Automated band selection for Bayesian FFT modal identification Based on RetinaNet

Chen Wang[1a], Zhilin Xue[2], Yipeng Su[1b], Binbin Li[1c*]

[1]ZJU-UIUC Institute, Zhejiang University, Haining, 314400, China
[2]Department of Civil Engineering, Dalian University of Technology, Dalian, 116023, China

[1a]chenw.18@intl.zju.edu.cn, [2]zhilin_xue@mail.dlut.edu.cn, [1b]Yipeng.18@intl.zju.edu.c ,[1c*]bbl@zju.edu.cn

Abstract. Bayesian FFT algorithm is a popular method to identify modal parameters, e.g., modal frequencies, damping ratios, and mode shapes, of civil structures under operational conditions. It is efficient and provides the identification uncertainty in terms of posterior distribution. However, in utilizing the Bayesian FFT algorithm, it is tedious to manually select frequency bands and initial frequencies. This step requires professional knowledge and costs most of time, which prevents the automation of Bayesian FFT algorithm. Regarding the band selection as an object detection problem, we design a band selection network based on the RetinaNet to automatically select frequency bands and a peak prediction network to predict the initial frequencies. The designed networks are trained using the singular value spectrum of measured ambient vibration data and verified by various data sets. It can achieve the human accuracy with much less operation time, and thus provides a corner stone for the automation of Bayesian FFT algorithm.

Keywords-Modal identification; object detection; AI automation; uncertainty quantification

1. Introduction

Modal Identification is an essential aspect of the Civil Engineering field. It could be used to identify building damage, structural system load, and structural system dynamic features. The traditional modal analysis method is mainly experimental modal analysis, which requires artificial excitation of the structure to analyze the modal parameters (such as frequency, damping ratio, and vibration mode) by establishing the input-output relationship of the structure. However, for large civil engineering structures, a sizeable external excitation is often required to obtain sufficient structural response, and the time-consuming and labor-intensive testing process dramatically limits its usefulness. In contrast, operational modal analysis can identify the modal parameters of a structure by simply measuring the random vibration response of the structure under operational conditions, which has gained wide attention and rapid development in the field of civil engineering in recent years.

In operational modal analysis, the external excitation is unknown and is usually assumed to be a stationary random process. Therefore, compared with experimental modal analysis, operational modal analysis is more complex for the identified modal parameters have great uncertainty. Bayesian operational modal analysis converts the modal identification problem into solving the posterior probability estimation of the modal parameters and achieves the uncertainty measure of the modal parameters simultaneously. The Bayesian FFT algorithm is a method of Bayesian operational modal analysis [1], which balances robustness and computational efficiency in the modeling assumption process compared with other Bayesian algorithms. Frequency domain Bayesian FFT algorithm requires two kinds of input to function well, frequency bands containing nature frequencies and the initially nature frequencies predicted by us. Those values are chosen based on professional knowledge, which limits the efficiency of the algorithm and make the automation of identification impossible.

Object detection aims to identify the type of object in the image and locate the object's position. It is a relatively mature direction in deep learning, and a considerable number of object detection algorithms, including Faster-RCNN, YOLO, and RetinaNet have been proposed [2]. Retinanet has good performance in solving the problem of imbalance between background and foreground categories of images. RetinaNet has the advantages of simple structure and high detection accuracy, and we utilize it to realize to enable the choosing of bounding bands. For only one-dimension matters in bounding frequency, the proposed band selection network is a one-dimensional CNN, which reduces computational complexity without losing bounding accuracy. The frequency bands selected by this network could be fed to the Bayesian FFT algorithm directly, significantly reduces the time and labor costs of applying the algorithm.

2. Theoretical background

2.1 Bayesian FFT Algorithm

Bayesian FFT Algorithm is the latest method of operational modal analysis. The following part is a brief introduction of the algorithm. Suppose that the time-history response data of a structure under environmental excitation is $\{\hat{\boldsymbol{y}}_j \in \mathbb{R}^n : j = 0,1,\dots,N-1\}$, ($\{\hat{\boldsymbol{y}}_j\}$for short), where N is the number of samples of each channel, and n is the number of data channels. Regard the structural response signal as a stationary stochastic process, then its Fast Fourier transform (FFT) can be expressed as:

$$\hat{\boldsymbol{\mathcal{F}}}_k = \sqrt{\Delta t/N} \sum_{j=0}^{N-1} \hat{\boldsymbol{y}}_j e^{-2\pi \mathbf{i} jk/N} \tag{1}$$

Δt is the sampling time interval, and $\mathbf{i}$ is the unit of the imaginary number. If $k \leq \mathrm{int}[N/2]$, $\hat{\boldsymbol{\mathcal{F}}}_k$ is the FFT data of frequency corresponding $\mathrm{f}_k = k/N\Delta t$ (Hz), which is used to identify modal parameters of the measured structure.

