Adversarial Spiral Learning Approach to Strain Analysis for Bridge Damage Detection

Takaya Kawakatsu[1](✉), Akira Kinoshita[2], Kenro Aihara[2], Atsuhiro Takasu[2], and Jun Adachi[2]

[1] The University of Tokyo, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan
kat@nii.ac.jp

[2] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan

Abstract. When a vehicle passes over a bridge, the bridge distorts in response to the vehicle's load. The response characteristics may change over time if the bridge suffers damage. We consider the detection of such anomalous responses, using data from both traffic-surveillance cameras and strain sensors. The camera data are utilized to treat each vehicle's identified properties as explanatory variables in the response model. The video and strain data are transformed into a common feature space, to enable direct comparisons. This space is obtained via our proposed spiral learning method, which is based on a deep convolutional neural network. We treat the distance between the video and strain data in the space as the anomaly score. We also propose an adversarial unsupervised learning technique for removing the influence of the weather. In our experiments, we found anomalous strain responses from a real bridge, and were able to classify them into three major patterns. The results demonstrate the effectiveness of our approach to bridge damage analysis.

Keywords: Structural health monitoring · Deep learning
Multimedia analysis

1 Introduction

Many road bridges built in Japan in the 1960s have deteriorated and now require substantial inspection. We approach the health monitoring problem by installing a number of inexpensive sensors on a bridge as shown in Fig. 1. We aim to identify small signs of bridge deterioration by developing a fully data-driven inspection technique. Our research started from a simple question. Could the bridge strain response caused by a vehicle be predicted from the surveillance video, using an *encoder–decoder* [1] approach? If we set up such a predictor during the bridge's construction, we could capture small signs of deterioration later by monitoring the frequency of *anomalous* vehicles. Unfortunately, we encountered a difficulty, namely that the video data could not supply information about a vehicle's axle weights.

C. Ordonez and L. Bellatreche (Eds.): DaWaK 2018, LNCS 11031, pp. 49–58, 2018.
https://doi.org/10.1007/978-3-319-98539-8_4

In general, the strain response will depend on the moving axle loads, so measuring axle weights directly seems an obvious approach. To obtain the axle weights, the options are to install a pavement sensor or use a bridge weighing in motion (BWIM) technology [2]. The former is fragile, hard to retrofit to existing bridges, and limits the traveling speeds. The latter requires that strain response characteristics be obtained in advance, which is not applicable to the anomaly detection problem. We, therefore, abandoned the axle weight approach.

Instead, we developed a new sensor fusion approach that directly compares the vehicle image and strain response in a common feature space. This approach is somewhat similar to the Siamese network [3], except that our approach utilizes two different neural networks for the video and strain data. Moreover, unlike the Siamese network, the two networks are trained to predict vehicle speeds and loci individually. As we reported in a previous paper [4], vehicle speeds and loci may be predictable from both video and strain data, and they affect the shape of the strain response significantly. By learning these two tasks, the two networks can acquire feature spaces that seem to comprise *common factors* of the video and strain data. Finally, the bases of the two spaces can be matched by minimizing the distance between two related elements in the respective spaces. We call this approach *spiral learning*. Some video–response pairs caused by the same vehicle may be inconsistent in the common space. We treat such an event as an *anomaly*, making the assumption that such a response may be caused by bridge structural damage. It should be noted that bad weather, e.g., a snowstorm or heavy rain, may disrupt the video signal. In such a situation, an event may be misidentified as an anomaly, even if the strain response was normal. Therefore, we proposed adding an *adversarial learning mechanism* as a countermeasure.

We evaluated our proposed approach using real observations recorded over a six-month period. We then classified the observed anomalies into some patterns to demonstrate the weather resistance. The results show the effectiveness of our approach to bridge damage analysis.

2 Bridge Analysis: The Influence Line

Studies of bridge damage detection can be classified into two approaches, namely steady-state analysis [5] and transient-response analysis [6,7]. The former detects damage by detecting changes in natural frequencies and time constants, whereas the latter aims to detect damage by focusing on temporary events such as passing vehicles. These approaches require the vehicle speed, locus (traveling position in the lane), and axle positions, as explanatory variables in the response model.

