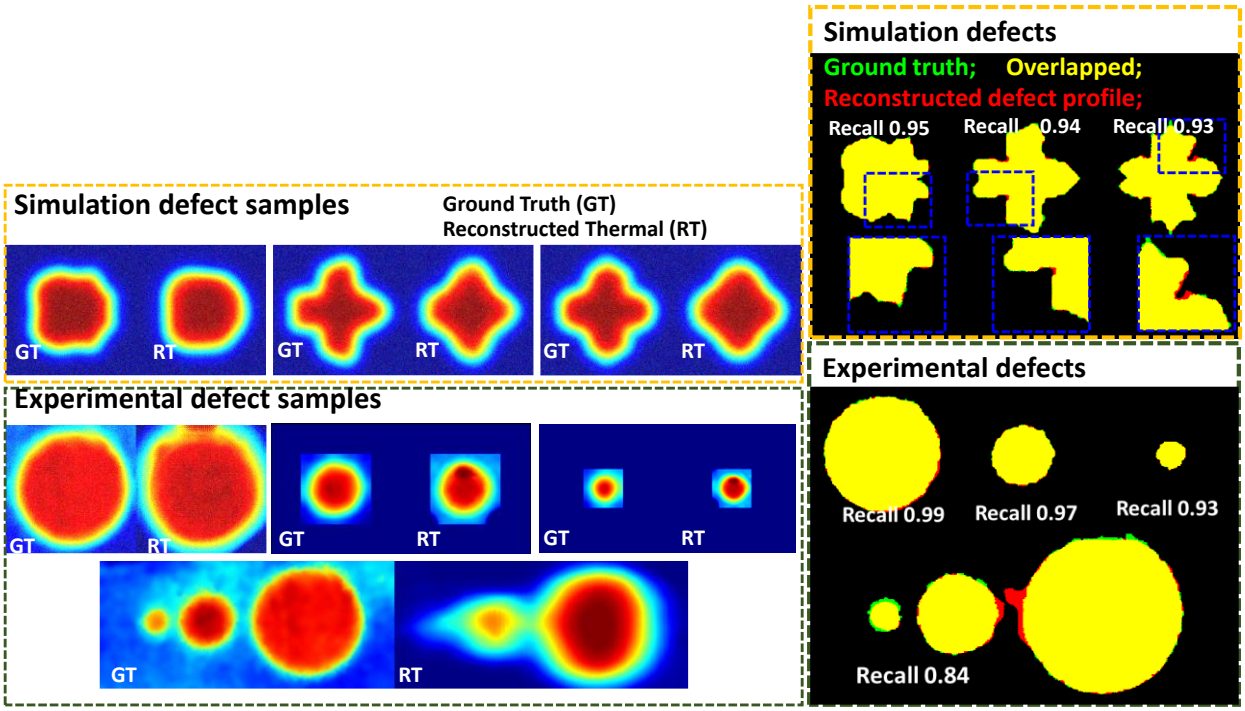
Figure 8. Comparison between ground truth and reconstructed results. (a) Ground Truth (GT) and Reconstructed Thermal (RT) image; (b) Defect profiles overlap plot;

The FID and Recall scores of the iterations of network self-enhancement are presented in TABLE II. For the FID score, the lower score indicates a closer generative thermography image to its corresponding ground truth. The Recall score graphs measure the overlay ratio (e.g. range of [0, 1]) of the defect profile to its ground truth. The trend of the FIDs suggests an improving network performance in the starting iterations, then a degrading performance in the later iterations. Finally, it concludes with an optimal generative performance at the 4th iteration with an FID of 54.2057. The 2nd column in TABLE II shows the Recalls counting graph of 9 testing defects in each iteration. All graphs demonstrate a high inspection capability of defect profile with all defects with a Recall score over 0.92. It also shows an enhanced performance in generating binary profile images, which has more defects inspected with higher Recall values over 0.95 at the 4th iteration. Both scores effectively demonstrate that the proposed network has a significant self-enhancement feature with an optimal iteration step, which presents that the network has a convergent learning capability to balance the benefits of data augmentation and the iteration degradation from network imperfection. It should be noted that the optimal iteration step may vary on the inspection for different materials & depth types.

