QIRT'2022
10.21611/qirt.2022.2022
16th Quantitative InfraRed Thermography Conference

Undercomplete Autoencoder for dimensional reduction applied to Pulsed thermography

by S.Ebrahimi*, J.R.Fleuret*, C.Ibarra-Castanedo*,**, M.Klein**, X.P.V.Maldague*

* *Laval University, Computer Vision and Systems Laboratory (CVSL), Multipolar Infrared Vision team (MIVIM), Quebec City, Canada*
** *Visiooimage Inc. Infrared Thermography Testing Systems, Quebec City, Canada*

Abstract

Non-destructive testing and evaluation techniques are essential in structural health monitoring and safety control in industry and aerospace. Among the different NDT techniques, pulsed thermography has demonstrated to be an effective approach for the inspection of carbon fiber-reinforced polymer (CFRP). A signal processing technique called undercomplete autoencoder (UAE) is explored in this work as a means to extract meaningful information for anomaly detection. The high dimensionality typical of thermal data sequences have pushed researchers to study innovative approaches to reduce thermal sequences to lower-dimensional data sequences that highlight the hidden anomalies in materials. The proposed approach is a dimensionality reduction method that reveals anomalies and provides better visibility. In a comparative study with other dimensionality reduction approaches such as PCT, UAE presents promising results on pulse thermography data.

1. Introduction

Carbon fiber-reinforced polymers (CFRPs), due to their excellent mechanical properties, are extensively applied in structural engineering and aerospace. Features such as high strength and rigidity, low density, excellent damping properties, and high resistance to impact make CFRP interesting in reducing the structural weight while durability and strength are increased; and therefore, improving the structural efficiency. Since health monitoring of CFRP components and damage identification plays a crucial role in service function and safety of the systems, an accurate and non-invasive prediction of the remaining life service is always interesting. In-situ on-line health monitoring can provide safety and reliability of the structural components and prevent catastrophic failure. Among non-destructive testing techniques, infrared thermography, which involves mapping the surface temperatures, could be used to assess the damage state.

Pulsed thermography (PT) is a non-contact and full-field Infrared Non-Destructive Testing (IRNDT) approach based on thermal heat transfer analysis during the cooling period. After external excitation, which consists of a very short thermal pulse (down to $\sim 10^{-3}s$), the incident radiation delivered to the surface becomes a thermal wave that propagates by conduction through the material. The infrared camera records the temperature decay during the cooling period. Subject to the presence of discontinuity, depending on its material and thermal properties and depth, defects will be revealed at different times. The deeper defects appear later with lower thermal contrast. In order to obtain quantitative information from thermal data, several approaches have been proposed. Manipulating thermal data makes active thermography an attractive and powerful approach for industrial control and maintenance purposes.

Between different IRNDT approaches, thermographic signal reconstruction (TSR) [1], pulsed phase thermography (PPT) [2], and principal component thermography (PCT) [3, 4] have gained wide popularity in many fields. Thermal data often contain high levels of noise, outliers, and distortions which can reduce the quality of the enhanced images. Having an approach that is less sensitive to noise and applicable to other IRNDT approaches in order to improve defect detection is always interesting. Saeed et al. [5] developed an automatic flaw detection and depth estimation via a convolutional neural network (CNN) in tandem with a Deep Feed Forward Neural Network (DFF-NN) algorithm. The method is applied to thermal images obtained from a pulsed thermography setup from a CFRP sample and can detect the defects accurately, but the algorithm requires a complex and computationally intensive training process. Luo et al. [6] introduced a combined transient and spatial deep learning architecture for defect detection. Deep cross-learning strategy-based segmentation models to improve the contrast between defective and non-defective regions. Jie et al. [7] proposed a sparse moving window principal component thermography (SMWPCT) method to enhance the distinction between regions with defect and defect-free regions and eliminate the influence of noise and non-uniform heating generated by PT using CFRP sample. The authors claimed that the proposed method is superior to several alternatives. Zhang et al. [8] introduced a neural learning-based approach to blind source separation for defect detection on additively manufactured (AM) materials in pulsed thermography. They reported that the internal calibrated defects of different sizes and depths were detected by merging artificial intelligence with photothermic. Azizinasab et al. [9] proposed a method called the phase of transient response using the local reference pixel vector (PTR-LRPV) to process pulsed thermography data for defect detection and depth estimation in CFRP. The authors stated that the method works effectively in detection and depth estimation tasks and is comparable with state-of-the-art algorithms. Prosvirin et al. [10] used the deep undercomplete denoising autoencoder for estimating the nonlinear function of the system under normal operating conditions in their method for a blade rub-impact fault. Kraljevski et al. [11] employed unsupervised machine learning with autoencoders for ultrasonic testing of adhesive bonding. Wei et al. [12] present an approach based on