A bridge bends as vehicles pass over it. If we assume a linear-response model, the strain is proportional to the vehicle weight and closeness between the sensor and the vehicle. The bending moment $m(t)$ at time t will vary depending on the vehicle position, x. $m(t)$ can be estimated by the *influence line* $i(x)$ and strain measurements $s(x,t)$ given in Eq. (1):

$$\hat{m}(t) = \int_0^l w(x,t)i(x)dx \approx \int_0^l ES(x)s(x,t)dx, \tag{1}$$

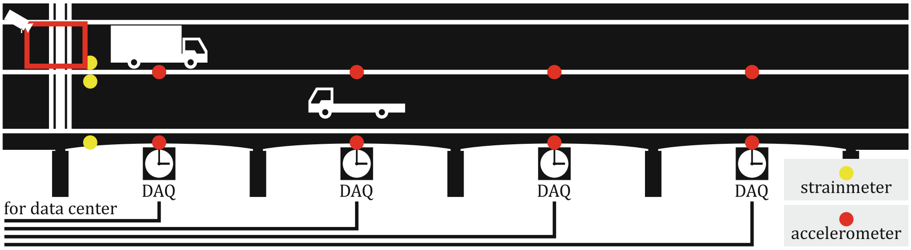


Fig. 1. Sensor positions.

where $\hat{m}(t)$ is the predicted moment and $w(x,t)$ is the axle weight at x. E is the elasticity modulus and $S(x)$ is the section modulus.

The influence line characterizes an individual bridge but may change its shape over time owing to structural damage to the bridge. A number of researchers [6,7] evaluated a damage detection technique using influence lines. To obtain an ideal influence line, we should run a test vehicle with known axle weights, speed, and locus, excluding the influence of other vehicles. However, that can prove difficult, particularly for bridges carrying heavy traffic.

Zaurin and Catbas [8] proposed a novel sensor fusion approach for monitoring bridges in service. They installed a traffic-surveillance camera and strain sensors on a miniature bridge. The camera was used for axle detection. The axle positions were obtained by tracking axles and were used for calculating the surface load $w(x,t)$. The authors [8] assumed that axle weights could be measured via BWIM. The influence line was estimated for every vehicle. By clustering the estimated influence lines, the potential damage was identified. Their approach has a serious weakness in that BWIM requires obtaining the influence line in advance. Once the bridge becomes degraded, the influence line may change, which means that the BWIM system may be unreliable for their purpose.

3 The Spiral Learning Approach

Spiral learning is a technique whereby a pair of samples from two data sources are fed into two separate networks. The networks are independent of each other, except for a final linear layer that shares the same weight matrix. Each network learns from its respective dataset and acquires a feature mapping that contains information about vehicle properties such as speed and locus. The samples refer to the same vehicle as a latent variable.

Table 1 shows the neural network architectures. $[k/d, c]$ denotes a c-channel convolution layer with a kernel width of k, which downsizes the sample to $1/d$-th of the original. Each stacked matrix denotes a plain residual block [9]. CamNet receives 50 grayscale video frames (taken over two seconds), which are resized to 224 × 224 pixels, and outputs the estimated speed and locus of the target vehicle. SigNet receives four-second batches of raw strain data sampled at 200 Hz. Each strain sample was rescaled so that its maximum and minimum were normalized to

Table 1. Network architectures for video and strain data.