The number of information used for modal identification and the risk of model error should be considered comprehensively. Generally, only the narrowband containing the concerned modes is selected for the modal identification[3]. Assume that there are N_f data points in the selected frequency band, ($\{\hat{\boldsymbol{\mathcal{F}}}_k\}$ for short), and at the same time:

$$\hat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{\mathcal{F}}_k + \boldsymbol{\varepsilon}_k \tag{2}$$

$\boldsymbol{\mathcal{F}}_k$ represents the result of theoretical calculation for a structural dynamic response, $\boldsymbol{\varepsilon}_k$ represents the prediction error between theoretical result and the measurements of modal respond, including the model error, environmental noise, and so on. If there is an m-orders modal in the selected band, according to the mode superposition method, $\boldsymbol{\mathcal{F}}_k = \boldsymbol{\Phi}\boldsymbol{\eta}_k$, and $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_m] \in \mathbb{R}^{n \times m}$ is the matrix of a modal shape corresponding to the selected sensor location points. $\boldsymbol{\eta}_k \in \mathbb{C}^{m \times 1}$ is the FFT of modal response at the frequency of f_k. Moreover, if $\boldsymbol{p}_k \in \mathbb{C}^{m \times 1}$ represents the model forces at the frequency of f_k, there is $\boldsymbol{\eta}_k = \boldsymbol{h}_k \boldsymbol{p}_k$, then there is:

$$\hat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{\Phi}\boldsymbol{h}_k\boldsymbol{p}_k + \boldsymbol{\varepsilon}_k \tag{3}$$

$\boldsymbol{h}_k = \mathrm{diag}(h_{1k}, h_{2k}, \dots, h_{mk})$ is diagonal matrix, where the diagonal elements are made up of frequency-response function. In addition, the frequency-response function of i-order mode can be denoted as:

$$h_{ik} = \frac{(2\pi \mathrm{i} \mathrm{f}_k)^{-q}}{(1-\beta_{ik}^2) - \mathrm{i}(2\zeta_i \beta_{ik})} \quad \beta_{ik} = \frac{f_i}{\mathrm{f}_k} \quad q = \begin{cases} 0, accleration\ data \\ 1, \quad velocity\ data \\ 2, diplacement\ data \end{cases} \tag{4}$$

f_i (Hz) is the frequency and ζ_i is the damping ratio of the i-order.

During the Operational Modal Analysis, due to the lack of actual excitation data of the structure, modal loads $\boldsymbol{p}_k$ can only be approximated by statistical modeling. Assume the loads meet the steady conditions like zero-mean, and the length of data is long enough ($N_f \gg 1$), then $\boldsymbol{p}_k$ are approximately complex Gaussian distribution and they are statistically independent of each other at different frequencies. The covariance matrix (power spectral density) of them at a specific frequency domain is an unknown constant $\boldsymbol{S} \in \mathbb{C}^{m \times m}$, which means $\boldsymbol{p}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{S})$. Based on the Maximum Information Entropy Theory, giving the predicted error $\boldsymbol{\varepsilon}_k$ meets the complex Gaussian distribution, of which the zero-mean covariance is $S_e \boldsymbol{I}_n$ ($\boldsymbol{I}_n \in \mathbb{R}^{n \times n}$ represents the unit matrix) [4]. By further presuming that $\boldsymbol{p}_k$ and $\boldsymbol{\varepsilon}_k$ are mutually independent, it can be inferred that $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ also obeys the zero-mean complex Gaussian distribution, and its covariance matrix is:

$$\boldsymbol{E}_k = \boldsymbol{\Phi} \boldsymbol{H}_k \boldsymbol{\Phi}^{\mathrm{T}} + S_e \boldsymbol{I}_n \tag{5}$$

The covariance matrix $\boldsymbol{E}_k$ plays a significant role in modal identification, as it contains all unknown parameters, $\boldsymbol{\theta} = \{\boldsymbol{f}, \boldsymbol{\zeta}, \boldsymbol{\Phi}, \boldsymbol{S}, S_e\}$, ($\boldsymbol{f} = [f_1, f_2, \dots, f_m]^{\mathrm{T}}$, $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \dots, \zeta_m]^{\mathrm{T}}$).
Based on the above modal assumptions, it could be further assumed that the prior probability is a uniform distribution. Then, according to the Bayes' theorem, the posterior probability distribution of the unknown parameter $\boldsymbol{\theta}$ is:

$$p(\boldsymbol{\theta} | \{\widehat{\boldsymbol{\mathcal{F}}}_k\}) \propto p(\{\widehat{\boldsymbol{\mathcal{F}}}_k\} | \boldsymbol{\theta}) = \frac{\pi^{-nN_f}}{\prod_k |\boldsymbol{E}_k|} \exp\left[-\sum_k \widehat{\boldsymbol{\mathcal{F}}}_k^* \boldsymbol{E}_k^{-1} \widehat{\boldsymbol{\mathcal{F}}}_k \right] \tag{6}$$

