AUTOMATED GENERATION OF TRAINING DATASET FOR CRACK DETECTION IN NUCLEAR POWER PLANT COMPONENTS

Zhouxiang Fei, Graeme M. West*, Paul Murray and Gordon Dobie
Department of Electronic and Electrical Engineering
University of Strathclyde
204 George Street, Glasgow, G1 1XW, U.K.
zhouxiang.fei@strath.ac.uk, graeme.west@strath.ac.uk, paul.murray@strath.ac.uk and gordon.dobie@strath.ac.uk

ABSTRACT

Inspection for crack-like features in nuclear power plant components is vital to maintain safe continued operation. The traditional manual-based inspection can discover such features but could be time-consuming and repetitive. Deep learning technique such as convolutional neural network (CNN) offers improved efficiency by automating the inspection of cracks in images and videos. However, a significant overhead of the CNN implementation is the preparation of the large training dataset for training the classification system. The traditional manual-based labelling process costs intensive labour and could become prone to inconsistent labelling standard to annotate cropped patches from inspection videos. As a result, the CNN system may learn irrelevant features for decision-making. This paper presents an automated labelling technique to efficiently generate crack training dataset with consistent labelling standard. The proposed labelling technique is based on binary masks and can generate sufficient labelled patches with customised resolution. A showcase study of automated patch labelling of crack-like features in superheater tube plate upper radius is provided. The result shows that the proposed labelling technique is suitable for the fast creation of training dataset to build autonomous crack detection systems. The only overhead is the generation of binary crack masks which weights much less than the manual labelling burden.

Key Words: Automated labelling technique; crack detection; nuclear power plant inspection; remote visual inspection support

1 INTRODUCTION

Nuclear power plant components are regularly inspected to ensure that safe continued operations are maintained. During the inspection process, certain anomalies such as corrosion and cracks are detected and evaluated regarding the risks of triggering abnormal operations. The visual inspection process could vary depending on examined components. For example, nuclear reactor cores are inspected by filming core surfaces and then analysing the videos, whereas superheaters are visually examined via a boroscope and analysed simultaneously. However, the traditional manual-based visual inspection process is well-known for its time-consuming and repetitive features. This is due to the large amount of data to be analysed (a typical visual inspection of superheaters can last between one ~ two hours).

Automated detection techniques for detecting anomalies were explored to support manual-based visual inspections in various domains. To illustrate some nuclear engineering applications, automated detection of corrosion in used nuclear fuel dry storage canisters [1] was implemented using residual neural networks (ResNets) [2], a specific convolutional neural network (CNN) structure. In [3] crack regions in 360° panoramic images of fuel channels (known as chanoramas in [4], [5]) were detected using the Bag-of-Visual-Words method. A customised CNN system was developed to automatically detect cracks in the mock-up surface of reactors in [6]. On the civil engineering aspect, CNN frameworks were used to automate the detection of concrete cracks [7] and pavement cracks [8].

Amongst various techniques for automated anomaly detection, CNN is a popular candidate due to its wide versatility. For example, in addition to engineering applications, cancer diagnosis was supported using the CNN technique to detect mucosal lesions in [9]. This is due to the data-driven feature of the CNN framework. However, a sufficiently large number of training images of each category are needed before CNN learns to classify. The creation of the training dataset could be an excessively laborious process, as each training image is manually labelled.

To generate the training dataset for crack detection tasks, the current approach is via extracting patches from inspection video frames and manually annotating each patch as “crack” or “non-crack background”. Note that patches rather than full frames are prepared to train the CNN classification system, as patches contain more prominent features of cracks. The patch resolution is chosen to be the required input image size of the first convolutional layer of the CNN system. In this way, upsampling or downsampling which alters original images can be avoided before CNN operations.

