NGHIÊN CỨU NHẬN DIỆN HƯ HỎNG BẤT THƯỜNG TRONG DẦM THÉP BẰNG VARIATIONAL AUTOENCODER (VAE)

STUDY ON ANOMALY DETECTION OF STEEL BEAM USING VARIATIONAL AUTOENCODER (VAE)

Lưu Thanh Tùng[1,2], Lê Quý Phương[1,2], Nguyễn Đức Thiên Ân[1,2], Đặng Long Khang Huy[1,2]
[1]Khoa Cơ khí, Trường Đại học Bách Khoa, Đại học Quốc gia Thành phố Hồ Chí Minh (HCMUT)
[2]Đại học Quốc gia Thành phố Hồ Chí Minh (VNU-HCM)

TÓM TẮT

Trong lĩnh vực bảo trì và bảo dưỡng, việc tiến hành khảo sát và điều tra để xác định và phân tích các hư hỏng đóng vai trò quan trọng trong quá trình sản xuất. Gần đây, đã có một lượng lớn nghiên cứu về các kỹ thuật phát hiện bất thường nhằm tránh các hạn chế liên quan đến chi phí và nhãn dữ liệu. Do đó, chúng tôi đề xuất sử dụng một phương pháp học máy không giám sát được biết đến là Variational AutoEncoder (VAE) để chẩn đoán các hư hỏng của dầm, một chi tiết cơ bản trong lĩnh vực cơ khí. Bằng cách tiến hành thực nghiệm trên các mô hình dầm, chúng tôi có thể đánh giá và biểu diễn các hư hỏng một cách hiệu quả trong không gian tiềm ẩn. Qua đó, nghiên cứu này đã cho thấy tính hiệu quả và khả thi của không gian ẩn trong mô hình VAE, có thể ứng dụng với nhiều loại hư hỏng và nhiều chi tiết khác trong công nghiệp.

Từ khóa: *Hư hỏng của dầm; Nhận diện bất thường; VAE; Không gian tiềm ẩn.*

ABSTRACT

In the field of maintenance and upkeep, conducting surveys and investigations to identify and analyze faults plays a crucial role in the production process. Recently, there has been a significant amount of research on anomaly detection techniques to overcome limitations related to cost and data labeling. Therefore, we propose the use of an unsupervised machine learning method known as Variational AutoEncoder (VAE) for diagnosing beam faults, a fundamental component in the mechanical industry. By conducting experiments on beam models, we can effectively evaluate and represent faults in the latent space. This research demonstrates the effectiveness and feasibility of the latent space in the VAE model, which can be applied to various types of faults and different components in the industry.

Keywords: *Damage of beam; Anomaly detection; VAE; Latent space.*

1. ĐẶT VẤN ĐỀ

Việc điều tra, khảo sát hư hỏng luôn là mối quan tâm trong lĩnh vực bảo trì và bảo dưỡng máy móc, trong đó việc khảo sát hư hỏng do mỏi thường khó để nhận biết sự xuất hiện của các vết nứt trong vật liệu. Những năm gần đây, các phương pháp chẩn đoán sử dụng trí tuệ nhân tạo (AI) thường liên quan đến tác vụ phân loại hoặc hồi quy (classification and regression) [1], [2], tuy nhiên để thu thập dữ liệu hư hỏng đòi hỏi công sức và chi phí đáng kể. Các dữ liệu cho các trạng thái hỏng hiếm và đôi khi không thể có được [3]. Trong khi đó, việc phân loại mức độ hư hỏng do các yếu tố khác nhau trong kết cấu cũng khó khăn. Dữ liệu được gán nhãn thường không có sẵn, khiến các thuật toán không hiệu quả [4].

