Neural Computing and Applications (2023) 35:11207–11221
https://doi.org/10.1007/s00521-023-08293-7

ORIGINAL ARTICLE

Classification of barely visible impact damage in composite laminates using deep learning and pulsed thermographic inspection

Kailun Deng[1] · Haochen Liu[1] · Lichao Yang[1] · Sri Addepalli[2] · Yifan Zhao[1]

Received: 11 October 2022 / Accepted: 6 January 2023 / Published online: 31 January 2023
© The Author(s) 2023

Abstract
With the increasingly comprehensive utilisation of Carbon Fibre-Reinforced Polymers (CFRP) in modern industry, defects detection and characterisation of these materials have become very important and draw significant research attention. During the past 10 years, Artificial Intelligence (AI) technologies have been attractive in this area due to their outstanding ability in complex data analysis tasks. Most current AI-based studies on damage characterisation in this field focus on damage segmentation and depth measurement, which also faces the bottleneck of lacking adequate experimental data for model training. This paper proposes a new framework to understand the relationship between Barely Visible Impact Damage features occurring in typical CFRP laminates to their corresponding controlled drop-test impact energy using a Deep Learning approach. A parametric study consisting of one hundred CFRP laminates with known material specification and identical geometric dimensions were subjected to drop-impact tests using five different impact energy levels. Then Pulsed Thermography was adopted to reveal the subsurface impact damage in these specimens and recorded damage patterns in temporal sequences of thermal images. A convolutional neural network was then employed to train models that aim to classify captured thermal photos into different groups according to their corresponding impact energy levels. Testing results of models trained from different time windows and lengths were evaluated, and the best classification accuracy of 99.75% was achieved. Finally, to increase the transparency of the proposed solution, a salience map is introduced to understand the learning source of the produced models.

Keywords CFRP · Barely visible impact damage · NDT · Deep learning · Damage classification

1 Introduction

Composite laminates have gained tremendous popularity due to their superior performance compared to traditional materials in the past few decades, especially in aviation and automotive applications. The high Strength-to-Weight ratio is always a significant advantage in improving the product strength and energy efficiency by widely applying them to load bearing structures. CFRP are commonly rated as one of the most widely used composite materials owing to their outstanding Strength-to-Weight ratio [1]. Previous studies have claimed that the proper use of CFRP materials in vehicles can lead to a weight reduction of 40–65% [2]. Although the benefits of CFRP materials are overwhelming, such as high load-bearing capacity, high chemical resistance and them being lightweight, their behaviour when it comes to impact events results in poor mechanical performance [3]. Impact events can cause cracking, fibre breakage and delamination of CFRP, leading to catastrophic results, especially in the aerospace industry [4]. In practical applications, bird strikes in aviation are high-risk hazards to passengers and stakeholders.

Among the damage that CFRP can generate in an impact event, Barely Visible Impact Damage (BVID) is a typical type which has drawn significant attention [5]. BVIDs refer to subsurface damage usually caused by low-velocity impact and cannot be easily detected by regular visual

✉ Yifan Zhao
yifan.zhao@cranfield.ac.uk

[1] Centre for Life-cycle Engineering and Management, School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedford MK43 0AL, UK

[2] Centre for Digital Engineering and Manufacturing, School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedford MK43 0AL, UK

inspections, causing structural and functional failures [6]. Under this premise, many efforts have been made to detect and study BVIDs [7–9]. Considering most CFRP-based structures are of high value and are costly and laborious to be replaced, Non-Destructive Testing (NDT) is always recommended for the inspection of BVIDs due to its favoured nature of examining test objects without interfering with their integrity [10, 11].

Thermography is evolving into a promising branch in NDT techniques because of its robustness, capability to inspect large areas with intuitive imagery and short asset downtimes [12, 13]. Thermography inspection can generally be divided into passive and active modes [14]. Active thermography techniques, such as Pulsed Thermography and Laser Thermography have a proven track record to inspect CFRP materials [15–17]. Pulsed Thermography has gained tremendous popularity where it employs a homogeneous heat pulse as a stimulation to map the sub-surface structure and thereby identifies BVIDs [18]. Besides, the technique is also sensitive to the detection of delamination in BVIDs [19].

In recent years, with the rapid development of Artificial Intelligence (AI) technologies, researchers in various fields made many attempts to transfer AI-based techniques to their specialised areas. Some studies have applied AI technologies to identify and study composite materials' defects using thermography. Mamani et al. [20] proposed a machine learning-based approach for defect classification in CFRP materials. In this research, three exponential-model-based features related to the depth of defects were fed into the selected machine learning classifier for training. The generated decision forest model achieved more than 99% defect classification accuracy. Additionally, two Artificial Neuro Networks (ANN) were proposed to implement damage severity quantification and location. This research used simulated frequency shift data on composite materials generated by Finite Element Analysis (FEA) to feed the proposed ANNs and achieved a prediction accuracy of up to 95% [21]. Another study from Tavares et al. [22] employed both simulation and experimental data to perform damage detection on CFRP structures. The obtained Frequency Response Functions (FRFs) and time signals from inspections were processed with the K-means Clustering and Multivariate Anomaly Detection and thus created a defects detection model. Some other research focussed on improving the automation of defect detection in CFRP. Saeed et al. [23] combined a Convolutional Neural Network (CNN) and a Deep Forward neural network and achieved defects detection and depth measurement using thermograms. A deep learning-based impact damage segmentation method was brought forward by Wei et al. [24]. In this research, impact damage on curved surfaces of CFRP specimens were inspected by Pulsed Thermography. The obtained thermal images were then processed by Principal Component Thermography (PCT) and Empirical Orthogonal Functions (EOF) before being used for training in a U-Net. The trained models achieved F1 scores of over 87% on both middle and long-wave infrared data. Oliveira et al. [25] tested forty CFRP impact damage samples using Lock-in thermography, and the acquired images were used to train a U-Net model to perform damage segmentation. Zhou et al. [19] presented a framework to extract impact damage contours using image processing methods and then used three features, including area, perimeter and major axis length, for further shallow machine learning-based damage classification according to different impact energy levels. Another study conducted by Fotouhi et al. [26] also achieved impact energy-oriented classification of BVID using a biomimetic tactile whisker and Support Vector Machine (SVM) classifier.

Based on the review above, it is noticed that most machine learning-based damage studies related to composites either use simulation data or a combination of simulation and experimental data to train the models. In contrast, others obtained crucial features for machine learning by using limited experimental specimens (specimen number ≤ 40 based on our review). Due to the limited samples, feature engineering associated with shallow machine learning is always used to understand impact damage. However, based on our review, there is limited research specialising in classifying experimental BVID in CFRP materials according to different impact energy levels using Deep Learning methods.

This research proposed a novel analytic framework, using a tailored deep learning approach, for characterising impact damage in CFRP laminates using only experimental data captured from the pulsed thermographic inspection of 100 specimens. It must be noted that the variations and uncertainties especially during the manufacture and the artificial creation of damage have a critical impact on this research. This study thus aims to understand the relationship between thermographic patterns of BVID in CFRP materials and their corresponding impact energy levels, which is significant for facilitating the impact energy prediction based on NDT inspections of damaged components. Compared with previous studies on this topic, which often employed image processing methods to extract morphological features and used shallow machine learning-based classification processes, the proposed method aims to learn features automatically for better classification accuracy using a deep learning approach. Another novelty of this work is using successive frames from the collected thermogram of each specimen to expand the training datasets. This study also explores how high classification accuracy is achieved to improve the transparency of the produced models. The impact of this study can be summarised as

follows: In real-world applications, this prediction of impact energy intensity can help infer the possible cause of impact incidents. On this basis, targeted predictive maintenance can be introduced to prevent potential hazards. The main contributions of this work can be summarised below:

(1) A new analytic framework for investigating the relationship between impact damage patterns and their corresponding impact energy levels in CFRP laminates using a deep learning network purely based on experimental data;
(2) A new data augmentation method of using successive thermal image frames from the inspection of each specimen to constitute training datasets for the selected deep learning network. This new approach is also enlightening for expanding datasets in other studies using thermographic inspection;
(3) A relatively large thermal imaging dataset from 100 CFRP specimens subject to low-velocity impact;
(4) The implementation of the Class Activation Map improves the transparency of the introduced deep learning method.

2 Methods

The proposed methodology is illustrated in Fig. 1. The first step of the study is data acquisition, including impact damage specimen preparation and implementation of pulsed thermographic inspection. The 16-bit raw data were then calibrated to RGB-colormap image frames and saved as 8-bit PNG images. The second step is the deep learning network selection. ResNet was selected due to its superior capability in image classification, and some fine-tuning was made for the network to fit our learning assignment [27]. In the following stage, different datasets were generated using saved thermal images and fed into the selected deep learning network for model training. The performance of the trained classification network was compared and analysed in the last section. More details for each part of the method are discussed below.

