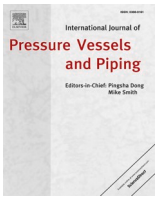
International Journal of Pressure Vessels and Piping 207 (2024) 105112

Contents lists available at ScienceDirect

International Journal of Pressure Vessels and Piping

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



CNN-based automated approach to crack-feature detection in steam cycle components

Zhouxiang Fei *, Graeme M. West, Paul Murray, Gordon Dobie

Department of Electronic and Electrical Engineering, University of Strathclyde, 204 George Street, Glasgow, G1 1XW, UK

ARTICLE INFO

Keywords:
Convolutional neural network
Crack detection
Nuclear power plant inspection
Pressure vessel inspection
Remote visual inspection support
Superheater inspection

ABSTRACT

Periodic manual inspection by trained specialists is an important element of asset management in the nuclear industry. Detection of cracks caused by stress corrosion is an important element of remote visual inspection (RVI) in power plant steam generator components such as boilers, superheaters and reheaters. Challenges exist in the interpretation of RVI footage, such as high degree of concentration for reviewing lengthy and disorienting footage due to narrow field of view offered by endoscope. Deep learning is considered useful to automate crack detection process for improved efficiency and accuracy, and has enjoyed success in related applications. This article utilises a new application of automated crack feature detection in steam cycle components to demonstrate a transferrable data-driven framework for a variety of anomaly inspections in such structures. Specifically, a case study of superheater (a type of reactor pressure vessel head) anomaly inspection is presented to automatically detect regions of crack-like features in inspection footage with a good accuracy of 92.97 % using convolutional neural network (CNN), even in challenging cases. Due to the black-box nature of the CNN classification, the explicability of the classification results is discussed to enhance the trustworthiness of the detection system.

1. Introduction

Various components in the nuclear sector (e.g., nuclear reactor cores [1–3]), civil infrastructures (e.g., concrete surfaces [4,5] and piping systems [6,7]) and oil and gas energy sectors (e.g., subsea pipelines [8]) benefit from regular inspections by trained specialists. This provides an understanding of their current and future conditions and to ensure they are able to perform the functions for which they are designed. However, internal structural inspection of pressure vessels is commonly a complex task due to the constraint of inlet/outlet sizes in such structures and the associated difficult accessibility.

Taking superheater inspection as example: as superheaters are enclosed structures, remote visual inspection is performed by inserting an endoscope into the superheater via the access point shown in Fig. 1 (a). While inspecting the real-time footage from the endoscope, engineers need to seek a range of defects such as spalling, oxides and cracks. Amongst these defects, one important category is the stress corrosion crack features at the tube plate upper radius (i.e., the circumferential area at the superheater bottom) in Fig. 1(b). This is because cracks are a pivotal feature for assessing the structural integrity of superheaters. The observed crack features are recorded in the logbook which is then checked by metallurgists to decide whether the logged features are genuine stress corrosion cracks. In addition to remote visual inspection via endoscopes, other techniques via electromagnetic wave [9] and ultrasound [10] probing are also possible for crack inspection in certain pressure vessel scenarios. This article focuses on the detection of superheater crack-features in the remote visual inspection phase, and is hoped to be inspiring and versatile to deal with anomaly inspection in other pressure vessel types.

In practice, during the remote visual inspection process, hours of real-time footage need to be manually examined. As a result, this process is time-consuming and labour-intensive. Therefore, it is beneficial to support the conventional visual inspection process with an automated crack-feature detection system for improved efficiency. The supported anomaly inspection will offer useful valuable information to evaluate the remaining useful life of pressure vessels in nuclear power plants.

Image processing techniques have been widely applied to detect cracks in various scenarios. Different edge-detection techniques (i.e., the fast Haar transform, fast Fourier transform, Sobel and Canny edge detectors) were investigated for their efficacy on crack detection in bridge deck surfaces [11]. Gabor filters is a tool to perform convolution operation with images to extract texture features with certain frequency.

* Corresponding author.
E-mail address: zhouxiang.fei@strath.ac.uk (Z. Fei).

https://doi.org/10.1016/j.ijpvp.2023.105112
Received 20 October 2023; Received in revised form 1 December 2023; Accepted 10 December 2023
Available online 13 December 2023

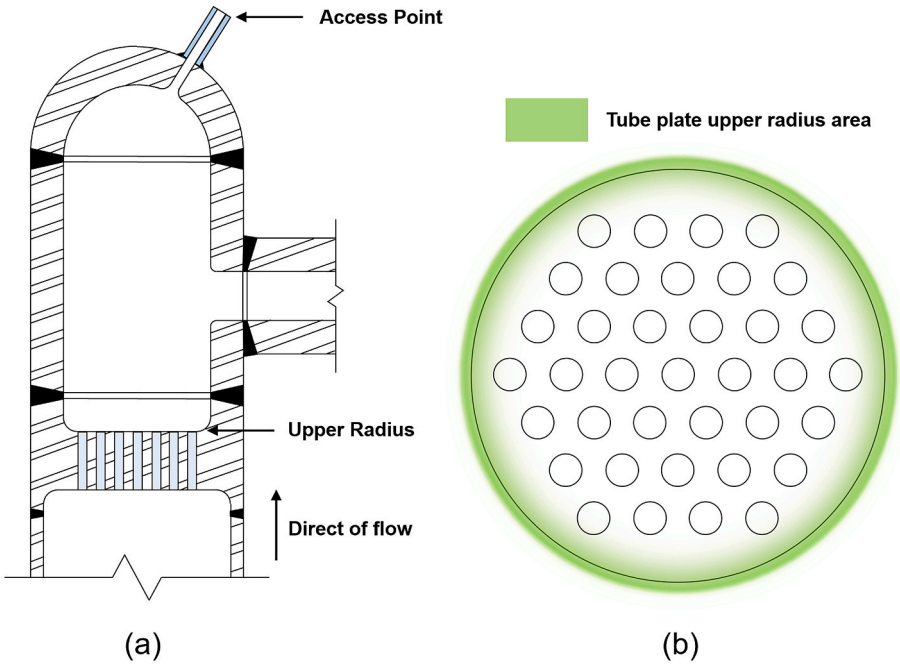


Fig. 1. Detail of the superheater structure. (a) Cross-section view of the superheater. (b) Bird's-eye view of the tube plate. Superheater is one major type of pressure vessel assets used in nuclear power plants to constitute the boiler units and convert saturated steam into overheated steam for the electricity generation process.

This technique was applied to extract features from crack images to build crack classifiers [12]. Frangi filter is a Hessian matrix-based technique originally proposed to detect tubular structures in medical images [13]. This technique was adopted to detect thin and bright cracks in steel beams [14]. An improved Hessian-matrix-based technique was proposed in [15] to segment concrete crack pixels with crack characteristics (e.g., length and width in pixel measurements) quantified. Precise and efficient pixel-level crack segmentation was addressed in [15] as crack segmentation was only needed for targeted crack regions flagged through the faster region-based convolutional neural network (Faster R–CNN) technique [16]. Jahanshahi et al. identified discriminative features suitable for crack feature extraction, and on this basis investigated the performance of different types of classifiers (e.g., Support Vector Machine and Neural Network, trained with extracted features) for crack detection [17].