Trong lĩnh vực trí tuệ nhân tạo (AI), nhận diện bất thường (anomaly detection) đã được chú ý đến nhờ ưu điểm tránh những hạn chế của các phương pháp đã nêu trên [5]. Xirui Ma và cộng sự đã chỉ ra rằng học không giám sát có thể chỉ ra những hư hỏng bất thường có thể phát hiện từ các đặc trưng trích xuất từ dữ liệu dao động mà không cần phải dán nhãn [6]. Ở nghiên cứu này, chúng tôi đưa ra một phương pháp có thể nhận biết được hư hỏng sử dụng học máy không giám sát (unsupervised learning) có tên là Variational AutoEncoder (VAE) [7]. Các thiết kế, tính toán về quy trình lấy mẫu, xử lý số liệu bằng Variational Mode Decomposition (VMD) [8] và thông số mạng cũng được trình bày. Ngoài ra, miền không gian ẩn (latent space) trích xuất từ dữ liệu rung động thông qua VAE sẽ được phân tích, đánh giá và biểu diễn để thể hiện hư hỏng so với dầm thép nguyên vẹn.

2. PHƯƠNG PHÁP

2.1. Dữ liệu thu thập

Tín hiệu liên tục từ cảm biến được đọc và chuyển thành tín hiệu số thông qua card âm thanh của máy tính bằng phần mềm, từ đó được lưu dưới dạng tệp có đuôi mở rộng *.wav. Các thông số lấy mẫu được cài đặt ngay trên phần mềm như tần số lấy mẫu, độ phân giải của tín hiệu. Tần số lấy mẫu được chọn là 16000 Hz.

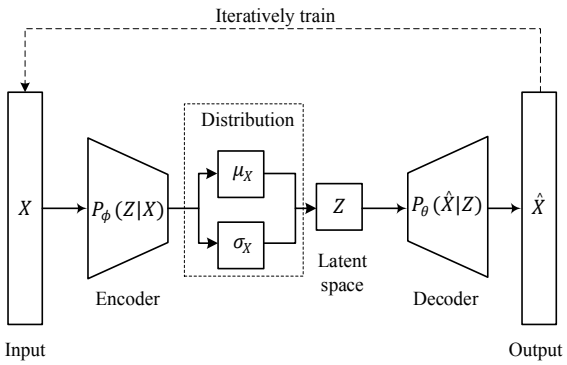
Có hai vấn đề cần quan tâm khi chọn tần số lấy mẫu. Đầu tiên là tần số lấy mẫu phải lớn hơn gấp hai lần tần số cao nhất của tín hiệu. Thứ hai, tần số lấy mẫu không quá lớn vì nếu quá lớn đồng nghĩa máy tính sẽ xử lý nhiều hơn cho mỗi giây. Vì vậy để chọn tần số lấy mẫu cần cho chạy thử với các tần số khác nhau ứng với dãy tần số giới hạn của cảm biến. Sau đó thông qua phổ FFT của tín hiệu và khả năng xử lý của máy tính (đo bằng thời gian) để chọn ra tần số thích hợp.

2.2. Xử lý tín hiệu

Đối với dạng tín hiệu dạng sóng, kỹ thuật trích xuất các hàm có tên là Intrinsic Mode Function [9] (IMF), theo đó một tín hiệu bao gồm nhiều mode dao động, mỗi dao động là một hàm IMF mang ý nghĩa là các dao động riêng lẻ hợp thành tín hiệu, mỗi phần mang một đặc trưng là tần số và biến đổi biên độ riêng. Do đó, ở nghiên cứu này, phương pháp Variational Mode Decomposition (VMD) [8] được sử dụng để trích xuất những mode tín hiệu khác nhau.

Thuật toán VMD là một quá trình tối ưu hóa thông qua các vòng lặp. Các IMF xác định bằng cách tìm các tần số trung tâm, thông qua việc tối thiểu hóa hàm chi phí được tính toán thông qua chuẩn Euclid (L2) giữa tín hiệu gốc và tín hiệu được ước tính. Các tham số sẽ được cập nhật thông qua phép nhân tử Larrange. ☞

2.3. Variational AutoEncoder



Hình 1. Sơ đồ nguyên lý mạng VAE. Dữ liệu vào X được đưa qua bộ mã hóa $P_\phi(Z|X)$ để đưa về miền ẩn rồi từ đó truyền qua bộ giải mã $P_\theta(\hat{X}|Z)$ để sinh ra dữ liệu tái tạo $\hat{X}$.