Due to the large number of training images to be manually annotated, the labelling process is intrinsically laborious. For example, the training dataset of cracks in [6] consisted of 5,326 annotated patches before data augmentation. In [7], a total of 32,000 patches of cracks and non-crack background were manually cropped from 277 full images and labelled. The pavement crack detection system in [8] was trained using 21,410 crack patches. Therefore, it is necessary to develop an automated labelling technique to efficiently generate the training dataset of cracks. Thus, the overall implementation of developing a crack detection system can be significantly accelerated. In addition to the shortened labelling process, the automated labelling technique can ensure that a consistent labelling standard is used to label crack patches. This is particularly important when labelling patches containing ambiguous crack features (see Fig. 1 for example). Whether these patches are labelled as cracks depends subjectively on engineers and the labelled results could vary significantly. Even if the recorded time log of crack features in the inspection video is provided, manual labelling with a consistent standard could still not be guaranteed.

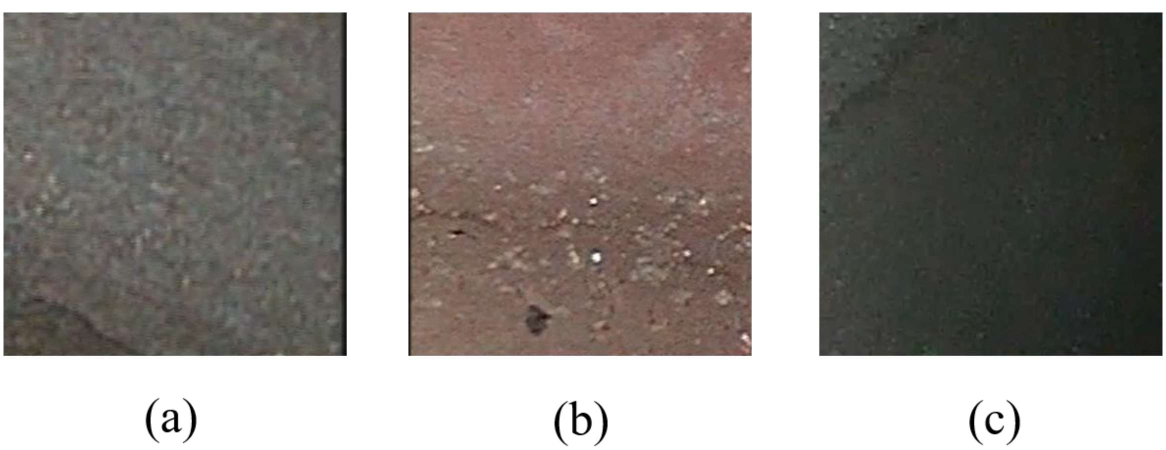


Figure 1. Patches containing ambiguous crack features in superheaters. (a) Crack features only occupying a small part of the patch and not centred. (b) Crack features with light trace. (c) Crack features with dark illumination.

Another reason for developing the automated labelling technique is that the generation of training dataset with a different patch resolution can be automatically performed if needed, without needing to repeat the tedious manual labelling process. This is particularly useful when training a new CNN framework with a different input image size. To demonstrate the variation of the input image size of different networks: the classification systems in [6] and [7] used the input image sizes of 120×120×3 and 256×256×3 pixels, respectively. Therefore, evaluating the crack detection performances of different CNN structures can be accelerated significantly using the automated labelling technique. Note that alternatively it is possible to change the patch resolution by upsampling or downsampling to suit the input image size of the CNN framework. However, such operation alters the original information in the patch.

In summary, an automated labelling technique with a consistent labelling standard is beneficial to creating the training dataset of automated crack detection systems. In this paper, an automated labelling technique based on binary masks is proposed to automatically generate the training dataset to train autonomous crack detection frameworks. To the best of authors' knowledge, little research work had focused on this field and our proposed technique is the first successful one to significantly accelerate the construction of the crack training dataset. This paper presents a showcase study of the automated training dataset generation for stress corrosion crack feature detection in the tube plate upper radius of nuclear power plant superheaters. Superheaters are used in the thermal side of nuclear power plants to overheat the saturated steam. The tube plate upper radius is the circumferential area at the bottom of the superheater. Please see Fig. 2 for the structural view of the superheater. The remainder of this paper is organised as follows: Section 2 outlines the detailed methodology of the automated labelling process. Section 3 presents the results and discussion of the training dataset generated using the automated labelling technique. The conclusions of this paper are given in Section 4.