Recent literature demonstrates an increasing trend of adopting data-driven techniques to autonomously detect cracks. For example, the Bag-of-Visual-Words approach generates histograms of visual word (i.e., physical features) occurrence for labelled images and a classifier is trained using these histograms to categorise new images. This technique was applied to develop a detection system of cracks in panoramas view of fuel channels [3]. Amongst different data-driven techniques, the convolutional neural network (CNN) approach draws particular focus for its good accuracy and the ability to automatically learn the features of interest for image classification. For example, a CNN system was developed to automate the detection of cracks at patch level in concrete surfaces [18]. To obtain more refined crack location, a U-Net-based model called CrackUnet was proposed to detect concrete cracks at pixel level [19]. Real-time pixel-wise segmentation of concrete cracks on complex backgrounds was implemented through a semantic damage detection network (SDDNet) in [20] and a semantic transformer representation network (STRNet) in [21]. The efficacy of various state-of-the-art CNN models on detecting cracks in concrete structures was compared in [22], reporting that the balance between inference time and accuracy of a model needs to be considered for real-time applications. In the domain of nuclear plant inspection, a CNN-based automated system was presented in [23] to detect cracks in the mock-up reactor surfaces. In the oil and gas transmission sector, a CNN system for detecting pitting and cracks in oil and gas pipelines can be found in [24].

However, the inspection environment of pressure vessels poses several challenges to understanding the performance of autonomous data-driven approaches for detecting crack features, and superheater is no exception. Typical challenging factors include highly reflective metallic surfaces, varying illuminations and rich presence of irrelevant intensive texture features in the structure. Therefore, a heuristic application of the novel CNN approach for superheater inspection advances the knowledge of data-driven technique capability for automated anomaly detection in the complex pressure vessel environment, as addressed in this article. To build a CNN-based crack detection system, the first common step is to generate the training dataset by cropping patches with a specified resolution (i.e., typically the same as the input image size of the CNN) from camera-taken images [18] or video frames [23]. Then, each cropped patch is manually labelled as "crack" or "non-crack background". Here, small-sized patches rather than full images are preferred as patches can demonstrate better prominent crack features, whereas full images may contain a large proportion of irrelevant features. However, CNN systems typically require a sufficient number of labelled images for training, and the associated manual labelling process can be excessively laborious. In addition to the intensive labour cost for labelling enough patches, another challenge of manual labelling is to ensure that a consistent labelling standard is used. This becomes particularly challenging when the crack feature has small size or is in low contrast with respect to the background [25]. To overcome these challenges, an automated labelling approach governed by a consistent labelling standard was introduced in [25] to efficiently generate the training dataset.

This article utilises a new application of automated crack feature detection in superheaters to demonstrate a transferrable data-driven framework for various anomaly inspection in complex pressure vessel structures. Specifically, the training, validation and testing datasets of crack features in superheaters are prepared using the automated labelling technique [25]. On this basis, an autonomous system is developed via the transfer learning technique to detect crack features in the tube plate upper radius of superheaters used in nuclear power plants. Due to the black-box nature of the CNN operation, the explicability of the

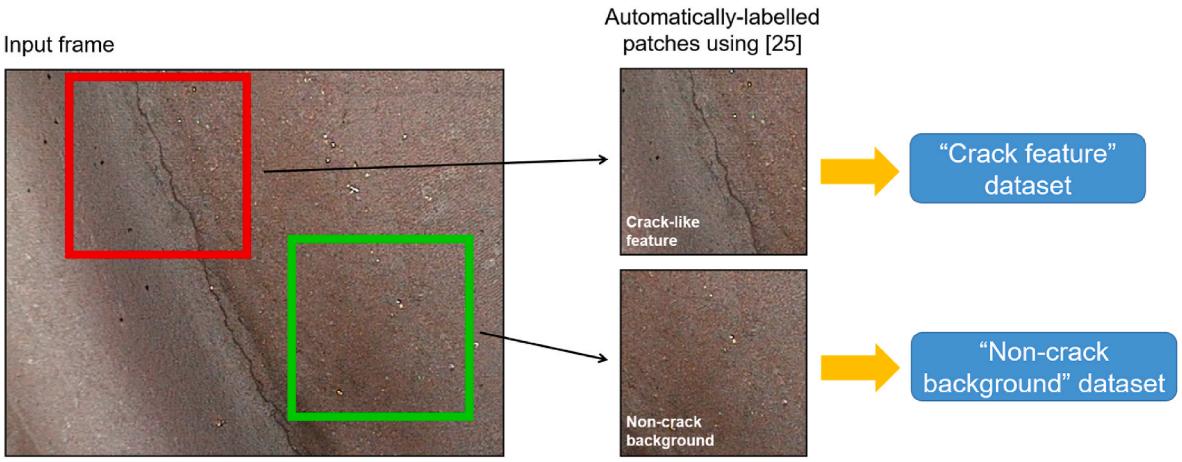


Fig. 2. Illustration of the frame-to-patch process to constitute the datasets.

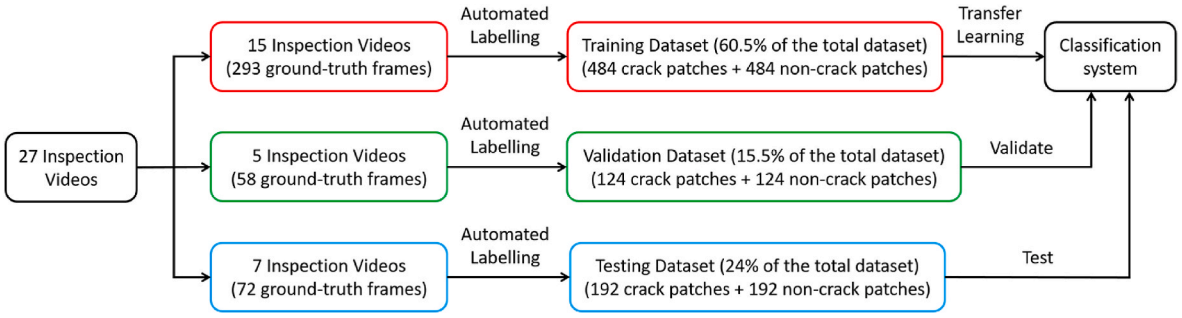


Fig. 3. Constitution of each dataset and its use.

classification mechanism for crack features is investigated to demonstrate the confidence of classification results to an extent. The remainder of this article is organised as follows: Section 2 presents the detailed configuration of the training, validation and testing datasets. Section 3 describes the framework of the CNN classification for crack features via transfer learning. The performance of the crack-feature detection system is presented and discussed in Section 4. The conclusions of this article are given in Section 5.

2. Dataset preparation

Data preparation is important as it facilitates the training, validation and testing processes and even affects the feature types of interest learnt by the CNN system for decision-making. The automated labelling technique proposed in [25] is applied to generate the training, validation and testing datasets. The implementation of the automated labelling technique is outlined in [25] and therefore not repeated herein for brevity.

In our study, a total of 423 ground-truth frames containing crack features at the superheater tube plate upper radius are extracted from 27 inspection videos. These videos were recorded during the real-time inspections at different nuclear power plant sites in the UK. Each video is accompanied with an inspection logbook which records the timing and shapes of crack features in the video. These records are used to obtain the indices of the ground-truth crack feature frames in the inspection videos. The patch resolution is set to 224 × 224 × 3 pixels, the same as the input image size of the CNN system (i.e., GoogLeNet [26]) used in our study. An illustration of the frame-to-patch process to constitute the "crack feature" and "non-crack background" datasets is given in Fig. 2.

To prevent data leakage between the datasets, the ground-truth frames from a specific video are only used in one of the training, validation or testing datasets. Specifically, 15 inspection videos which contribute 293 ground-truth frames are used to construct the training dataset. Another 5 videos which are the sources of 58 ground-truth frames are used to prepare the validation dataset. The remaining 7 videos which produce 72 ground-truth frames are used to constitute the testing dataset. The videos for each dataset are carefully selected to present a wide variety of crack features that are of interest to inspection engineers in practice.

Table 1
Size and number of patches in the training, validation, and testing datasets.

