Composite Structures 337 (2024) 118030

Contents lists available at ScienceDirect

Composite Structures

journal homepage: www.elsevier.com/locate/compstruct

Attention mechanism enhanced spatiotemporal-based deep learning approach for classifying barely visible impact damages in CFRP materials

Kailun Deng [a], Haochen Liu [a,b], Jun Cao [a], Lichao Yang [a], Weixiang Du [a,c], Yigeng Xu [a], Yifan Zhao [a,*]

[a] School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedford MK43 0AL, UK
[b] Research Center for Applied Mechanics, Xidian University, Xi'an 710071, China
[c] Gansu Province Special Equipment Inspection & Testing Research Institute, Lanzhou, Gansu 730030, China

ARTICLE INFO

Keywords:
Machine Learning
Attention mechanism
Barely visible impact damage (BVID)
Carbon fibre reinforced polymer (CFRP)
Pulsed thermography (PT)

ABSTRACT

Most existing machine learning approaches for analysing thermograms mainly focus on either thermal images or pixel-wise temporal profiles of specimens. To fully leverage useful information in thermograms, this article presents a novel spatiotemporal-based deep learning model incorporating an attention mechanism. Using captured thermal image sequences, the model aims to better characterise barely visible impact damages (BVID) in composite materials caused by different impact energy levels. This model establishes the relationship between patterns of BVID in thermography and their corresponding impact energy levels by learning from spatial and temporal information simultaneously. Validation of the model using 100 composite specimens subjected to five different low-velocity impact forces demonstrates its superior performance with a classification accuracy of over 95%. The proposed approach can contribute to Structural Health Monitoring (SHM) community by enabling cause analysis of impact incidents based on predicting the potential impact energy levels. This enables more targeted predictive maintenance, which is especially significant in the aviation industry, where any impact incidents can have catastrophic consequences.

1. Introduction

Composite materials have gained tremendous popularity in the modern industry over the past few decades due to their numerous advantages, including a high strength-to-weight ratio, excellent corrosion resistance to environmental factors, and durability [1]. Despite their remarkable toughness, composite materials tend to be brittle when subjected to impact forces [2], where low-velocity impacts catch the most attention. They can cause barely visible impact damages (BVID) in composite materials [3] that cannot be easily detected by typical visual inspection but can be disastrous to aircraft performance and threaten the safety of complex systems, such as aircraft [4]. Under this premise, Non-Destructive Testing (NDT) techniques are employed to inspect composite structures for BVID. In the case of composite materials, the most commonly used NDT techniques include ultrasonic testing, eddy current testing, radiography and thermographic inspection [5]. Ultrasonic testing utilises high-frequency sound waves to detect subsurface defects in materials [6]. The sound waves are transmitted into materials and reflected from defects, giving information about the location or geometric data of the defects. Eddy current testing can detect surface and near-surface defects in composite materials [7]. It involves inducing an electric current in materials and measuring the resultant magnetic field. Some composite materials include conductive fibres (like carbon fibres) or elements. Eddy current testing can be effective in these cases, as the eddy currents interact with these conductive components, allowing for detecting defects or changes in the conductive paths. It can detect delamination, voids, or other defects that affect composites' electrical conductivity. Radiography passes a controlled amount of X-ray radiation through materials and captures the resulting image on a film or digital sensor [8]. This technique is based on the differential absorption of X-rays by different internal conditions within the material. Defects such as voids, cracks, or delamination in composites present a different density than the intact material, leading to a contrast in the X-ray image. The typical radiography used in the composite inspection is X-ray examination and Computed Tomography (CT). Thermographic inspection involves using external heat sources and infrared cameras to detect

* Corresponding author.
E-mail addresses: K.Deng@cranfield.ac.uk (K. Deng), liuhaochen@xidian.edu.cn (H. Liu), j.cao.3@bham.ac.uk (J. Cao), Lichao.Yang@cranfield.ac.uk (L. Yang), Yigeng.Xu@cranfield.ac.uk (Y. Xu), yifan.zhao@cranfield.ac.uk (Y. Zhao).

https://doi.org/10.1016/j.compstruct.2024.118030
Received 20 November 2023; Received in revised form 20 February 2024; Accepted 11 March 2024
Available online 14 March 2024

temperature evolution on the surface of materials, which can indicate internal defects [9]. This research adopts Pulsed thermography (PT), a popular thermographic inspection technique, due to its rapid inspection process and relatively simple deployment [10]. Pulsed thermography uses optical excitation to heat the sample, by which subsurface anomalies in the material can be detected with the help of an infrared camera. [10].

While the cooperation of NDT and corresponding post-processing techniques such as Thermographic Signal Reconstruction (TSR), wavelet transform, and Principal Component Analysis (PCA) have proven effective in detecting subsurface anomalies in composites [11,12], one major drawback limits its widespread use in real-world applications [13]. The data post-processing intrinsically requires a high level of domain knowledge, and the generation of final inspection results heavily depends on the operators' expertise. Artificial Intelligence (AI) has been introduced in Thermographic NDT (TNDT) to mitigate the influence of human factors and enhance the reliability of decision-making. Oliveira et al. [14] studied impact damage segmentation in carbon fibre-reinforced polymer (CFRP) specimens using a U-Net model that outperformed four traditional image processing methods in recognising damage profiles. Liu et al. [15] proposed a defect profile segmentation method using a Mask Region-based Convolution Neural Network (Mask-RCNN). Simulation data was used to train the network, and the proposed model achieved an intersection over union (IoU) metric of more than 90 %. Bang et al. [16] utilised the Faster-RCNN model to detect defects in thermal images, where the region proposal method was utilised to analyse thermal patterns associated with defects. Wei et al. [17] introduced another deep learning-based impact damage segmentation method for CFRP specimens with curved surfaces using PT. The obtained thermal images were processed by Principal Component Thermography (PCT) and Empirical Orthogonal Functions (EOF) before being fed to a U-Net model that achieved F1 scores exceeding 87 %. Deng et al. [18] successfully characterised Barely Visible Impact Damage (BVID) defects by employing multi-frame thermal imaging derived from pulsed thermography of BVID samples, in conjunction with a ResNet50 model.

Unlike the abovementioned thermal image-based machine learning applications that primarily process spatial information from TNDT, other research focuses on analysing temporal information. These studies leverage machine learning models' superior pattern recognition capability to distinguish variations in the temperature degradation process between defective and sound areas or different defective areas with varying damage depths [19]. Marani et al. [20] extracted three input features from the temperature decay curve of the specimen surface, which were then fed into multiple machine learning-based classifiers for defect detection. Among these classifiers, a decision forest comprising 30 trees achieved the highest standard accuracy of 99.47 % and a balanced accuracy of 86.9 %. In another study, Cao et al. [21] introduced a novel two-stream CNN model for defect localisation in CFRP samples using data from Lock-in thermography. A pair of 1D temperature decay signals from defective or sound areas were processed separately in a two-stream CNN feature extractor. The extracted features were then fused to calculate a similarity index that leads to the final defect detection. Fang and Maldague [22] proposed a defect depth measuring technique employing a Gated Recurrent Unit (GRU) network, where simulated flat bottom holes defects with varying depths embedded in CFRP materials were used. The differences in temperature decay patterns for defects with different depths were employed to predict the defect depth.