Variational Autoencoder [8], hay gọi tắt là VAE, là một loại mạng nơron nhân tạo được sử dụng trong học máy không giám sát. Dữ liệu ban đầu thông qua VAE sẽ được nén về một miền dữ liệu ẩn (latent space) bằng phương pháp ánh xạ qua bộ mã hóa (encoder) với số chiều nhỏ hơn dữ liệu gốc. Sau đó, một mạng nơron giải mã (decoder) sẽ ánh xạ từ latent space về miền dữ liệu ban đầu, cụ thể nguyên lý hoạt động của VAE được thể hiện qua Hình 1. Ưu điểm của mô hình neural network này là latent được nén về biểu diễn nhỏ hơn, nhưng vẫn giữ được các đặc tính quan trọng của dữ liệu gốc. Ngoài ra, ta có thể quyết định phân phối mà latent sinh ra nhờ Kullback-Leibler (KL) divergence [10].

Đối với miền dữ liệu rung động không có quá nhiều sự khác biệt giữa dữ liệu hư hỏng và dữ liệu nguyên vẹn. Do đó, trong nghiên cứu này, chúng tôi sử dụng dữ liệu ở trạng thái ban đầu của dầm (không bị hư hỏng) để trích xuất latent dưới dạng một phân phối chuẩn $\mathcal{N} \sim (\mu_X, \sigma_X^2)$ như đã nêu trên Hình 1. Các điểm trong miền ẩn ở trạng thái hư hỏng sẽ được đối chiều với các điểm ở trạng thái nguyên vẹn trong miền ẩn. Theo nghiên cứu [6], hàm mất mát của VAE được tính toán từ Kullback Leibler divergence (KL) dùng để ràng buộc phân phối chuẩn cho latent space thể hiện ở công thức (1).

$$D_{KL}(p(x)||q(x)) = \sum_{x\in\chi} p(x)\log\left(\frac{p(x)}{q(x)}\right) \quad (1)$$

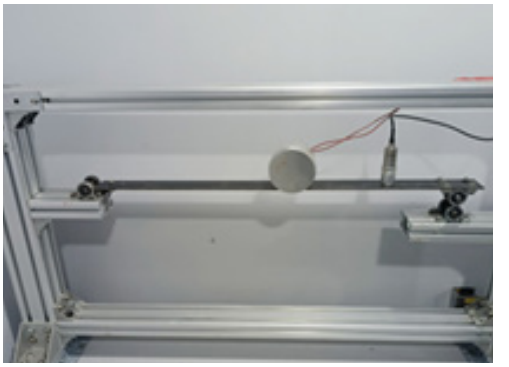
Với x là dữ liệu ban đầu, p(x) và q(x) là hai phân phối do encoder và decoder tạo ra. Như vậy, hàm mất mát cuối cùng của VAE trình bày ở công thức (2).

$$\mathcal{L}_{VAE} = \lambda||x-\hat{x}|| + E_x D_{KL}(p(z|x)||p(z)) \quad (2)$$

Với $x, \hat{x}$ lần lượt là dữ liệu gốc và dữ liệu khôi phục, p(z|x) là phân phối của latent sinh ra từ dữ liệu gốc bởi encoder và p(z) là phân phối của latent mà ta mong muốn trở thành.

3. THÍ NGHIỆM

Mô hình thí nghiệm gồm 03 phần chính: dầm, khung đỡ dầm (bao gồm liên kết với dầm) và động cơ rung, trong đó mô hình dầm có kích thước là 90 x 50 x 5 (mm).



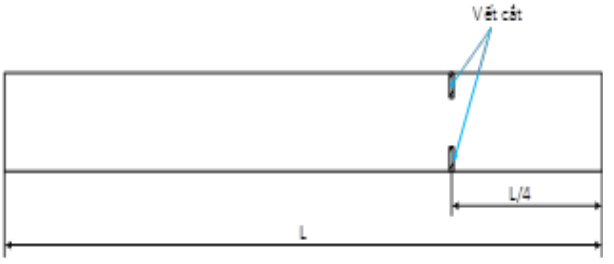
Hình 2. Mô hình thí nghiệm dầm.