Dataset	Size (RGB channels)	Total number	Crack patches	Non-crack patches
Training	224 × 224 × 3	968	484	484
Validation	224 × 224 × 3	248	124	124
Testing	224 × 224 × 3	384	192	192

The detail of the dataset constitution is summarised in Fig. 3 and Table 1. In total, 800 "crack feature" patches and 800 "non-crack background" patches are obtained from all the 423 ground-truth frames. The balanced training, validation and testing datasets account for 60.5 %, 15.5 % and 24 % of all the 1600 patches, respectively. Examples of the labelled "crack feature" and "non-crack background" patches are shown in Figs. 4 and 5, respectively.

3. CNN implementation

3.1. Brief background of CNN

The convolutional neural network (CNN) [27] approach has been widely used in automated anomaly detection in numerous fields such as building defect inspection [18] and cancer diagnosis [28]. This is due to

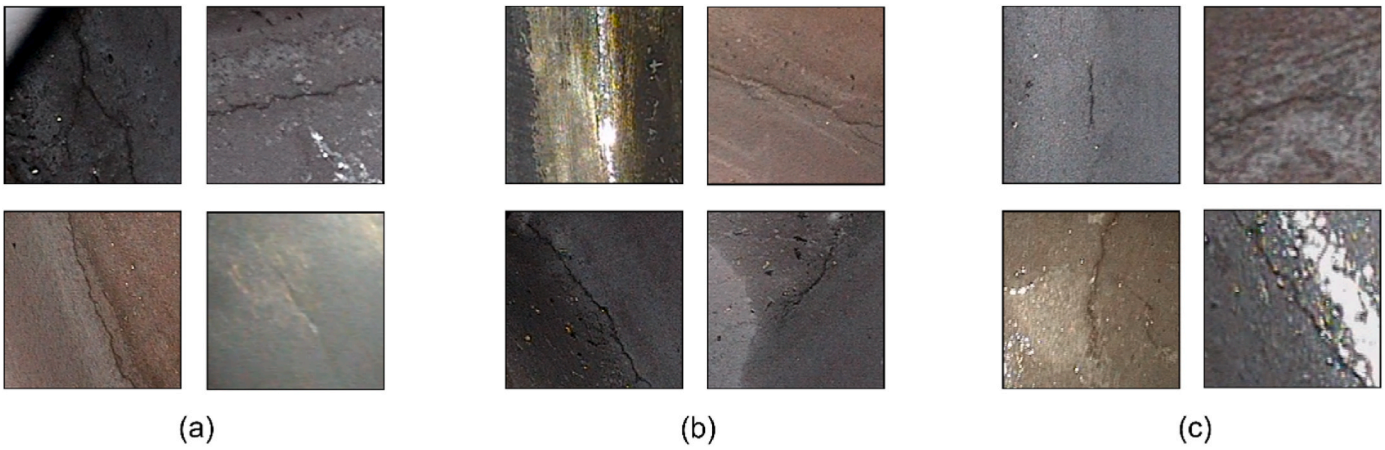


Fig. 4. Examples of "crack-feature" patches in: (a) the training dataset, (b) validation dataset and (c) testing dataset. A wide range of crack feature types with different background textures and variant illumination conditions are represented in each dataset.

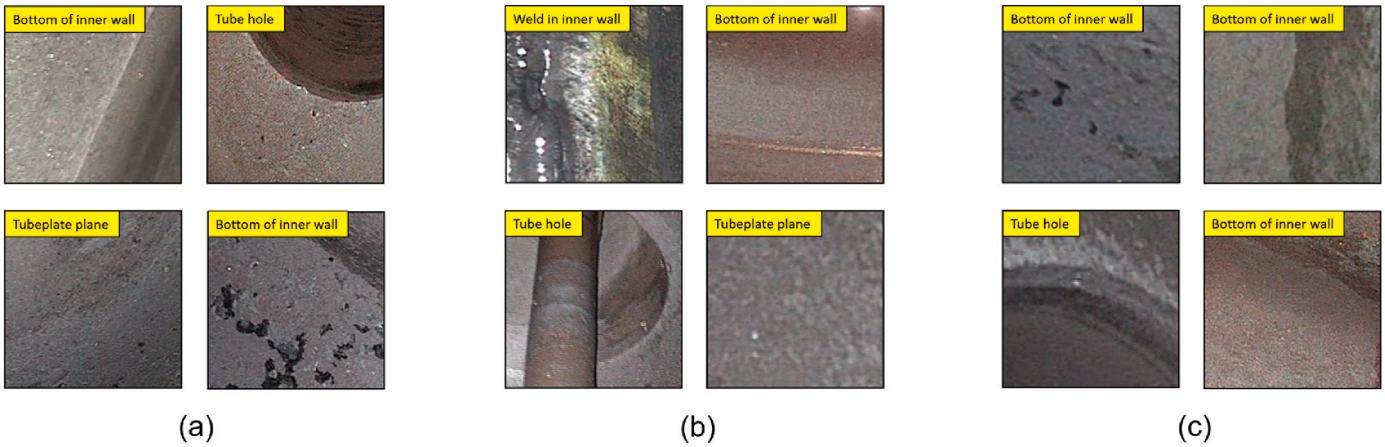


Fig. 5. Examples of "non-crack background" patches in: (a) the training dataset, (b) validation dataset and (c) testing dataset. Note that the "non-crack background" patches with complex line features are included in the training dataset to help prevent False-Positive detections. Such features are included in the validation and testing datasets to comprehensively evaluate the classification system performance. The annotations show the locations of these "non-crack background" examples in the superheaters.

the data-driven feature of the CNN approach and its capability of autonomously learning the features of interest from training images for decision-making [18].

Generally speaking, the CNN system uses the Convolution Layer and Pooling Layer to extract features of interest and deploys the Fully-Connected Layer and Softmax Layer for classification. In the Convolution Layer, a set of filters are used to extract specific features from the input image. Here, each filter is a matrix to perform element-wise products with subparts of the input image in a manner controlled by strides, resulting in a feature map. Then, all the feature maps are processed using the rectified linear unit (ReLU) activation function which set negative elements in the feature maps to zeros, in order to accelerate the training process [18]. After the ReLU operation, the Pooling Layer is used to reduce the size of the feature maps via either the max pooling or mean pooling option. The feature maps from the Pooling Layer are then flattened into a one-dimensional array which becomes the input of the Fully-Connected Layer. The Fully-Connected Layer combines extracted features from previous layers and outputs a one-dimensional array of N elements where N is the classification category number. Every output element of the Fully-Connected Layer is contributed by all the input elements multiplied by different weights. These output elements are passed to the Softmax Layer to calculate the score of each category which is used in the Classification Layer to predict the category of the input image. The summation of all the category scores is one, with each score ranging between zero and one. Note that other layer types such as the Batch Normalisation Layer [29] and Dropout Layer [30] could be seen in some CNN structures to improve the performance.

Deep CNN systems can be constructed using repetition and combination of the aforementioned layer types to improve the performance. Given a specific classification task, one can choose to either develop a customised CNN structure or use a pre-trained deep CNN system such as GoogLeNet [26], VGG [31] and ResNet series [32]. This is because the first few layers in these pre-trained deep networks are used to extract elementary features (e.g., edges) from images and can be shared across different classification tasks. The transfer learning technique [33] needs to be applied to perform new classification tasks using pre-trained networks. This technique usually requires less training data, and its implementation only requires small modifications of the pre-trained network. In our study, the detection of crack features in the superheater tube plate upper radius is implemented using transfer learning based on the GoogLeNet.

