Unsupervised Knowledge Transfer for Structural Damage Detection with Limited Data

Xutong Zhang[1], Xinqun Zhu[1*], Jianchun Li[1] and Yang Yu[2]
[1]*School of Civil and Environmental Engineering, the University of Technology Sydney, Ultimo, NSW 2007, Australia*
[2]*School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia*
Email : Xutong.Zhang-1@student.uts.edu.au; Xinqun.Zhu@uts.edu.au; Jianchun.Li@uts.edu.au; yang.yu12@unsw.edu.au

ABSTRACT: Deep learning (DL) has proven effective in extracting damage-sensitive features by training neural network models with extensive numerical and experimental data. However, these models typically perform well only when the test samples come from data with the same distribution as the training data, which is often unrealistic in practical scenarios due to numerical modeling errors and operational variations. Furthermore, collecting data on the damage state is challenging due to the rarity and irreversibility of structural damage in real-world situations. To address these challenges, a novel method called maximum discrepancy adversarial domain adaption (MDAD) is proposed by jointly aligning the distributions of damage-sensitive features across different domains at both the class and domain levels. It consists of a feature generator, two classifiers, and one discriminator. The MDAD method employs a domain discriminator and feature generator to merge the distributions at the domain level, overcoming the problem of insufficient data from real structures. Additionally, the generator and two classifiers align at the class level by leveraging the classification discrepancy between them.

KEY WORDS: *Structural damage detection, knowledge transfer, domain adaptation, limited measurement data*

1 INTRODUCTION

Vibration-based methods are commonly used to detect structural damage or deterioration in civil infrastructure. These methods can be categorized into model-based and data-driven approaches [1]. Model-based approaches rely on accurate numerical modeling but are affected by uncertainties [2]. Data-driven approaches, especially deep learning (DL) based methods, have gained popularity due to their ability to automatically extract structural damage features [3]. Deep learning models, such as convolutional neural networks (CNNs), recurrent neural networks, recursive neural networks, and unsupervised pre-trained networks (UPNs), have been developed for structural damage detection. UPNs and CNNs are commonly used in vibration-based methods. Researchers have used deep CNNs to extract damage features and identify damage locations from sensor data with noise effects [4]. Deep autoencoders have been employed for effective feature learning and mapping input modal information to output structural stiffness parameters for damage identification. The effectiveness of DL-based damage detection methods relies on two assumptions [5]: having training data with the same distribution as the test data and having a large amount of labelled training data. However, obtaining real structural damage data is challenging, and training data are often generated from numerical models with different damage scenarios [6,7]. Due to uncertainties such as modelling errors, operational and environmental variations, and measurement noise, the generated data may not align with real measurement data [8]. Consequently, when applying DL models trained based on generated data to real structures, their performance may significantly degrade [9]. In this study, a transfer learning-based method is proposed for structural damage detection with limited measurements. The goal is to overcome the challenges of inconsistent data distributions and limited real measurement data to enhance the performance of DL models for practical applications.

Transfer learning (TL) is a valuable technique in machine learning that addresses the challenge of limited training data [10]. It involves leveraging knowledge learned from one domain and applying it to a related but different domain, even when the target domain lacks labeled data [11]. Domain adaptation (DA) is a specific area of transfer learning that focuses on aligning the distributions between the source and target domains [12]. DA has been successfully applied in various real-world tasks such as video analysis and natural language processing to improve performance [13,14,15].

In the context of machinery diagnosis, transfer learning has also been utilized. For example, in bearing fault diagnosis, unsupervised deep transfer learning has been explored to enhance diagnostic accuracy. Adversarial training is one technique that aids in aligning domain shifts and improving the effectiveness of transfer learning models [5]. Researchers have investigated the application of transfer learning in bearing diagnosis, considering aspects such as source domain data selection, data transformation techniques, and the selection of appropriate transfer learning models [16]. This opens up possibilities for utilizing information from numerical models to address the challenge of limited real-world damage data in structural health monitoring and damage detection.

