Proceedings of the Seventeenth International Conference on
Civil, Structural and Environmental Engineering Computing
Edited by: P. Iványi, J. Kruis and B.H.V. Topping
Civil-Comp Conferences, Volume 6, Paper 5.2
Civil-Comp Press, Edinburgh, United Kingdom, 2023
doi: 10.4203/ccc.6.5.2

An Unsupervised Crack Detection Approach Based on Dliding Window Variational Autoencoder

Y.H. Wei[1,2] and Y.Q. Ni[1,2]

[1]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[2]Hong Kong Branch of Chinese National Engineering Research Center on Rail Transit Electrification and Automation, Hong Kong

Abstract

The current investigation presents a novel approach to detect cracks using the variational autoencoder (VAE). In this method, the input image is first divided into multiple segments using sliding windows and then fed into the VAE sequentially. The use of sliding windows effectively limits the number of neural nodes in the input layer of the VAE, which enhances the method's robustness. Additionally, the sliding window technique allows for the image information to be viewed as a time series, with cracks being treated as anomalies in the time series. By using the sliding window VAE (SW-VAE) with robust properties, such anomalies can be discarded during the reconstruction process. As a result, the detection of cracks can be achieved by comparing the difference between the input and output of the SW-VAE. Notably, this technique does not require positive sample training or learning image features specific to cracks, thus avoiding the challenge posed by the lack of training data or imbalanced datasets.

Keywords: crack detection, variational autoencoder, sliding windows, serialized input, anomaly detection, robustness, unsupervised learning.

1 Introduction

The existence of cracks within a structural system may lead to a reduction in its stiffness, ultimately resulting in significant hazards [1, 2]. Thus, the detection and identification of cracks are crucial topics in the domain of structural health monitoring (SHM). Traditional manual methods have demonstrated limitations in identifying cracks in a timely manner. In light of this, given the contemporary era of rapid advancement in artificial intelligence, it is imperative to propose automatic crack identification techniques based on deep learning.

Generally speaking, deep learning based crack detection approaches can be classified into two main categories: supervised learning and unsupervised learning. The supervised learning based crack identification methods have demonstrated outstanding performance when a sufficiently large dataset is available for model training [3]. Nonetheless, in situations where training data is inadequate, or there is a substantial imbalance between positive and negative samples, supervised learning based crack identification methods tend to face difficulties in achieving satisfactory outcomes [4, 5]. Therefore, the development of unsupervised techniques for crack detection holds significant value for various application scenarios.

Studies have shown that unsupervised learning can be effectively utilized for identifying outliers in time series data. The presence of outliers can disrupt the temporal structure between data points in a time series. However, a probabilistic model with robust properties can mitigate the impact of occasional outliers [6]. Moreover, due to environmental factors and the influence of data acquisition equipment, noise often accompanies the collected time series. Therefore, treating the collected time series as random variables and modeling them with a probabilistic model is more reasonable.

Based on the analysis presented above, this research aims to transform images into sequence information that mimics time-series data. As a result, cracks present in the images can be treated as outliers and eliminated using a robust probabilistic model through unsupervised learning.

As a form of deep learning algorithm, the variational autoencoder (VAE) [7] exhibits both unsupervised learning capabilities and probabilistic modeling properties. Furthermore, the presence of Kullback-Leibler (KL) divergence in the loss function confers excellent regularization properties upon the VAE model, rendering it robust to outliers. Hence, this study employs VAE to perform probabilistic modeling of input image information. To serialize image information, we propose a sliding window strategy that segments image data into fragments, which are subsequently input into the VAE model in a sequential manner. This way, crack information in the image can be treated as an outlier in the time series and discarded by the robust sliding window VAE (SW-VAE). Finally, crack information can be revealed by comparing the input image with the reconstructed image output by the SW-VAE.