stacked denoising autoencoder (DAE) to enhance the defective regions using pulsed thermography data. Application of deep autoencoder for data enhancement and defect detection are presented in several research studies [13, 14]. Also, Kaur et al. [15] developed a novel autoencoder-based thermography approach for subsurface defect detection and stated that their approach provides better defect detection in terms of signal-to-noise ratio than PCT. This article proposes to utilize autoencoders for the compression of thermal data. Autoencoders are a type of neural network in which the trained network replicates the input. The trained autoencoder compresses and reconstructs input data based on a given training dataset. The rest of this article is organized to progressively describe the application of an undercomplete autoencoder in pulsed thermography data. Section 2 briefly explains the mathematical basis of autoencoders and the parameters which define the performance.

2. Method and Materials

2.1 Undercomplete-autoencoder

Autoencoders [16] are an unsupervised neural network which are part of a larger family of representation learning methods that can automatically leverage the neural networks to learn features from unlabeled data. These methods are evolved to map the input data to an internal latent representation, then employed to produce output data similar to the input data. The autoencoders were first developed for nonlinear dimensionality reduction. An autoencoder consists of two primary elements, an encoder function $h = f(x)$ and a decoder function which reproduces a reconstruction $x\prime = g(h)$ of the input. Figure 1 shows the general schematic structure of an autoencoder. The encoder takes input data and maps it into a latent encoding space, and then the decoder tries from this latent code vector to reconstruct the input data as accurately as possible. Compared with the input data's dimension, the latent space's dimensionality can be either the same, larger or smaller. In the latter case, the encoder carries out a dimensionality reduction during the encoding step. Such an autoencoder is named an undercomplete-autoencoder. Dependency between input features is important for dimensionality reduction and reconstruction; the correlation between input features can be learned, and the data from the encoder is projected into lower-dimensional space. During the training of the autoencoders, the weights of the neurons are adjusted so that the loss function will yield a minimum, and consequently, the output will represent the input in the best feasible way. The loss function L, such as the mean squared error (MSE), penalizes $g(f(x))$ for being dissimilar from $\mathbf{x}$, and also the model learns the most important and relevant attribute of the input data. Mathematically speaking, h and r define as Equation 1 and the goal is shown in Equation 2 [17]:

$$\mathbf{h} = f(\mathbf{W}\mathbf{x}); \mathbf{r} = g(\mathbf{V}\mathbf{h}) \tag{1}$$

$$\min_{\mathbf{W},\mathbf{V}} \frac{1}{2N} \sum_{n=1}^{N} \|\mathbf{x}^{(n)} - \mathbf{r}^{(n)}\|^2 \tag{2}$$

If f and g are linear :

$$\min_{\mathbf{W},\mathbf{V}} \frac{1}{2N} \sum_{n=1}^{N} \|\mathbf{x}^{(n)} - \mathbf{V}\mathbf{W}\mathbf{x}^{(n)}\|^2 \tag{3}$$

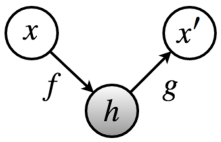


Fig. 1. *The general schematic structure of an autoencoder mapping an input x to reconstruction $x\prime$ via latent code h. The encoder f maps the input x to h and the decoder g maps h to $x\prime$.*

The undercomplete autoencoder uses a linear activation function in combination with the mean squared error (MSE) loss function and learns to span the same subspace as PCA. Depending on the application, all output of the decoder may not be interesting and salient properties from the data are preferred. One of the possible ways is to employ undercomplete autoencoders.

2.2 Principal Component Thermography

Rajic et al. [3] based on PCA, which tends to maximize the variance and projected onto a lower-dimensional linear space, is called a principal subspace. The principal scores or empirical orthogonal functions (EOF) are representations of complex input signals and present the most critical data variability.