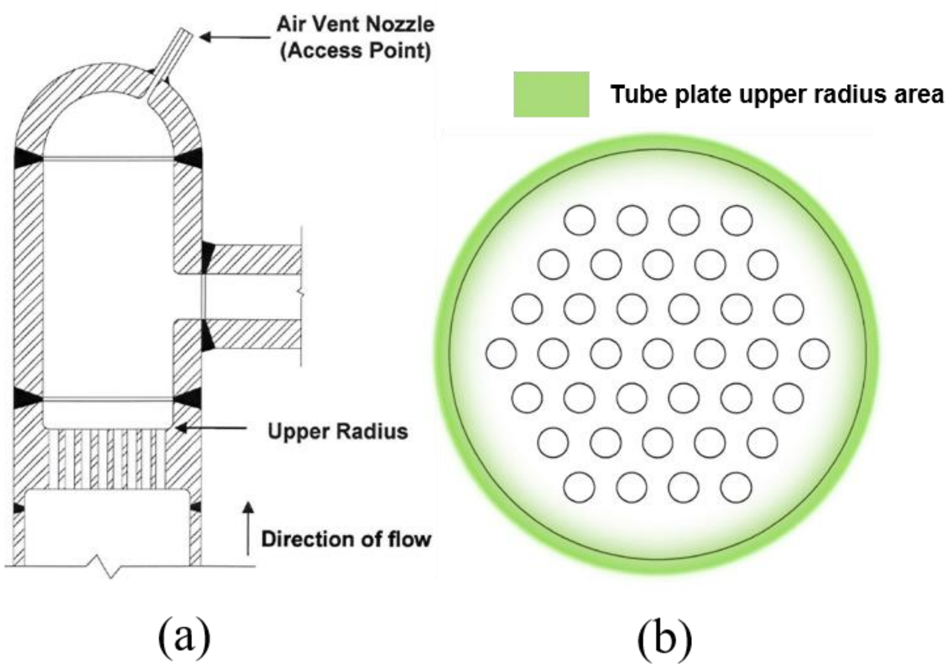


Figure 2 [10]. Detailed structural view of the superheater. (a) The cross-sectional view of the superheater. (b) The top view of the tube plate, with the upper radius area highlighted in green.

2 AUTOMATED LABELLING METHODOLOGY

As discussed in Section 1, the conventional construction of the crack training dataset is subject to excessive labour cost and inconsistent labelling standards. To overcome these challenges, an automated process of creating the training dataset needs to be developed to reduce the manual labour cost involved in the implementation of the automated crack detection system. This section presents a detailed description of the automated labelling methodology. Note that in this paper, potential crack indications are referred to as crack-like features. This is because potential cracks are identified and recorded as crack-like features by engineers during the remote visual inspection process. The records are then handed to mechanical specialists and metallurgists to decide if the recorded crack-like features are actual cracks. This paper focuses on the crack-like features in the remote visual inspection phase.

The automated labelling technique is based on binary masks and can be decomposed into the following five steps:

Step 1: Identify and extract frames containing ground-truth crack-like features from a selection of inspection videos. Note that some inspection videos need to remain untouched in this step for the validation purpose during the classification phase. The classification phase is beyond the scope of this paper.

Step 2: Perform pixel-level labelling to create binary crack masks for the extracted ground-truth frames in Step 1. During the pixel-level labelling of each frame, one needs to label the pixel constituting the ground-truth crack-like feature as "1", and annotate the pixels outside the crack-like indication as "0".

After performing pixel-level labelling, the resulting frame is referred to as the binary crack mask of the ground-truth frame. An example of the ground-truth frame and its corresponding binary crack mask are given in Fig 3.

Step 3: Generate a large number of patches with specific resolution at random or predefined positions in each extracted ground-truth frame. Generating patches at random positions is recommended as randomness can be introduced into the training dataset to develop robust detection systems.