TABLE II
NETWORK SELF-ENHANCEMENT PERFORMANCE

Iterations	FID of 9 test samples	Recall counts
1st	82.25	
2nd	78.75	
3rd	65.43	
4th **(Best)**	**54.20**	
5th	73.16	
6th	67.16	
7th	73.21	
8th	86.49	
Summary	 Optimal FID means best thermal generation quality	The 4th iteration has most defects characterised with the Recall score of over 0.95.

4.2. Experimental defect profile characterisation



(a) (b)

Figure 9. Comparison result using synthetic dataset. (a) Ground Truth (GT) and Reconstructed Thermal (RT) image; (b) Defect profiles overlap plot;

In this section, the proposed network results using a synthetic dataset (a combination of simulation & experimental data) are discussed. The ratio of training and testing data was set at 4:1, which means 50 and 13 images in the training and testing set, respectively. Additionally, the thermal image containing all 3 defects presented in Figure 5 was also used in the testing set.

Figure 9 shows the comparison between the network results and the ground truth (3 simulation defects and 3 experimental defects). Figure 9(a) shows that the network can also provide high-quality thermal images for defects using a synthetic dataset. It not only manifests promising defect thermal similarity but also generates some thermal experimental noise patterns around defects. In Figure 9(b), the defect profiles are correctly reconstructed, evidenced by the Recall scores of over 0.93. Furtherly, even though the training dataset only contains samples of a single defect, it still shows the powerful defect reconstruction capability in the image containing all 3 defects with a Recall score of 0.84. Only a small misalignment in the interference area between two defects is observed. All these results demonstrate the feasibility of the proposed method. It also demonstrates that the network learned the complexity from the simulation data and the realistic noise level from experimental data.

It should be noted that the proposed method shows a higher accurate capability in defect profile reconstruction than thermal image generation. In the thermal image shown in Figure 9(a), there is a slight thermal pattern distortion in shape and kurtosis. And there is a failed reconstruction for the smallest defect in the 3-defect image. There are two main reasons for the missed thermal generation. First, with three defects in one image, the big defects compress the thermal contrast of the minor defect, making it hard to be learned

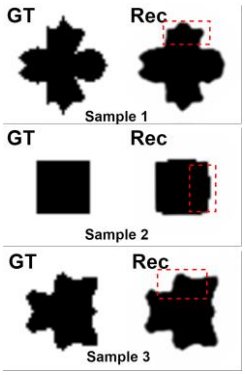
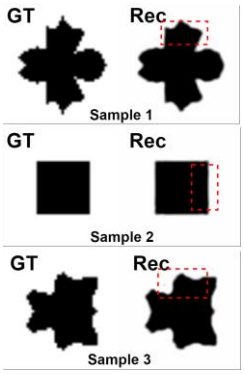
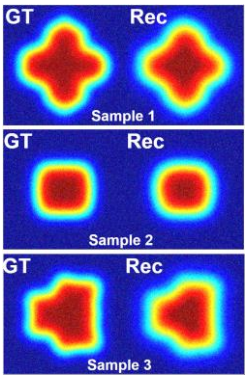
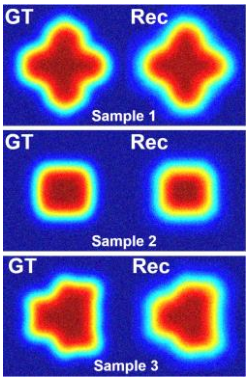
by the model. Second, this is a typical issue of small object learning challenge for neural network when the model deals with multiple targets in an image. The network loses the small target information through a large number of convolution layers, especially for this ascending dimension-like problem (from binary profile image to complex thermal kurtosis image generation). Both reasons joint result in the missed generation issue. In further work, the object is to quantitively investigate the network layers to improve its attention to small target generation and clarify the contrast changing influence, respectively. It is also considered that a super-resolution algorithm will be deployed before the network to enhance the spatial representation of the small defects. In addition, there is only one type of experimental defect (round shape defect) in the current dataset. It is believed that this will be much improved when more variations of experimental defects are included in the initial training dataset.