The CNN system needs to be trained before being used for classification tasks. During the training process, the parameters of the CNN (e.g., the filters in the Convolution Layer and the weights in the Fully-Connected Layer) are optimised using the training images via the stochastic gradient descent (SGD) technique [34], in order to minimise the deviation between the predicted and actual categories. The training process is iterated for multiple times until a satisfactory classification

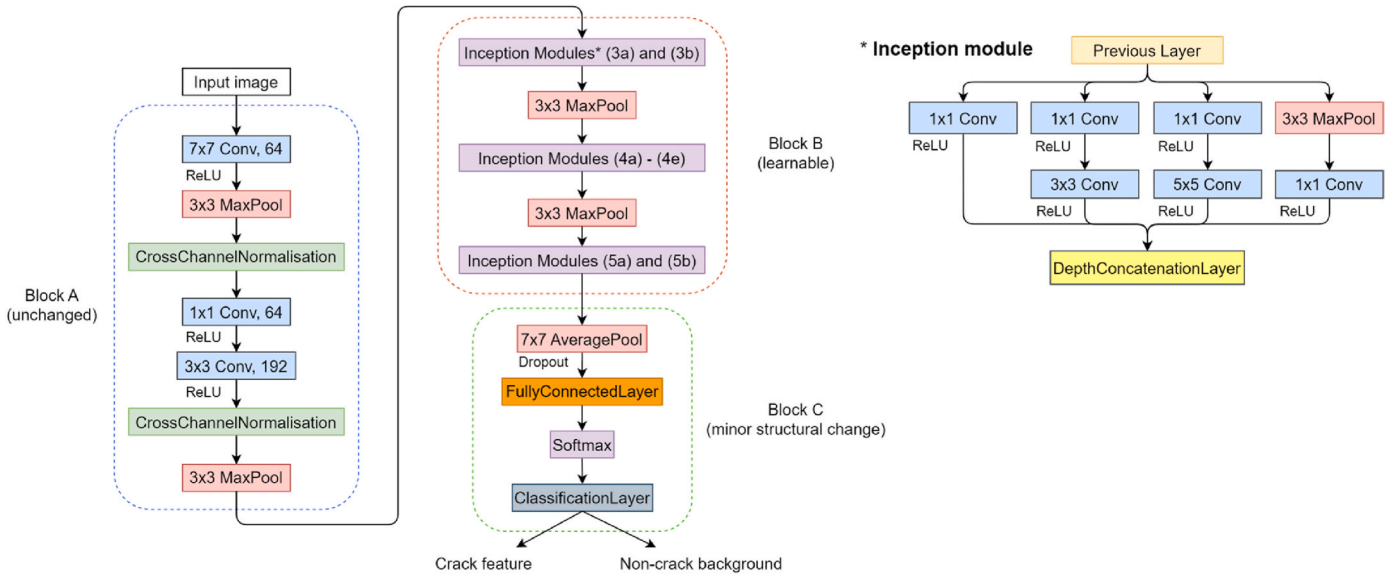


Fig. 6. Simplified classifier framework based on GoogLeNet provided in the MATLAB environment. The number of filters in the Convolution Layers of each Inception Module may vary.

accuracy on the validation dataset is achieved. Please refer to [18,34] for a comprehensive mathematical operation of the CNN system and its training process.

3.2. Transferable classifier implementation

The classifier architecture is established based on the GoogLeNet architecture provided in the MATLAB environment. The GoogLeNet system was originally developed to classify 1,000 different categories in the ImageNet dataset [35]. The classifier architecture can be divided into three main blocks shown in Fig. 6. Specifically, the Convolution Layers in *Block A* are frozen during the training process. The learnable parameters (e.g., weight and bias) in *Blocks B* and *C* are updated during the training process using the crack-feature training dataset, in order to tailor the classifier for the crack-feature detection task. As the classifier is designed to take an image patch as the input and produce a binary label (either "crack feature" or "non-crack background") for the content of the patch, minor structural change is made to *Block C* by setting the output number of the Fully-Connected Layer and Classification Layer to two. This presented methodology of classifier implementation can be directly applicable to new domains of automated anomaly inspection in other pressure vessel types.

3.3. Training process

Data augmentation is performed on the training patches to prevent overfitting and improve the variety of the training dataset. Different data augmentation options are applied, including random horizontal reflection, horizontal and vertical translation up to 30 pixels, and horizontal and vertical scaling up to 10 % of the original size. For each

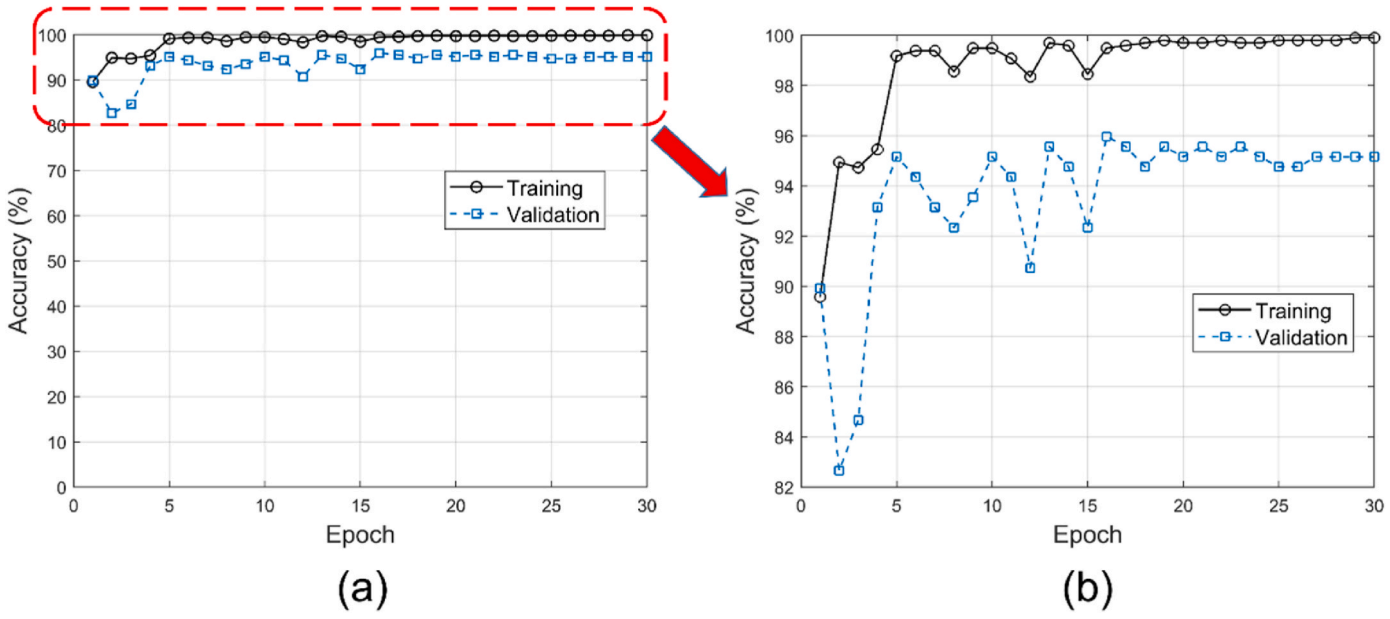


Fig. 7. Training and validation accuracy rates at each epoch. (a) Original plot with *Y*-axis in the range of [0, 100 %]. (b) Detailed "zoom-in" plot of the training and validation accuracy curves between [82 %, 100 %].