2.1 Data acquisition

One hundred specimens of CFRP samples with the size of 150 × 100 × 3 mm were manufactured in the laboratory before subjecting to drop-impact test. The samples were cut from 4 larger plates (750 × 750 × 3 mm) marked with P1, P2, P3 and P4, manufactured with the same CFRP material specified in Table 1 in Appendix. Then, the drop-impact experiment was conducted on each sample with pre-set impact energy levels of 4 J, 6 J, 8 J, 10 J and 12 J. The impact was performed by the free fall of a hemispherical indenter with an exact weight of 2.281 kg. The precise impact energies was achieved by subtly setting the free fall distance and the force of impact from a force transducer. The 100 prepared samples were equally divided into five groups, and each group was exposed to one particular level of impact energy. Thus 100 specimens (from S001 to S100) were produced. Figure 2 displays the front and back sides (only the Region of Interest) of five specimens with different corresponding impact energy levels. The drop-impact was applied on the front side of each sample (see the first row of Fig. 2), and no impact damage can be detected by visual inspection from either of the sample surfaces ranging from no visible features to barely visible features (Fig. 2). In this investigation, pulsed thermographic inspection was conducted on the back side of each specimen to reveal the subsurface impact damage.

The principle of pulsed thermography inspection is illustrated in Fig. 3a. In this case, an artificial temperature gradient was induced by a high-energy heat pulse produced by flash lamps. With time elapsing, the heat propagates within the specimen and displays unusual heat conduction behaviour when getting through damaged areas compared to sound regions [28]. The thermal camera records the whole process at an appointed frame rate, and these abnormalities can be detected and recorded in a sequence of frames which can be eventually interpreted into damage patterns.

For the thermography measurement, the Thermoscope II pulsed-active thermography system was used. The excitation source in the inspection is enabled by two capacitor banks-powered Xenon flash lamps that produced the heat

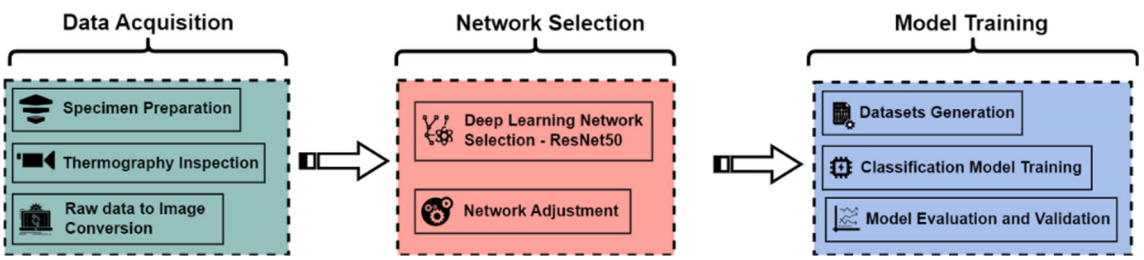

Fig. 1 Structure of the proposed methodology

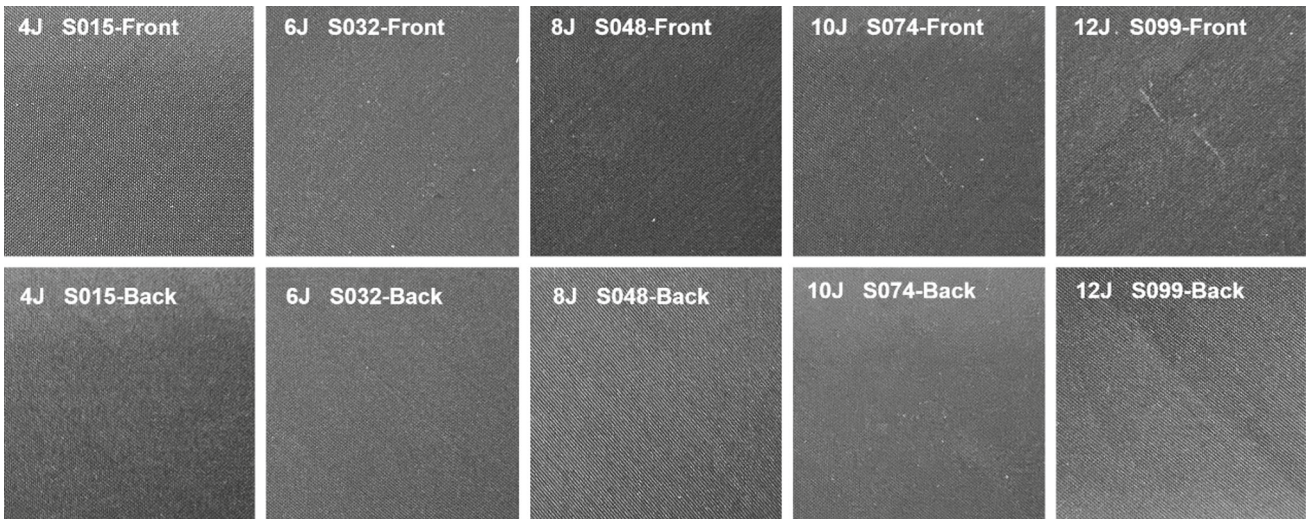

Fig. 2 Digital images of both front and back sides of 5 specimens for 4 J, 6 J, 8 J, 10 J & 12 J impacts

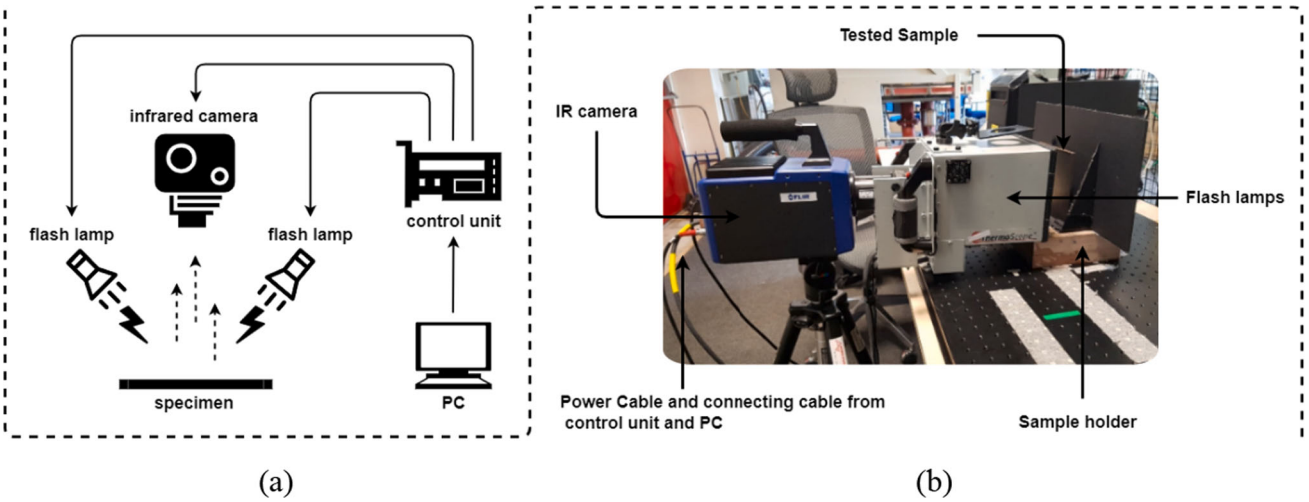

Fig. 3 **a** Illustration of Pulsed Thermography set-up; **b** A photo of the experimental set-up

pulse. The thermal camera utilised in this study is a cooled FLIR SC7000 series IR radiometer (see Fig. 3b) with a spatial resolution of 640 × 512 pixels. During each inspection, the specimen was placed in front of the IR camera at a distance of around 250 mm, and the surface of the specimen was kept perpendicular to the camera lens. The energy provided by the heat pulse in this inspection was around 2 kJ, and its effective area is 200 × 250 mm, which can cover the inspected specimen in full. The sampling rate of the camera was set to 50 Hz, which was determined after taking into account CFRPSs low thermal diffusivity property and the specimens' thickness into account.

The inspection lasted 20 s for each specimen, and a frame sequence containing 1000 thermal images was obtained. For each inspection, the stimulation flash was set at the 10th frame. The results showed that the 14th frame was the first frame from which impact damage features started to reveal themselves. Considering the facts above and the limited valid time of the induced heat propagation, we only employed the frames from the 14th to the 625th in each inspection. Each captured thermal image was cropped to 150 × 150 pixels, centred by the impacted area, to reveal the damage pattern. Figure 4 shows the raw thermal images at 0.3 s (the 15th frame) and 0.6 s (the 30th frame) after the pulse for each group.