Apart from studies either analysing image data or temporal information only, some studies incorporate both temporal and spatial information. Hu et al. [23] extracted temporal information from the obtained thermal sequence and compressed the sequence in the time domain into a single image. The synthetic image was then fed into a Faster-RCNN model for crack detection, demonstrating promising on natural cracks. Saeed et al. [24] proposed a hybrid model that simultaneously attained defects detection in CFRP samples by using pre-trained Cuda-Convnet and Alexnet models. Additionally, they utilised a Deep Feed Forward Neural Network (DFF-NN) algorithm to predict defect depth using time series information from thermograms. Another approach, proposed by Luo et al. [25], introduced a two-stream deep learning architecture for automatic thermography defects detection. The network can parallelly process spatial information in a single thermal image using a CNN module and temperature evolution data for each pixel in the time domain via a multi-Long Short-Term Memory (LSTM) network. The outcomes from both streams were then incorporated to make the final detection decision. Some customised 3D deep learning networks have also been proposed in the TNDT area. Hu et al. [26] proposed a simplified 3D CNN model integrating spatial information from every thermal image and temporal information of each pixel across the frame sequence, resulting in more robust and accurate results than several state-of-the-art defect segmentation algorithms. Dong et al. [27] presented a 3D CNN model consisting of spatial and temporal convolutional filters for defect detection and depth measurement tasks. It used thermal image sequences obtained from Lock-in thermography for CFRP-based specimens embedded with synthetic defects of varying shapes and depths. The results demonstrated that the proposed model outperformed other 2D CNN-based models applied in Lock-in thermography.

In summary, machine learning applications in the field of TNDT can be categorised into three groups according to data types: (a) spatial information-based models that are primarily based on 2D CNN models; (b) temporal data-driven models that employed 1D or 2D CNN models or RNN-based models such as GRU and LSTM; and (c) hybrid models that integrate both spatial and temporal information, often combining CNN and RNN-based models or using simplified 3D CNNs. It is evident that hybrid models represent the future direction, particularly considering that most TNDT techniques generate thermograms that inherently contain both spatial and temporal information. To the best of our knowledge, no spatiotemporal-based deep learning model has been developed and used to analyse impact damages in CFRP-based materials.

This research proposes a novel deep-learning analytic framework incorporating an attention module to study the relationship between impact damages and the corresponding impact energies exerted on the specimen. Compared with existing studies, this study's key innovation lies in utilising thermal image sequence (3D) as analysis objects, enabled by an end-end spatiotemporal deep learning network. Another novelty is the introduction of the attention mechanism, which is utilised for the first time to enhance the deep learning analysis of BVID. It allows the model to focus on the input data's most relevant features or regions rather than trying to learn all features equally [28]. The main contributions of this study are:

1. A spatiotemporal-based and attention mechanism enhanced deep learning network for investigating the relationship between thermographic patterns of BVID in CFRP materials and the corresponding impact energy levels.
2. A unique thermal sequence dataset collected from 100 meticulously crafted CFRP specimens subjected to 5 impact energies, which have a much larger number of samples than any other experiment-based datasets used in related fields.
3. A new data augmentation method for thermal image sequence data, employing downsampling with fixed frame intervals and sequence sliding within a small range.

2. Methodology

Fig. 1 depicts the methodology's complete workflow, comprising four modules: data acquisition, data generation, network construction, and performance evaluation. One hundred raw thermal image sequences, each consisting of 613 frames of thermal images with a resolution of 640 × 512 pixels, were obtained by conducting pulsed

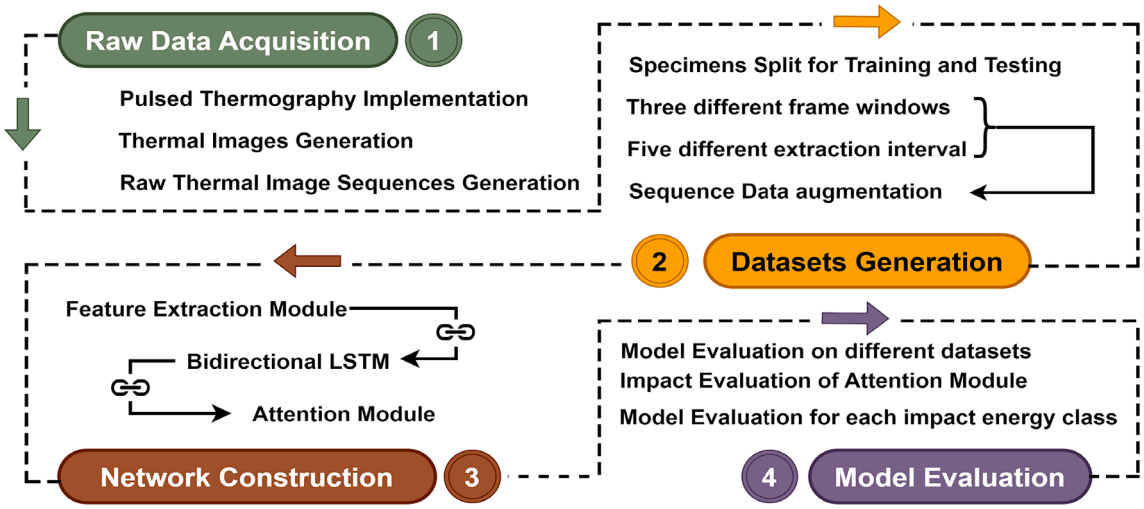

Fig. 1. The flow chart of the proposed methodology for this study.

thermography on 100 CFRP-based specimens. Subsequently, a down-sampling technique with fixed frame intervals was applied to the raw sequence data to create simplified sequences with significantly fewer frames. In order to expand the dataset for model training, a sliding-sequence data augmentation method was employed. Finally, an attention module-enhanced bidirectional-LSTM network was introduced to enable impact energy-based damage classification and discussion.

2.1. Raw data acquisition

Four carbon fibre-reinforced polymer (CFRP) plates were manufactured, all sharing the same material specifications and dimensions of 750 × 500 × 3 mm, serving as the base for creating impact damage specimens. Each plate was then divided into 25 smaller plates of 150 × 100 × 3 mm. This process yielded 100 CFRP specimens, which were further grouped into five sets, each containing 20 plates. The material specifications of the manufactured specimens can be found in Table A.1 in the Appendix.

Five different impact energies were employed for each sample set to conduct the low-velocity impact experiments: 4 J, 6 J, 8 J, 10 J, and 12 J. The low-velocity impact was induced by freely dropping a hemispherical indenter with a precise weight of 2.281 kg from a predetermined height. By adjusting the drop height, the desired impact energies are achieved.

Next, pulsed thermography was conducted on the 100 generated impact damage specimens. Fig. 2 (a) displays how the pulsed thermography system works. The process involves using two synchronised flash lamps to generate a heat pulse that is applied to the specimen under testing. This heat pulse causes a rapid temperature increase in the specimen. The impact damage alters the thermal transfer pattern within the material, resulting in uneven temperature distribution on the specimen's surface. An infrared camera captures these temperature variations, allowing for the detection of compromised areas in the specimen.