This paper introduces a novel approach for structural damage detection called Maximum Classifier Discrepancy Domain Adaptation (MDA). MDA addresses the challenge of significant domain discrepancy and limited high-quality data by using two-label classifiers to detect discrepancies at the class level. However, relying on class-level alignment may be limited when domain divergence is substantial and high-quality data is scarce. To overcome this, a joint approach called Maximum Classifier Discrepancy & Adversarial

Discriminative Domain Adaptation (MDAD) is proposed. MDAD combines domain-level distribution alignment, which calibrates domain divergence caused by modeling errors, with class-level alignment that addresses uncertainty between numerical simulations and experimental models. The main contributions of this paper are as follows:

1. The MDAD method, consisting of two classifiers and a discriminator, aligns distributions at both class and domain levels to adapt to divergence between numerical and experimental models.
2. Two-level distribution alignment enables the utilization of limited field data with unknown damage states in structural health monitoring tasks.
3. Two case studies are presented to validate the proposed method: knowledge transfer from one structure to another with uncertainty and from numerical to experimental models.

2 MAXIMUM DISCREPENCY ADVERSARIAL DISCRIMINATIVE DOMAIN ADAPTATION (MDAD) FOR STRUCTURAL DAMAGE DETECTION

2.1 Problem definition

This study focuses on the transfer of knowledge from a numerical model (source domain, D_S) to a real structure (target domain, D_T) for structural damage detection. The source domain consists of n labeled samples, denoted as $D_S = \{(\boldsymbol{x}_{si}, y_{si})|\, x_{si} \in X_S, y_{si} \in Y_S, i = 1,2, \dots, n\}$, where $X_S = \{x_{si}, i = 1,2, \dots n\}$ represents the input data and Y_S represents the output damage labels. The target domain, on the other hand, has limited measurement data, particularly for the structural damage states. It is defined as $D_T = \{(\boldsymbol{x}_{ti})|x_{ti} \in X_T, i = 1,2, \dots m\}$, where X_t represents the input data for the target domain.

The distributions of the source and target domains are represented by $P(X)$ and $Q(X)$, respectively, where X refers to a specific learning sample from either X_S or X_t. Additionally, there are conditional probability distributions, $P(Y|X)$ and $Q(Y|X)$, which represent the relationship between the input data and the corresponding damage labels. Due to uncertainties such as modeling errors, operational and environmental variations, and measurement noise, there exists a discrepancy between the data from the real structure and its numerical model. As a result, the distributions of the source and target domains are not identical. Consequently, a pre-trained deep learning network using the source domain data cannot perform well on the target domain data, and DL networks trained on data from the numerical model cannot directly predict the structural damage of the real structure. To address this challenge, the study proposes a new method based on maximum discrepancy discriminative adversarial domain adaptation for structural damage detection.

The MDAD framework for structural damage detection consists of a feature generator, two classifiers, and a discriminator, as shown in Figure 1. It addresses the challenge of limited data from the real structure by aligning distributions at both the domain and class levels. The feature generator and domain discriminator handle domain-level alignment, while the feature generator and classifiers align features at the class level. This alignment process helps eliminate modelling errors and uncertainties, allowing for the extraction and alignment of damage-sensitive features.

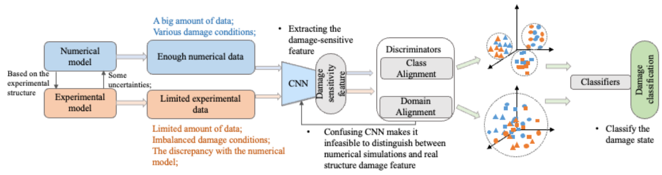


Figure 1. The architecture of the proposed framework

2.2 MDAD domain adaptation framework for structural damage detection

Aligning the distributions is essential for utilizing source domain knowledge to predict target domain data. However, calibrating the conditional distribution for different damage locations is challenging, limiting the ability to accurately predict damage locations. While previous work successfully calibrated conditional distributions for various damage severities in the same locations, predicting multiple damage locations in practice becomes difficult due to rough conditional probability matching between source and target domains.

Figure 2 illustrates a novel MDAD domain adaptation framework for structural damage detection. The framework consists of a generator G, two classifiers C_1 and C_2, and a domain discriminator D. To address both class and domain level discrepancies, two types of discriminators are incorporated [17,18]. The generator is trained to extract target features that minimize discrepancies at both levels. Through adversarial training, the generator aligns the source domain classifier boundary to generate damage-sensitive features from the target domain. Classifiers C_1 and C_2, along with domain discriminator D, distinguish each class and identify domain-invariant features to enhance classification accuracy (see Figures 3 (a) and (b)).