2.2 Deep learning network selection and optimisation

Convolutional Neural Networks (CNN) constitute a crucial part of Deep Learning and play an essential role in

computer vision tasks such as image segmentation and classification [29, 30]. ResNet is essentially a type of CNN, but its performance improved significantly due to its unique network architecture, which resolves deep CNN's gradient vanishing problem [27, 31–33]. Considering our research task is grouping thermal images into different categories, which is a typical image classification task, ResNet was selected as our dedicated deep learning tool. Regarding the layers of the selected neural network, ResNet50, which is 50 layers deep, was finally chosen as it had the best performance when compared with ResNet18 and ResNet34. More complex versions with more layers are not adopted after a balanced consideration of performance requirements for our tasks and training efficiency. Another reason for choosing ResNet50 is that transfer learning can be utilised using its pre-trained model, which has been trained on the ImageNet dataset [34]. Although 100 pieces of impact damage specimens are quite a large sample capacity within its specialised domain, it is still not enough to make up the dataset for training a deep learning model from scratch. The pre-trained model of ResNet50 can significantly reduce the dataset size requirement, making it feasible to train a classification model based on our limited specimens. Figure 5 shows the customised ResNet50 network architecture that our research used. Some modifications were made to the structure of a classic ResNet-50 to adapt the network to our specific research task, such as changing the output dimension of the final fully connected layer to 5, which means five categories of impact energy levels in our study. This study is essentially a multi-class classification problem. Cross Entropy is employed as the loss function for the selected deep learning network due to its good performance in classification models [35]. Adam is used in this study as an optimiser, as research confirms its strong usage in deep learning applications due to its capability of fast converging and good performance [36]. The initial learning rate was set as 0.001. Furthermore, the learning rate decay method has been employed to adjust the learning rate adaptively to ensure the optimisation process runs efficiently., After several tests, the number of epochs was set to 100, and the training batch size ranged from 16 to 256 according to different dataset sizes.

3 Model training and model evaluation

3.1 Dataset generation and training design

For comparing the classification performance of models in a sensitivity analysis, various datasets were generated with different combinations of thermal images captured in the pulsed thermographic inspection of the prepared impact damage specimens. The proposed methodology is based on two critical assumptions about the generated dataset. The first hypothesis is that the same frame of thermal images captured from each inspection of the specimens created by the same impact energy should have common features that can differentiate them from the same frame of these specimens exposed to a different impact energy level. The "same image frame" mentioned here means the thermal image was captured at the same time after the flash with an

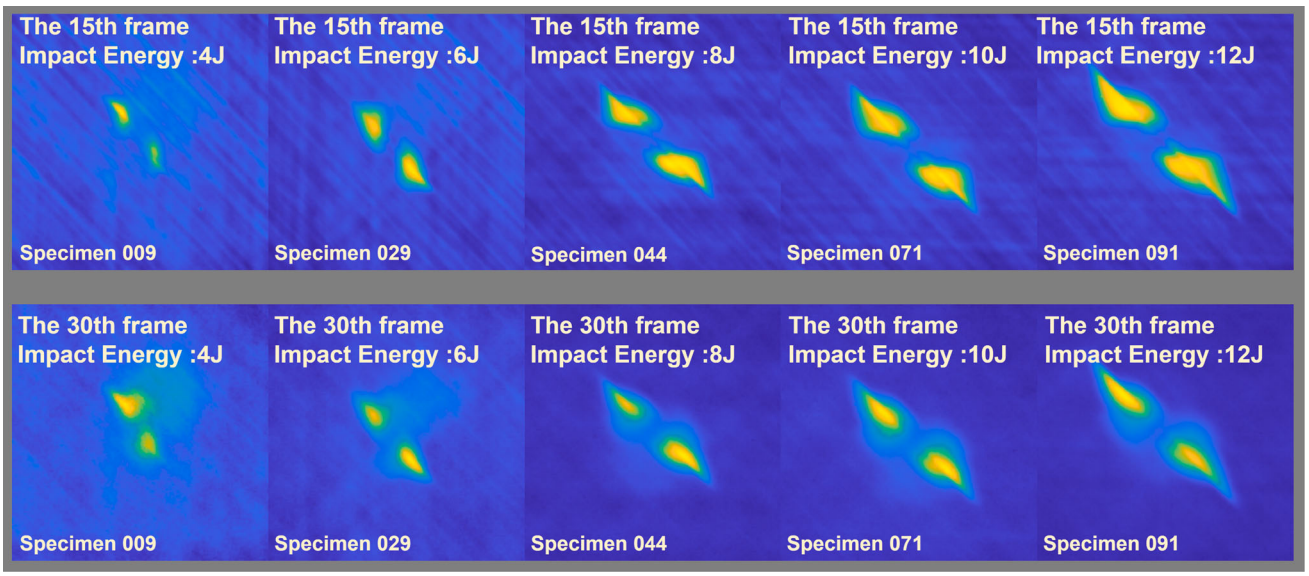

Fig. 4 Reconstructed thermal images of different specimens with the regulated size of 150 × 150 pixels

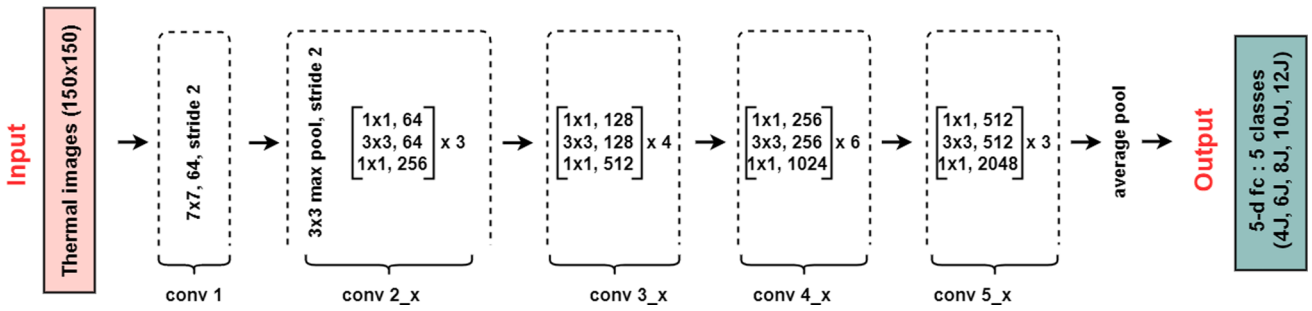

Fig. 5 Architecture of ResNet50 with customised input and output

identical inspection set-up. The identical inspection set-up includes the same inspecting equipment, excitation (including the same excitation source, energy level and pulse length for each inspection), inspection duration and capture frame rate. The data acquisition process for this research was manipulated to meet all the preconditions of the hypothesis mentioned above. An example can be taken from the data acquisition process to clarify the hypothesis. The 50th thermal image frame from the inspection of specimen A should have common features with the 50th thermal image frame from specimen B, where specimens A and B are subject to the same impact energy level, such as 8 J. Besides, these common features from the 50th frame of inspections of specimens A and B should differ from other common features from the 50th frame of inspections of specimens C and D that are subject to different impact energy. Based on this hypothesis, a dataset used for classification model training in this research can be generated by taking the same specific thermal image frames from each pulsed thermographic inspection of 100 specimens. Initially, the 15th frame after the flash was selected since the most distinct damage pattern was detected around this frame for all 100 specimens. As a rule of thumb, more distinct patterns improve classification accuracy in deep learning practices. For the convenience of the follow-up training result comparison based on different datasets, this dataset formed from the 15th frame of each thermal image sequence of the 100 specimens is labelled with Dataset15.

Although the utilisation of transfer learning makes it possible for model training on small datasets such as Dataset15, which only has 100 images, increasing the dataset size is still the most straightforward way to improve training performance in deep learning tasks. This research proposes a novel method for obtaining more training data from limited impact damage specimens. The second hypothesis is that the series of thermal image frames coming from one inspection of one specific specimen should share some common features which are related to their corresponding impact energy level, while these common features can differentiate them from corresponding frames of other specimens created by different impact energies. An example to explain this hypothesis is that the 15th frame and the 20th frame from the same thermal image sequence of specimen A should have some characteristics in common, and these characteristics should be different from those common features found in the 15th and 20th frames from the inspection of specimen B with a different impact energy level. In the light of this hypothesis, the size of the datasets for our research can be enlarged by importing more thermal image frames instead of one single frame from each inspection. In the first attempt, five successive frames from the 14th frame to the 18th frame were selected from the thermal images of each specimen, and thus a new dataset containing 500 thermal images was formed and marked with Dataset14-18. The reason for initially choosing these five frames is that these frames display the most noticeable damage patterns. For finding the best five frames that can contribute to the best classification accuracy using machine learning, instead of the choice of 14–18 frames by intuitive feeling, a series of datasets were generated with consecutive 5-frame windows sliding from the 14th to the 623rd frame. These 5-frame-based datasets are denoted as Dataset14-18, Dataset19-23, Dataset24-28, …, and Dataset619-623.

The same strategy was applied to 10-frame windows, 20-frame windows and even 60-frame windows to study the correlation between the classification accuracy and the frame number constituting the dataset. The corresponding datasets are Dataset14-23, ……, Dataset614-623, Dataset14-33, ……, Dataset594-613 and Dataset14-73, ……, Dataset554-613. Figure 6 illustrates the dataset generation process in this research.