2 Deriving the loss function of SW-VAE

The VAE algorithm treats images $\boldsymbol{x}$ as random variables generated through latent variables $\boldsymbol{z}$. According to the Bayesian law, the posterior of latent variables can be obtained through the following:

$$p(\boldsymbol{z}\mid\boldsymbol{x})=\frac{p(\boldsymbol{x}\mid\boldsymbol{z})\,p(\boldsymbol{z})}{p(\boldsymbol{x})}. \tag{1}$$

The prior $p(\boldsymbol{z})$ in VAE is set as the standard normal distribution. However, since the marginal likelihood $p(\boldsymbol{x})$ and the likelihood $p(\boldsymbol{x}\mid\boldsymbol{z})$ are both unknown, the posterior is intractable in many cases [7]. In view of this, an additional distribution $q(\boldsymbol{z}\mid\boldsymbol{x})$ is introduced for approaching the true posterior $p(\boldsymbol{z}\mid\boldsymbol{x})$. Take the additional $q(\boldsymbol{z}\mid\boldsymbol{x})$ into the log marginal likelihood $p(\boldsymbol{x})$, and we can get:

$$\begin{aligned}\log p(\boldsymbol{x}) &= \int q(\boldsymbol{z}\mid\boldsymbol{x})\log p(\boldsymbol{x})d\boldsymbol{z}\\ &= \int q(\boldsymbol{z}\mid\boldsymbol{x})\log\frac{p(\boldsymbol{x},\boldsymbol{z})q(\boldsymbol{z}\mid\boldsymbol{x})}{q(\boldsymbol{z}\mid\boldsymbol{x})\,p(\boldsymbol{z}\mid\boldsymbol{x})}\\ &= \int q(\boldsymbol{z}\mid\boldsymbol{x})\log\frac{p(\boldsymbol{x},\boldsymbol{z})}{q(\boldsymbol{z}\mid\boldsymbol{x})}d\boldsymbol{z}+KL(q(\boldsymbol{z}\mid\boldsymbol{x})\,\|\,p(\boldsymbol{z}\mid\boldsymbol{x})).\end{aligned} \tag{2}$$

Omit the nonnegative KL term on the right hand side (RHS), then:

$$\begin{aligned}\log p(\boldsymbol{x}) &\geq \int q(\boldsymbol{z}\mid\boldsymbol{x})\log\frac{p(\boldsymbol{x},\boldsymbol{z})}{q(\boldsymbol{z}\mid\boldsymbol{x})}d\boldsymbol{z}\\ &\geq \int q(\boldsymbol{z}\mid\boldsymbol{x})\log p(\boldsymbol{x}\mid\boldsymbol{z})d\boldsymbol{z}+\int q(\boldsymbol{z}\mid\boldsymbol{x})\log\frac{p(\boldsymbol{z})}{q(\boldsymbol{z}\mid\boldsymbol{x})}d\boldsymbol{z}\\ &\geq E_{q(\boldsymbol{z}|\boldsymbol{x})}[\log p(\boldsymbol{x}\mid\boldsymbol{z})]-KL(q(\boldsymbol{z}\mid\boldsymbol{x})\,\|\,p(\boldsymbol{z})).\end{aligned} \tag{3}$$

The terms on the RHS are called evidence lower bound (ELBO). Combining equation (2) with inequality (3), it can be observed that maximizing the ELBO can push the estimate $q(\boldsymbol{z}\mid\boldsymbol{x})$ towards the true posterior $p(\boldsymbol{z}\mid\boldsymbol{x})$. The VAE combines the encoding process with $q(\boldsymbol{z}\mid\boldsymbol{x})$ and the decoding process with $p(\boldsymbol{x}\mid\boldsymbol{z})$, resulting in:

$$\log p(\boldsymbol{x})\geq E_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}\mid\boldsymbol{z})]-KL(q_{\boldsymbol{\phi}}(\boldsymbol{z}\mid\boldsymbol{x})\,\|\,p(\boldsymbol{z})). \tag{4}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ respectively represent the parameters of the encoder and decoder in the VAE. To accommodate the properties of stochastic gradient descent in deep learning, the loss function needs to take the negative of the ELBO and minimize it:

$$L=-E_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}\mid\boldsymbol{z})]+KL(q_{\boldsymbol{\phi}}(\boldsymbol{z}\mid\boldsymbol{x})\,\|\,p(\boldsymbol{z})). \tag{5}$$