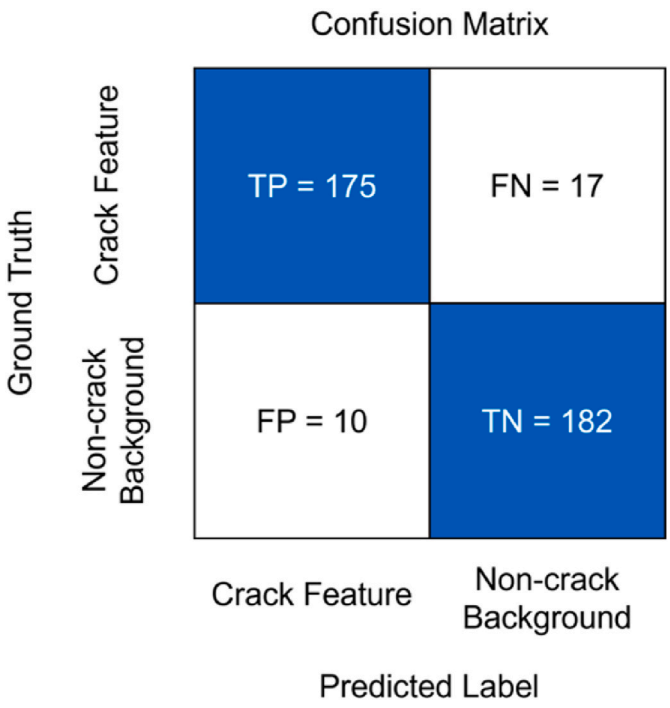


Fig. 8. Confusion matrix on the testing dataset.

training epoch, the training patches are transformed using random combinations of the data augmentation options. As a result, the training dataset varies slightly for different epochs.

The training process of our classification system is performed in the MATLAB environment using the SGD technique, based on a CPU of 2.8 GHz and RAM of 32 GB, with the mini-batch size of 10, maximum epoch number of 30 and momentum of 0.9. The global learning rate is initially set to 0.001 and drops every two epochs by multiplying the decaying factor of 0.8. Here, the piecewise decreasing learning rate is used to facilitate the efficient optimisation of the learnable parameters [36]. The learning rate of the final Fully-Connected Layer is set to 10 times the global learning rate to accelerate the learning of parameters in this layer.

4. Results and discussion

4.1. Patch-level classification

The training and validation accuracy rates at each epoch are shown in Fig. 7. Here, the accuracy rate is defined as the ratio of the correctly classified patch number to the total patch number in the dataset. The training accuracy is calculated at the end of each epoch using the original training patches without data augmentation changes. As shown in Fig. 7, the converged training and validation accuracy rates at the 30th epoch are 99.90 % and 95.16 %, respectively. This performance is achieved using a training time of 65 min. The trained network at the 30th epoch is used as the classification system for the analysis hereafter.

The testing dataset is then used to extensively evaluate the classification system performance. The values of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are given in the confusion matrix in Fig. 8. Accuracy is calculated to be 92.97 % with a recall of 91.15 % and precision of 94.59 %, resulting in F1-Score of 92.84 %.

Note that the testing accuracy of 92.97 % is not as high as the training accuracy (99.90 %) or validation accuracy (95.16 %) at the 30th epoch. This could be partly due to certain crack-feature types in the testing dataset which are not well represented in the training dataset. This is because the crack features at the superheater tube plate upper radius could take various forms and the number of the provided inspection videos is limited. As a result, certain crack-feature forms are not commonly shared between the datasets. Further provision of superheater inspection videos from different sites is deemed useful to increase the crack-feature diversity in the training dataset and therefore improve

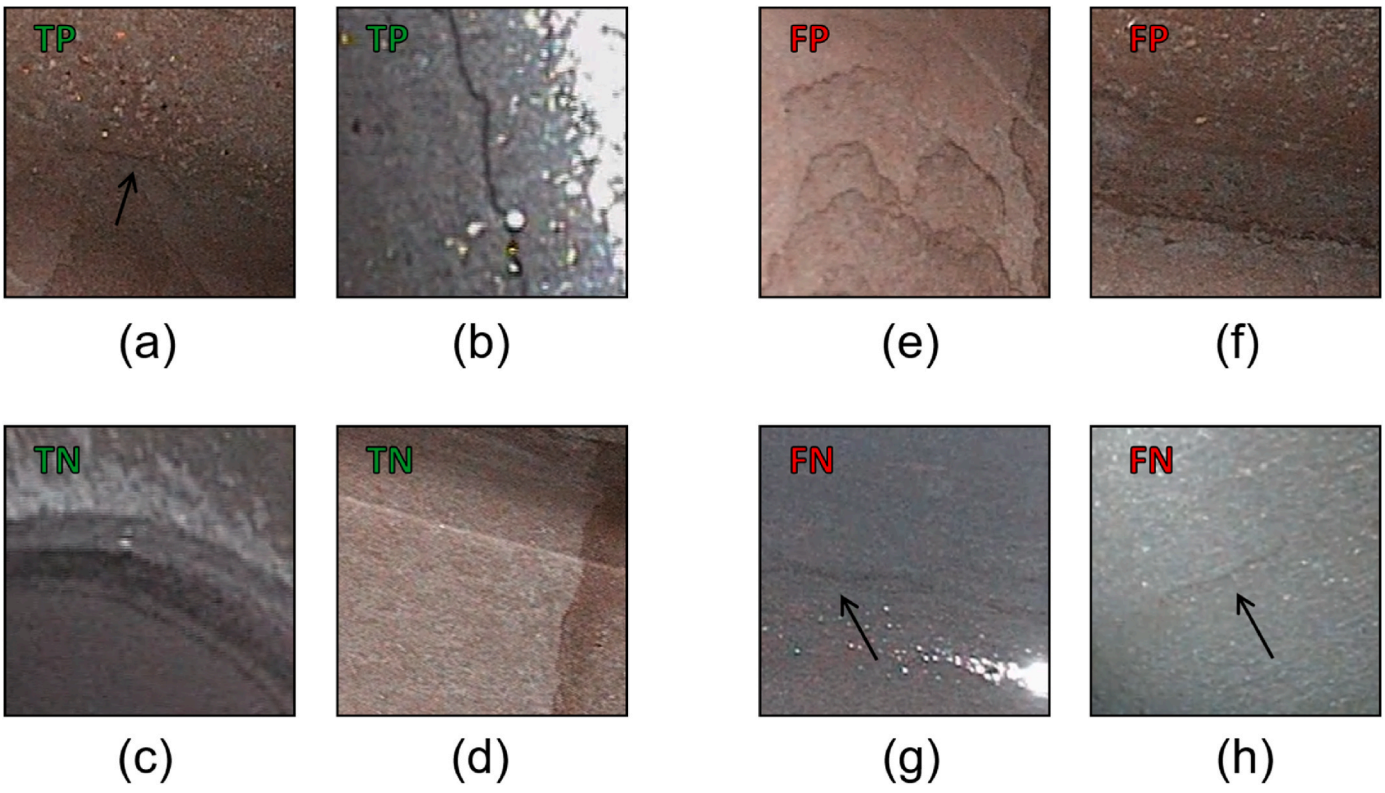


Fig. 9. Examples of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) classification results.