5. Discussion

5.1. Network loss structure

TABLE III

IDENTITY LOSS NECESSITY COMPARISON

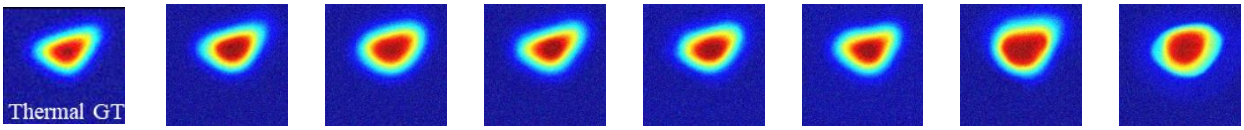
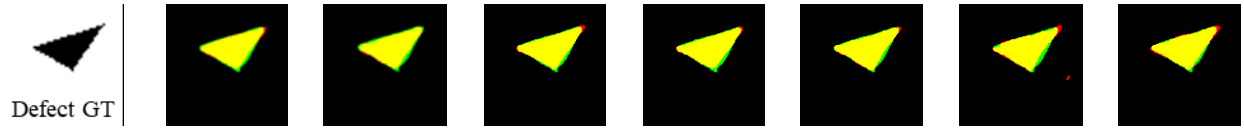
Item	**Loss type (only $\mathcal{L}_{cyc}$, $no\ \mathcal{L}_{ide}$)**	**Loss type ($\mathcal{L}_{cyc}$, $\mathcal{L}_{ide}$)**
Mean FID	146.98	**103.07**
Mean Recall	0.87±0.05	**0.94±0.03**
Defect profile inspection		
Thermal image generation		

In this section, the network performance comparison between different loss functions is discussed. Shown

in Eq. (2), the loss function includes 4 parts where the consistency loss $\mathcal{L}_{cyc}$ and identification loss $\mathcal{L}_{ide}$ are the two keys controlling overall generative performance. Compared to original cycle-GAN, it is evident that the existence of the identity loss $\mathcal{L}_{ide}$ is necessary to guarantee thermographic data generation (see the comparison in TABLE III). The case without identity loss $\mathcal{L}_{ide}$ has a higher FID and lower recall, indicating a low quality of dual-type image generation. It results in rough defect profile reconstruction (all 3 defect samples reconstructed in row 3, and degraded contour and shape distortion in the generated thermal image (all 3 thermal images reconstructed in row 4. In comparison, the case with both consistency loss $\mathcal{L}_{cyc}$ and identification loss $\mathcal{L}_{ide}$ exhibits better performance on both complex profile reconstruction (especially in the sharp edges and corners) and fidelity of the generated thermal image.

Furtherly, we compared the performance of different loss types between the options of $norm\ L_1$ as absolute difference, $norm\ L_2$ as squared difference and $norm\ smooth\ L_1$ (SL_1) as a combination of L_1 and L_2. The FID and Recall scores (average of all testing samples) and the imagery comparison (thermal and profile of one test sample) from the 1st round training network (not self-enhanced yet) are shown in TABEL IV. For a self-enhancement network, it is critical to design a better loss function for the 1st round of training, otherwise, the self- enhancement process will degrade with a dramatically rough learning network while iterating.