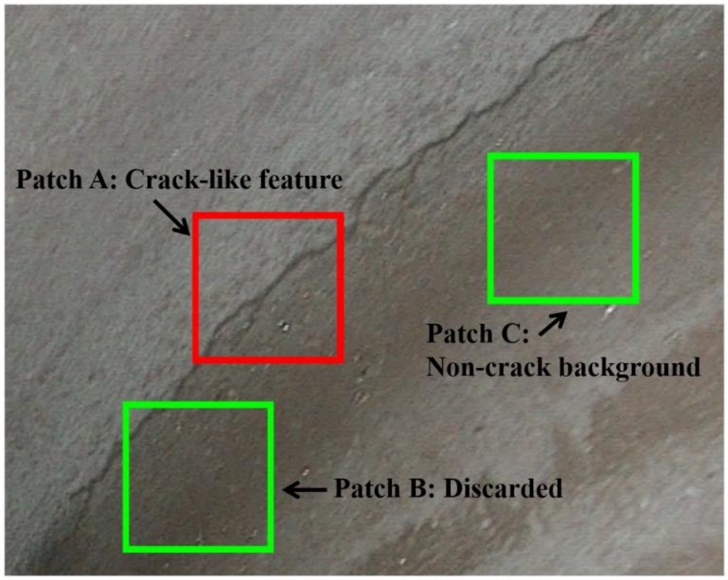
Step 4: For each patch, obtain the number of crack-labelled pixels (i.e., labelled as "1") within the same patch region in the corresponding binary crack mask.

The number of crack-labelled pixels can be used to calculate a measure of the crack-like feature intensity in the patch. As the patch resolution could vary for different CNN structures, one option is to calculate the ratio of crack-labelled pixel number to the overall pixel number in the patch and compare this ratio to a predefined threshold.

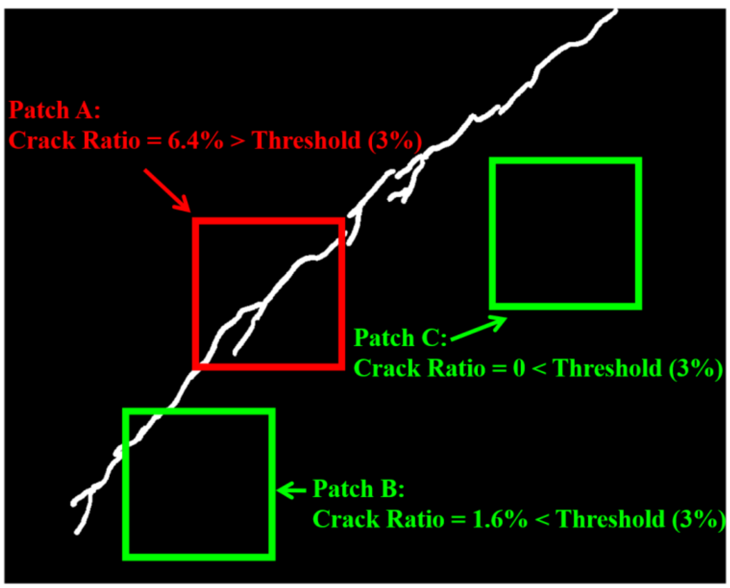
This predefined threshold is given by the minimum measure of the crack-like feature intensity that presents a prominent crack-like indication in the patch.

Step 5: Label the patch as "crack-like feature" if the threshold is exceeded. Annotate the patch as "non-crack background" if no crack-labelled pixels are within the patch region. Otherwise the patch is discarded. Please see Fig. 3 for examples of patch labelling using a predefined threshold of 3%.

In summary, the benefits of this automated labelling technique are threefold. Firstly, a sufficient number of patches can be automatically generated and labelled to construct the training dataset. Secondly, a consistent labelling standard is applied to label each patch. Thirdly, the generation of the training dataset with a different resolution can be performed efficiently. The only overhead of this automated labelling technique is the creation of binary crack masks, and costs far less than the manual labelling burden.



(a)



(b)

Figure 3. (a) Example of a ground-truth frame at the superheater tube plate upper radius. Patches are generated at random positions and automatically labelled as "crack-like feature" or "non-crack background" based on the crack ratios in Fig. 3(b). (b) The binary crack mask of the ground-truth frame in Fig. 3(a). The pixels labelled as "1" constitute the crack-like indication (i.e., the white region). The non-crack background (i.e., the black region) consists of the pixels labelled as "0". The labelling of each patch is given by comparing the crack ratio of the patch to a predefined threshold of 3%.