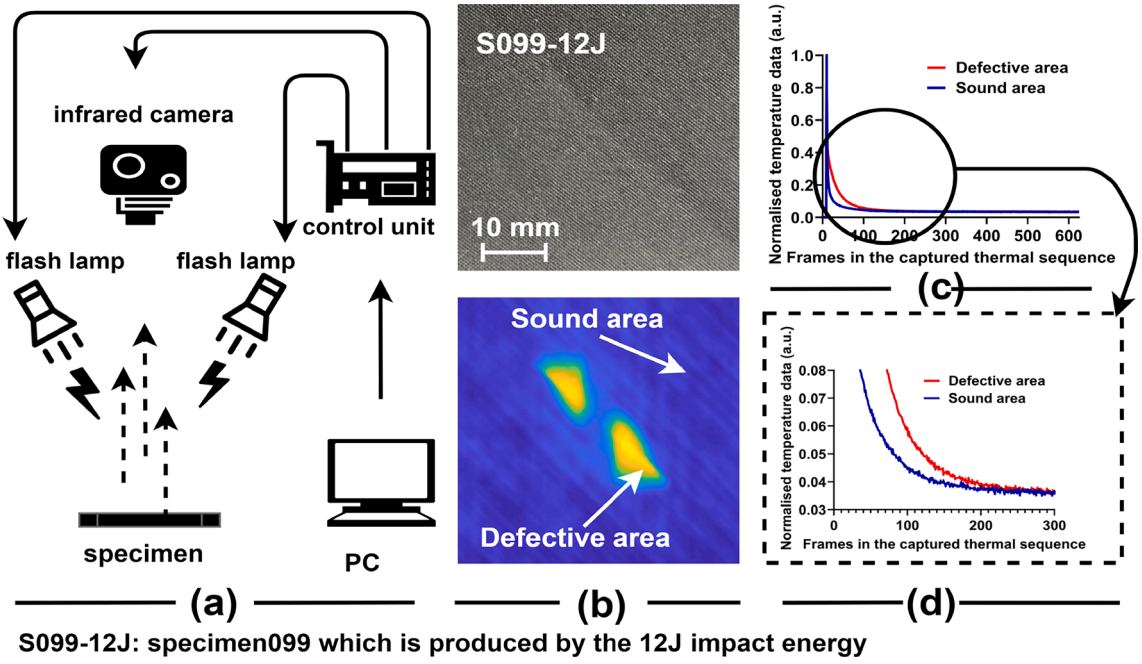

Fig. 2. (a) The experiment setup of Pulsed Thermography. (b) Photos of a specimen (S099) and the corresponding thermal image (the 15th frame after the flash) captured by Pulsed Thermography. (c) The temperature decay curves representing the temperature evolution process of the defective and non-defective areas on the specimen surface. (d) Highlight of the temperature decay curves within the first 300 frames range of the captured thermal image sequence.

The Thermoscope II pulsed-active thermography system was used in this study. Two capacitor banks-powered Xenon flash lamps generated heat pulses with around 2KJ of energy per flash. An infrared camera from the FLIR SC7000 series with a spatial resolution of 640 x 512 pixels was employed. The camera captured thermal images at a frequency of 50 Hz for 20 s. The flash time was set at the 10th frame after the capture started. The thermal image sequence consisted of 1000 frames, but only 13 to 625 were considered. The whole temperature degradation process is included within this range based on observation of the recorded temperature data. Each thermal image was cropped to a 150 x 150 region (pixel) containing the whole impact damage pattern (see Fig. 2(b)). This produced a thermal image sequence with a 150 x 150 x 613 dimension for each specimen.

2.2. Thermal sequence datasets generation

The original thermal sequence for each specimen comprises 613 frames. As shown in Fig. 2(c)-(d), the most noticeable temperature changes and damage manifestation occur within the initial 200 frames. Subtle and gradual temperature changes were observed between frames 200 and 300. Beyond the 300th frame, no significant temperature variation was observed visually. However, upon analysing the temperature data from the infrared camera, a gradual and slight temperature drop was discovered between frames 300 and 625. To assess the model's performance across various time windows, three sub-sequences were cropped from the original sequence, starting at the 13th frame and spanning 200 frames (the 13th-212th frame-span, referred to as SP-200), 300 frames (the 13th-312th frame-span, referred to as SP-300), and 600 frames (the 13th-612th frame-span, referred to as SP-600), respectively.

The thermal pattern changes in the sequence occur continuously without sudden shifts. On this basis, extracting a smaller number of frames uniformly from the original sequence is hypothesised to represent the whole original sequence effectively. Based on this hypothesis, downsampling was applied to SP-200, SP-300, and SP-600 using fixed frame intervals. This downsampling approach significantly reduces the data size, thereby reducing computational requirements during model training. To examine the impact of different downsampling intervals on model performance, five intervals of 5, 10, 20, 30, and 60 frames (referred to as INT-5, INT-10, INT-20, INT-30, and INT-60, respectively) were tested in this study.

This study generated fifteen datasets on the three different SP-200, SP-300, and SP-600 frame spans, with five different downsampling intervals of INT-5, INT-10, INT-20, INT-30 and INT-60. The dataset generated on SP-200 and with INT-5 is labelled as Dataset-200–5. The other datasets used in this study are named accordingly. As an example, Fig. 3 illustrates the generation process of Dataset-200–5. The new sequence begins with the 13th frame, followed by frames extracted every five frames until the 208th frame (The extracted frames are limited on SP-200, which ranges from the 13th to the 212th frame). This procedure generates 100 simplified image sequences from the 100 specimens. However, using deep learning approaches, 100 sequences are insufficient for a 5-class classification task. This study introduces a sliding-sequence extraction method to expand the dataset. This method involves shifting the first simplified sequence by one frame along the original sequence, extracting images from these new positions, and constructing a new sequence. This operation is repeated on the newly generated sequence data to create the next new sequence. To ensure all the restructured sequences cover almost the whole range of the original sequence, the maximum sliding step was set to 10. Consequently, ten reconstructed sequence data were obtained from each specimen, enlarging the data number from 100 to 1000.

2.3. Network construction

Previous studies have demonstrated that thermal images of impact damages carry discriminative information associated with the specific impact energy level experienced by the specimen [29]. Furthermore, it has been observed and proved that defective areas exhibit distinct temperature decay curves compared to non-defective areas (refer to Fig. 2(c)(d)). Consequently, a hybrid model incorporating spatial and temporal information is proposed to capture these characteristics effectively.

The proposed neural network starts with a CNN-based feature extractor that captures high-dimensional spatial features from individual images within the sequence data. It also includes a time series model that handles the temporal information of the sequence. Additionally, a temporal-based attention module is incorporated to prioritise important frames in the sequence data. The network architecture is depicted in Fig. 4.

2.3.1. Backbone model proposal for analysing image sequence

Building upon the advancements in deep learning applications for text classification and sentiment analysis in the field of Natural Language Processing (NLP), we propose to use a Bidirectional LSTM (BiLSTM) network as the core model for this research. Similar to the way that LSTM processes language or text, we treat the image sequence as a sequence of inputs, where each image corresponds to a specific timestep in the model. By adopting this approach, the proposed BiLSTM model can effectively capture both the spatial information from the thermal images and the temporal information presented in the sequence. To ensure compatibility with the BiLSTM model, the images in the sequence undergo feature extraction and transformation before being fed into the network.