TABLE IV
COMPARISON OF DIFFERENT LOSS STRUCTURE

Loss type ($\mathcal{L}_{cyc}, \mathcal{L}_{ide}$)	L_1, L_1	$\boldsymbol{L_1, SL_1}$	$\boldsymbol{SL_1, L_1}$	SL_1, SL_1	L_2, L_1	L_2, L_2	L_2, SL_1
Mean FID	108.56	**103.78**	**104.94**	108.13	144.09	147.77	153.36
Mean Recall	0.95±0.02	**0.94±0.02**	**0.93±0.03**	0.88±0.09	0.86±0.09	0.89±0.10	0.91±0.08
Thermal GT							
Defect GT							

The results show a clear performance variance with different loss calculation methods. It is recommended that the combination of $norm\ L_1$ and SL_1 (bold marked in TABEL IV) energise the network with a higher learning ability to generate high-quality thermal and defect profile images. The lowest average FIDs and relatively high Recall suggest the optimal loss selection for this dual generative network. In contrast, the loss function using $norm\ L_2$ brings an obvious performance drop to the network (shown in the last 3 selections). The generated thermal image has a severe distortion and the inspected defect profile images show a bigger deviation in the Recall scores. This analysis demonstrates that the norm types of loss (L_1 or SL_1) can not only alleviate the learnt signal noise but also constrain the generative variations of the network, especially for thermal images in this thermography NDT application. The constraint of the GAN variation or diversifying capability is important to this method for NDT because this method values generation accuracy more than

diversity. It has a high impact on defect characterisation and its self-enhancement performance. In addition, the imagery (thermal or defect profile) in this application is single-style imagery compared to other imagery applications, so it has lower requirements of generative diversity.

5.2. Scalability of the proposed method

In this section, a scalability investigation of the proposed method for different materials and defect types is discussed. Several composite laminate samples with impact damage are inspected using the proposed method. In composite samples, the impact damage consists of delamination, cracks, fibre fracture, etc. At this stage, this attempt aims to verify its reconstruction of the delamination defect. The CFRP samples are 150 x 100 x 3 mm plate specimens, which are damaged by a hemispherical indenter weighing 2.281 kg, performing free drop under acceleration of gravity **Error! Reference source not found.**. The impact damage of CFRP specimens was induced with 10J drop energy. The synthetic dataset was established by both pulsed thermography experiments and FEM simulation [36] on impacted CFRP samples. The quantity of simulation and experimental samples are 30 (training) and 10 (7 training + 3 testing). It should be noted that the simulation samples are modelled with only 1 delamination in each specimen, but experimental impact damage would have not only one delamination. It is a good case to challenge the recognition and reconstruction capability of the network when defect quantity and distribution are unknown.