3 APPLICATIONS OF AUTOMATED LABELLING TECHNIQUE

In this study, a total of 362 ground-truth frames are extracted from 16 inspection videos (i.e., 60% of all the inspection videos) provided by the industrial collaborator. Other inspection videos are reserved for the validation and testing of the autonomous detection system of crack-like features, which is beyond the scope of this paper. The indices of the ground-truth frames in the inspection video are obtained using the time log of crack-like features recorded in the associated inspection report. During the automated labelling process, it is important to keep the number of "crack-like feature" patches close to the number of "non-crack background" patches, in order to obtain a balanced training dataset. It is also beneficial to scatter the patch locations in the extracted ground-truth frames to reduce overlaps between patches.

The resolution of the inspection video could be either 720×576×3 pixels or 1920×1080×3 pixels, depending on each video. The patch resolution is set to 224×224×3 pixels which is the input image size of various CNN structures such as GoogLeNet [11] and VGG series [12]. A predefined threshold of 3% is used in Step 4 of the automated labelling process to distinguish between "crack-like feature" and "non-crack background" patches. The automated labelling algorithm is implemented in MATLAB, based on a CPU of 2.8 GHz and RAM of 32 GB. TABLE I shows the algorithm computation time and the detail of the training dataset generated using the automated labelling technique. Note that the computation time in TABLE I does not include the time spent on creating binary crack masks. In our experience, creating a binary crack mask takes around 20 seconds.

Table I. Detail of the training dataset generated using the automated labelling technique.

Patch size	Number of "crack-like feature" patches	Number of "non-crack background" patches	Computation time (s)
224 x 224 x 3	1995	1876	119

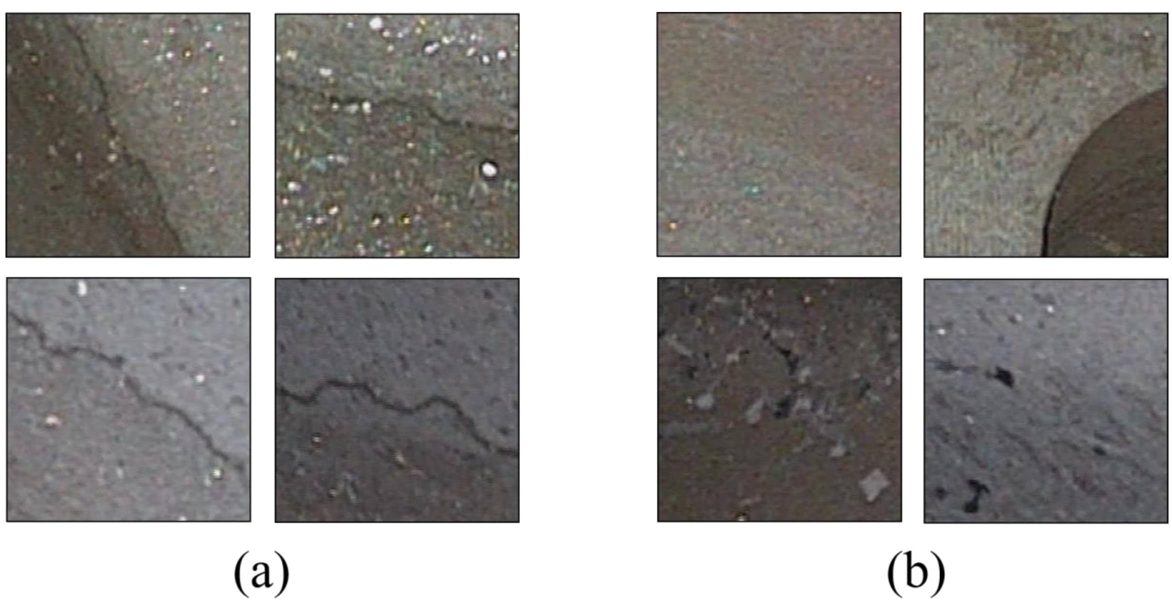


Figure 4. (a) Examples of patches automatically labelled as "crack-like feature". (b) Examples of patches automatically labelled as "non-crack background".