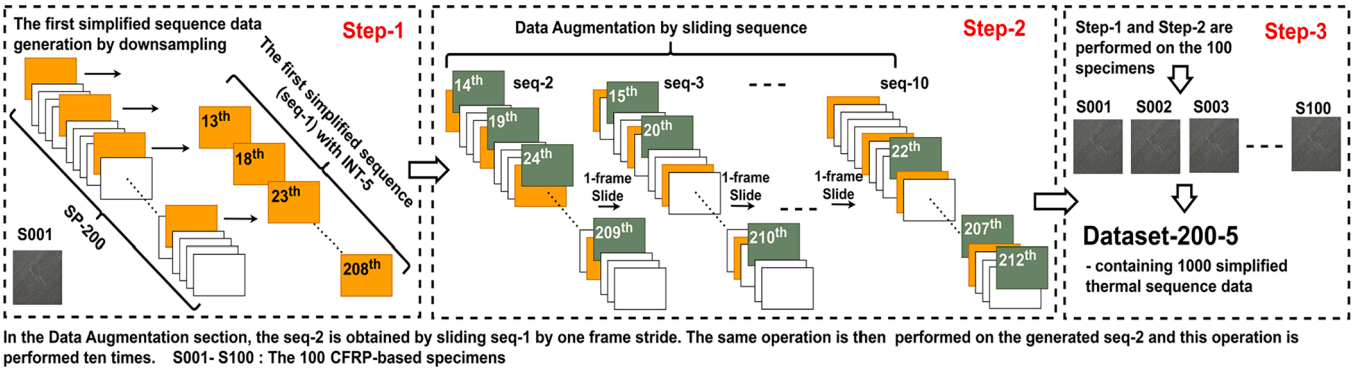

Fig. 3. The generation process of Dataset-200–5. On each specimen, the first simplified sequence data is created by a downsampling with the interval of 5 frames on the original thermal sequence span of SP-200. Then a sliding-sequence data augmentation operation is performed based on the first generated sequenced data, and 10 simplified sequences are obtained for each specimen. The abovementioned process is performed on the 100 specimens and a total of 1000 sequence data are gained to form this dataset.

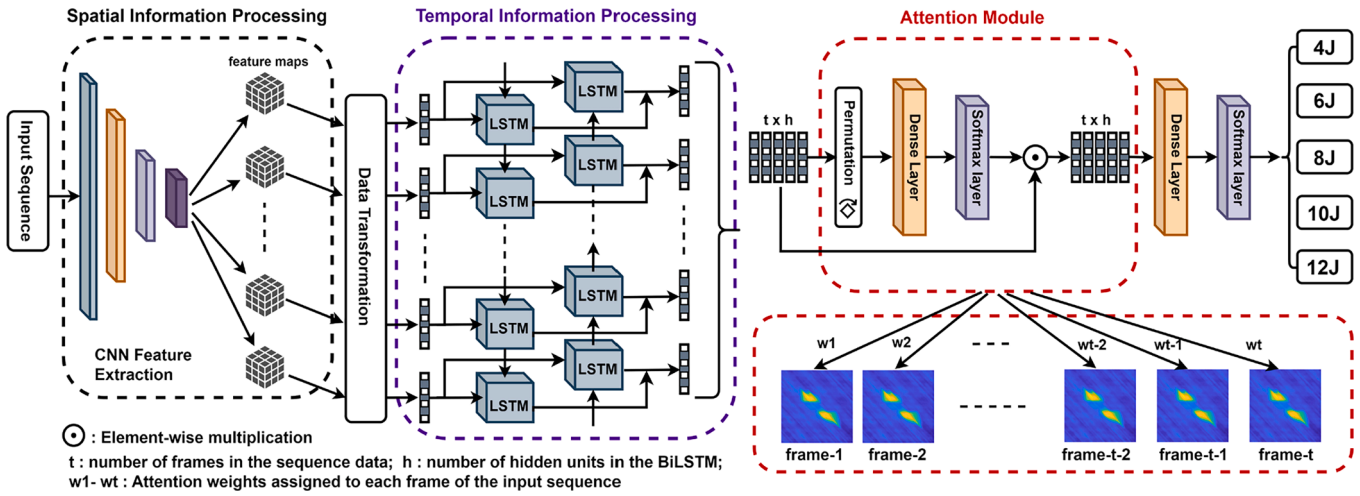

Fig. 4. The illustration of the proposed CNN + BiLSTM + Attention deep learning network. At the spatial information processing stage, each image in the input sequence goes through CNN layers for feature extraction. The extracted features are fed into a BiLSTM network to analyse the temporal correlations of different frames in the sequence data at the following stage. An temporal-based attention module is located after the BiLSTM and additional attention weights (w1, w2, …, wt) are assigned to each frame of the sequence data. The weighted output of the BiLSTM network is fully connected through a dense layer, and the final classification results are obtained after a softmax layer at the end of the proposed network.

2.3.2. *Spatial feature extraction and data transformation*

A CNN-based feature extractor initially processes The image sequence data, utilising a pre-trained Visual Geometry Group network (VGG-19) trained on ImageNet. Each thermal image in the sequence is transformed into a 7x7 feature map with 512 channels, denoted by $\boldsymbol{f}_i \in \mathbb{R}^{512\times7\times7}$. The feature extraction output of the whole sequence is marked as $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_t]$, where $\mathrm{F} \in \mathbb{R}^{t\times512\times7\times7}$ and t refers to the number of frames in the sequence data that is also the number of timesteps in the following BiLSTM network.

$\boldsymbol{f}_i$ is then transformed to $\boldsymbol{f}_i \in \mathbb{R}^{1\times25088}$ by rearranging its elements from a three-dimensional matrix to a one-dimensional sequence ($512 \times 7 \times 7 \rightarrow 1 \times 25088$). After this operation, $\mathrm{F} \in \mathbb{R}^{t\times25088}$ can be used as the input data of the following BiLSTM module.

2.3.3. *Attention mechanism utilisation in the time domain*

We introduce a temporal-based attention module to enhance the capability of the BiLSTM network by assigning trainable weights to each timestep. This attention mechanism enables the model to emphasise specific frames within the sequence that significantly impact determining the data category.

The BiLSTM network has 128 hidden units, resulting in an output dimension of 128 for each timestep, referring to $\mathbf{A} \in \mathbb{R}^{t\times128}$. The function of the subsequent attention module can be represented using the following formulas.

First, the transposed matrix of $\mathbf{A}$ is obtained:

$$\mathbf{A}^T = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_{128}] \in \mathbb{R}^{128\times t}, \boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_{128} \in \mathbb{R}^{1\times t} \tag{1}$$

A full connection is performed on the last dimension of $\mathbf{A}^T$ and can be written as Eq. (2), where the output dimension of this operation is set to t.

$$\boldsymbol{b}_i = \boldsymbol{a}_i \mathbf{w}_{\mathbf{DS}} + \mathbf{b}_{\mathbf{DS}}, \boldsymbol{b}_i \in \mathbb{R}^{1\times t}, i \in [1, 128], \mathbf{w}_{\mathbf{DS}} \in \mathbb{R}^{t\times t}, \mathbf{b}_{\mathbf{DS}} \in \mathbb{R}^{1\times t}, \boldsymbol{b}_i = [b_i^1, b_i^2, ..., b_i^t] \tag{2}$$

where $\mathbf{w}_{\mathbf{DS}}$ represents the weight matrix of the dense layer in the proposed attention module, and $\mathbf{b}_{\mathbf{DS}}$ represents the corresponding bias term.

A SoftMax activation function is then performed on $\boldsymbol{b}_i$:

$$c_i^j = \boldsymbol{\sigma}(b_i^j) = \frac{b_i^j}{\sum_{j=1}^{t} b_i^j};\ j \in [1, t],\ c_i^j \in [0, 1],\ \boldsymbol{c}_i = [c_i^1, c_i^2, \cdots, c_i^t] \in \mathbb{R}^{1\times t};\ i \in [1, 128],$$

$$\mathbf{C} = [\boldsymbol{c}_1, \boldsymbol{c}_2, ..., \boldsymbol{c}_{128}] \in \mathbb{R}^{128\times t}, \boldsymbol{c}_1, \boldsymbol{c}_2, ..., \boldsymbol{c}_{128} \in \mathbb{R}^{1\times t} \tag{3}$$

where $\boldsymbol{\sigma}$ represents the SoftMax activation. The attention weights matrix $\mathbf{D}$ is acquired by a transposition of $\mathbf{C}$:

$$\mathbf{D} = \mathbf{C}^T \in \mathbb{R}^{t\times128} \tag{4}$$

The attention mechanism-optimised output of the BiLSTM network $\widetilde{\mathbf{A}}$ is generated by an element-wise multiplication of the original output of the BiLSTM network, $\mathbf{A}$ and the obtained attention weights matrix, $\mathbf{D}$ in Eq. (4):

$$\widetilde{\boldsymbol{A}} = \boldsymbol{A} \odot \boldsymbol{D}, \widetilde{\boldsymbol{A}} \in \mathbb{R}^{t\times128} \tag{5}$$

where $\odot$ represents the element-wise multiplication. The generated $\widetilde{\boldsymbol{A}}$ is then passed to the subsequent dropout and dense layers for the final classification.