ISSN 2615 - 9910 (bản in), ISSN 2815 - 5505 (online)
TẠP CHÍ CƠ KHÍ VIỆT NAM, Số 306, tháng 8 năm 2023
cokhivietnam.vn / tapchicokhi.com.vn

Dưới tác dụng của tải trọng, chi tiết dầm sẽ bị phá hủy sau một số chu kỳ nhất định.

Tuy nhiên, nếu để dầm mỏi một cách tự nhiên (không tác động thêm) trong quá trình mô phỏng hoạt động thì sẽ cần rất nhiều thời gian để dầm tiến đến trạng thái này. Do đó, dầm sẽ được cắt ban đầu để tạo nên ứng suất tập trung tại một vị trí cụ thể để giúp dầm phát triển vết nứt tế vi đó, từ đó có thể mô phỏng được quá trình hư hỏng của dầm.

Vị trí cắt là tại vị trí ¼ của dầm, dầm sẽ được rung liên tục và cứ cách mỗi phút sẽ lấy mẫu một lần. Dầm sẽ lấy với 10 mức độ hư hỏng, ứng với mỗi mức độ thì chiều dài vết cắt sẽ tăng thêm 2mm. Giữa các mẫu đều có thời gian nghỉ vì khi hoạt động, động cơ nóng sinh nhiệt ảnh hưởng đến kết quả đo. Các mẫu thí nghiệm phải được đánh số thứ tự và thời gian và điều kiện lấy mẫu.



Hình 3. Tạo vết cắt kích thích hư hỏng cho dầm.

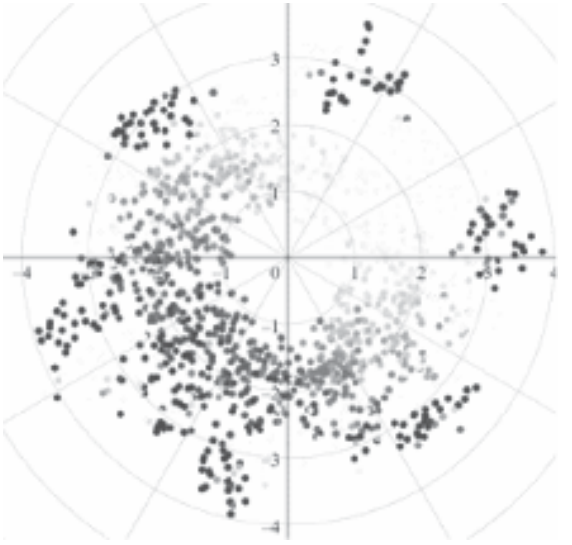
4. KẾT QUẢ VÀ THẢO LUẬN

Qua quá trình xử lý tín hiệu thu được số lượng từ các kênh IMF thông qua quá trình xử lý bằng VMD, cụ thể VMD tối đa 05 kênh IMF. Để tín hiệu được trích xuất một cách hiệu quả, các IMF được đánh giá thông qua hệ số tương quan giữa các kênh với tín hiệu gốc được thể hiện ở Bảng 1.

Bảng 1. Hệ số tương quan của các kênh IMF so với tín hiệu gốc thu được từ các dầm

IMF	1	2	3	4	5
Dầm 1	0.023	0.035	0.042	0.108	0.996
Dầm 2	0.012	0.029	0.028	0.106	0.998
Dầm 3	0.043	0.061	0.088	0.069	0.993
Dầm 4	0.045	0.070	0.177	0.098	0.980

Những kênh có hệ số tương quan lớn hơn giá trị 0,1 sẽ được chọn để xử lý. Các thành phần dao động cưỡng bức mang đặc tính của dao động cưỡng bức, nếu có sự thay đổi bất kỳ nào thì biên độ của các thành phần này cũng sẽ thay đổi. Do đó, sau khi chọn được các kênh IMF chủ đạo thì sử dụng biên độ tức thời của chúng để làm thông số đầu vào cho quá trình chẩn đoán hư hỏng bằng VAE.