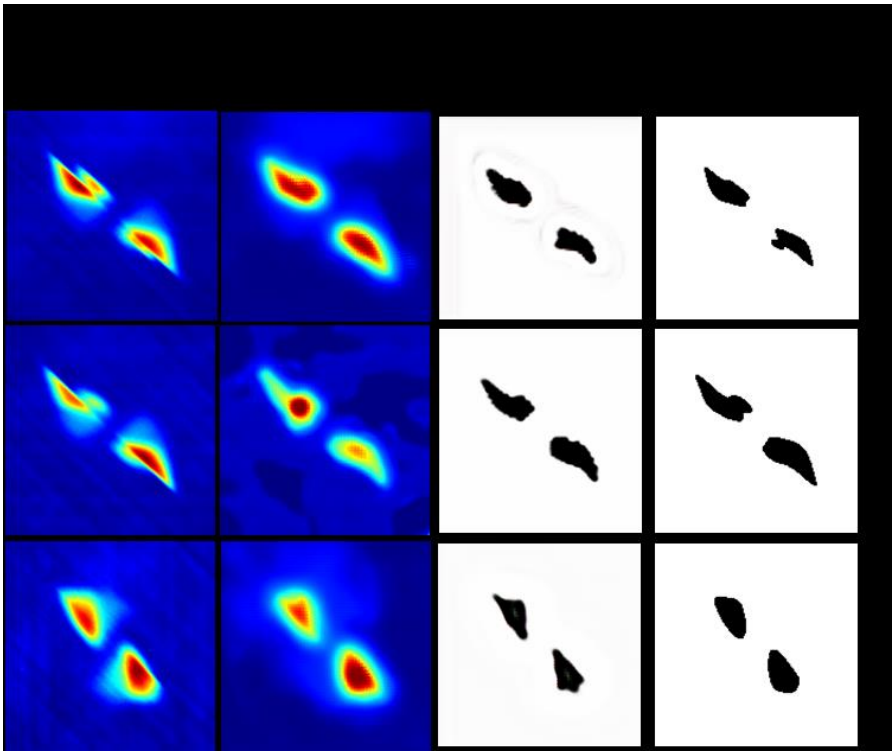


Figure 10. Network results of impact delamination in CFRP.

The results of 3 CFRP impact damage samples are shown in Figure 10. The thermal image generation effectively reflects the thermal pattern at the fully delaminated location shown in the 2nd column in Figure 10. The generated thermal images well present the quantity, distribution and shape of the delamination compared

to the ground truth. Even though there is still imperfection of thermal reconstruction in corners and edges, this still demonstrates its capability of thermal data augmentation that can support NDT demands.

The inspected results also reconstructed a fully delaminated defect profile, shown in the 3rd column in Figure 10. This is compared with delamination results processed by a method of Stationary Wavelet Transform Dissection (SWTD) and confidence map [36]. The results present that the proposed technique can achieve close performance with state-of-the-art post-processing methods. It demonstrates that the proposed method has extensibility to other materials with similar defect shapes, even though the current network is not trained by an ideal synthetic dataset, which contains simulated flat-bottom holes and experiments CFRP inspection results. It is confident that this method can be used for CFRP impact damage characterisation when the simulation and experimental are used. It presents a promising vision to extend the proposed method to other scenarios both for inspection and data augmentation with adequate realistic data.

Defect reconstruction based on the inversion strategy is another effective approach contributing the quantitative characterisation of internal structures and interfacial delamination defects [37-38]. The distinct advantages of inversion methods are: first, a big amount of inspection data and defect sample diversity are not needed. In an alternative view, the relationship between defect and inspection signal is already built in the forward problem modelling. Second, within the pre-set constraints of inspection modelling, the inversion methods seek the "true" defect in a process under clearer searching and divergence directions. Differently, by learning the uncertainties from sufficient inspection signals, the neural learning networks have stronger stability and resistance to noise, specimen and environmental deviation. Once trained, the network models will quickly output the reconstructed decision, which are more applicable in the on-site inspection.

As the two sides of a coin, the disadvantages of two methods are obvious. For instance, in an inversion strategy, an accurate forward modelling method for complex structures and defects is difficult to build and hard to be flexible to similar inspection structures. In neural network methods, a large amount of data with enough defect diversity is costly in time and budget to collect. There is no there is no superiority or inferiority between the two types of methods, and it should be wisely selected in different applications.

6. Conclusion

In this paper, a new DL-based technique for thermography NDT defect characterisation with self-enhancement capability is proposed to address two challenges simultaneously: the shortage of representative experimental samples in real applications, and the inaccurate and distortion of complex defect profile characterisation in thermographic NDT. Aiming at these two challenges, it is designed based on Cycle-GAN that can achieve dual-direction generations for both thermal inspection image augmentation and defect profile reconstruction. By using a synthetic dataset of both simulation and experimental samples, the network not only can train its learning of complex defect profile diversity from numerical simulation thermography data, but

also can learn the realism uncertainties of noise level and real thermal pattern from the experimental data. More importantly, it can continuously self-enhance by iteratively using its generated thermal image to augment the dataset until it reaches an optimal dual generative capability for defect characterisation.