The attention module influences the final model prediction results by assigning additional weights $\mathbf{D}$ to the output of each timestep in the BiLSTM network.

2.3.4. *Network configuration and implementation*

The model was implemented on TensorFlow 2.10.0 backend. To train the model, we employed the Categorical-crossentropy loss function, which is commonly used in multi-class classification tasks. We chose RMSprop as the optimiser with an initial learning rate of 0.001. The dataset was split into 80 % for training and 20 % for testing. It should be noted that the testing data consisted of specimens that were not included in the training process, ensuring the rigorous evaluation of the model's ability to generalise to unseen samples.

The training was performed using a batch size of 32 and a maximum of 60 epochs per training round. Model testing was conducted at the end of each epoch, and the best test accuracy was recorded. The training and testing processes were repeated six times on each dataset, with a random allocation of specimens for training and testing in each repetition.

3. Results and discussion

The performance of the proposed model was evaluated on the fifteen datasets with simplified thermal image sequences. The sensitivity of selecting frame spans and downsampling intervals for generating simplified sequence data for model training was analysed. Meanwhile, to evaluate the impact of the attention module on model performance, both models with and without the attention module were constructed and compared on each dataset. For ease of reference, the model equipped with the attention module is referred to as AT-model, and the model without the attention module is referred to as NAT-model.

Since all datasets used in the study are balanced across the five impact energy categories, accuracy is selected as the evaluation metric for the model performance. The classification accuracies achieved by the AT-model and NAT-model on each dataset are presented in Fig. 5.

3.1. Model Validation

The results depicted in Fig. 5(a)–(e) demonstrate that the AT-model consistently outperformed the NAT-model in terms of classification accuracy across all fifteen datasets, which suggests the superior performance of the introduced attention module. Additionally, we observed that the attention module accelerated the model's fitting process. Fig. 5 (f) presents the training accuracy curves of both the AT-model and NAT-model on Dataset-200–10. The curves show that the AT-model achieves convergence around epoch 7, whereas the NAT-model reaches convergence around epoch 12. This observation suggests that the attention module effectively reduces the model's convergence time. This pattern has been consistently observed across the other fourteen datasets.

The NAT-model achieved test accuracies ranging from 0.75 to 0.9 across all datasets, as shown in Fig. 5. Considering the model's task of performing a 5-class multi-classification, this level of accuracy serves as strong evidence that the proposed CNN extractor + BiLSTM network structure can effectively process thermal image sequence data and extract discerning features from them. Furthermore, the validity of the NAT-model supports the hypothesis proposed in the methodology section that a reduced number of uniformly extracted frames can effectively approximate the original sequence in our proposed analytical method.

To validate the superiority of the proposed spatiotemporal model over the existing thermal image-based spatial model, ResNet-50 was utilised to perform the same classification task using the thermal images from SP-200, SP-300, and SP-600. The performance comparison between the proposed spatiotemporal models (NAT and AT model) and state-of-the-art is shown in Table 1. The results demonstrate that the proposed spatiotemporal model outperforms the spatial model across all frame spans (SP-200, SP-300, and SP-600). No comparison was made between the proposed model and a temporal-based model as the classification task of impact damages cannot be effectively performed using solely temporal information.

Performance comparisons were also made between the proposed

Table 1
Classification accuracy comparison between the proposed method and state-of-the-art.

Frame span	SP-200	SP-300	SP-600
Model			
ResNet-50 (spatial-based)	0.761	0.728	0.753
3D ResNet-18 (spatiotemporal-based)	0.780	0.765	0.710
3D ResNet-50 (spatiotemporal-based)	0.730	0.705	0.680
NAT-model (spatiotemporal-based)	0.862	0.883	0.894
Our AT model (spatiotemporal-based)	**0.986**	**0.978**	**0.956**

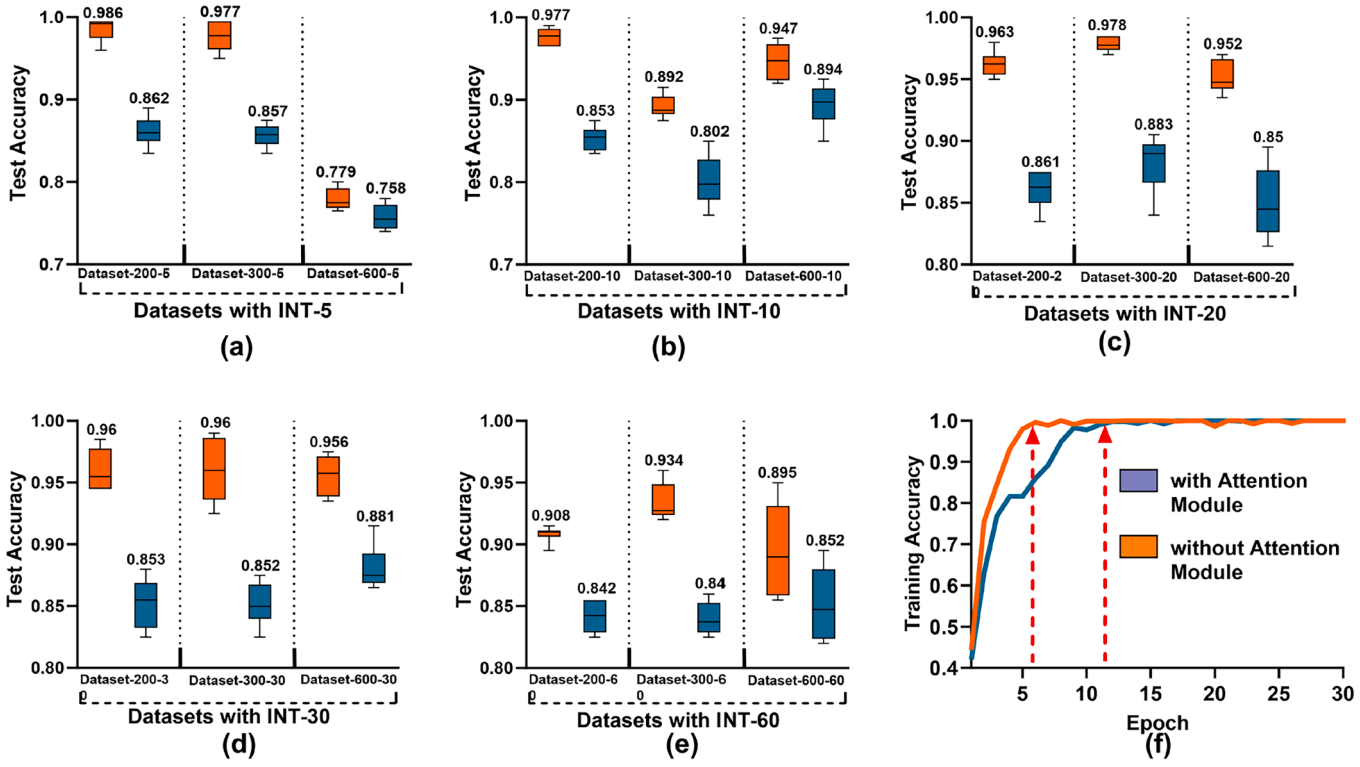

Fig. 5. (a)–(e) Classification accuracies achieved by the AT-model and the NAT-model on the generated fifteen datasets on different frame spans of SP-200, 300 or 600 and with different downsampling intervals of INT-5, INT-10, INT-20, INT-30 or INT-60. (f) The training accuracy curves of both the AT-model and the NAT-model on Dataset-200-5.

spatiotemporal model and 3D ResNet with 18 and 50 layers (shown in Table 1). The highest classification accuracy of 0.78 was achieved by 3D ResNet-18 (18 layers) on SP-200, with a downsampling interval of 5. It is evident that 3D ResNet is not as capable as the proposed model in this case. The possible reason is that 3D ResNet considers more detailed spatiotemporal correlations (at kernel size level), which may result in overfitting and lead to performance degradation.