Layers	CamNet	SigNet
Conv 1	$\begin{bmatrix}3/2, 50\\3/2, 50\end{bmatrix}$	$\begin{bmatrix}25/4, 64\end{bmatrix}$
Conv 2	$\begin{bmatrix}3/1, 50\\3/1, 50\end{bmatrix}$	$\begin{bmatrix}7/1, 64\\7/1, 64\end{bmatrix}$
Conv 3	$\begin{bmatrix}3/2, 50\\3/2, 50\end{bmatrix}$	$\begin{bmatrix}7/4, 64\\7/4, 64\end{bmatrix}$
Conv 4	$\begin{bmatrix}3/1, 50\\3/1, 50\end{bmatrix}$	$\begin{bmatrix}3/1, 64\\3/1, 64\end{bmatrix}$
Conv 5	$\begin{bmatrix}3/2, 50\\3/2, 50\end{bmatrix}$	$\begin{bmatrix}3/1, 64\\3/1, 64\end{bmatrix}$
Conv 6	$\begin{bmatrix}3/1, 50\\3/1, 50\end{bmatrix}$	$\begin{bmatrix}3/4, 64\\3/4, 64\end{bmatrix}$
Conv 7	$\begin{bmatrix}3/2, 50\\3/2, 50\end{bmatrix}$	$\begin{bmatrix}3/1, 64\\3/1, 64\end{bmatrix}$
Conv 8,9	$\begin{bmatrix}3/1, 50\\3/1, 50\end{bmatrix}$	$\begin{bmatrix}3/1, 64\\3/1, 64\end{bmatrix}$
Linear 1	Output: 100 × 1	
Linear 2	Output: Speed and locus	

1 and 0, respectively for effective learning. We used ReLU [10] for the activation functions in each layer of the two networks, except for the output layers.

We selected speed and locus as the outputs because these two properties have the next strongest effect on the signal shape of the strain response, after the axle loads. We can expect the two networks to generate the same feature vector as a common factor, through learning the two prediction tasks and sharing the same output layer. The loss function for CamNet is defined in Eq. (2):

$$\mathcal{L}_{\mathrm{MSE1}}(f, h) = \frac{1}{N}\left\{\sum_{n=1}^{N}\frac{\left[h_{\mathrm{s}}(f(\boldsymbol{x}_n)) - s_n\right]^2}{\mathrm{Var}(s)} + \sum_{n=1}^{N}\frac{\left[h_{\mathrm{l}}(f(\boldsymbol{x}_n)) - l_n\right]^2}{\mathrm{Var}(l)}\right\}, \tag{2}$$

where $\boldsymbol{x}_n$ is the n-th video sample. f and h denote the feature extraction through CamNet and the Linear 2 layer, respectively. s and l are ground truth annotations for the speed and locus prediction tasks. The Linear 2 layer is shared by the two models, enabling the outputs of the Linear 1 layer to be treated as feature vectors in a common feature space. Both networks learn the correlation between the two sources by drawing their feature vectors together. The attracting mechanism can be described as the anomaly loss $\mathcal{L}_{\mathrm{MSE3}}$ defined in Eq. (3).

$$\mathcal{L}_{\mathrm{MSE3}}(f, g) = \frac{\lambda}{N}\sum_{n=1}^{N}\left\|\frac{f(\boldsymbol{x}_n)}{\|f(\boldsymbol{x}_n)\|_2} - \frac{g(\boldsymbol{y}_n)}{\|g(\boldsymbol{y}_n)\|_2}\right\|_2^2, \tag{3}$$

where λ is a weight of $\mathcal{L}_{\mathrm{MSE3}}$ and is set as 10. g denotes the feature extraction through SigNet and $\boldsymbol{y}_n$ is the n-th strain sample. Consequently, the optimization problem for the combined network, named SpiNet, can be described in terms of multitasking [11], as given in Eq. (4).

$$\mathcal{L}_{\mathrm{TMSE}}(f, g, h) = \mathcal{L}_{\mathrm{MSE1}}(f, h) + \mathcal{L}_{\mathrm{MSE2}}(g, h) + \mathcal{L}_{\mathrm{MSE3}}(f, g). \tag{4}$$

Equation (4) minimizes five individual losses, namely speed and locus prediction from the video data, speed and locus prediction from the strain data, and the L^2 norm between the feature vectors in CamNet and SigNet. Because these two networks share the output layer, Eq. (3) plays the role of matching the correlative elements from the two feature spaces. As a result, the video and strain feature spaces will coalesce after a long training period, as two *cannibal black holes* forming a spiral trajectory. The video and strain data may *differ* from each other, even though they cover the same target vehicle. This can be identified by monitoring the L^2 norm of the subtraction between $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$. We, therefore, define $\mathcal{L}_{\mathrm{MSE3}}$ as an *anomaly score*, setting λ as 1. The outliers of the distribution will be identified as *anomalous* vehicles.