Mathematically, consider an observation matrix $\{\mathbf{x}_n\}$ where $n = 1, ..., N$, and $\mathbf{x}_n$ is a Euclidean variable with dimensionality D [18]. The goal is to project the data onto a M dimension space where $M < D$. For the projection onto a one-dimensional space ($M = 1$), the D-dimensional vector $\mathbf{u}_1$ can be considered. Each pixel $\mathbf{x}_n$ is projected onto a scalar value $\mathbf{u}_1^T$. The variance of the projected data is given by:

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\{\mathbf{u}_1^T\mathbf{x}_n - \mathbf{u}_1^T\overline{\mathbf{x}}\}^2 = \mathbf{u}_1^T\mathbf{S}\mathbf{u}_1 \tag{4}$$

where $\mathbf{S}$ is the covariance matrix:

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T \tag{5}$$

The appropriate constraint comes in order to prevent $\|\mathbf{u}_1\| \to \infty$ for $\mathbf{u}_1^T\mathbf{S}\mathbf{u}_1$ maximization. To solve this problem, a Lagrange multiplier is employed, which is denoted by λ_1 and makes an unconstrained maximization of :

$$\mathbf{u}_1^T\mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^T\mathbf{u}_1) \tag{6}$$

$\mathbf{u}_1$ must be an eigenvector of $\mathbf{S}$, and to maximize the variance, $\mathbf{u}_1$ is set to the eigenvector with the largest eigenvalue λ_1. Briefly, PCA consists in evaluating the mean ($\overline{\mathbf{x}}$) and the covariance matrix ($\mathbf{S}$) of input data and finding the M eigenvectors of the covariance matrix corresponding to the M largest eigenvalues.

In the application of PCA over thermograms, defective and non-defective pixels provide different contributions to the same EOF, and hence, it can be used to highlight the anomaly information.

2.3 Thermographic Signal Reconstruction (TSR)

Thermographic Signal Reconstruction is a popular method for the enhancement and analysis of thermographic sequences. This method involves fitting the thermogram by a logarithmic polynomial of degree n (Equation 7) and the computation of the first and second logarithmic time derivatives (without generating additional noise). Shepard, in 2001 [19] proposed to transform the problem to the logarithmic domain in order to compensate for some troublesome parameters such as the presence of temporal noise due to the IR camera and data saturation immediately after the flash. The one-dimensional surface temperature solution of the heat equation for a semi-infinite slab stimulated with a Dirac pulse can be expressed as [20]:

$$T = \frac{Q}{e\sqrt{\pi t}} \tag{7}$$

where t, Q and e are the time, the energy density at the surface, and the material effusivity, respectively. After applying the natural logarithm from Equation 7, the logarithmic behavior is described by two terms; the first term that depends on e and Q, and second one that depends on t, and acts as a straight line with slope -0.5 in the logarithmic space. In practice, logarithmic data may deviate from ideal behavior due to different reasons e.g. nonlinear camera response and background radiation and convection. Still, sound areas decrease nearly linearly, whilst defective regions depart from this behaviour at a particular time (related to their depth). Equation 7 can be reformulated as [19]:

$$\ln T(t) = a_0 + a_1\ln(t) + a_2\ln(t)^2 + ... + a_n\ln(t)^n \tag{8}$$

While TSR reduces the high-frequency noise in data like a low pass filter, the time derivatives of logarithmic time history provide discrimination between defect and sound areas. As mentioned earlier, time derivatives can considerably reduce not only the background noise but also the blurring effect and preserve the shape of the defects. Moreover, if either the thermal diffusivity of the sample or the flaw depth is defined, the time at which the second derivative peak occurs can be used to estimate the non-defined e value [21].