Upon understanding the process of generating various datasets in our research, it is not difficult to notice that the number of images correlating to each impact energy level is the same in each dataset. In this study, thermal images in each dataset were divided into training, testing and validation subsets with a proportion of 60%, 20% and 20%,

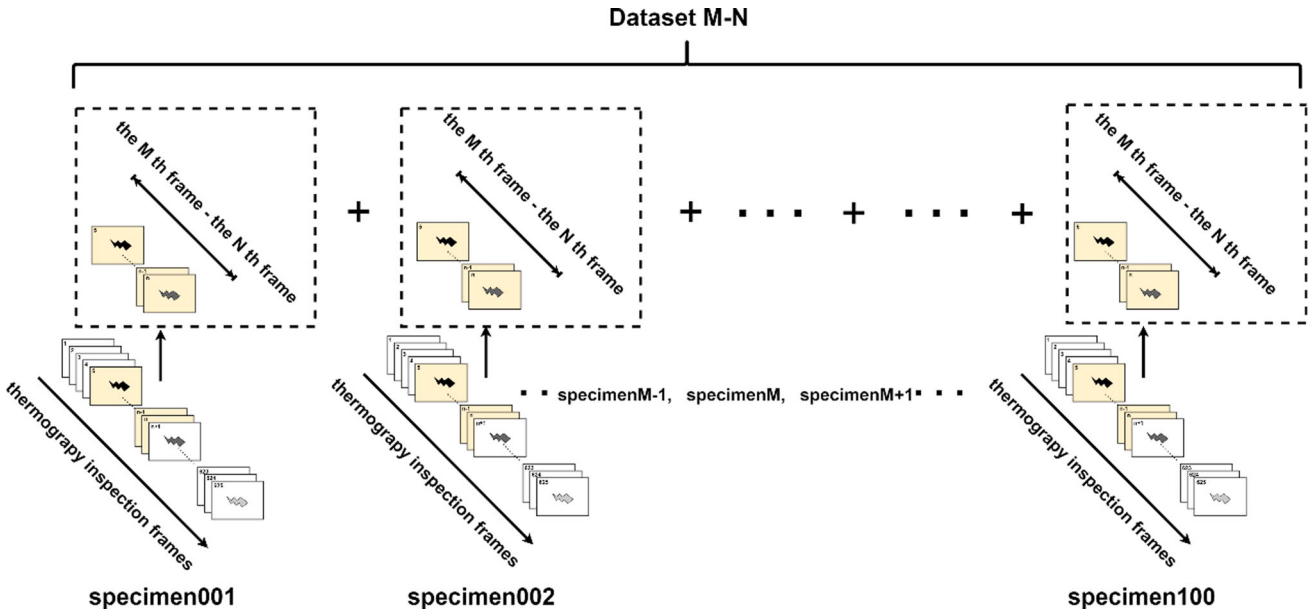

Fig. 6 Illustration of the generation of Dataset M–N

respectively. And in each subgroup, the number of images corresponding to different impact energy levels is identical.

3.2 Deep learning network fine-tuning and model evaluation

Considering the fact that the captured datasets are still relatively small compared with the typical dataset used for deep learning training, a pre-trained version of ResNet-50 was loaded for model training. The loaded network has already been trained on more than one million images from the ImageNet database and can classify images into one thousand categories.

The parameters of the pre-trained network have already been optimised on the pre-trained dataset, making it suitable for small datasets that cannot provide adequate data for the deep learning network starting from nothing. In the training process, images in each dataset were randomly redistributed three times, and the training was performed three times. This procedure aims to gain an average accuracy that can genuinely reflect the model's performance.

In the model testing process, a Confusion Matrix was employed to reveal detailed information about the model performance. The model evaluation index, Accuracy was selected to indicate the overall model performance. Accuracy is one of the most critical evaluation metrics in classification tasks in the deep learning domain, and it is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (1)$$

where TP = TruePositive, TN = TrueNegative, FP = FalsePositive and FN = FalseNegative.

In multi-class classification tasks, the numerator TP + TN represents the number of elements which are correctly classified to their respective true categories, while the denominator TP + TN + FP + FN includes both correctly classified elements and incorrectly classified ones. Accuracy calculation is straightforward for multi-class classification using the Confusion Matrix. The sum of numbers appearing on the matrix's main diagonal is the total number of correctly classified elements, that is TP + TN in Eq. (1). The total amount of elements in the matrix forms the denominator of the right side of Eq. (1).

In the last section of this study, the Class Activation Map (CAM) was adopted to validate the generated models after training. By utilising this technique on any training or tested image, the discriminative object parts on the image can be detected, and the predicted class scores are visualised. In this way, the critical areas which contribute the most to image classification in any image are revealed, and thus the model's validity was tested [37].

4 Results and discussion

4.1 Training profile

Figure 7 shows the loss and accuracy curves throughout the training process on three different datasets. The training loss and validation loss from each graph have a similar tendency that decreases from the beginning and level off as

the training epoch increases. Meanwhile, both accuracies from each training process share the opposite pattern from losses, which rise to the peak at the early stage and then converge around a specific value. Observed from the loss and accuracy charts, it is clear that models trained on various datasets all have quick convergence at around the 40th epoch (the total epochs are 100). Both the training accuracy and validation accuracy reached a relatively high level, with more than 70% on Dataset15 and over 90% on Dataset14-18 and Dataset14-33. In addition, similar accuracies were noticed between training and validation in each epoch, especially after the model achieved convergence, which suggests a high consistency between training and validation indicating that the trained models have neither underfitting nor overfitting problems. In brief, the loss and accuracy curves suggest that ResNet50 is suitable for our research task, and the network parameters for the model training are appropriate.

4.2 Classification performance

The test accuracy can represent whether the network effectively learns the relationship between the damage pattern and its class (impact energy level). Starting from the result from the single frame dataset Dataset15, which contains the 15th frame from each specimen's inspection and 100 thermal images included, an average accuracy of 70% was achieved. For Dataset14-18, which consists of 5 adjacent frames from the 14th to the 18th and has a dataset size of 500, the accuracy increased dramatically to 94.67%. With the dataset size expanding, higher classification performance was achieved on Dataset14-23, Dataset14-33 and Dataset14-73, with average accuracies of 98.67%, 99.25% and 99.75%, respectively.

The result based on a single frame dataset (Dataset15) proved the first hypothesis of this research. The test accuracy of 70% means that more than two third thermal images in this dataset can be correctly classified according to their corresponding impact energy level. This result proved that thermal images from the same frame of different specimens have some impact energy-related features that can be extracted and utilised by deep learning networks to perform impact energy-oriented classification.

The second hypothesis of our study is also proved by the classification performance on datasets with multi-frame windows like Dataset14-18 and Dataset14-24. The achieved high accuracies on these datasets confirm that common features exist in each image frame of the same inspection, which is associated with the impact energy the

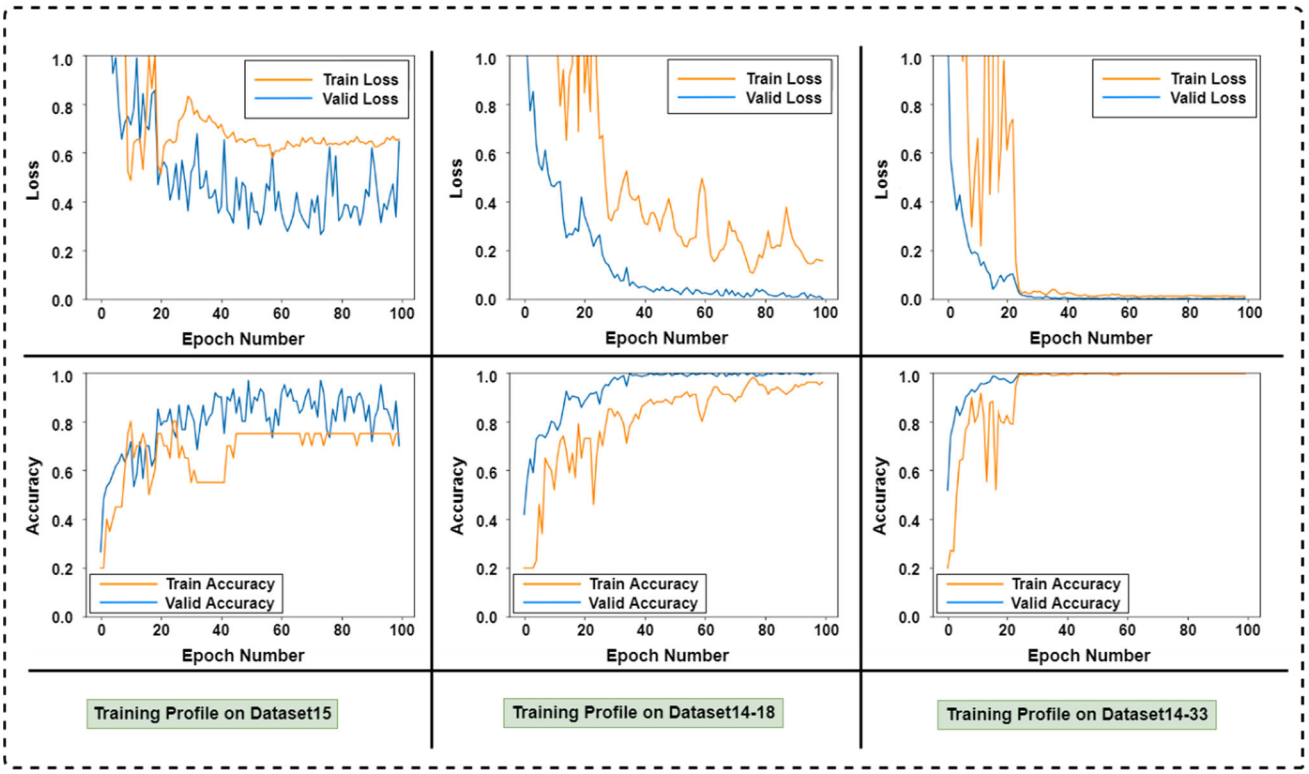

Fig. 7 Training loss and accuracy profiles of 3 different datasets

specimen was subjected to. ResNet50 can learn these features and use them for classification.