The anomalous vehicle detection based on spiral learning depends on features extracted from the surveillance video recordings. The video can be disturbed by environmental conditions such as weather, traffic jams, pedestrians, and vehicles in the opposite lane. Such disturbances may cause mistaken anomaly detections. The *adversarial spiral learning* can reduce these errors, particularly those caused by bad weather. It may enable video features to correlate less with the weather conditions, including heavy rain, snowstorms, deep snow, and morning haze.

We, therefore, combined the adversarial learning concept [12] with our spiral learning proposal for this purpose. Adversarial learning has been introduced for image generation, utilizing a discriminator and a generator implemented by two separate neural networks. The discriminator finds fake images from a given image set that includes real and generated images. The generator creates images that are exactly like the real ones, which the discriminator may then misjudge as real images. The training mechanism can be described as a discriminator aiming to minimize the discrimination error whereas the generator aims to maximize it.

One of the simplest methods to achieve weather resistance is to use a weather discriminator. The discriminator tries to find videos recorded under bad weather conditions by examining video features carefully and in detail. To implement this function, we need to append weather tags to the traffic dataset. The loss function for SpiNet can be described as in Eq. (5), using the mean cross entropy $\mathcal{L}_{\mathrm{MCE}}$:

$$\mathcal{L}_{\mathrm{spin}}(f, g, h) = \mathcal{L}_{\mathrm{TMSE}}(f, g, h) - \mathcal{L}_{\mathrm{MCE}}(p, q), \tag{5}$$

where p and q denote the discriminator and the weather tag, respectively. After a long training period, the discriminator can no longer find faults in the obtained video features.

Initially, we tried to tag each vehicle by consulting the historical climate data archived by the government, but we encountered a major difficulty. The weather conditions at the bridge did not always correspond with the historical data.

This was because weather conditions were recorded at the nearest observation station, a few kilometers away from the target bridge. We then tried to tag the vehicles manually by watching the video, but encountered another difficulty. Because of the complicated weather situations, it was hard to formulate a robust policy for weather annotation. For example, should there be a cloudy tag in addition to a sunny tag, and is there a boundary between cloudy and light-rain conditions, or between snowfall and snow accumulation? Sometimes, the video lost focus owing to morning haze, twilight, or a fogged lens. Additionally, not only precipitation but also the intensity of solar radiation, which also has a strong impact on the video quality, should be noted.

We, therefore, abandoned the weather-annotation plan and developed a fully unsupervised approach. Here, the issue was that a vehicle might be mistakenly judged anomalous because of bad weather, even though the strain response was normal. This was a situation where an *anomalous strain response* could be found without consulting the strain feature, but by consulting the video feature alone instead. This might be a problematic situation, considering the main purpose of the strain characteristics analysis. We therefore defined the ground truth tag for the discriminator as given in Eq. (6):

$$q(\boldsymbol{x}, \boldsymbol{y}, f, g) = \mathbb{H}\left\{\|f(\boldsymbol{x}) - g(\boldsymbol{y})\|_2^2 - \mathcal{L}_{\mathrm{MSE3}}^{\mathrm{train}}(f, g)\right\}, \tag{6}$$

where $\mathbb{H}$ is the step function and $\mathcal{L}_{\mathrm{MSE3}}^{\mathrm{train}}$ is the anomaly loss (3) for the training dataset. The initial value of $\mathcal{L}_{\mathrm{MSE3}}^{\mathrm{train}}$ for the first epoch was 0.1. The adversarial network, named TwiNet, was defined as a perceptron whose hidden layer has 10 dimensions.