The amplitude constraint of the mode of vibration is usually introduced to make the modal parameters identifiable, so we assume $\boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{\phi}_i = 1$.

With the condition of long data length, the posterior probability distribution of the unknown parameter $\boldsymbol{\theta}$ can be approximated by the Gaussian distribution:

$$p(\boldsymbol{\theta} | \{\widehat{\boldsymbol{\mathcal{F}}}_k\}) \approx (2\pi)^{-n_\theta/2} |\widehat{\boldsymbol{C}}|^{-1/2} \exp\left[-\frac{1}{2} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^{\mathrm{T}} \widehat{\boldsymbol{C}}^{-1} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \right] \tag{7}$$

$\widehat{\boldsymbol{\theta}}$ is a maximal posterior estimate. $\widehat{\boldsymbol{C}}$ is covariance matrix equal to logarithmic posterior probability distribution function at the value $\widehat{\boldsymbol{\theta}}$ of the inverse Hessian matrix. It means the posterior uncertainty of the parameter [5].

Maximum posterior estimation of $\widehat{\boldsymbol{\theta}}$ and covariance matrix $\widehat{\boldsymbol{C}}$ are based on measured data, so both can be determined when FFT data $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ is acquired. However, the process of getting the result is not easy. Because of the large number of unknown parameters and the covariance matrix $\boldsymbol{E}_k$ close to the singular value, direct optimization is not practical (e.g., maximum gradient method, Newton method). At the same time, there is a huge challenge for the convergence of algorithms under the condition of closely spaced modes. Against these difficult points, we develop the algorithm based on the Expectation Maximum(EM) principle. Our algorithm regards the modal respond $\boldsymbol{\eta}_k$ as a hidden variable. The analytical solutions of the expected step and the maximization step are derived at the same time. As a result, the calculation speed is extremely fast. The detailed introduction of the algorithm is in the reference [5].

In practice, however, the EM algorithm mentioned above needs preset value of two boundaries for each band and initial frequencies. The boundaries need to be selected manually which is tedious and time-costing, preventing the EM algorithm from analyzing long-time-span data. Solution for this could be found in object detection in computer vision field.

2.2 Object Detection

Most of the object detection neural network is designed to bound and identify objects in images. As shown in figure 1, the triangle, hexagon and diamond represent three kinds of targets respectively. The image-based target detection algorithm needs to locate the position of the target in the image by the coordinates of the upper-left and lower-right corners of the target border, and at the same time identify the kind to which the target belongs [6-7].

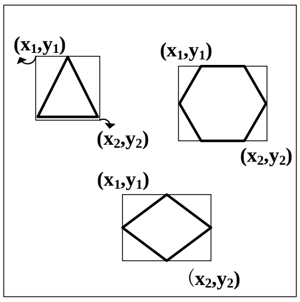


Fig.1 Image Object detection

As shown in figure 2, the frequency band to be selected can be determined with only two parameters: each band's left and right boundary. The two-dimensional objects are simplified to one-dimensional ones, which reduces computational complexity by having fewer anchors and fewer network parameters to evaluate. The problem of automatic frequency band selection could be regarded as a one-dimensional target detection task.

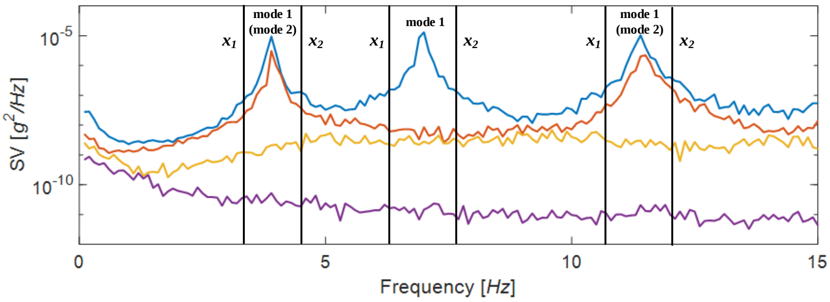


Fig.2 Frequency bands detection in SV

3. Proposed Method

3.1 Overview of Proposed Method

As shown in figure 3, there are mainly two steps in the proposed method: the first step is to select bands by the designed band selection network which is employed to identify the band coordinates and number of modes; the second step is to predict the initial frequencies in selected bands based on the designed peak prediction network and the results of the band selection network.