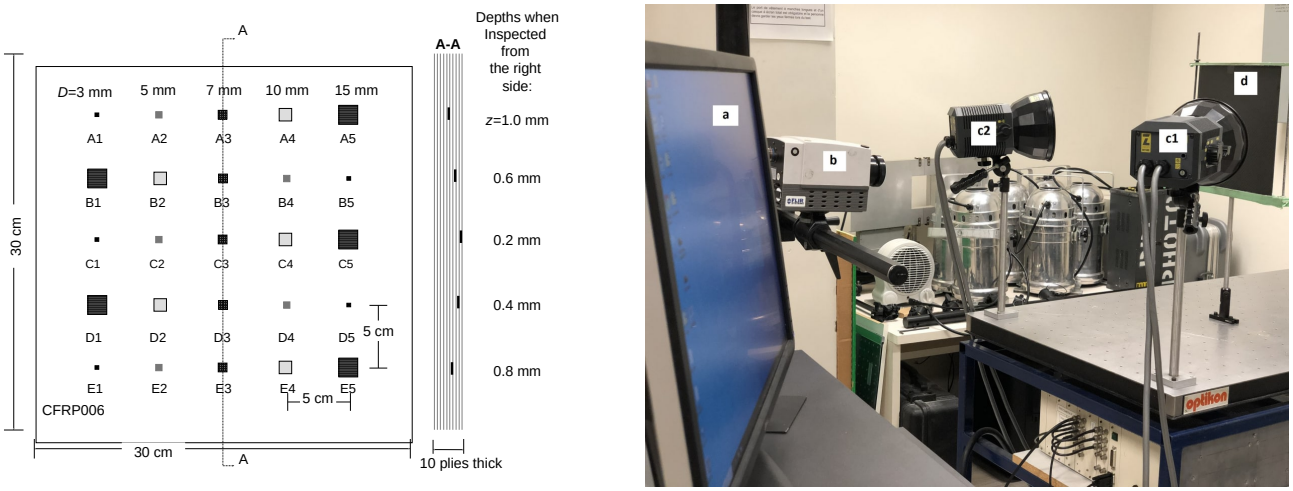
2.4 Data Acquisition

The experiments were carried out on an academic (CFRP) plate (30 cm × 30 cm × 2 mm) with 25 defects of quadrangular Teflon inserts. In this academic sample the Teflon sheets inserted between plies varies from 0.2 to 1 mm, while the lateral size varies from 3 to 15 mm. The defects specification are presented in Table 1, and depicted in the sample schematic in Figure 2a.

In PT [22] inspection, the test specimen is exposed to a thermal pulse for $5ms$ using two flash lamps ($6.4KJ$/flash (Balcar, France)); the thermal flux emanating from the specimen surface is captured by a cooled infrared camera (FLIR Phoenix (FLIR Systems, Inc., Oregon, USA), InSb, midwave, 3–5 μm, Stirling Cooling) and spatial resolution of 640 × 512 pixels. A frame-rate of 145 Hz was used to record the temperature profile in reflection mode (Figure 2b). Infrared images were taken from a distance of 70 cm by the IR camera, which was perpendicular to the plate surface.

Table 1. *Defect specifications for the CFRP Plate.*

Defect	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
Depth (mm)	1	1	1	1	1	0.6	0.6	0.6	0.6	0.6	0.2	0.2	0.2	0.2	0.2	0.4	0.4	0.4	0.4	0.4	0.8	0.8	0.8	0.8	0.8
Lateral size (mm)	3	5	7	10	15	3	5	7	10	15	3	5	7	10	15	3	5	7	10	15	3	5	7	10	15



(a) **(b)**

Fig. 2. *(**a**) CTA CFRP plate, where Z is the defect depth, and labels are used to identify the location of each defect; (**b**) pulsed thermography setup. a, PC; b, IR camera; c1 and c2, left and right flashes; d, CFRP specimen.*

2.5 Contrast-to-Noise Ratio (CNR)

The results of the IRNDT approach show the anomalies as a contrast or color variation in the thermograms which are obtained from the methods. The contrast-to-noise ratio (CNR) is one of the suggested methods to quantify the results to evaluate defect detectability. CNR is similar to signal-to-noise ratio (SNR), which is mostly used for electrical signal description to evaluate the detectability or undetectability of defects. The CNR presents the contrast between the surface regarding the defective region and its surroundings. Between different definitions of CNR, the more robust formulation against noise and image enhancement operation [23] is used. The CNR can be determined by using Equation 9:

$$CNR = \frac{S}{N} = \frac{\mid \mu_{S\_area} - \mu_{N\_area} \mid}{\sqrt{\frac{(\sigma_{S\_area}{}^2 + \sigma_{N\_area}{}^2)}{2}}} \tag{9}$$

where S_area and N_area are the signal and noise regions above a defect and its neighborhood correspondingly. μ_{S_area} and μ_{N_area} are the average levels of contrast in signal and noise areas, respectively; σ_{S_area} and σ_{N_area} are the standard deviation of the contrast in S_area and N_area, respectively.