In order to identify the 5-frame window for generating a dataset on which the best classification accuracy can be achieved, the deep learning training was conducted on a series of temporal-shifting window datasets through the whole thermal sequence (Dataset14-18, Dataset19-23, Dataset24-28, …, Dataset619-623). The classification accuracies of different time-window datasets are shown in Fig. 8. It can be observed that the average test accuracy keeps declining from more than 95% (with a peak of 99.33%) and converges around 78%. The turning point of the curve from declining emerges at the frame window around Dataset214-218, which suggests that the crucial features that can contribute most to the classification accuracy can only be found on the images before this turning point. There are some impact energy-related features for those images after the turning point, but the network cannot build high-performance models only based on these images. The best accuracy (99.33%) can be observed on the frame window from 19th to 23rd (on Dataset19-23), which indicates that the thermal images within this period in inspection sequences contain the most prominent defect features and representation of damage severity.

Then, to figure out the most appropriate duration of the time-window abovementioned, the same training program was implemented for a 10-frame dataset, a 20-frame dataset and a 60-frame dataset. The accuracy shifting curve for each of them can be found in Figs. 13, 14, 15 in Appendix. Figure 9 is a joint illustration of these shifting curves of accuracy. If the length of the frame window for generating the dataset is fixed, the window starting from the 14th frame would be the best one for sourcing training data. Moreover, a dataset using images from earlier frame windows leads to better classification accuracy. This observation can help select proper thermal data from thermography inspections for deep learning training.

Apart from the similar variation trend of each curve, another significant pattern was observed. In any specific frame window, such as 164th to 264th, the accuracy of the model trained from the 60-frame dataset ranked at the top, followed by the model trained from the 20-frame datasets. The models trained from the 10-frame dataset took the third position, outperforming the models trained from the 5-frame dataset. The observed ranking indicates that within a specific frame window of a thermographic inspection, more successive frames of thermal images are employed to constitute the dataset, and better classification accuracy will be achieved. This conclusion is confirmed by analysing the results on different datasets. Figure 10 lists the test accuracies of trained models on ten different datasets. On the top row, an accuracy of 96.33% is displayed, which is achieved on the 60-frame dataset Dataset254-313. Another three model accuracies are placed on the middle row, which is gained by the models trained on three 20-frame datasets (Dataset254-273, Dataset274-293 and Dataset294-313). The test accuracies based on the 10-frame datasets from Dataset254-263 to Dataset304-313 occupied the bottom line. Essentially, all the images from Dataset254-273, Dataset274-293 and Dataset294-313 are identical to all the images from Dataset254-313. The average model accuracy trained on Dataset254-273 is 89.92%, while 89.83% for Dataset274-293 and 89.92% for Dataset294-313. These three accuracies are much lower than the accuracy of 96.33% achieved on Dataset254-313. This phenomenon can also be interpreted as the model accuracy will increase if thermal images from successive frames are added to either Dataset254-273, Dataset274-293 or

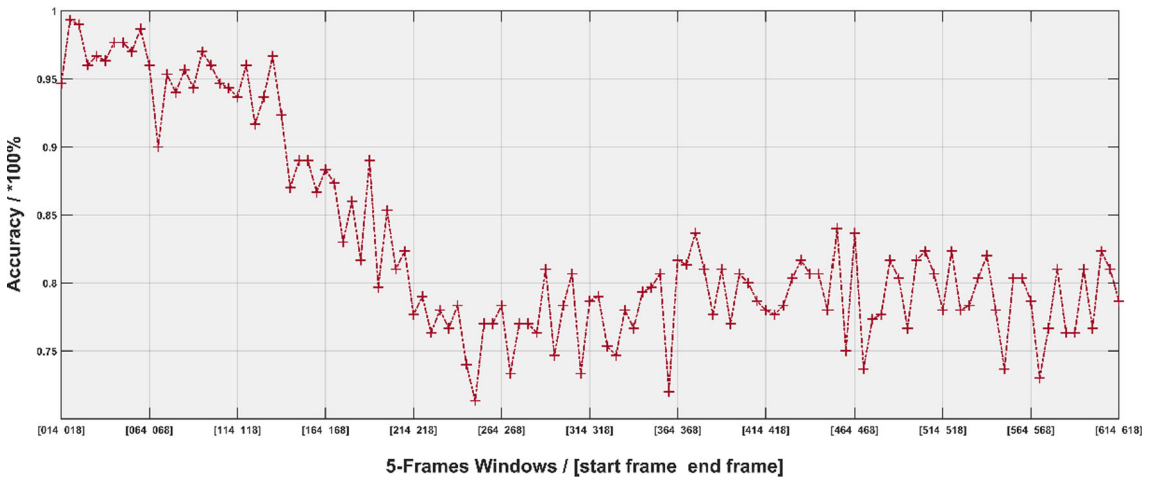

Fig. 8 The curve of classification accuracy with a temporal-shifting window of 5-frame datasets

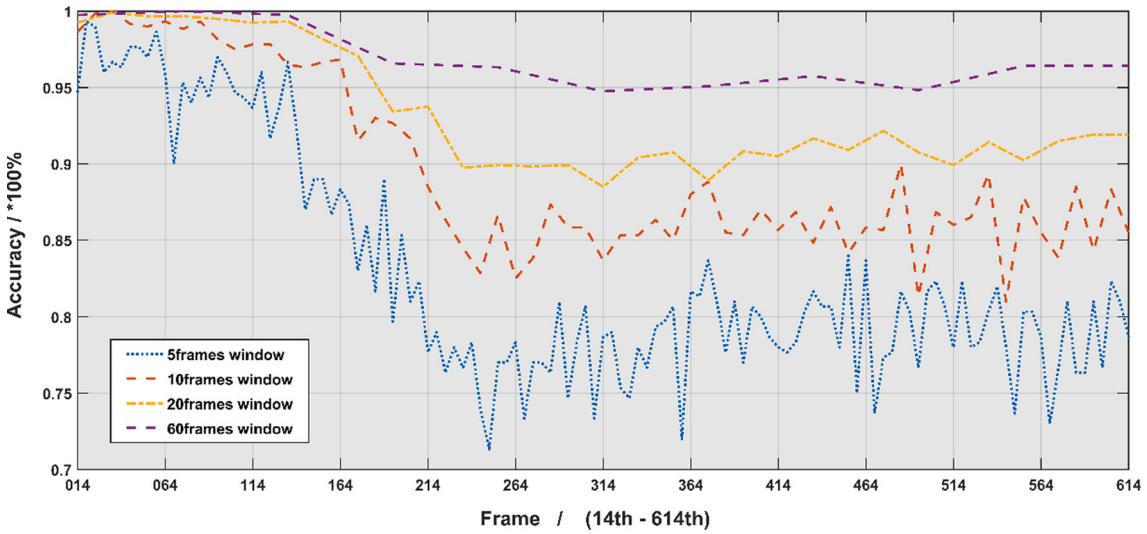

Fig. 9 The curves of classification accuracy with a temporal-shifting window of datasets with different frame numbers

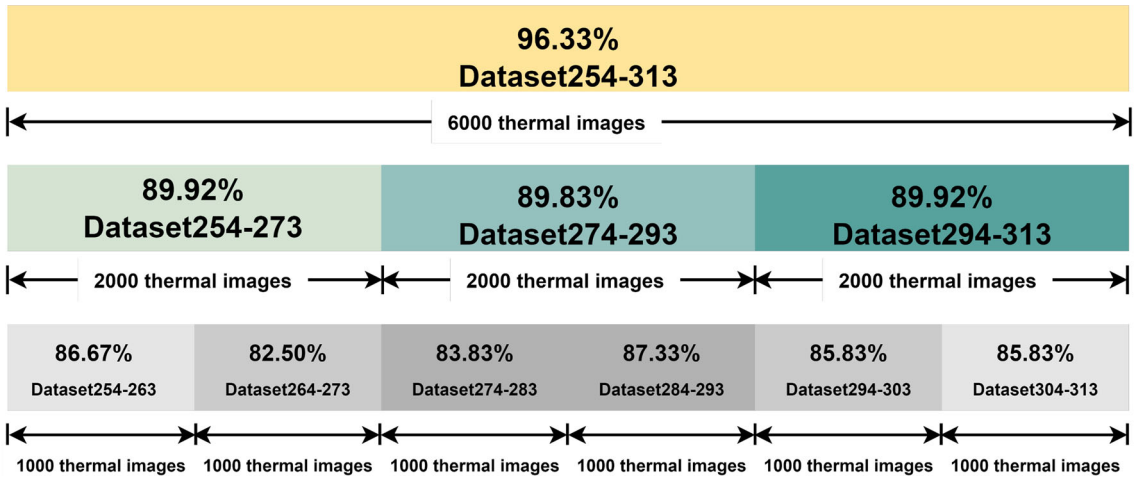

Fig. 10 Classification accuracy comparison of different datasets generated using thermal images on the window from 254 to 313th frame

Dataset294-313. The same conclusion can be drawn by comparing the data between the figure's middle and bottom rows. Since the deep learning model training is only conducted on datasets sourcing from no more than 60 successive frames in this study, this inference can only be valid within this scope.