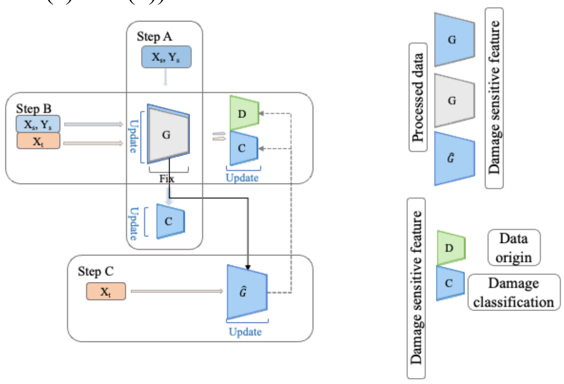
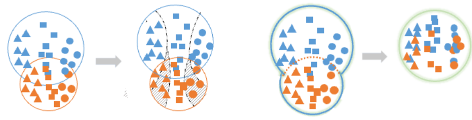


Figure 2. Maximum Discrepancy Discriminative Adversarial (MDDA) domain adaptation architecture



(a) At Class level (b) At Domain level

Figure 3. Class and domain discrepancy alignments

3 CASE STUDY

3.1 Knowledge transfer between two structures with modelling errors

In this section, we explore the transfer of knowledge between structures of different sizes. We utilize a three-storey building model shown in Figure 6 as a representative sample. Model 2 refers to an existing model that possesses labelled damage data, while Model 1 is our proposed model aimed at predicting damage. The objective is to leverage the knowledge gained from Model 2 to enhance the damage prediction capabilities of Model 1.

Structural damage is characterized by a reduction in stiffness. In this study, damage scenarios are simulated in different locations of the structure, including undamaged and single damage locations on each floor. Model 2 is used to simulate damage severities by reducing the structural stiffness from 0-30%. Each damage level is tested with one hammer excitation on each floor, resulting in a total of 243 data samples per damage case. The details are summarized in Table 4. The undamaged case also has 243 repeated test samples. In total, there are 972 labelled samples covering the undamaged case and three damage scenarios, classified as 0, 1, 2, 3. The target domain data are collected from Model 1. The MDAD network is evaluated for single-damage identification, which aligns with real engineering practices. The lumped model simulation and data processing were conducted using Matlab 2022b.

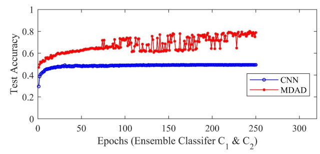
i. *Identically damage dataset (Identical damage label spaces)*

The target domain data consists of 972 samples collected from numerical Model 1, with identical damage severities as the source domain data. Initially, the networks were trained on the labeled source data from Model 2 without knowledge transfer. However, when the trained CNN model was applied directly to the target domain without domain adaptation (DA), the classification accuracy of the two classifiers was significantly reduced, as shown in Figure 4. This indicates that the model's performance suffers when applied to the target domain.

To visualize the feature extraction process, Principal Component Analysis (PCA) was performed on the source and target domains, specifically on PCA components 1, 2, and 3. Figure 5(a) presents the distribution of each PCA component for the source and target domains, highlighting the divergence in the main features. For clarity, only one excitation location was selected from each model. The x-axis represents the PCA value, while the y-axis represents the damage scenario class. The legend indicates the number of distributions for each class. It can be observed that, before domain adaptation, the global distribution of each PCA component in both domains does not align well. Additionally, the local distribution for each class is separate and unaligned between the source and target domains.

In Figure 4(a), the testing classification accuracy of the two classifiers in the target domain is represented by the red line. It can be observed that the accuracy improves over epochs, indicating the learning process of the model. In Figure 4(b), the damage classification results of the target domain are displayed. It shows that the damage location classification achieves an accuracy of 83%, which is a significant improvement compared to the CNN model trained only with the source domain data. This demonstrates the effectiveness of the proposed approach in improving the classification performance in the target domain.

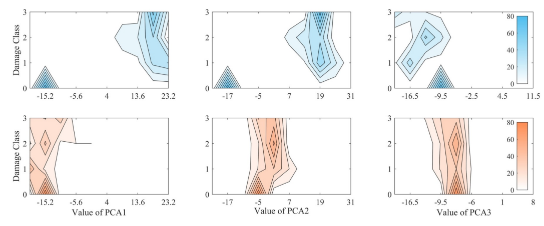
Figure 5(b) illustrates the impact of domain adaptation (DA) on the PCA distribution. With the MDAD method, the global distribution of PCA components between the source and target domains becomes well-aligned, as opposed to the misalignment observed in Figure 5(a) without DA. The local distributions for each class also undergo calibration through knowledge transfer and feature merger. This successful alignment of both global and local distributions enhances transfer learning performance and improves classification accuracy in the target domain.