3. Analysis

The thermal maps are captured at 145 Hz sampling frequency for a duration of 15 seconds, leading to 3-D thermal data that contains T= 2203 thermograms of M by N images (T$\times$ M$\times$ N). For further processing, the 3-D data sequence is recomposed as a to a 2-D (T$\times$ M*N) matrix. Each row in this matrix is a thermal profile of each image pixel. As described in Figure 3a, the autoencoder was trained over TSR data and in order to have comparative results, PCT approach was applied over the pre-processed data. As far as implementation details regarding the training of the autoencoder are concerned, the training of the autoencoder model is implemented using Pytorch library [24] on a PC with 56GB of RAM memory and an Intel(R) Core(TM) i7-4820K control processing unit. Nonlinear activation functions–SELU have been employed at the encoder, and the model is trained using mean-square-error (MSE) as the loss function. For the optimizer algorithm, we selected the AdaGrad algorithm with a learning increment of 0.1 to adjust the learning rate. Other parameters included a batch size of 128 images and a number of epochs equal to 20. A mono-dimensional batch normalization layer is utilized to normalize the data before being fed into the next layer, before training the data standardized by removing the mean and scaling to unit variance.

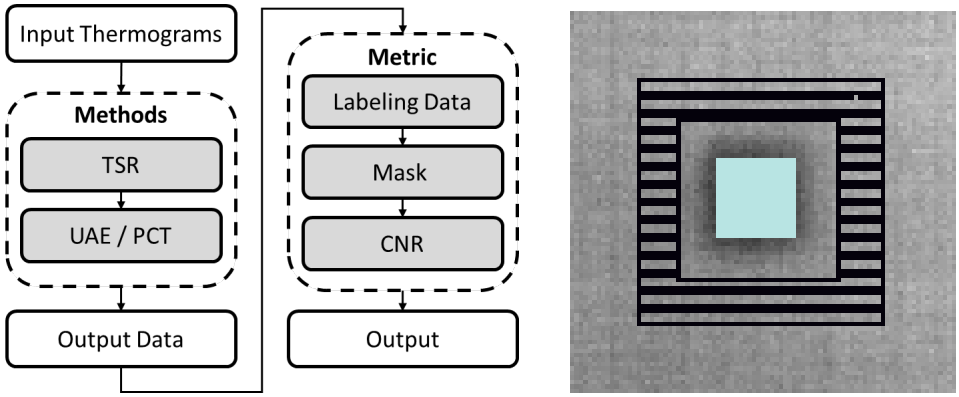


Fig. 3. *(**a**) Research flow-graph ; (**b**) Examples of reference and defect regions. The boundaries of the reference region is a region with black stripes, whilst the defective region is the light blue area.*

To evaluate the method quantitatively, the defective areas were labeled using LabelMe © [25]. From the border of the defective region, 15 pixels are considered as a transient region, and from the boundaries of this area, 20 pixels are automatically counted as a non-defective or sound area. Figure 3b illustrates the aforementioned regions so as to estimate the CNR score. The average and standard deviation values for all labeled regions are obtained from Equation 9.

4. Results and Discussions

Using the proposed approach described in Section 2, raw thermographic data was pre-processed by TSR (significantly reducing noise), and then the output data was used to train the UAE model. The first three obtained components from UAE and PCT are shown in Figure 4. It is clearly evident from the figure that the defect visibility in all components from UAE increased compared with PCT. The defects located at different depths reach the level of contrast to be clearly visible in a single component. Contrast is a conventional metric to evaluate the saliency of an object in a scene. A region with higher contrast compared to its surrounding is more distinguishable. Figure 5 shows the intensity variations along different horizontal lines of an image that is chosen arbitrarily. The Figures indicate that the components can reveal most of the defects.