4.3 Confusion matrix

Apart from the overall test accuracy, the Confusion Matrix from the test process was also provided to reveal the models' classification performance for each class/impact energy. Six confusion matrixes are displayed in Fig. 11, in which the results of the first row are generated from the 20-frame datasets, and the results of the second row are obtained from the 60-frame datasets. Except for individual test accuracy for each class of impact damage with different impact energy levels, some other interesting characteristics can also be found in the confusion matrixes. Impact damage generated with 4 J's impact energy are inclined to confuse with damage from 6 J, and the same phenomenon is also observed between 8 and 10 J. For example, four elements in the first column which should belong to class 6 J, are classified into class 4 J, which results in the most significant proportion of misprediction for class 4 J. In the third and fourth columns, seven elements from class 8 J are classified as class 10 J, and another seven elements from class 10 J are classified as class 8 J, almost occupying all the wrong predictions in these two categories. Similar behaviours have been observed in the other five confusion matrixes in Fig. 11. The observed confusion tendency between damage patterns

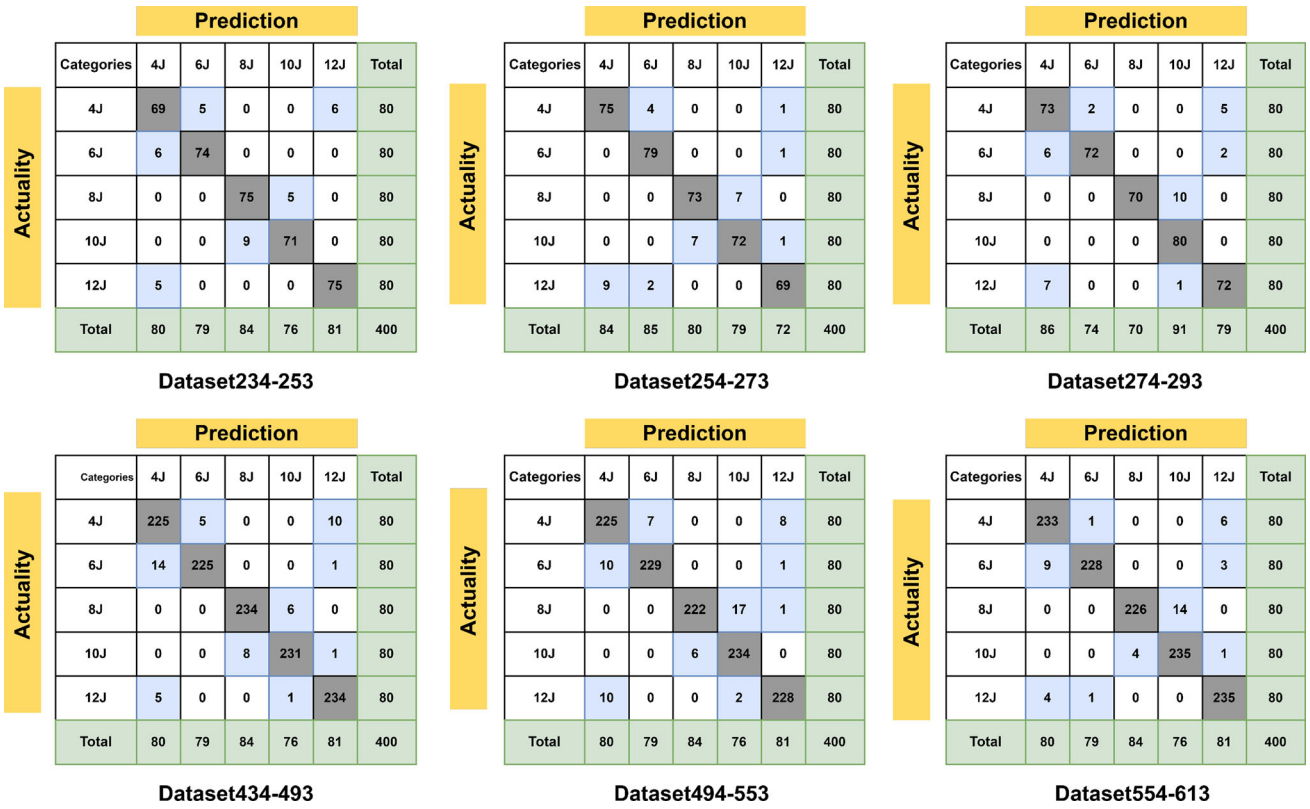

Fig. 11 Confusion Matrixes of testing results on six different datasets

caused by adjacent impacts energy levels such as 4 J and 6 J, 8 J and 10 J may be explained by the minor impact energy differences leading to similar damage patterns.

4.4 Class activation map (CAM)

This study employed Class Activation Map (CAM) to better understand how the trained models work. This technique can visualise the model's learning process on one specific image. More specifically, a class activation map can reveal the critical areas of the tested image from where the model learns discriminative features that contribute to the classification. Figure 12 shows class activation maps based on 24 tested thermal images from Dataset14-33. A translucent heat map is generated by CAM and overlaps with the original image. In the map, the regions in red play more critical roles than those in blue regions in identifying a specific class of the original image. Looking into each picture in Fig. 12, almost all the red areas are centred on or around the impact damage areas enclosed with the dotted line. This observation suggests that the models learned features from damage patterns instead of other parts of the images. The classification was based on features closely related to impact damage patterns.

5 Conclusions

This paper presents a study to understand the relationship between barely visible impact damage in CFRP materials and the corresponding impact energy level using a deep learning-based classification method. Thermal imaging of impact damage was collected using Pulsed Thermography, and the revealed damage patterns were then utilised to construct datasets for the training of classification models. The conclusions below are drawn by comparing classification accuracies trained on different datasets.

(1) For impact damage patterns captured in the same specific frame (e.g. the 15th frame) of inspections for all specimens, impact energy-related common features exist among these images and can distinguish them from any other damage patterns caused by different impact energy levels.
(2) For impact damage patterns captured in different image frames of an inspection of one specific specimen, impact energy-related common features