As the image $\boldsymbol{x}$ is partitioned into several parts $\sum_1^n \boldsymbol{x}_w^i$ with a sliding window, the integrated loss function is represented as:

$$L=\sum_1^n \{-E_{q_{\phi}(z_w^i|x_w^i)}[\log p_{\theta}(x_w^i \mid z_w^i)] + KL(q_{\phi}(z_w^i \mid x_w^i) \| p(z_w^i))\} . \tag{6}$$

3 Crack Detection with SW-VAE

In this study, SW-VAE is employed to detect cracks in the rotating binaural of the high-speed railway catenary system. The grayscale image of the rotating binaural component is shown in Figure 1. Sliding the window down the lengthwise direction of the image x, moving it down one pixel at a time. Assuming the length of the window is L_w and the width of the image is W, the size of the image extracted by the window in each step is $x_w^i = L_w \times W$. After being extracted by the sliding window, x_w^i is first compressed into latent variables through the encoder, and then, after sampling, the latent variables are input into the decoder to output the reconstructed $\hat{x}_w^i$. Each pixel of the reconstructed $\hat{x}_w^i$ is represented by an individual normal distribution. For convenience, only mean values $\mu_{\hat{x}_w^i}$ of $\hat{x}_w^i$ are used to reconstruct the image. Since there is a lot of overlap between the reconstructed $\mu_{\hat{x}_w^i}$, we only combine the middle line of each $\mu_{\hat{x}_w^i}$ into the reconstructed image $\hat{x}$. The parts at approximately one-half of the L_w length from the beginning and end of the image are scanned less frequently by the sliding window mechanism than other parts. The parts about half the length of the sliding window at the beginning and end of the image are scanned less frequently by the sliding window mechanism than other parts. Therefore, these parts can be filled with zeros or left for unimportant components.

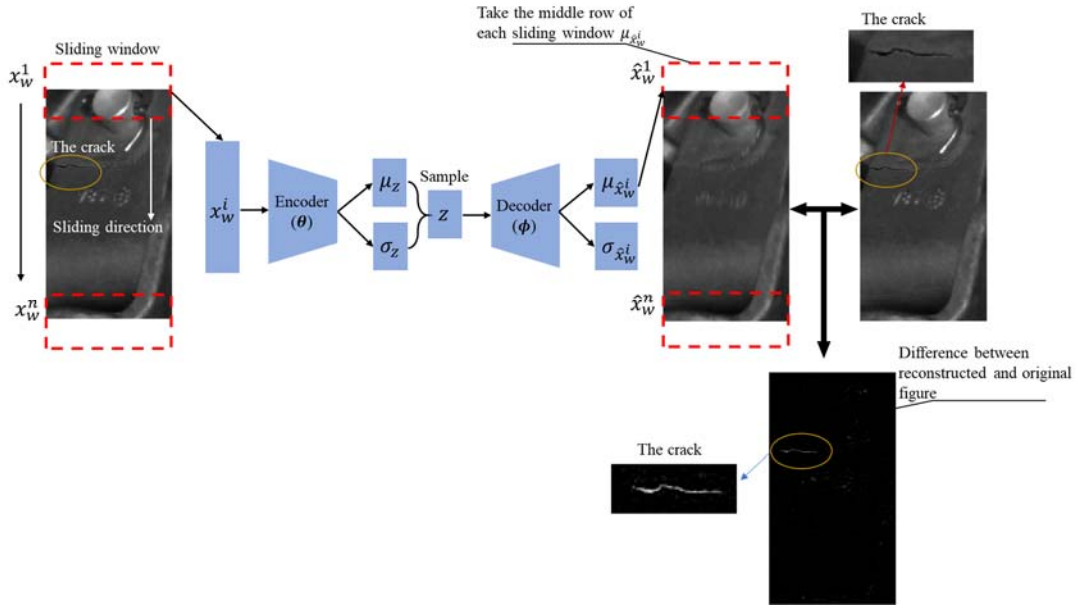

Figure 1: SW-VAE for crack detection on rotating binaural.