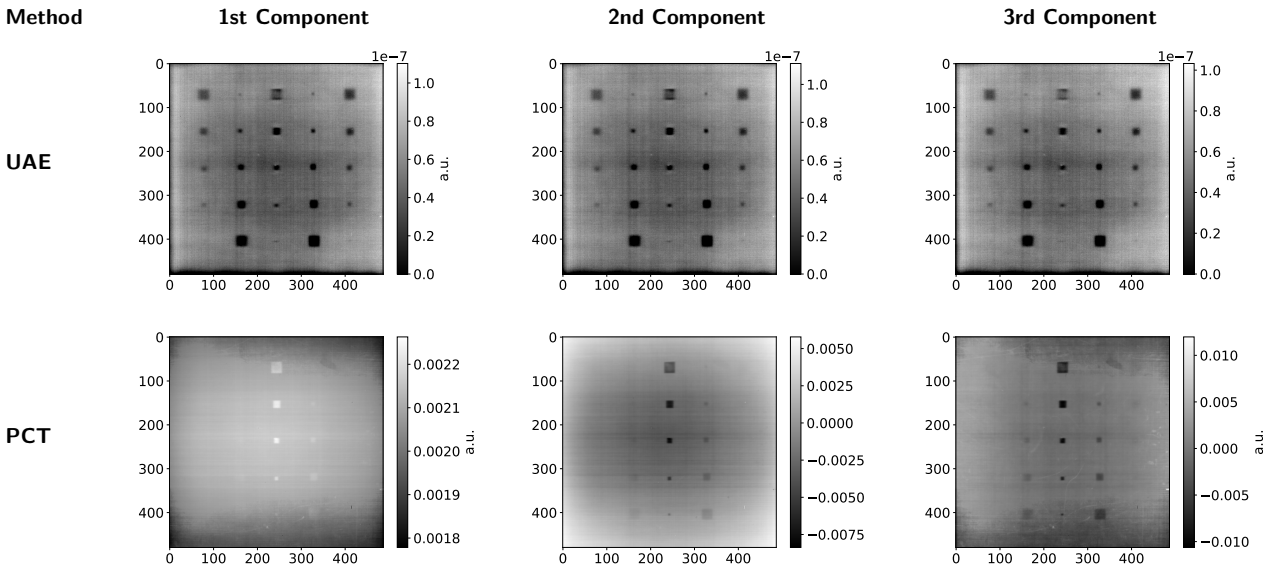


Fig. 4. *(1st row) These images present the first three components of undercomplete autoencoder. (2nd row) These images present the first three components of PCT on raw data*

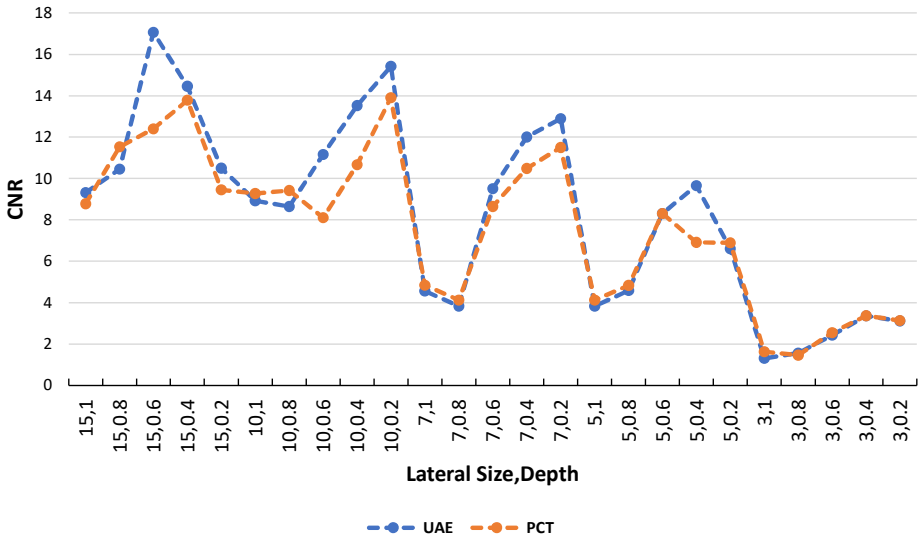


Fig. 5. *Maximum CNR for defects as a function of defect depth for all data sequences*