The FID and Recall scores demonstrate that the network performance has optimal self-enhancement convergency that balances the benefits from data augmentation and iteration degradation from network imperfection. In addition, the results of the synthetic dataset demonstrate the feasibility of the proposed network with high-quality thermography augmentation and accurate defect profile characterisation with a Recall score of over 0.92. In the network structure, comparison results prove that the cycle consistency loss and identity loss are both important to constrain the GAN transfer capability, ensuring the characterisation accuracy of defect profile and thermal image fidelity. A performance comparison between different loss function selections was implemented. It indicates the SL_1 and L_1 norm loss set is optimal for this complex profile characterisation scenario. Finally, the scalability discussion of the proposed technique on the delamination defect in impacted CFRP specimens manifests a promising vision for the inspection of different materials and defect types. The proposed method will facilitate the thermal imagery-based NDT inspection for different materials and damages. It forwards a solid step on the applicability of AI for NDT and reveals a bright perspective of FEM+GANs in NDT. For further work, considering that the deviation of composite layers and fibre directions affect the reconstruction accuracy. Different types of composites will be counted in training and quantitatively consider the structure deviation as a weight parameter in the network model. In the future, the influence of thermal images at different time windows on network performance will be studied to optimise its versatility.

The proposed technique will facilitate the thermal imagery-based NDT defect reconstruction for different materials and damages. It forwards a solid step on the applicability of ML for NDT and reveals a bright perspective of FEM+GANs in NDT. It pushes the Cycle-GAN to the accurate application of NDT image quantification rather than the image-style translation. In future, the influence of thermal images at different time windows on network performance will be studied. And the extension work on the composite material with compound damages will be investigated using the proposed method. s

Acknowledgement

This work was supported by the UK EPSRC Platform Grant: Through-life performance: From science to instrumentation (Grant number EP/P027121/1). Data supporting this study are available from Cranfield Online Research Data at https://doi.org/10.17862/cranfield.rd.22559989 .

Reference

[1] Y. Fu, Y. Shang, W. Hu, B. Li, and Q. Yu, “Non-contact optical dynamic measurements at different ranges: a review,” *Acta Mechanica Sinica*, vol. 37, no. 4, pp. 537–553, 2021.

[2] S. He, J. Chen, Z. Chen, and G. Song, “An exploratory study of underwater bolted connection looseness detection using percussion and a shallow machine learning algorithm,” *Acta Mechanica Sinica*, vol. 39, no. 4, Apr. 2023.

[3] S. Xiang, A. M. Omer, M. Li, D. Yang, A. Osman, B. Han, Z. Gao, H. Hu, C. Ibarra-Castanedo, X. Maldague, Q. Fang, S. Sfarra, H. Zhang, and Y. Duan, “A reliability study on automated defect assessment in optical pulsed thermography,” *Infrared Phys. Technol.*, vol. 134, p. 104878, Nov. 2023.

[4] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa, “Infrared machine vision and infrared thermography with deep learning: A review,” *Infrared Physics & Technology*, vol. 116, p. 103754, 2021.

[5] L. Shen, Y. Liu, and M. Hui, “Dynamic thermo-mechanical behaviors of SME TiNi alloys subjected to shock loading,” *Acta Mechanica Sinica*, vol. 36, no. 6, pp. 1336–1349, 2020.

[6] X. Meng, F. Wang, Z. Liang, X. Zhang, J. Weng, Z. Li, J. Liu, M. Chen, Y. Wang, H. Yue, "Research on the detection of silica/phenolic composite surface cracks using instantaneous high-power xenon lamp-induced chirp-pulsed radar thermography", *Measurement*, vol. 209, pp. 112494, 2023.

[7] Q. Luo, B. Gao, W. Woo and Y. Yang, "Temporal and spatial deep learning network for infrared thermal defect detection", *NDT & E International*, vol. 108, p. 102164, 2019.

[8] C. Hu, Y. Duan, S. Liu, Y. Yan, N. Tao, A. Osman, C. Ibarra-Castanedo, S. Sfarra, D. Chen, and C. Zhang, “LSTM-RNN-based defect classification in honeycomb structures using infrared thermography,” *Infrared Physics & Technology*, vol. 102, p. 103032, Nov. 2019.