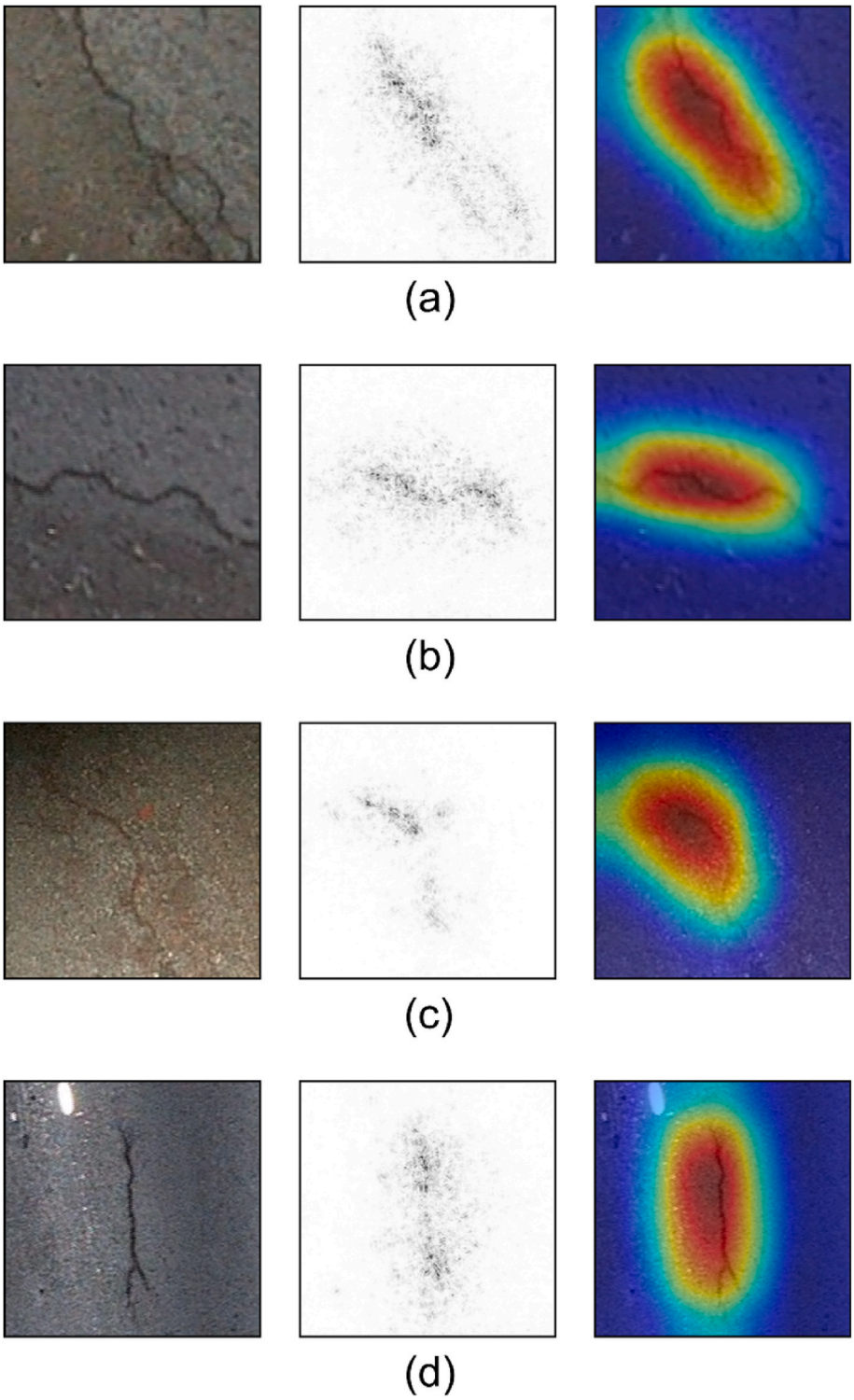


Fig. 10. Left column: four TP classification examples. Middle column: corresponding saliency maps for the classified "crack-feature" category, produced using the Saliency Map technique based on guided backpropagation. Right column: corresponding localisation maps to show the most important regions (highlighted in red) for the classified "crack-feature" category, produced using the Grad-CAM technique.

the classification system performance.

Examples of TP, TN, FP and FN classifications are shown in Fig. 9 to demonstrate the wide diversity of crack-feature types, background textures and lighting conditions in the testing patches. Specifically, Fig. 9 (a) presents a TP example with small crack indication in the dim illumination condition. In contrast, an example of correct classification with large crack size in strong brightness is given in Fig. 9(b). Therefore, the classification system is not susceptible to the variant crack-feature sizes and illumination conditions. Fig. 9(c) and (d) present two correctly classified "non-crack background" patches containing deceptive line features. It is clear that the classification system is able to differentiate crack features from some complex background textures. However, the classification system may see certain line features as crack features, as illustrated by the two FP examples in Fig. 9(e) and (f). Such

misclassifications may be due to the close resemblance between the crack features and other line features formed by discolouration marks. Finally, Fig. 9(g) and (h) present two FN examples which contain light trace of crack features. Although being not easily identifiable due to small size and low contrast, such light crack features are still reportable to engineers. The correct classification in this scenario is challenging but could be improved by increasing the representations of light crack features in the training dataset. To continuously widen the range of classifiable crack-feature types, a human-in-the-loop feedback mechanism needs to be developed. In this feedback mechanism, misclassified or unseen images from new inspections will be fed into the training dataset to update the classification system. It should be noted that while we have demonstrated the utility of our approach for crack feature detection in superheaters using the GoogLeNet structure, the same approach with other network structures could be straightforwardly adopted by other researchers for their own applications on detecting image regions containing defects, anomalies or other regions of interest. The possible candidates for adoption include but are not limited to: VGG [31], ResNet [32], AlexNet [37], and other data-driven techniques such as the Bag-of-Visual-Words approach [3]. A complete comparison of different networks and approaches is beyond the scope of this article.

4.2. Explicability of decision-making

The CNN classification is conventionally considered a black-box operation. Therefore, it is necessary to understand the decision-making mechanism of our classification system. This understanding is useful for engineers to obtain insights into the decision-making process and gain confidence in the classified results. Such interpretability is exceptionally imperative in the highly-regulated nuclear power generation industry. Techniques such as Saliency Map [38] and Gradient-Weighted Class Activation Mapping (Grad-CAM) [39] were proposed to help understand why the CNN makes a specific classification for the input image.

Specifically, the Saliency Map technique is used to produce a pixel-level resolution map of the input image to show which pixels are most important to contribute to the predicted category of the input image. This resolution map is obtained by computing the derivative of the classification score for the predicted category with respect to each pixel in the input image. On the other hand, a localisation map can be produced using the Grad-CAM technique to show what regions of the input image most affect the classification score for the predicted category. The Grad-CAM technique is performed by calculating the gradients of the classification score for the predicted label with respect to the activations of the final feature maps.

In this section, the Saliency Map technique based on guided back-propagation [40] and the Grad-CAM technique are applied to the four TP examples to visualise which regions of the patches are most impactful on the "crack-feature" classification results. The results are given in Fig. 10. Taking Fig. 10(a) for example, the left of Fig. 10(a) shows a testing patch correctly classified as "crack feature". The corresponding saliency map in the middle of Fig. 10(a) reveals which pixels are most impactful to contribute to the "crack-feature" decision. As can be seen, the most impactful pixels correspond to the crack-feature region in the testing patch. On the right of Fig. 10(a), the localisation map is placed on top of the testing patch to show the most important region contributing to the classified "crack-feature" category. As can be seen, the crack-feature region highlighted in red has the greatest impact on the classified "crack-feature" decision. In this example, both the Saliency Map and Grad-CAM techniques suggest that the trained network focuses on the correct region to make the "crack-feature" classification. Such consistency is seen in other TP examples from Fig. 10(b)–10(d). Therefore, it is suggested that the classification system uses the crack-feature regions in patches to make "crack-feature" decisions, which enhances the trustworthiness of classification results.

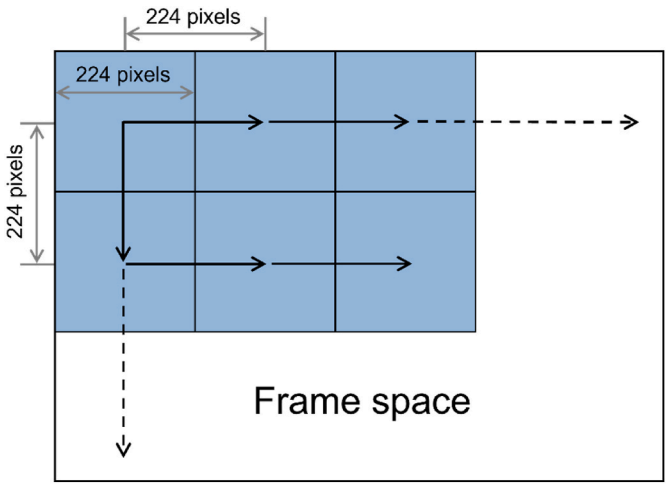


Fig. 11. Full-grid scanning plan of the frame using a detection window sliding in horizontal and vertical directions with the step size of 224 pixels.