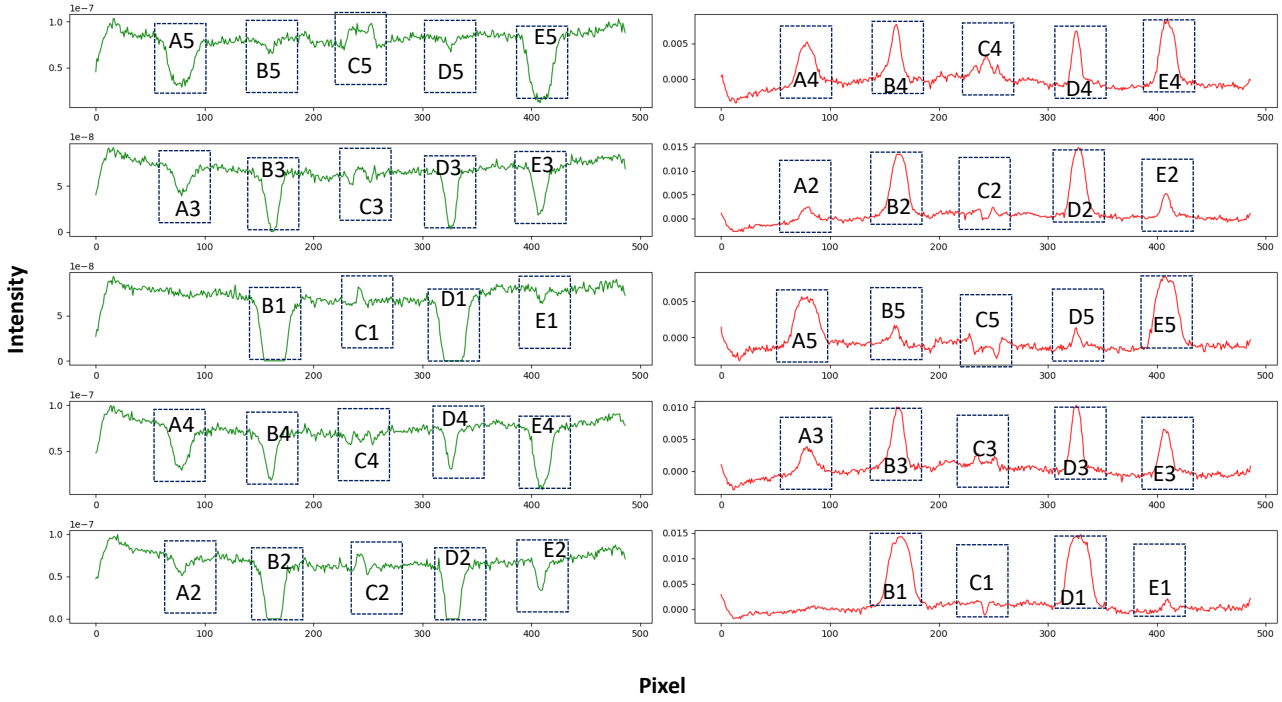


Fig. 6. *Gray level value variations in the UAE and PCT thermograms in the third component along the different lines. The green and red lines show intensity values in UAE and PCT, respectively.*

A quantitative comparison between UAE and PCT is presented in Figure 6 as a function of defect depth. The graph indicates that UAE's CNR scores in 13 over 25 cases (52%) have a higher value than the PCT results. The ratio of the

improvement varies from 0.01 % to 39.67%. In 48 % of the cases PCT results provide an improvement which varies 0.18% to 19.47%. In comparison to UAE results, PCT results provide an improvement which varies... This proposed processing approach can be extended to other approaches of thermography methods in the future and and other properties can also be investigated in order to make more improvements in terms of defect detection.

5. Conclusions

Enhancing defect contrast is always helpful in improving anomaly detection in pulse thermography. In this study, the application of UAE on PT data was investigated. In the proposed approach, the extracted features from data are employed to reconstruct new images with better anomaly visibility. Since UAE is a dimensional reduction method, a comparative study with PCT was done to show the potential of the method in IRNDT methods. Results showed that the UAE provide CNR scores higher than PCT, in 52% cases.

6. Funding

We acknowledge the support of the Natural Sciences and Engineering Council of Canada (NSERC), CREATE-oN DuTy Program [funding reference number 496439-2017], the NSERC DG program, the Canada Research Chair in Multipolar Infrared Vision (MIVIM), and the Canada Foundation for Innovation.

References

[1] Steven M Shepard. Flash thermography of aerospace composites. In *IV Conferencia Panamericana de END Buenos Aires*, volume 7, 2007.

[2] Xavier Maldague and Sergio Marinetti. Pulse phase infrared thermography. *Journal of applied physics*, 79(5):2694–2698, 1996.

[3] Nik Rajic. Principal component thermography for flaw contrast enhancement and flaw depth characterization in composite structures. *Composite Structures*, 58:521–528, 12 2002.

[4] Nikolas Rajic. Principal component thermography. Technical report, DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION VICTORIA (AUSTRALIA ..., 2002.