3.2. Model performance comparison on datasets with different frame spans and downsampling intervals

Fig. 6(a)-(c) depicts the classification accuracies obtained by the AT-model across the fifteen generated datasets. The highest accuracy of 0.9858 was achieved on Dataset-200–5. Additionally, a consistent decline in accuracy was observed for the 200-frame span as the downsampling intervals increased from 5 to 60 frames. This finding can be attributed to the temperature decay process. The rapid temperature changes occur within the 200-frame span, as evident from Fig. 2(c)-(d). Drastic temperature variations can occur within a few frames during this timeframe. Increasing the downsampling interval results in sparser and fewer thermal images in a single sequence, reducing the information content. Considering the rapid temperature changes during this period, it is highly likely that valuable information relevant to model classification will be discarded. Consequently, the model's performance deteriorates with an increased downsampling interval within the 200-frame span.

The highest classification accuracy of 0.9858 was achieved on Dataset-200–5, and the lowest accuracy of 0.779 was obtained on Dataset-600–5. Compared to Dataset-200–5, Dataset-600–5 includes additional frames ranging from 312 to 625 of the original thermal sequences. However, the accuracies obtained from Dataset-600–10 (0.947), Dataset-600–20 (0.952), Dataset-600–30 (0.956), and Dataset-600–60 (0.895), which also include additional images from the 312th frame to the 625th frame, are significantly higher than that of Dataset-600–5. This difference can be attributed to the 5-frames downsampling interval, which introduces more thermal images from the original sequences span of the 312th to 625th frame than those datasets with INT-10, 20, 30 and 60. The denser information from the original frame span of 312 to 625 leads to a detrimental effect on the model's performance.

A notable observation from Fig. 6(b) is the unexpected decrease in accuracy observed on Dataset-300–10. Comparing the accuracy on Dataset-300–10 with datasets that have the same downsampling interval of INT-5 but smaller or larger frame span (SP-200 or SP-600), a significant decline of accuracy from 0.9767 on Dataset-200–10, and 0.947 on Dataset-600–10 to 0.8917 on Dataset-300–10 has been observed. A similar pattern of accuracy decline has also been observed when comparing Dataset-300–10 with datasets that have the same frame span of SP-300 but smaller or larger downsampling intervals of INT-5 or INT-20.

To explore the underlying cause of this observation, the confusion matrices in Fig. 6(d)(e) were utilised to gain detailed insights into the classification results of Dataset-300–10 and the other four datasets mentioned above. A distinct feature was found in the confusion matrix of Dataset-300–10 by comparing the five confusion matrices. Approximately half (107 out of 240) of the impact damages with the impact energy of 10 J were mispredicted as belonging to the 8 J impact energy group, which results in the relatively low classification accuracy obtained on Dataset-300–10. A tentative conclusion is drawn from this finding: On the INT-10 basis, the simplified thermal sequence data of impact damages caused by the 10 J impact energy exhibit very similar feature patterns to the sequence data from damages with 8 J. These similar features heavily influence the model's decision-making.

3.3. Classification performance in each energy category

To fully reveal the model's performance, confusion matrices were adopted to visualise the model's predictions on the five different impact energy levels. Fig. 7 illustrates the confusion matrices of the AT-model and NAT model's classification on the five different datasets on the 200-frame-span with extraction frame intervals of 5, 10, 20, 30 and 60, respectively. A bar chart showing the corresponding statistics of different adjacent confusion types is under each matrix.

A consistent pattern emerges from the overall analysis of the confusion matrices and bar charts in Fig. 7. Most mispredictions occur

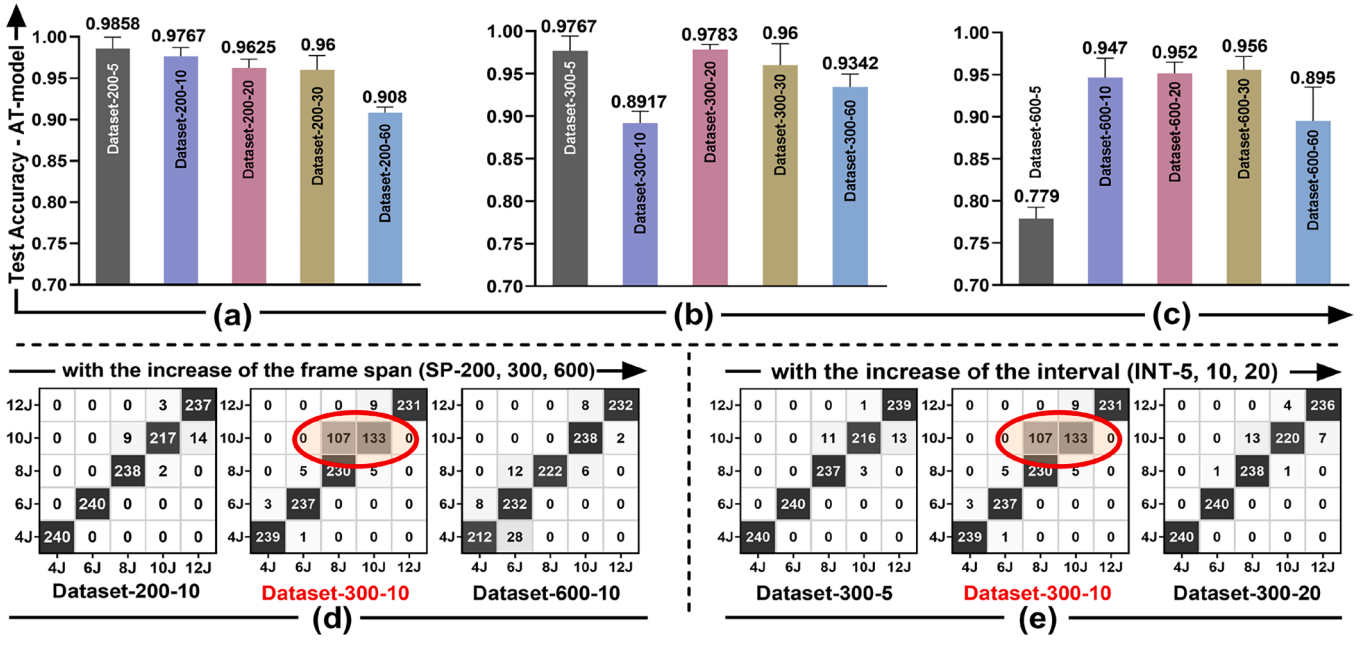

Fig. 6. (a)–(c) Classification accuracy of the AT-model on the fifteen datasets. (d) The confusion matrices comparison on the three different datasets with the same downsampling interval of INT-10 and different frame span of SP-200, SP-300 and SP-600. (e) The confusion matrices comparison on the three different datasets with the same frame span of SP-300 and different downsampling interval of INT-5, INT-10 and INT-20.