4 Training and Evaluation Data

We conducted our experiments on a 300-m-long prestressed concrete bridge in Japan. The bridge had four spans and two lanes as shown in Fig. 1, and has suffered damage caused by snowy weather. We have deployed a highly sensitive strainmeter and a surveillance camera on the bridge. The strain sensor observed the horizontal strain of its deck slab, sampled at 200 Hz.

To make a traffic dataset for the experiments in Sect. 5, we improved the traffic-surveillance system (TSS) [13]. After a vehicle enters the bridge, a camera installed at the entrance captures the vehicle. TSS detects the bounding box for the vehicle image, by using Faster R-CNN [14]. TSS outputs data about vehicles one by one, including properties such as lane, speed and locus. The mechanism for estimating vehicle speeds and loci was as follows. First, the TSS identifies all possible pairs of two axles between two consecutive video frames. Next, the TSS calculates the amount of movement for all pairs. Finally, the TSS estimates the traveling speed from the median of the movement amounts, and the locus from the average bottom position of the frontmost axles. It should be noted that the coordinates are transformed so that the distance in pixels is proportional to the distance in meters.

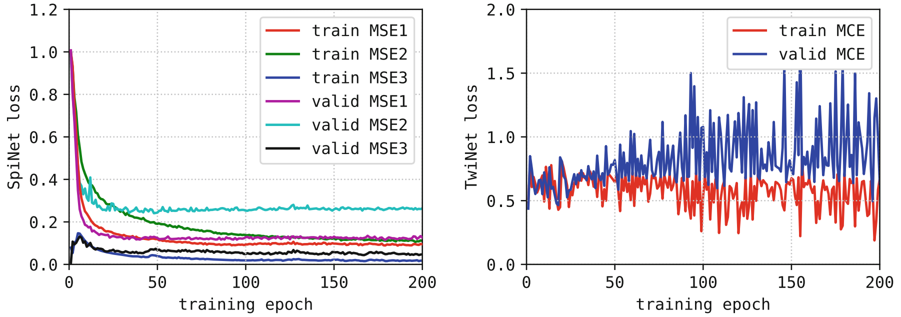


Fig. 2. Time evolution of training and validation losses.

By using TSS, we prepared a ground truth dataset named DS601 from videos recorded daily from 08:00 to 16:00 between November 2016 and April 2017. The dataset contained information about 1,014,083 vehicles. We prepared the *trainval* and *evaluation* datasets by randomly dividing DS601 in half. 80% of the *trainval* vehicles were assigned as *training* data, with the remaining 20% being *validation* data. Each time the validation loss $\mathcal{L}_{\mathrm{TMSE}}^{\mathrm{valid}}$ reached a new minimum, the model was saved. Finally, the evaluation processes were performed, in the same fashion as the early stopping approach [15].

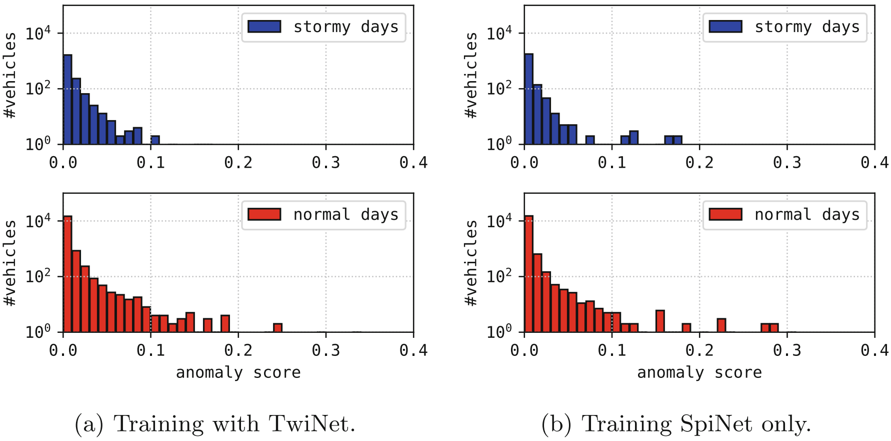
Each vehicle record was described as a tuple $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{t})$. The video input $\boldsymbol{x}$ was a sequence of length 50, and the strain input $\boldsymbol{y}$ was a sequence of length 800. $\boldsymbol{t}$ is the ground truth of the vehicle speed and locus. These datasets used only the left-to-right (LtoR) subset of vehicles. Vehicles with fewer than three axles were ignored because bridge experts are interested mainly in large vehicles. Vehicles crossing the target bridge using the opposite lane were ignored, to stabilize the training process. We also ignored video data from January 6th to 15th because the sensor data during this period was lost because of a fault in the observation environment. In the end, the *trainval* and *evaluation* datasets contained 17,757 and 17,967 vehicles, respectively.