[5] Numan Saeed, Nelson King, Zafar Said, and Mohammed A. Omar. Automatic defects detection in cfrp thermograms, using convolutional neural networks and transfer learning. *Infrared Physics & Technology*, 102:103048, 2019.

[6] Qin Luo, Bin Gao, Wai Lok Woo, and Yang Yang. Temporal and spatial deep learning network for infrared thermal defect detection. *Ndt & E International*, 108:102164, 2019.

[7] Jing Jie, Shiqing Dai, Beiping Hou, Miao Zhang, and Le Zhou. Defect detection in composite products based on sparse moving window principal component thermography. *Advances in Polymer Technology*, 2020, 2020.

[8] Xin Zhang, Jafar Saniie, and Alexander Heifetz. Neural learning based blind source separation for detection of material defects in pulsed thermography images. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, pages 112–116. IEEE, 2020.

[9] Bahadin Azizinasab, Reza PR Hasanzadeh, Saeid Hedayatrasa, and Mathias Kersemans. Defect detection and depth estimation in cfrp through phase of transient response of flash thermography. *IEEE Transactions on Industrial Informatics*, 18(4):2364–2373, 2021.

[10] Alexander E Prosvirin, Farzin Piltan, and Jong-Myon Kim. Blade rub-impact fault identification using autoencoder-based nonlinear function approximation and a deep neural network. *Sensors*, 20(21):6265, 2020.

[11] Ivan Kraljevski, Frank Duckhorn, Martin Barth, Constanze Tschoepe, Frank Schubert, and Matthias Wolff. Autoencoder-based ultrasonic ndt of adhesive bonds. In *2021 IEEE Sensors*, pages 1–4. IEEE, 2021.

[12] Ziang Wei, Henrique Fernandes, Jose Ricardo Tarpani, Ahmad Osman, and Xavier Maldague. Stacked denoising autoencoder for infrared thermography image enhancement. In *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, pages 1–7. IEEE, 2021.

[13] V Gopi Tilak, VS Ghali, A Dilip Kumar, K Bala Sai Sankar, and VSNS Sharanya. Deep autoencoder for automatic defect detection in thermal wave imaging. *Journal of Green Engineering*, 10:13107–13118, 2020.

[14] Mingkai Zheng, Kaixin Liu, Nanxin Li, Yuan Yao, and Yi Liu. Deep autoencoder for non-destructive testing of defects in polymer composites. In *2021 8th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*, pages 91–95. IEEE, 2021.

[15] Kirandeep Kaur, Ravibabu Mulaveesala, and Priyanka Mishra. Constrained autoencoder based pulse compressed thermal wave imaging for sub-surface defect detection. *IEEE Sensors Journal*, 2021.

[16] Walter Hugo Lopez Pinaya, Sandra Vieira, Rafael Garcia-Dias, and Andrea Mechelli. Chapter 11 - autoencoders. In Andrea Mechelli and Sandra Vieira, editors, *Machine Learning*, pages 193–208. Academic Press, 2020.

[17] Zhaomin Chen, Chai Kiat Yeo, Bu Sung Lee, and Chiew Tong Lau. Autoencoder-based network anomaly detection. In *2018 Wireless Telecommunications Symposium (WTS)*, pages 1–5. IEEE, 2018.

[18] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[19] Steven M. Shepard. Advances in pulsed thermography. In Andres E. Rozlosnik and Ralph B. Dinwiddie, editors, *Thermosense XXIII*, volume 4360, pages 511 – 515. International Society for Optics and Photonics, SPIE, 2001.

[20] Jaeger J.C. Carslaw H.S. *Conduction of Heat in Solids*. Oxford University Press, nov 1959.

[21] G. Giorleo, C. Meola, and A. Squillace. Analysis of defective carbon-epoxy by means of lock-in thermography. *Research in Nondestructive Evaluation*, 12(4):241–250, Dec 2000.

[22] Xavier PV Maldague. *Nondestructive testing handbook. 4. Infrared and thermal testing*. American Society for Nondestructive Testing, 2022.

[23] R. Usamentiaga, Clemente Ibarra-Castanedo, and X. Maldague. More than fifty shades of grey: Quantitative characterization of defects and interpretation using snr and cnr. *Journal of Nondestructive Evaluation*, 37, 06 2018.

[24] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[25] Kentaro Wada. labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme, 2016.