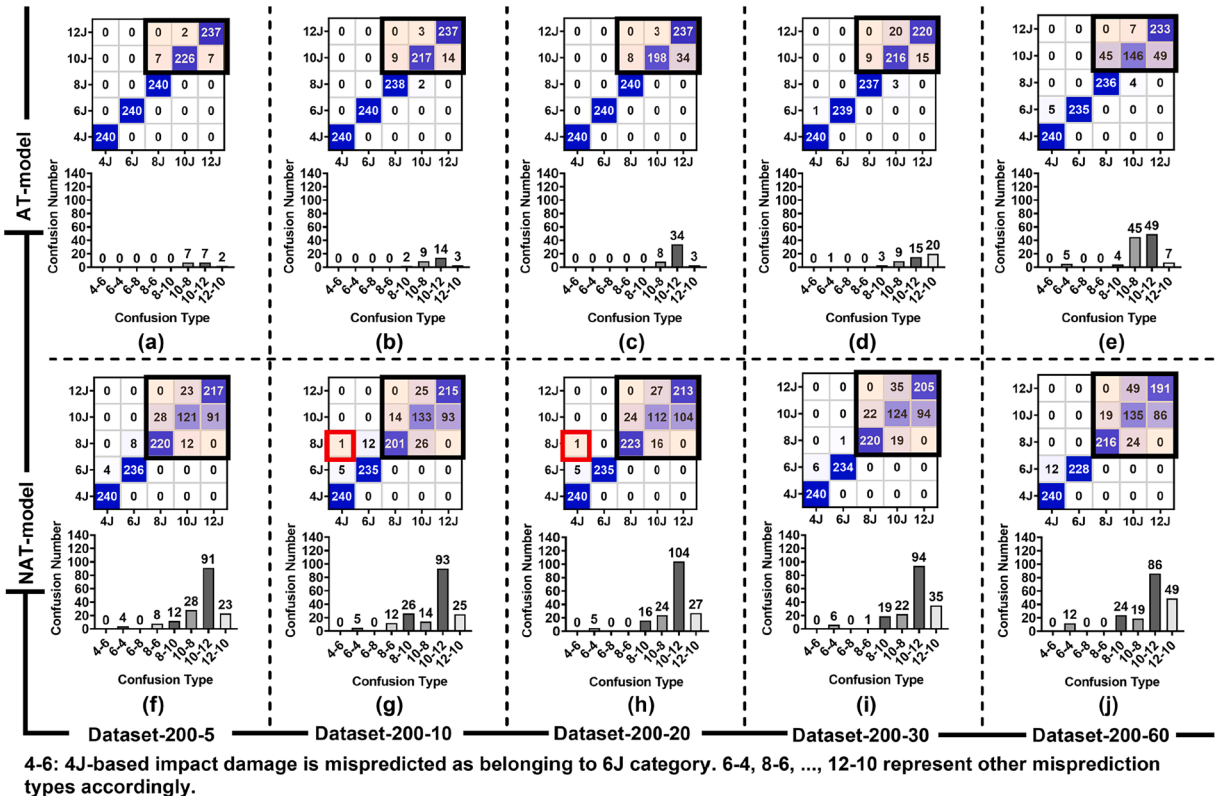

Fig. 7. (a)–(e) The confusion matrices and corresponding bar charts illustrating the AT-model's classification performance on the five datasets generated on SP-200 with five different downsampling intervals of INT-5, INT-10, INT-20, INT-30 or INT-60. For each energy category, 240 sequence data were used for model testing.(f)-(j) The confusion matrices and corresponding bar charts illustrating the NAT-model's classification performance on the corresponding datasets.

among impact energy levels of 8 J, 10 J, and 12 J, while the accuracies remain high for 4 J and 6 J impact damages. The bar charts clearly display the frequency of mispredictions across different energy levels. Specifically, the most common misprediction pattern is between 10 J and 12 J, followed by 12 J-10 J and 10 J-8 J. Previous studies on low-velocity impact damages in composites have suggested that certain damage properties, such as size or depth, are correlated with the impact energy within a certain threshold [30]. Based on these findings, a hypothesis is formed to explain the higher occurrence of confusion between higher energy levels, such as 10 J and 12 J. It is assumed that key properties of the thermal sequence data, on which the model relies for classification, are proportional to the impact energy level. Consequently, a larger ratio between two energy levels indicates higher separability between impact damages. For instance, the ratio of 6 J to 4 J is 1.5, greater than that of 12 J to 10 J, which is 1.2. As a result, impact damages with energy levels of 4 J and 6 J are more easily distinguishable than those of 10 J and 12 J.

Comparing the bar charts in Fig. 7(a)-(e) with Fig. 7(f)-(j), it is evident that the model with the attention module significantly reduces mispredictions. For example, on Dataset-200–5, the number of 10 J-12 J mispredictions is reduced from 91 to 7 with the addition of the attention module. Furthermore, the attention module eliminates nonadjacent energy level confusion. In Fig. 7(g)-(h), two 8 J-4 J confusions are highlighted in a red box, while no such confusions have been observed in Fig. 7(a)-(e). This elimination of nonadjacent confusion demonstrates the superior discriminative power of the attention module-enhanced model. The confusion matrices presenting the model's performance on other datasets generated on SP-300 and SP-600 can be found in Fig. A.1 and A.2 in the Appendix, where similar patterns have been observed.

4. Conclusion

This paper presents an attention mechanism-enhanced spatiotemporal deep learning model for classifying BVID in CFRP material based on impact energy levels. Unlike existing works that analyse either thermal images or the temperature curve for a specific pixel, this study focuses on simplified thermal image sequences that contain both spatial and temporal information. Based on the results and discussions, the following conclusions can be drawn.

1. The proposed model can effectively categorise barely visible impact damages into their respective energy categories, with or without the attention module.
2. The employed attention module significantly enhances the performance of the proposed model on all datasets, reinforcing the model's discriminative power.
3. The crucial information determining impact energy categories lies in the frame span of the captured thermal sequence where temperature sharply declines. Within this range, higher accuracy is achieved by utilising sequence data generated at smaller downsampling intervals.
4. The smaller the ratio between a higher and a lower impact energy level, the higher the likelihood of misclassification between them.

It should be noted that only five impact energy levels within a limited range have been studied in the research, indicating that the model's application scope is limited. Furthermore, the proposed model lacks the ability to predict energy levels that were not included in the training data, such as 5 J and 7 J. Future work could involve developing a more versatile model capable of handling a wider range of impact energy levels and recognising unlearned energy levels.

5. Statements and declarations

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.

CRediT authorship contribution statement

Kailun Deng: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Haochen Liu:** Writing – review & editing, Methodology, Formal analysis. **Jun Cao:** Writing – review & editing, Formal analysis. **Lichao Yang:** Writing – review & editing. **Weixiang Du:** Writing – review & editing. **Yigeng Xu:** Writing – review & editing. **Yifan Zhao:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Acknowledgement

This work was partially supported by the Royal Academy of Engineering Industrial Fellowship [#grant IF2223B-110], and partially supported by the Science and Technology Department of Gansu Province Science and Technology Project Funding, 22YF7GA072.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compstruct.2024.118030.