5 Experimental Results

We implemented the proposed models on Chainer[1] 4.0, and accelerated them by using CUDA[2] 8.0. We used the AMSGrad [16] optimizer. The batch size was 10. Figure 2 shows the time evolution of the losses over 200 epochs. Figure 3 shows the logarithmic histograms of the anomaly score. As we anticipated, a small number of responses were identified as anomalous. In this paper, we do not consider the proper thresholds for anomaly detection, but simply investigate the *anomalous* responses in detail and classify them into three classes as follows.

[1] https://chainer.org.

[2] https://developer.nvidia.com/cuda.



(a) Training with TwiNet.

(b) Training SpiNet only.

Fig. 3. Histograms for anomaly scores under bad and good weather conditions.

1. Some extremely slow vehicles may be misidentified as anomalies, in particular under snowstorm conditions. In such a situation, a vehicle takes an excessive amount of time to arrive at the sensor installation position, exceeding the four seconds allowed. SigNet will, therefore, fail to capture the strain peak caused by the vehicle. We also found cases where SigNet identified false peaks caused by other vehicles appearing just before the target vehicle arrived at the bridge entrance. To deal with such extremely slow vehicles, the input length of SigNet should be extended to enable capture of the peaks for all targets. It should be also noted that a vehicle caught in a traffic jam might not be able to drive at a steady speed, which may also affect the strain response.
2. In some vehicles, cargo may be moving about, with the resulting mechanical shock being captured by the strain sensor. This may cause ripples in the strain peak, causing the signal to resemble that of a vehicle with additional wheels. Such a situation may be observed in particular with light-loaded vehicles, but was hardly detectable from an image with a resolution of 224 × 224. Because the road surface was flat, a heavily loaded cargo bed was unlikely to vibrate.
3. Some vehicles may be misidentified as anomalous due to errors in the traffic dataset. We found that some cars and motorbikes with two axles were included in the large-vehicle dataset, and were then classified as anomalies. Some cars would have more than two axles if they were towing trailers or other vehicles. Such vehicles may be identified as anomalous in this class, not because their appearance was rare, but because of their small impact on the strain response.

6 Discussion

The anomalous response detection based on spiral learning depends on features extracted from the traffic-surveillance video. Video features can be disturbed by environmental conditions, including weather, traffic jams, pedestrians, and other vehicles in the opposite lane. For the target bridge, pedestrians were rare, and we did not observe cases where a vehicle was mistakenly detected as anomalous because of pedestrians. Vehicles in the other lane were also not a problem because the large traffic volume provided SpiNet with sufficient opportunities for learning such situations. Traffic jams may be caused by construction work, snowstorms, or traffic signals. For example, the far-side lane was closed on November 9, 2016 due to construction work. Under such conditions, the bridge behaved abnormally because many vehicles were required to reverse over the bridge. We need not be concerned with such one-off events. However, traffic signals can be a big problem in general. This problem will not be addressed in this paper, because there were no signals near our target bridge.