Besides cracks, we noticed that the SW-VAE may also remove the highlight of the grayscale image as outliers. However, as the pixel values of cracks are typically low

and those of the highlighted parts are high, the difference between the reconstructed and original images is positive in the crack area but negative in the highlighted area. To effectively display the location and shape of the cracks, we retain only nonnegative values during interpolation. It is found from Figure 1 that the crack of the rotating binaural is displayed by subtracting the original image from the reconstructed image.

The size of the sliding window in SW-VAE has a significant impact on the reconstructed image. If the window size is too small, the overlap between windows is reduced and the correlation between them is weakened, which affects the validity of treating them as a time series. On the other hand, if the window size is too large, the number of parameters in the input layer of the VAE increases, which may lead to overfitting of image information and make it harder for cracks to be identified as outliers and automatically removed.

Figure 2 displays the reconstruction performance of SW-VAE with various sliding window sizes. Since the reconstruction is carried out by selecting the middle row of each sliding window to form the image, the window sizes are set to odd numbers. It is evident that SW-VAE fails to achieve the desired performance when the window size is either too small or too large. Setting the window size to 1/3~1/5 of the overall image size produces satisfactory results for SW-VAE. The performance of SW-VAE with Lw=101, which is shown in Figure 1, demonstrates the effectiveness of this window size setting.

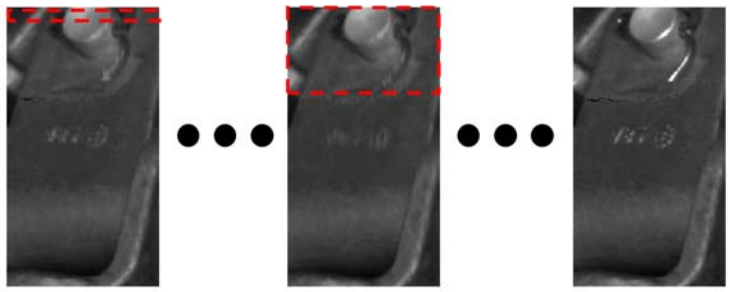

Figure 2: Reconstruction performance of SW-VAE with various window sizes.

4 Conclusions and Contributions

This article uses sliding windows to transform image information into time series data, thereby treating cracks in the image as outliers in the sequence. Due to the robustness of SW-VAE, the crack information in these serialized images is treated as outliers and automatically removed. As a result, the cracks can be revealed by comparing the reconstructed image from SW-VAE with the original image.

SW-VAE has the advantage of not requiring the use of datasets containing cracks for model training, nor does it require the model to learn the normal pattern of images

through positive samples without crack information. As a result, SW-VAE provides a viable option for crack detection when training datasets are scarce.

Acknowledgements

The authors would like to appreciate the funding support by the Innovation and Technology Commission of Hong Kong SAR Government to the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center (Grant No. K-BBY1).

References

[1] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, D.J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks", Journal of Sound and Vibration, 388, 154-170, 2017.

[2] B.F. Spencer, V. Hoskere, Y. Narazaki, "Advances in computer vision-based civil infrastructure inspection and monitoring", Engineering, 5(2), 199-222, 2019.

[3] Z. Fan, Y. Wu, J. Lu, W. Li, "Automatic pavement crack detection based on structured prediction with the convolutional neural network", arXiv preprint arXiv:1802.02208, 2018.

[4] X. Cui, Q. Wang, J. Dai, Y. Xue, Y. Duan, "Intelligent crack detection based on attention mechanism in convolution neural network", Advances in Structural Engineering, 24(9), 1859-1868, 2021.

[5] Y.H. Wei, Y.Q. Ni, (2019, January). "Variational autoencoder-based approach for rail defect identification", in "12th International Workshop on Structural Health Monitoring: Enabling Intelligent Life-Cycle Health Management for Industry Internet of Things", DEStech Publications Inc., Lancaster, USA, 2818-2824, 2019.

[6] Y.H. Wei, Y.W. Wang, Y.Q. Ni, "Online railway wheel defect detection under varying running-speed conditions by multi-kernel relevance vector machine", Smart Structures And Systems, 30(3), 303-315, 2022.

[7] D.P. Kingma, M. Welling, "Auto-encoding variational bayes", arXiv preprint arXiv:1312.6114, 2013.