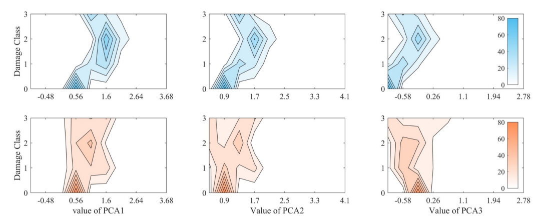
	0	1	2	3	Accuracy
0 (243)	243	0	0	0	100%
1 (243)	45	198	0	0	82%
2 (243)	33	2	208	0	86%
3 (243)	37	62	0	144	59%
Overall	/	/	/	/	82%

(a) Testing accuracy over Epochs　　(b) Confusion Matrix of Classification results

Figure 4. MDAD method performance for Identical damage datasets



(a)



(b)

Figure 5. Comparisons of PCA data distribution without DA (a) and with DA (b) for identical damage dataset (damage class 1-4 indicate the label 0, 1, 2 and 3)

3.2 Knowledge Transfer from the Numerical Simulation to the Real Structure

The limited availability of data and the discrepancy between numerical models and real structures pose challenges in transferring knowledge from numerical models to actual structures. To address this, we applied the proposed MADA method to predict damage locations on a laboratory-tested three-storey building model. The model consists of three identical beams connected by two columns, with specific dimensions and material properties.

The laboratory-tested building model (Figures 6 and 7) comprises three identical beams, with the dimension of 394 mm length, 50 mm width, and 30 mm thickness, connected by two columns, with the dimension of cross-section of 50 mm × 3 mm and the length of 900 mm. The material used has a Young's modulus of 435 MPa and a density of 7.5×10^3 kg/m^3.

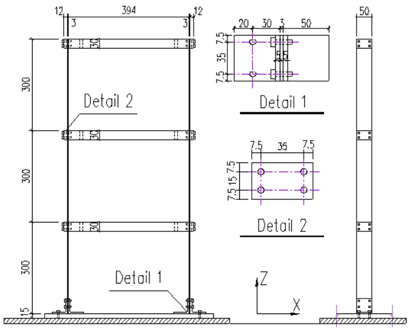


Figure 6. Configuration of building model (Front view and side view)

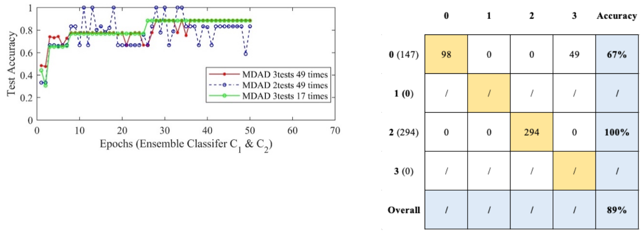
Figure 7. Experimental arrangement

The source domain data for training the CNN model consisted of 972 samples. Undamaged scenarios were generated by simulating impact hammers on three floors, each repeated 81 times. Damage scenarios included 58 severity levels on each floor. The labels for the source data indicated the damage location: 0 (undamaged), 1 (first floor damage), 2 (second floor damage), and 3 (third floor damage). Target domain contains only scenarios 0 and 2 (10% and 20%) from laboratory-tested building model with 49 times repeat tests for each floor.

The CNN model was initially trained on the source domain data without domain adaptation (DA). The batch size was set to 256, and the model was trained for 50 epochs. The test accuracy on the target domain, without knowledge transfer, was 69%.

After applying the MDAD method for knowledge transfer, the damage prediction accuracy improved to 89% (Figure 8). The MDAD method showed a significant increase in classification accuracy of around 20% using the limited experimental data.

The study also analysed the effect of sample numbers on classification accuracy. It was observed that reducing the number of repeated tests and the augmentation times had an impact on accuracy. The proposed MDAD method demonstrated better performance in transferring knowledge between numerical and experimental models. It effectively dealt with boundary conditions and experimental uncertainties, which are often overlooked when building simulation models.