[9] Y. Duan, S. Liu, C. Hu, J. Hu, H. Zhang, Y. Yan, N. Tao, C. Zhang, X. Maldague, Q. Fang, C. Ibarra-Castanedo, D. Chen, X. Li, and J. Meng, “Automated defect classification in infrared thermography based on a neural network,” *NDT & E International*, vol. 107, p. 102147, 2019.

[10] Z. Wei, H. Fernandes, H. G. Herrmann, J. R. Tarpani, and A. Osman, “A deep learning method for the impact damage segmentation of curve-shaped CFRP specimens inspected by infrared thermography,” *Sensors*, vol. 21, no. 2, pp. 1–18, 2021.

[11] J. Xie, C. Xu, G. Chen, and W. Huang, “Improving visibility of rear surface cracks during inductive thermography of metal plates using Autoencoder,” *Infrared Physics & Technology,* vol. 91, pp. 233–242, 2018.

[12] R. Marani, D. Palumbo, U. Galietti, and T. D’Orazio, “Deep learning for defect characterisation in composite laminates inspected by step-heating thermography,” *Optics and Lasers in Engineering*, vol. 145, p. 106679, 2021.

[13] C. Schmidt, T. Hocke, and B. Denkena, “Artificial intelligence for non-destructive testing of CFRP prepreg materials,” *Production Engineering,* vol. 13, no. 5, pp. 617–626, 2019.

[14] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, and A. Ahmad, “Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning,” *Solar Energy*, vol. 198, no. November 2019, pp. 175–186, 2020.

[15] A. Nasiri, A. Taheri-Garavand, M. Omid, and G. M. Carlomagno, “Intelligent fault diagnosis of cooling radiator based on deep learning analysis of infrared thermal images,” *Applied Thermal Engineering*, vol. 163, no. January, p. 114410, 2019.

[16] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, “A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines,” *Neurocomputing*, vol. 272, pp. 619–628, 2018.

[17] T. Fang, J. An, Q. Chen, Y. He, H. Wang, and X. Zhang, “Progress and comparison in nondestructive detection, imaging and recognition technology for defects of wafers, chips and solder joints,” *Nondestructive Testing and Evaluation*, pp. 1–56, Dec. 2023.

[18] X. Chen, G. Tian, J. Wu, C. Tang, and K. Li, “Feature-Based Registration for 3D Eddy Current Pulsed Thermography,” *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6998–7004, Aug. 2019.

[19] C. Jian, K. Yang, and Y. Ao, “Industrial fault diagnosis based on active learning and semi-supervised learning using small training set,” *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104365, 2021.

[20] Q. Fang, C. Ibarra-castanedo, and X. Maldague, “Automatic defects segmentation and identification by deep learning algorithm with pulsed thermography: Synthetic and experimental data,” *Big Data and Cognitive Computing.*, vol. 5, no. 1, pp. 1–21, 2021.

[21] Y. Gao, L. Gao and X. Li, "A Generative Adversarial Network Based Deep Learning Method for Low-Quality Defect Image Reconstruction and Recognition”, *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3231-3240, May 2021.

[22] G. Zhang, K. Cui, T. Hung, S. Lu, “Defect-GAN: High-Fidelity Defect Synthesis for Automated Defect Inspection”, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2524-2534, 2021.

[23] J. Zhao, J. Zhang, Z. Li, J.-N. Hwang, Y. Gao, Z. Fang, X. Jiang, and B. Huang, “DD-cyclegan: Unpaired image dehazing via double-discriminator cycle-consistent generative Adversarial Network,” *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 263–271, 2019.

[24] Y. Gao, X. Liu and J. Xiang, "FEM Simulation-Based Generative Adversarial Networks to Detect Bearing Faults”, *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4961-4971, July 2020.

[25] S. Wu, C. Dong, and Y. Qiao, “Blind Image Restoration Based on Cycle-Consistent Network,” *IEEE Transactions on Multimedia*, vol. 25, pp. 1111–1124, 2023.