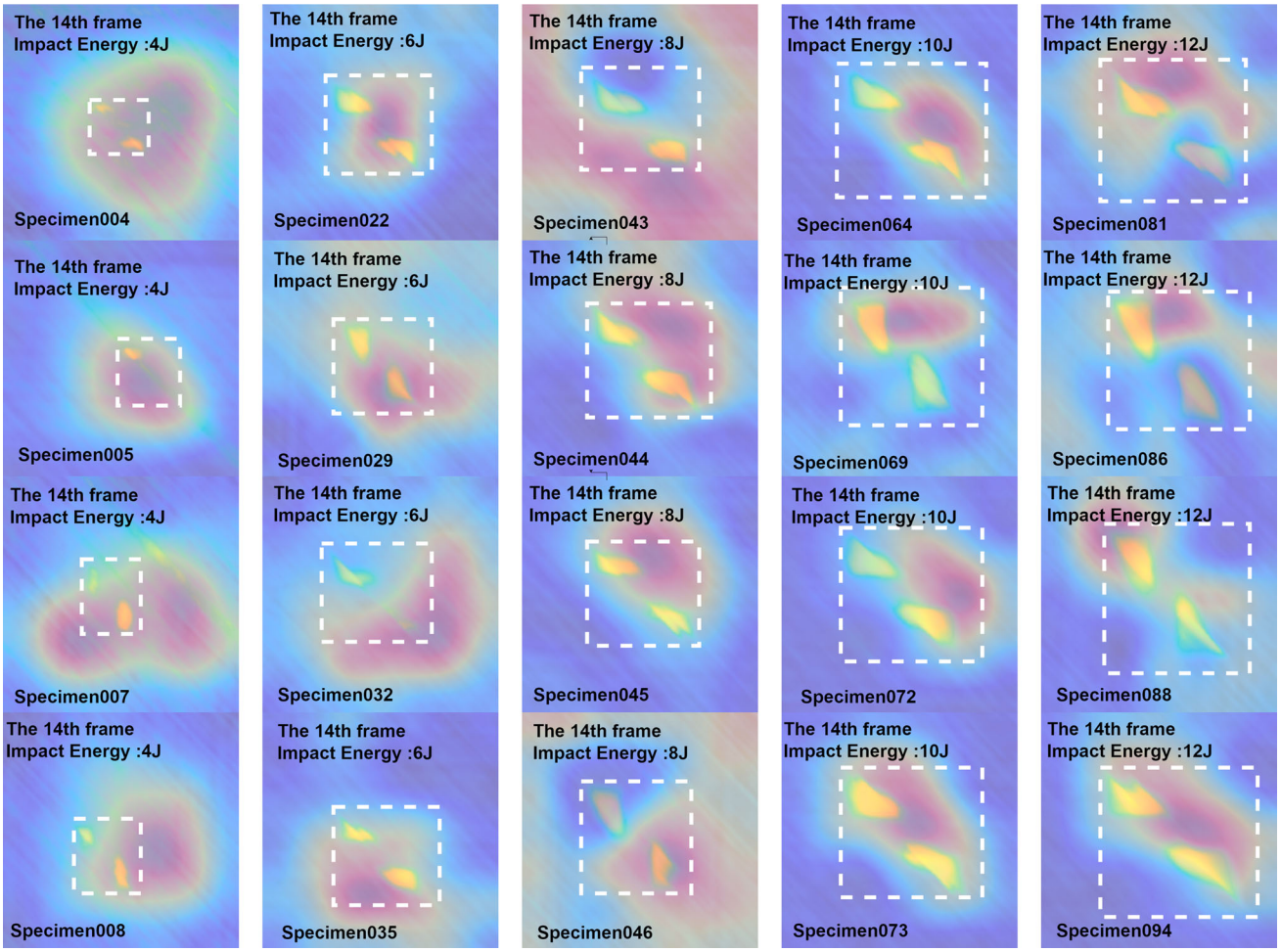

Fig. 12 Class Activation Maps of 20 testing images of different impact energies on Dataset14-33

exist among these images and can distinguish them from any other damage patterns caused by different impact energy levels.

(3) If the number of successive image frames constituting the training dataset is fixed (e.g. 5-frame dataset: Dataset14-18), the window on which the trained model achieves the highest classification accuracy was found at an early period of the inspection process of the specimens. This observation is subject to the depth of the damage.

(4) Within a particular frame window of a thermography inspection, datasets consisting of more successive frames of thermal images can lead to higher classification accuracy.

(5) Damage patterns introduced by 4 J's impact energy can be more easily confused with those created by 6 J's energy, the same is found to be true in the case of 8 J and 10 J samples as well. The possible reason is that slight energy differences between adjacent impact energy levels cause similar damage patterns.

This study potentially inspires the analysis of possible causes of impact damage in composite components by investigating the impact energy level they are subject to. Extended studies can contribute to the predictive maintenance of composite structures in certain areas, especially in the aerospace industry.

Appendix

See Figs. 13,14,15 and Table 1.

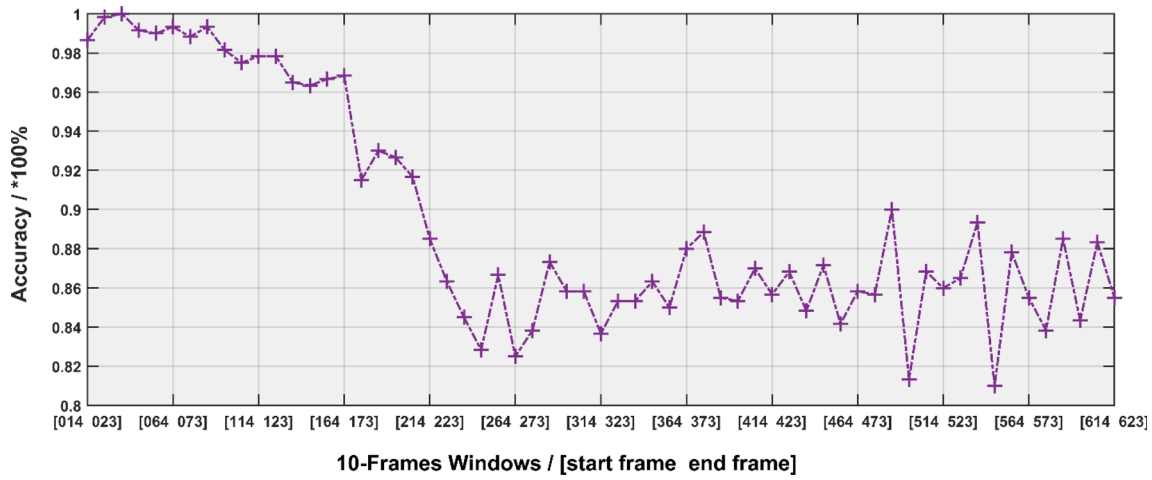

Fig. 13 The classification accuracy shifting curves with temporal-shifting window of 10-frame datasets

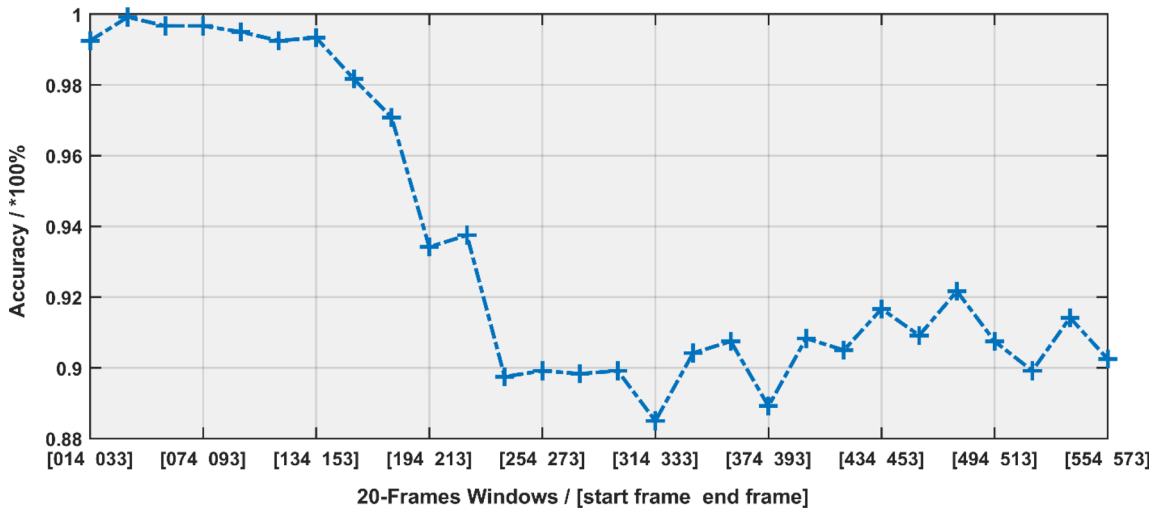

Fig. 14 The classification accuracy shifting curves with temporal-shifting window of 20-frame datasets

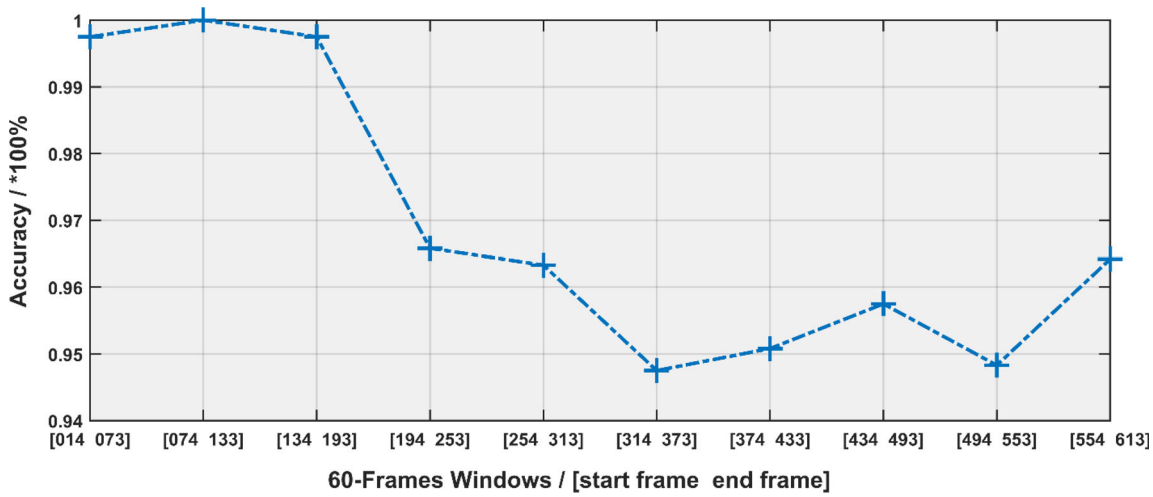

Fig. 15 The classification accuracy shifting curves with temporal-shifting window of 60-frame datasets

Table 1 Material specification of specimens

Material name	Epoxy resin-reinforced SHD composites MTC510 CFRP 300GSM
Stacking sequence	[45°, 0°, − 45°, 90°, 45°]
Size	150 × 100 mm^2 (final size)
Volume fraction	59.94 vol%
Thickness	~ 3 mm (meets the requirements of ASTM D7136)

Authors' contribution Conceptualisation: KD, YZ; Methodology: KD, YZ; Validation: SA; Formal analysis and investigation: KD, YZ; Writing—original draft preparation: KD; Writing—review and editing: YZ, HL, LY; Supervision: YZ.