(a) Testing accuracy and error over epochs

(b) Confusion Matrix of Classification results

Figure 8. MDAD method performance for single-damage damage datasets

4 CONCLUSIONS

The proposed method uses an MDAD-based approach for knowledge transfer and merging in structural damage detection. It involves a CNN generator, two classifiers, and a discriminator. By pre-training a deep learning model on one structure, the method allows for damage detection on other structures without labelled data. It addresses limited damage classes and data samples by combining local and global divergence levels to merge sensitive features across structures. The network is trained to detect damage by aligning and extracting damage-sensitive features. The effectiveness of the method is demonstrated in transfer learning scenarios between simulation models with uncertainty and from simulation to experimental structures, for the transfer learning between two numerical models with uncertainty, the damage is enabled to be detected with their damage locations accurately. For the transfer learning between the simulation and experimental model, the damage also can be localized even if the experimental data is limited, and classification is incomplete. The purposed method overcoming data limitations in practical applications.

ACKNOWLEDGMENTS

This study was funded in part by the Australia Research Council (ARC) through Discovery Project (DP) No.: 23010806.

REFERENCES

[1] Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M. and Inman, D.J. 2021, 'A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and

Deep Learning applications', Mechanical Systems and Signal Processing, 147, 107077.

[2] Hou, R. and Xia, Y. 2021, 'Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019', *Journal of Sound and Vibration*, 491, 115741.

[3] Azimi, M., Eslamlou, A.D. and Pekcan, G. 2020, 'Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review', *Sensors*, 20(10), 2778.

[4] Lin, Y.Z., Nie, Z.H. and Ma, H.W., 2017. 'Structural damage detection with automatic feature-extraction through deep learning.' *Computer-Aided Civil and Infrastructure Engineering*, *32*(12), 1025-1046.

[5] Zhao, Z.B., Zhang, Q.Y., Yu, X.L., Sun, C., Wang, S.B., Yan, R.Q. and Chen, X.F. 2021, 'Applications of unsupervised deep transfer learning to intelligent fault diagnosis: a survey and comparative study.' *IEEE Transactions on Instrumentation and Measurements*, 70, 3525828.

[6] Chen, Z., Wang, C., Wu, J., Deng, C. and Wang, Y. 2023, 'Deep convolutional transfer learning-based structural damage detection with domain adaptation', *Applied Intelligence*, 53, 5085-5099.

[7] Jiao, J., Lin, J., Zhao, M. and Liang, K. 2020, 'Double-level adversarial domain adaptation network for intelligent fault diagnosis', *Knowledge-Based Systems*, 205, 106236.

[8] Lin, Y.Z., Nie, Z.H. and Ma, H.W. 2022, 'Dynamics-based cross-domain structural damage detection through deep transfer learning', *Computer-Aided Civil and Infrastructure Engineering*, 37, 24-54.

[9] Zhang, Z., Sun, C. and Guo, B. 2022, 'Transfer-learning guided Bayesian model updating for damage identification considering modelling uncertainty', *Mechanical Systems and Signal Processing*, 166, 108426.

[10] Wang, M. and Deng, W. 2018, 'Deep visual domain adaptation: A survey', *Neurocomputing*, 312, 135-53.

[11] Farahani, A., Voghoei, S., Rasheed, K. and Arabnia, H.R., 2021. 'A brief review of domain adaptation.' Proceedings of the 16th International Conference on Data Science and the 19th International Conference on Information & Knowledge Engineering, pp.877-894.

[12] Kouw, W.M. and Loog, M., 2019. 'A review of domain adaptation without target labels.' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3), 766-785.

[13] Vasili Ramanishka, Abir Das, Dong Huk Park, Subhashini Venugopalan, Lisa Anne Hendricks, Marcus Rohrbach, and Kate Saenko. Multimodal video description. In Proceedings of the 24th ACM international conference on Multimedia, pages 1092–1096, 2016.

[14] Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine transla- tion. arXiv preprint arXiv:1406.1078, 2014.

[15] Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685, 2015.

[16] Chen, X.H., Yang, R., Xue, Y.H., Huang, M.J., Ferrero, R. and Wang, Z.D. 2023. 'Deep transfer learning for bearing fault diagnosis: a systematic review since 2016.' *IEEE Transactions on Instrumentation and Measurement*, 72, 3508221.

[17] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M. and Lempitsky, V., 2016. 'Domain-adversarial training of neural networks.' *The Journal of Machine Learning Research*, 17(1), 2096-2030.

[18] Tzeng, E., Hoffman, J., Saenko, K. and Darrell, T., 2017. 'Adversarial discriminative domain adaptation.' *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167-7176.