[26] L. Tian, Z. Wang, W. Liu, Y. Cheng, F. Alsaadi and X. Liu, "A New GAN-Based Approach to Data Augmentation and Image Segmentation for Crack Detection in Thermal Imaging Tests", *Cognitive Computation*, vol. 13, no. 5, pp. 1263-1273, 2021.

[27] L. Cheng, Z. Tong, S. Xie and M. Kersemans, "IRT-GAN: A generative adversarial network with a multi-headed fusion strategy for automated defect detection in composites using infrared thermography", *Composite Structures*, p. 115543, 2022.

[28] K. Liu, Y. Tang, W. Lou, Y. Liu, J. Yang, and Y. Yao, “A thermographic data augmentation and signal separation method for defect detection,” *Measurement Science and Technology*, vol. 32, no. 4, p. 045401, 2021.

[29] R. Ali and Y.-J. Cha, “Attention-based generative adversarial network with internal damage segmentation using thermography,” *Automation in Construction*, vol. 141, p. 104412, 2022.

[30] L. Ruan, B. Gao, S. Wu, and W. L. Woo, “DeftectNet: Joint loss structured deep adversarial network for thermography defect detecting system,” *Neurocomputing*, vol. 417, pp. 441–457, 2020.

[31] Y. Pang, J. Lin, T. Qin, and Z. Chen, “Image-to-Image Translation: Methods and Applications,” *IEEE Transactions on Multimedia*, vol. 24, pp. 3859–3881, 2022.

[32] H. Tang, H. Liu, D. Xu, P. H. S. Torr, and N. Sebe, “Attention-GAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1972–1987, 2023.

[33] H. Han, L. Hao, D. Cheng and H. Xu, "GAN-SAE based fault diagnosis method for electrically driven feed pumps", *PLOS ONE*, vol. 15, no. 10, p. e0239070, 2020.

[34] K. Zhang, Y. Zhang and H. Cheng, "Self-Supervised Structure Learning for Crack Detection Based on Cycle-Consistent Generative Adversarial Networks", *Journal of Computing in Civil Engineering*, vol. 34, no. 3, p. 04020004, 2020.

[35] H. Liu, C. Pei, S. Xie, Y. Li, Y. Zhao and Z. Chen, "Inversion Technique for Quantitative Infrared Thermography Evaluation of Delamination Defects in Multilayered Structures," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4592-4602, July 2020.

[36] H. Liu, W. Du, H. Yazdani Nezhad, A. Starr and Y. Zhao, "A dissection and enhancement technique for combined damage characterisation in composite laminates using laser-line scanning thermography", *Composite Structures*, vol. 271, p. 114168, 2021.

[37] L. Zhuo, Y. Xu, J. Zhu, C. Li, C. Liu, and F. Yi, “Quantitative Evaluation of Interfacial Defect Size and Pattern by Solving a 3D Inverse Problem Based on Step Heating Thermography,” *Experimental Mechanics*, vol. 64, no. 2, pp. 245–259, Dec. 2023.

[38] L. Zhuo, X. Yang, J. Zhu, Z. Huang, J. Chao, and W. Xie, “Size determination of interior defects by reconstruction of subsurface virtual heat flux for step heating thermography,” *NDT & E International*, vol. 133, p. 102734, Jan. 2023.



https://dspace.org/

A cyclic self-enhancement technique for complex defect profile reconstruction based on thermographic evaluation

Liu, Haochen; Wang, Shuozhi; Zhao, Yifan; Deng, Kailun; Chen, Zhenmao

2025-05
Springer
https://doi.org/10.1007/s10409-024-24076-x
Liu H, Wang S, Zhao Y, et al., (2025) A cyclic self-enhancement technique for complex defect profile reconstruction based on thermographic evaluation. Acta Mechanica Sinica, Volume 41, Issue 5, May 2025, Article number 424076

Downloaded from a DSpace repository. https://dspace.org/