Data availability Data are available on request.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Pramanik A et al (2017) Joining of carbon fibre reinforced polymer (CFRP) composites and aluminium alloys—A review. Composit Part A: Appl Sci Manuf. https://doi.org/10.1016/j.compositesa.2017.06.007
2. Das S (2001) The cost of automotive polymer composites:a review and assessment of Doe'S light weight materials composites research. Oak Ridge National Laboratory, 2001.
3. Tay AK, Wilson DA, Demirdogen AC, Houghton JR, Wood RL (1990) Microdamage and optical signal analysis of impact-induced fracture in smart structures. In Fiber Opt Smart Struct Skins III. https://doi.org/10.1117/12.24844
4. Sohn MS, Hu XZ, Kim JK, Walker L (2000) Impact damage characterisation of carbon fibre/epoxy composites with multilayer reinforcement. Compos B Eng. https://doi.org/10.1016/S1359-8368(00)00028-7
5. Davies G, Irving P (2015) Impact, post-impact strength and post-impact fatigue behaviour of polymer composites. Polym Compos Aerosp Ind. https://doi.org/10.1016/B978-0-85709-523-7.00009-8
6. Goossens S et al (2021) Practicalities of BVID detection on aerospace-grade CFRP materials with optical fibre sensors. Compos Struct. https://doi.org/10.1016/j.compstruct.2020.113243
7. Chi Hau C, Ultrasonic and Advanced methods for nondestructive testing and material characterization. 5 Toh Tuck Link, Singapore: World Scientific Publishing Co. Pte. Ltd., 2007. https://doi.org/10.1142/9789812770943.
8. Vandendriessche J, Orta AH, Verboven E, van Paepegem W, van den Abeele K, Kersemans M (2022) Probabilistic ultrasound C-scan imaging of barely visible impact damage in CFRP laminates. Compos Struct. https://doi.org/10.1016/j.compstruct.2022.115209
9. Tsoi KA, Brooks C, Rajic N, Antolis C (2021) Lowenergy sonic thermographic inspection of impact damage in aerospace composites. Mater Res Proc. https://doi.org/10.21741/9781644901311-7
10. Jolly M et al (2015) Review of non-destructive testing (NDT) techniques and their applicability to thick walled composites. Procedia CIRP. https://doi.org/10.1016/j.procir.2015.07.043
11. Saeedifar M, Najafabadi MA, Zarouchas D, Toudeshky HH, Jalalvand M (2018) Barely visible impact damage assessment in laminated composites using acoustic emission. Compos B Eng. https://doi.org/10.1016/j.compositesb.2018.07.016
12. Bates D, Smith G, Lu D, Hewitt J (2000) Rapid thermal non-destructive testing of aircraft components. Compos B Eng. https://doi.org/10.1016/S1359-8368(00)00005-6
13. Gholizadeh S (2016) A review of non-destructive testing methods of composite materials. Procedia Struct Integr. https://doi.org/10.1016/j.prostr.2016.02.008
14. Wiecek B (2005) Review on thermal image processing for passive and active thermography. In Annual International Conference of the IEEE Engineering in Medicine and Biology—Proceedings, 2005, vol. 7 https://doi.org/10.1109/iembs.2005.1616506.
15. Pohl J Active thermographic testing of CFRP with ultrasonic and flash light activation' 2016. [Online]. Available: http://creativecommons.org/licenses/by-nd/3.0/
16. Liu H, Du W, Yazdani Nezhad H, Starr A, Zhao Y (2021) A dissection and enhancement technique for combined damage characterisation in composite laminates using laser-line scanning thermography. Compos Struct. https://doi.org/10.1016/j.compstruct.2021.114168
17. Ciampa F, Mahmoodi P, Pinto F, Meo M (2018) Recent advances in active infrared thermography for non-destructive testing of aerospace components. Sensors (Switzerland). https://doi.org/10.3390/s18020609
18. Hu J et al (2020) Enhanced infrared sparse pattern extraction and usage for impact evaluation of basalt-carbon hybrid composites by pulsed thermography. Sensors (Switzerland). https://doi.org/10.3390/s20247159
19. Zhou J, Du W, Yang L, Deng K, Addepalli S, Zhao Y (2021) Pattern recognition of barely visible impact damage in carbon composites using pulsed thermography. IEEE Trans Ind Inform. https://doi.org/10.1109/tii.2021.3134184
20. Marani R, Palumbo D, Renò V, Galietti U, Stella E, D'Orazio T (2018) Modeling and classification of defects in CFRP laminates by thermal non-destructive testing. Compos B Eng. https://doi.org/10.1016/j.compositesb.2017.10.010

21. Oliver GA, Ancelotti AC, Gomes GF (2021) Neural network-based damage identification in composite laminated plates using frequency shifts. Neural Comput Appl. https://doi.org/10.1007/s00521-020-05180-3
22. Tavares A, di Lorenzo E, Peeters B, Coppotelli G, Silvestre N (2021) Damage detection in lightweight structures using artificial intelligence techniques. Exp Tech. https://doi.org/10.1007/s40799-020-00421-5
23. Saeed N, King N, Said Z, Omar MA (2019) Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning. Infrared Phys Technol 102:103048
24. Wei Z, Fernandes H, Herrmann HG, Tarpani JR, Osman A (2021) A deep learning method for the impact damage segmentation of curve-shaped cfrp specimens inspected by infrared thermography. Sensors (Switzerland). https://doi.org/10.3390/s21020395
25. Oliveira BCF, Seibert AA, Borges VK, Albertazzi A, Schmitt RH (2021) Employing a U-net convolutional neural network for segmenting impact damages in optical lock-in thermography images of CFRP plates. Nondestruct Test Eval. https://doi.org/10.1080/10589759.2020.1758099
26. Fotouhi S, Khayatzadeh S, Pui WX, Damghani M, Bodaghi M, Fotouhi M (2021) Detection of barely visible impact damage in polymeric laminated composites using a biomimetic tactile whisker. Polymers (Basel). https://doi.org/10.3390/polym13203587
27. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-December. https://doi.org/10.1109/CVPR.2016.90
28. Shepard SM (2005) Advances in pulsed thermography. Quality. https://doi.org/10.1117/12.421032
29. Milosevic N (2020) Introduction to convolutional neural networks. Berkeley, CA: Apress, 2020. https://doi.org/10.1007/978-1-4842-5648-0.
30. Dhaka VS et al (2021) A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. Sensors. https://doi.org/10.3390/s21144749
31. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 9908 LNCS. https://doi.org/10.1007/978-3-319-46493-0_38
32. Sarwinda D, Paradisa RH, Bustamam A, Anggia P (2021) Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. Procedia Comput Sci. https://doi.org/10.1016/j.procs.2021.01.025
33. Ma L, Shuai R, Ran X, Liu W, Ye C (2020) Combining DC-GAN with ResNet for blood cell image classification. Med Biol Eng Comput. https://doi.org/10.1007/s11517-020-02163-3
34. Zhuang F et al (2021) A comprehensive survey on transfer learning. Proc IEEE. https://doi.org/10.1109/JPROC.2020.3004555
35. Cao J, Su Z, Yu L, Chang D, Li X, Ma Z (2018) Softmax cross entropy loss with unbiased decision boundary for image classification. Proc Chin Autom Congress CAC 2018:2019. https://doi.org/10.1109/CAC.2018.8623242
36. Kingma DP, Ba JL (2015) Adam: a method for stochastic optimisation. In 3rd International conference on learning representations, ICLR 2015—Conference track proceedings, 2015
37. B Zhou, A Khosla, A Lapedriza, A Oliva, A Torralba (2016) Learning deep features for discriminative localisation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition https://doi.org/10.1109/CVPR.2016.319.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

https://dspace.org/

Classification of barely visible impact damage in composite laminates using deep learning and pulsed thermographic inspection

Deng, Kailun; Liu, Haochen; Yang, Lichao; Addepalli, Sri; Zhao, Yifan

2023-05-01
Springer
https://doi.org/10.1007/s00521-023-08293-7
Deng K, Liu H, Yang L, et al., (2023) Classification of barely visible impact damage in composite laminates using deep learning and pulsed thermographic inspection. Neural Computing and Applications, Volume 35, Issue 15, May 2023, pp.11207-11221

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