References

[1] Ngo T-D. 'Introduction to composite materials', in *composite and nanocomposite materials - from knowledge to industrial applications*. IntechOpen 2020. https://doi.org/10.5772/intechopen.91285.
[2] Wang H, et al. Stress mapping reveals extrinsic toughening of brittle carbon fiber in polymer matrix. Sci Technol Adv Mater 2020;21(1):267–77. https://doi.org/10.1080/14686996.2020.1752114.
[3] Zhang X, et al. CFRP barely visible impact damage inspection based on an ultrasound wave distortion indicator. Compos B Eng 2019;168:152–8. https://doi.org/10.1016/j.compositesb.2018.12.092.
[4] Shi Y, Pinna C, Soutis C. Low-velocity impact of composite laminates: damage evolution. In: Dynamic Deformation, Damage and Fracture in Composite Materials and Structures. Elsevier; 2023. p. 89–119. https://doi.org/10.1016/B978-0-12-823979-7.00005-3.
[5] Łukaszuk R, Chady T. Nondestructive examination of carbon fiber-reinforced composites using the Eddy current method. Materials 2023;16(2):pp. https://doi.org/10.3390/ma16020506.
[6] Sinha A, Mata Vaishno S, Sharma A, Sinha AK. 'Ultrasonic testing for mechanical engineering domain: present and future perspective', *international journal of research*. Ind Eng 2018;7(2).
[7] Cheng J, Qiu J, Xu X, Ji H, Takagi T, Uchimoto T. Research advances in eddy current testing for maintenance of carbon fiber reinforced plastic composites. Int J Appl Electromagn Mech 2016;51(3):pp. https://doi.org/10.3233/JAE-150168.
[8] Jones TS, Polansky D, Berger H. Radiation inspection methods for composites. NDT Int 1988;21(4). https://doi.org/10.1016/0308-9126(88)90341-0.
[9] Deane S, et al. Application of NDT thermographic imaging of aerospace structures. Infrared Phys Technol 2019;97. https://doi.org/10.1016/j.infrared.2019.02.002.
[10] Panella FW, Pirinu A. Application of pulsed thermography and post-processing techniques for CFRP industrial components. J Nondestr Eval 2021;40(2):52. https://doi.org/10.1007/s10921-021-00776-8.
[11] Liu G, et al. Low-velocity impact damage detection in CFRP composites by applying long pulsed thermography based on post-processing techniques. Nondestructive Testing and Evaluation 2023. https://doi.org/10.1080/10589759.2023.2284248.
[12] Liang T, Ren W, Tian GY, Elradi M, Gao Y. Low energy impact damage detection in CFRP using eddy current pulsed thermography. Compos Struct 2016;143. https://doi.org/10.1016/j.compstruct.2016.02.039.
[13] Balageas DL, Roche J-M, Leroy F-H, Liu W-M, Gorbach AM. The thermographic signal reconstruction method: a powerful tool for the enhancement of transient thermographic images. Biocybern Biomed Eng 2015;35(1):1–9. https://doi.org/10.1016/j.bbe.2014.07.002.
[14] Oliveira BCF, Seibert AA, Borges VK, Albertazzi A, Schmitt RH. Employing a U-net convolutional neural network for segmenting impact damages in optical lock-in thermography images of CFRP plates. Nondestructive Testing Evaluation 2021;36(4):440–58. https://doi.org/10.1080/10589759.2020.1758099.
[15] Liu H, Li W, Yang L, Deng K, Zhao Y. Automatic reconstruction of irregular shape defects in pulsed thermography using deep learning neural network. Neural Comput Appl 2022;34(24):21701–14. https://doi.org/10.1007/s00521-022-07622-6.
[16] Bang H-T, Park S, Jeon H. Defect identification in composite materials via thermography and deep learning techniques. Compos Struct 2020;246:112405.
[17] Wei Z, Fernandes H, Herrmann H-G, Tarpani JR, Osman A. A deep learning method for the impact damage segmentation of curve-shaped CFRP specimens inspected by infrared thermography. Sensors 2021;21(2):395. https://doi.org/10.3390/s21020395.
[18] Deng K, Liu H, Yang L, Addepalli S, Zhao Y. Classification of barely visible impact damage in composite laminates using deep learning and pulsed thermographic inspection. Neural Comput Appl 2023;35(15):11207–21. https://doi.org/10.1007/s00521-023-08293-7.
[19] Sirikham A, Zhao Y, Mehnen J. Determination of thermal wave reflection coefficient to better estimate defect depth using pulsed thermography. Infrared Phys Technol 2017;86:1–10. https://doi.org/10.1016/j.infrared.2017.08.012.
[20] Marani R, Palumbo D, Renò V, Galietti U, Stella E, D'Orazio T. Modeling and classification of defects in CFRP laminates by thermal non-destructive testing. Compos B Eng 2018;135:129–41. https://doi.org/10.1016/j.compositesb.2017.10.010.
[21] Cao Y, Dong Y, Cao Y, Yang J, Yang MY. Two-stream convolutional neural network for non-destructive subsurface defect detection via similarity comparison of lock-in thermography signals. NDT and E Int 2020;112:102246. https://doi.org/10.1016/j.ndteint.2020.102246.
[22] Fang Q, Maldague X. A method of defect depth estimation for simulated infrared thermography data with deep learning. Appl Sci 2020;10(19):6819. https://doi.org/10.3390/app10196819.
[23] Hu J, et al. Pattern deep region learning for crack detection in thermography diagnosis system. Metals (Basel) 2018;8(8):612. https://doi.org/10.3390/met8080612.
[24] Saeed N, King N, Said Z, Omar MA. Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning. Infrared Phys Technol 2019;102:103048. https://doi.org/10.1016/j.infrared.2019.103048.
[25] Luo Q, Gao B, Woo WL, Yang Y. Temporal and spatial deep learning network for infrared thermal defect detection. NDT and E Int 2019;108:102164. https://doi.org/10.1016/j.ndteint.2019.102164.
[26] Hu B, et al. A lightweight spatial and temporal multi-feature fusion network for defect detection. IEEE Trans Image Process 2021;30:472–86. https://doi.org/10.1109/TIP.2020.3036770.
[27] Dong Y, Xia C, Yang J, Cao Y, Cao Y, Li X. Spatio-temporal 3-D residual networks for simultaneous detection and depth estimation of CFRP subsurface defects in lock-in thermography. IEEE Trans Industr Inform 2022;18(4):2571–81. https://doi.org/10.1109/TII.2021.3103019.
[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, 'CBAM: Convolutional Block Attention Module', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.
[29] Zhou J, Du W, Yang L, Deng K, Addepalli S, Zhao Y. Pattern recognition of barely visible impact damage in carbon composites using pulsed thermography. IEEE Trans Industr Inform 2022;18(10):7252–61. https://doi.org/10.1109/TII.2021.3134184.
[30] Safri SNA, Sultan MTH, Jawaid M, Abdul Majid MS. Analysis of dynamic mechanical, low-velocity impact and compression after impact behaviour of benzoyl treated sugar palm/glass/epoxy composites. Compos Struct 2019;226:111308. https://doi.org/10.1016/j.compstruct.2019.111308.

2024-03-14

Attention mechanism enhanced spatiotemporal-based deep learning approach for classifying barely visible impact damages in CFRP materials

Deng, Kailun

Elsevier

Deng K, Liu H, Cao J, et al., (2024) Attention mechanism enhanced spatiotemporal-based deep learning approach for classifying barely visible impact damages in CFRP materials. Composite Structures, Volume 337, June 2024, Article number 118030

https://doi.org/10.1016/j.compstruct.2024.118030