4.3. Frame-level detection

With the classification explicability investigated, the trained network is now used to detect crack-feature regions in the frames extracted from the testing videos. The frame-level detection consists of multiple patch-level classifications at different detection window locations in the frame. The full-grid scanning mechanism in Fig. 11 is used for our study. The detection window is of the same size as the patch. The scan starts from the top left of the frame and slides in horizontal and vertical directions with a step size of 224 pixels until the distance from the nearest frame border is smaller than the step size. Any regions classified as "crack feature" are highlighted using red borders.

Fig. 12 shows the examples of classified frames where the crack-feature regions are detected and highlighted. As shown in Fig. 12(a) and (b), the crack-feature regions in the frames can be accurately detected even in the presence of complex background textures formed by the tube hole contours or discolouration marks. Fig. 12(c) and (d) demonstrate that the classification system can correctly locate the crack-feature regions at reflective surface and in the dim lighting condition. More examples are shown from Fig. 12(e) to Fig. 12(o). Therefore, the classification system can be used to perform crack-feature detection at frame level via the full-grid scanning mechanism, even in challenging scenarios with the variant illumination condition and complex background textures. Though fixed bounding-box level detection is chosen in this article, recent research [20,21] demonstrates an increasing trend to investigate pixel-level crack segmentation, which could potentially be useful for our study by providing detailed morphology information of crack features. It is noted that the pixel-level crack segmentation is not pursued in this article. This is because the main objective is to provide decision support to the human-in-the-loop decision-making process, not extending to describe the morphology of crack features. This is done by flagging the presence of crack features through identifying the frames containing crack features in the video footage and reporting the observations of such findings to engineers. For this main target, the proposed patch-level classification is acceptable. Note that a refined search of crack features in frames is possible by reducing the step size of the scanning strategy, at the cost of an increased computational burden. Furthermore, more refined crack feature localisation is possible with bounding box level detection through Faster R–CNN techniques [16] to offer flexibility to the sizes of the bounding boxes. This can be particularly useful in the presence of various sizes and scales of input images. The success of the patch-level and frame-level detection lays a solid foundation to explore the accurate and efficient video-level detection of crack features as the future work. Furthermore, the successful

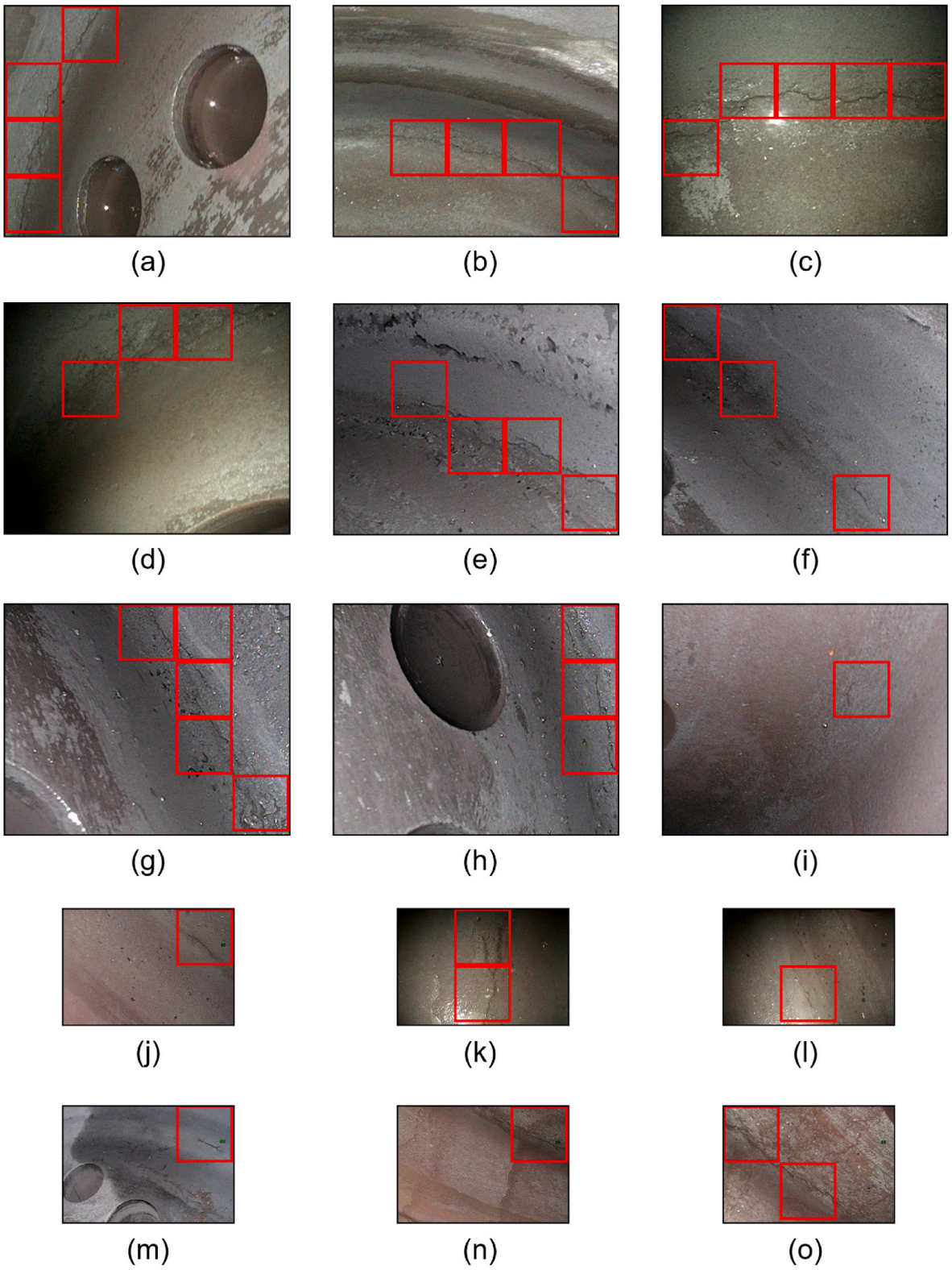


Fig. 12. Examples of classified frames with different resolution sizes, where the crack-feature regions are automatically detected using the classification system. Note that these frames are cropped from the original ones to remove the confidential site location and timing stamps.

development of the automated crack-feature detection system via the deep learning approach is a valuable asset to be transferred to other domains of anomaly detection in pressure vessels.

5. Conclusions

This article has introduced a data-driven framework to support automated remote visual inspection of anomalies in pressure vessel structures, showcased via successfully developing a new automated

detection system of crack features in nuclear power plant superheaters. Specifically, a good accuracy of 92.97 % has been achieved at the patch-level classification. On this basis, the frame-level classification has been implemented using the full-grid scanning strategy to accurately detect the crack-feature regions even in the presence of variant lighting conditions and complex background textures. The explicability of the classification results has been investigated to enhance the trustworthiness of the classification system. Note that whilst this article has used a CNN classifier combined with the sliding window technique to localise defected crack features, bounding box level detection is possible to realise more refined crack localisation, especially when the input images have different sizes and scales. Furthermore, pixel-level crack feature segmentation could also be useful to bring advance in providing more accurate crack feature localisation and its detailed morphology information, through image processing (e.g., Hessian-matrix-based technique) or deep learning based techniques. Comparative studies of different networks and approaches for a wider range of anomaly detection in steam cycle components will also be useful.