As shown in TABLE I, a training dataset consisting of 1,995 “crack-like feature” patches and 1,876 “non-crack background” patches from 16 inspection videos can be automatically generated in less than two minutes. In contrast, the manual labelling of patches for one inspection video typically exceeds one hour. Therefore, the training dataset for autonomous crack detection systems can be efficiently generated using the automated labelling technique. This advantage becomes particularly useful when a new training dataset with a different patch size needs to be generated. The only overhead of the automated labelling process is the creation of binary crack masks, which is negligible compared to the conventional manual labelling process. Fig. 4 shows the examples of training patches that are automatically generated and labelled using the automated labelling technique.

4 CONCLUSIONS

This paper has introduced a new automated labelling technique based on binary masks to generate training datasets for autonomous crack detection systems. The automated labelling of crack features in the superheater tube plate upper radius has been demonstrated as a showcase study. The results have shown that the training datasets of “crack-like feature” and “non-crack background” categories can be efficiently generated using the automated labelling technique. The only overhead is the creation of binary crack masks of the extracted ground-truth frames, but is negligible compared to the laborious and repetitive manual labelling cost. This automated labelling technique can be straightforwardly extended to construct the training dataset of other autonomous anomaly (e.g. corrosion) detection systems to support nuclear power plant inspection.

5 ACKNOWLEDGMENTS

This work was carried out as part of a collaborative research project supported by Doosan Babcock. The authors would like to thank Christopher Kelly from Doosan Babcock, for his invaluable discussion.

6 REFERENCES

1. T. Papamarkou, H. Guy, B. Kroencke, J. Miller, P. Robinette, D. Schultz, J. Hinkle, L. Pullum, C. Schuman, J. Renshaw, and S. Chatzidakis, “Automated detection of corrosion in used nuclear fuel dry storage canisters using residual neural networks,” *Nuclear Engineering and Technology*, **Vol.53**, no.2, pp.657-665 (2021).
2. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, Jun. 2016, pp.770-778.
3. M. G. Devereux, P. Murray, and G. M. West, “A new approach for crack detection and sizing in nuclear reactor cores,” *Nuclear Engineering and Design*, **Vol.359**, pp.1-11 (2019).
4. G. West, P. Murray, S. Marshall, and S. McArthur, “Improved visual inspection of advanced gas-cooled reactor fuel channels,” *International Journal of Prognostics and Health Management*, **Vol.6**, pp.1-11 (2015).
5. P. Murray, G. West, S. Marshall, and S. McArthur, “Automated in-core image generation from video to aid visual inspection of nuclear power plant cores,” *Nuclear Engineering and Design*, **Vol.300**, pp.57-66 (2016).
6. F. C. Chen and M. R. Jahanshahi, “NB-CNN: deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion,” *IEEE Transactions on Industrial Electronics*, **Vol.65**, no.5, pp.4392-4400 (2018).

7. Y. J. Cha, W. Choi, and O. Büyüköztürk, “Deep learning-based crack damage detection using convolutional neural networks,” *Computer-Aided Civil and Infrastructure Engineering*, **Vol.32**, no.5, pp.361-378 (2017).

8. B. Li, K. C. P. Wang, A. Zhang, E. Yang, and G. Wang, “Automatic classification of pavement crack using deep convolutional neural network,” *International Journal of Pavement Engineering*, **Vol.21**, no.4, pp.457-463 (2020).

9. L. Li, Y. Chen, Z. Shen, X. Zhang, J. Sang, Y. Ding, X. Yang, J. Li, M. Chen, C. Jin, C. Chen, and C. Yue, “Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging,” *Gastric Cancer*, **Vol.23**, no.1, pp.126-132 (2020).

10. C. Kelly, “Superheater header tube sheet, tube sheet radius and boiler tube inspections”, *EDF ENERGY internal report*, Technique Sheet Number: BI222/HRA-HYA/004, Revision: 007 (2018).

11. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going deeper with convolutions,” *Proceeding of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, Jun. 2015, pp.1-9.

12. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Proceeding of 2015 International Conference on Learning Representations*, San Diego, USA, May 2015, pp. 1-14.