Hình 4. Biểu diễn latent space trích xuất từ dầm nguyên vẹn và hư hỏng được thể hiện qua phép PCA. Các điểm màu xanh lục và xanh dương thể hiện dầm trong trạng thái hoạt động bình thường ở tập huấn luyện và tập kiểm chứng, các điểm màu đỏ biểu thị dầm bị hư hỏng trong tập kiểm chứng. ☞

Sau khi xử lý tín hiệu, ta trích xuất biểu diễn latent của mô hình đã được huấn luyện. Các đặc trưng trích xuất được thể hiện trong hình 5. Ta nhận thấy khi mức độ hư hỏng càng gia tăng thì giá trị latent có xu hướng tăng theo. Điều này chứng tỏ rằng encoder của VAE có khả năng chỉ ra những hư hỏng và có thể nhận biết một cách rõ ràng thông qua tỷ lệ phương sai.

Kết quả dự đoán hư hỏng của mô hình được thể hiện ở bảng (2).

Bảng 2. Kết quả dự đoán hư hỏng trên dầm.

	Độ chính xác
Dầm 1	85%
Dầm 2	89%
Dầm 3	85%
Dầm 4	89%

5. KẾT LUẬN

Trong nghiên cứu này, chúng tôi giới thiệu phương pháp VAE để nhận biết hư hỏng thông qua dữ liệu rung động. Kết quả chỉ ra rằng VAE có hiệu quả trong việc trích xuất những đặc tính của hệ dầm. Latent sinh ra từ encoder có thể phân biệt được một cách rõ ràng thông qua biểu diễn của nó trên đồ thị phân phối. Nó cũng làm nổi bật tầm quan trọng của các đặc trưng ẩn trong VAE nói riêng và mô hình generative deep learning nói chung.

Lời cảm ơn:

Nghiên cứu này được tài trợ bởi Trường Đại học Bách khoa, Đại học Quốc gia Thành phố Hồ Chí Minh trong khuôn khổ đề tài mã số SVKSTN-2022-CK-33. Chúng tôi xin cảm ơn Trường Đại học Bách khoa, ĐHQG-HCM đã hỗ trợ cho nghiên cứu này.❖

Ngày nhận bài: **15/5/2023**
Ngày phản biện: **26/6/2023**

Tài liệu tham khảo:

[1]. Zhao, B., Cheng, C., Peng, Z., Dong, X., & Meng, G. (2020). *Detecting the early damages in structures with nonlinear output frequency response functions and the CNN-LSTM model.* IEEE Transactions on Instrumentation and Measurement, 69(12), 9557-9567.

[2]. Chamangard, M., Ghodrati Amiri, G., Darvishan, E., & Rastin, Z. (2022). *Transfer Learning for CNN-Based Damage Detection in Civil Structures with Insufficient Data.* Shock and Vibration, 2022.

[3]. Wu, R., & Keogh, E. (2021). *Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress.* IEEE Transactions on Knowledge and Data Engineering.

[4]. Nguyen, K. T., & Medjaher, K. (2021). *An automated health indicator construction methodology for prognostics based on multi-criteria optimization.* ISA transactions, 113, 81-96.

[5]. An, J., & Cho, S. (2015). *Variational autoencoder based anomaly detection using reconstruction probability.* Special lecture on IE, 2(1), 1-18.

[6]. Ma, X., Lin, Y., Nie, Z., & Ma, H. (2020). *Structural damage identification based on unsupervised feature-extraction via Variational Auto-encoder.* Measurement, 160, 107811.

[7]. Kingma, D. P., & Welling, M. (2014, April). *Stochastic gradient VB and the variational auto-encoder. In Second international conference on learning representations,* ICLR (Vol. 19, p. 121).

[8]. Kingma, D. P., & Welling, M. (2013). *Auto-encoding variational bayes.* arXiv preprint arXiv:1312.6114.

[9]. Dragomiretskiy, K., & Zosso, D. (2013). *Variational mode decomposition.* IEEE transactions on signal processing, 62(3), 531-544.

[10]. Wang, G., Chen, X. Y., Qiao, F. L., Wu, Z., & Huang, N. E. (2010). *On intrinsic mode function. Advances in Adaptive Data Analysis,* 2(03), 277-293.

[11]. Goldberger, J., Gordon, S., & Greenspan, H. (2003, October). *An Efficient Image Similarity Measure Based on Approximations of KL-Divergence Between Two Gaussian Mixtures*. In ICCV (Vol. 3, pp. 487-493).