Three major aspects of the future work have been identified in this study. Firstly, due to the limited variety of crack features in the provided inspection videos, the future work will focus on developing a human-in-the-loop feedback mechanism to feed new crack-feature types or misclassified results into the training dataset and update the classification system. As a result, the crack-feature variety in the training dataset can be continuously increased to widen the adaptability of the classification system. The second aspect of the future work is the accurate and efficient video-level detection of crack features. Finally, the transfer learning process in this article will be translated to a broader scope of automated anomaly inspections of pressure vessels at nuclear power plants.

CRediT authorship contribution statement

Zhouxiang Fei: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing, Project administration. **Graeme M. West:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing - review & editing. **Paul Murray:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing - review & editing. **Gordon Dobie:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing - review & editing.

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 data that has been used is confidential.

Acknowledgement

The authors would like to thank the industrial partners from Altrad Babcock Ltd., for their valuable support and contributions. The funding support from the Advanced Nuclear Research Centre (ANRC) at the University of Strathclyde, United Kingdom is greatly acknowledged.

References

[1] G. West, P. Murray, S. Marshall, et al., Improved visual inspection of advanced gas-cooled reactor fuel channels, Int. J. Prognostics Health Manag. 6 (2015) 1–11.
[2] P. Murray, G. West, S. Marshall, et al., Automated in-core image generation from video to aid visual inspection of nuclear power plant cores, Nucl. Eng. Des. 300 (2016) 57–66.
[3] M.G. Devereux, P. Murray, G.M. West, A new approach for crack detection and sizing in nuclear reactor cores, Nucl. Eng. Des. 359 (2020) 1–9.
[4] H. Dow, M. Perry, J. Mcalorum, et al., Automation in construction skeleton-based noise removal algorithm for binary concrete crack image, Autom. ConStruct. 151 (2023) 1–10.
[5] G. Dobie, R. Summan, C. MacLeod, et al., Visual odometry and image mosaicing for NDE, NDT E Int. 57 (2013) 17–25.
[6] S. Hosseinzadeh, W. Jackson, D. Zhang, et al., A novel centralization method for pipe image stitching, IEEE Sensor. J. 21 (2021) 11889–11898.
[7] R. Summan, G. Dobie, G. West, et al., The influence of the spatial distribution of 2-D features on pose estimation for a visual pipe mapping sensor, IEEE Sensor. J. 17 (2017) 6312–6321.
[8] A. Stamoulakatos, J. Cardona, C. McCraig, et al., Automatic annotation of subsea pipelines using deep learning, Sensors 20 (2020) 1–15.
[9] K. Abbasi, N.H. Motlagh, M.R. Neamatollahi, et al., Detection of axial crack in the bend region of a pipe by high frequency electromagnetic waves, Int. J. Pres. Ves. Pip. 86 (2009) 764–768.
[10] S. Baby, T. Balasubramanian, R.J. Pardikar, Ultrasonic study for detection of inner diameter cracking in pipeline girth welds using creeping waves, Int. J. Pres. Ves. Pip. 80 (2003) 139–146.
[11] I. Abdel-Qader, O. Abudayyeh, M. Asce, et al., Analysis of edge-detection techniques for crack identification in bridges, J. Comput. Civ. Eng. 17 (2003) 255–263.
[12] E. Zalama, J. Gómez-García-Bermejo, R. Medina, et al., Road crack detection using visual features extracted by Gabor filters, Comput. Civ. Infrastruct. Eng. 29 (2014) 342–358.
[13] A.F. Frangi, W.J. Niessen, R.M. Hoogeveen, et al., Model-based quantitation of 3-D magnetic resonance angiographic images, IEEE Trans. Med. Imag. 18 (1999) 946–956.
[14] C.M. Yeum, S.J. Dyke, Vision-based automated crack detection for bridge inspection, Comput. Civ. Infrastruct. Eng. 30 (2015) 759–770.
[15] D. Kang, S.S. Benipal, D.L. Gopal, et al., Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning, Autom. ConStruct. 118 (2020) 1–13.
[16] Y.J. Cha, W. Choi, G. Suh, et al., Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, Comput. Civ. Infrastruct. Eng. 33 (2018) 731–747.
[17] M.R. Jahanshahi, S.F. Masri, Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures, Autom. ConStruct. 22 (2012) 567–576.
[18] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Comput. Civ. Infrastruct. Eng. 32 (2017) 361–378.
[19] L. Zhang, J. Shen, B. Zhu, A research on an improved Unet-based concrete crack detection algorithm, Struct. Health Monit. 20 (2021) 1864–1879.
[20] W. Choi, Y.J. Cha, SDDNet: real-time crack segmentation, IEEE Trans. Ind. Electron. 67 (2020) 8016–8025.
[21] D.H. Kang, Y.J. Cha, Efficient attention-based deep encoder and decoder for automatic crack segmentation, Struct. Health Monit. 21 (2022) 2190–2205.
[22] A.S. Rao, T. Nguyen, M. Palaniswami, et al., Vision-based automated crack detection using convolutional neural networks for condition assessment of infrastructure, Struct. Health Monit. 20 (2021) 2124–2142.
[23] F.C. Chen, M.R. Jahanshahi, NB-CNN: deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion, IEEE Trans. Ind. Electron. 65 (2018) 4392–4400.
[24] J. Yao, W. Liang, J. Xiong, Novel intelligent diagnosis method of oil and gas pipeline defects with transfer deep learning and feature fusion, Int. J. Pres. Ves. Pip. 200 (2022) 1–17.
[25] Z. Fei, G.M. West, P. Murray, et al., Automated generation of training dataset for crack detection in nuclear power plant components, In: Proceedings Of12th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies, USA (2021, June) 251–258.
[26] C. Szegedy, W. Liu, Y. Jia, et al., Going deeper with convolutions, in: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015, June, pp. 1–9.
[27] C.C. Aggarwal, Neural Networks and Deep Learning: A Textbook, first ed., Springer, New York, 2018.
[28] L. Li, Y. Chen, Z. Shen, et al., Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging, Gastric Cancer 23 (2020) 126–132.
[29] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: Proceedings of 32nd International Conference on Machine Learning, Lille, France, 2015, July, pp. 448–456.
[30] N. Srivastava, G. Hinton, A. Krizhevsky, et al., Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[31] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of 3rd International Conference on Learning Representations, San Diego, USA, 2015, May, pp. 1–14.
[32] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016, June, pp. 770–778.

[33] K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, et al., Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection, Construct. Build. Mater. 157 (2017) 322–330.
[34] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Proceedings of 19th International Conference on Computational Statistics, Paris, France, 2010, August, pp. 177–186.
[35] J. Deng, W. Dong, R. Socher, et al., ImageNet: a large-scale hierarchical image database, in: Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009, June, pp. 248–255.
[36] D.R. Wilson, T.R. Martinez, The need for small learning rates on large problems, in: Proceedings of 2001 International Joint Conference on Neural Networks, Washington, USA, 2001, July, pp. 115–119.
[37] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90.
[38] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: visualising image classification models and saliency maps, in: Proceedings of 2014 Workshop at International Conference on Learning Representations, Banff, Canada, 2014, April, pp. 1–8.
[39] R.R. Selvaraju, M. Cogswell, A. Das, et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2020) 336–359.
[40] J.T. Springenberg, A. Dosovitskiy, T. Brox, et al., Striving for simplicity: the all convolutional net, in: Proceedings of 2015 Workshop at International Conference on Learning Representations, San Diego, USA, 2015, May, pp. 1–14.