Figure 3(a) compares the distributions of $\mathcal{L}_{\mathrm{MSE3}}$ for two cases, with heavy snow or haze for several hours and when conditions were fair. We examined cases that clearly seemed to involve snowy or hazy conditions, paying attention to the image sharpness. The two graphs show the robustness of the adversarial–spiral method to changing weather conditions. However, it is not clear whether this approach is necessary to achieve *weatherproof* results. The DS601 database involves both sunny and snowy days, and SpiNet might be able to obtain weatherproof results without the adversarial method. Actually, Fig. 3(b) shows such results. On the other hand, Fig. 2 shows that TwiNet failed in learning an identification function that was generalized sufficiently to detect anomalous vehicles in both the training and validation datasets. This means that SpiNet might resist TwiNet to obtain a feature space from which it is difficult to distinguish anomalous vehicles. The issue of whether adversarial learning is beneficial will be investigated in future work. Another doubt about our results was the fact that some anomalies could be detected without consulting the video data. We are confident that the video feature is necessary as an explanatory variable for strain analysis. However, there might be some anomalies whose strain response appears strange at first glance, e.g., a strain response completely buried in white noise. Failures in sensors may also cause such situations. We need to investigate in detail what type of anomaly requires the spiral learning approach. This will be revealed in future experimental work by examining those cases where SigNet guessed anomalies detected by the spiral learning approach.

7 Conclusion

We have proposed a novel anomaly detection technique for bridge deterioration analysis. The spiral learning enables a direct comparison of vehicle appearance and bridge strain responses in a common feature space. The video feature may play the role of an explanatory variable in the response. The adversarial spiral learning proposal prevented our anomaly detector from being affected by adverse weather. We tested our proposals on real observation data and identified outliers of several identifiable types. In future experimental work, we will investigate the limits on the anomaly types detectable by our method. We believe our proposals will aid bridge damage detection by identifying anomalous strain responses whose characteristics are different from those observed during the construction period.

Acknowledgement. This work was supported by the cross-ministerial strategic innovation promotion (SIP) program (http://www8.cao.go.jp/cstp/gaiyo/sip/) of the Cabinet Office, Government of Japan.

References

1. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: NIPS (2014)
2. Yu, Y., Cai, C.S., Deng, L.: State-of-the-art review on bridge weigh-in-motion technology. Adv. Struct. Eng. **19**(9), 1511–1530 (2016)
3. Bromley, J., Guyon, I., LeCun, Y., Sackinger, E., Shah, R.: Signature verification using a "siamese" time delay neural network. In: NIPS (1993)
4. Kawakatsu, T., Kinoshita, A., Aihara, K., Takasu, A., Adachi, J.: Deep sensing approach to single-sensor bridge weighing in motion. In: EWSHM (2018)
5. Cao, M.S., Sha, G.G., Gao, Y.F., Ostachowicz, W.: Structural damage identification using damping: a compendium of uses and features. SMS **26**(4), 043001 (2017)
6. ZhiWei, C., QinLin, C., Ying, L., SongYe, Z.: Damage detection of long-span bridges using stress influence lines incorporated control charts. SCTS **57**(9), 1689–1697 (2014)
7. Huang, Y., Zhu, C., Ye, Y., Xiao, Y.: Damage detection of arch structure by using deflection influence line. In: SEEIE (2016)
8. Zaurin, R., Catbas, F.N.: Structural health monitoring using video stream, influence lines, and statistical analysis. SHM **10**(3), 309–332 (2011)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
10. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: AISTATS (2011)
11. Caruana, R.: Multitask learning. Mach. Learn. **28**(1), 41–75 (1997)
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NIPS (2014)
13. Kawakatsu, T., Kakitani, A., Aihara, K., Takasu, A., Adachi, J.: Traffic surveillance system for bridge vibration analysis. In: IICPS (2017)
14. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS (2015)
15. Yao, Y., Rosasco, L., Caponnetto, A.: On early stopping in gradient descent learning. Constr. Approx. **26**(2), 289–315 (2007)
16. Reddi, S., Kale, S., Kumar, S.: On the convergence of adam and beyond. In: ICLR (2018)