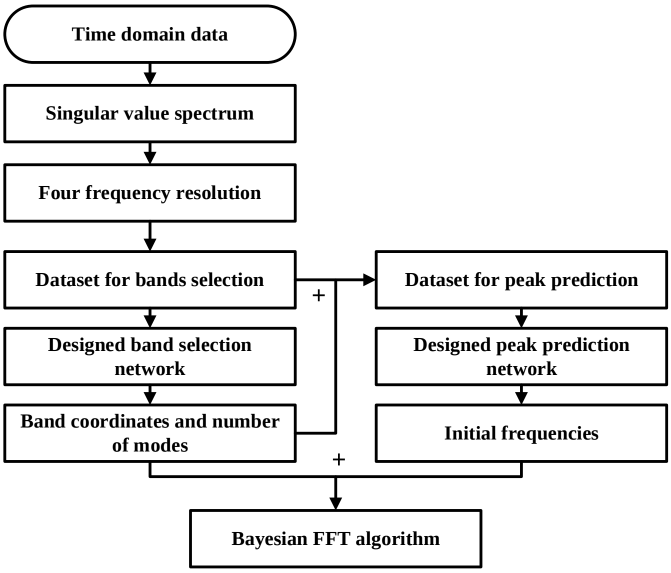
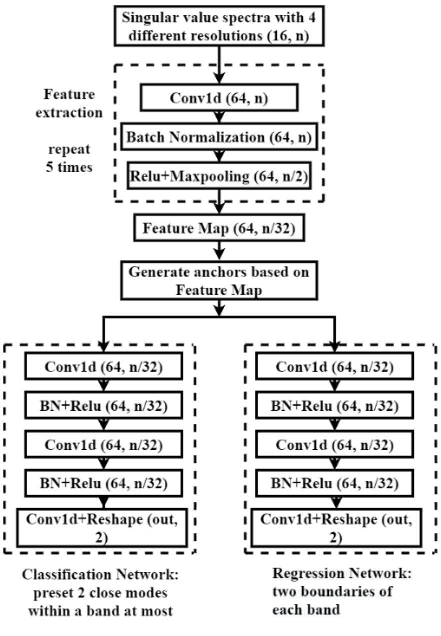
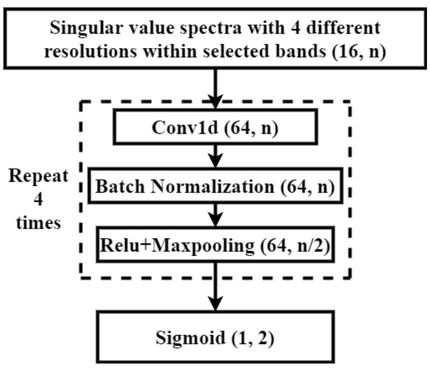


Fig.3 Overview of proposed method

3.2 Proposed Band Selection Network and Peak Prediction Network

The structure of the designed band selection network is shown in figure 4a, including a feature extraction network, a classification sub-network and a regression sub-network. We assumed there is a maximum of two close modes within a band, which is accurate enough in most cases. Two regression parameters are needed in the regression sub-network, i.e., the left and right boundary points of the selected band. There are only 9×10^4 parameters in the band selection network, considerably reduced the computation complexity. The structure of the peak prediction network is shown in figure 4b, which is composed of four one-dimensional convolution layers and a full connected layer. The full connected layer outputs the predicted initial frequencies in the selected bands.





a. Bands selection network b. Peak prediction network

Fig.4 Structure of the proposed band selection network and peak prediction network

4. Experimental verification

4.1 Laboratory Model and Dataset

In this section, the proposed band selection network together with 1D CNN is investigated with acceleration response data of a lab shear frame. The lab model is a 3-story shear frame, 5 kg per floor as shown in figure 5. The total measured duration time is $10\ min/set \times 54\ set = 3h$ at $256Hz$ and there are 4DOFs, xy at two corners on long side of top floor. 36 sets of lab data are fed into the network to train the model. The four frequency accuracies selected during the experiment were 0.01, 0.02, 0.05, 0.1Hz. The samples need to be labeled before they are normalized, and the labels need to be given two parameters, namely the starting frequency and the cutoff frequency. These two values are first selected artificially, and then the corresponding frequencies and damping ratios are identified by EM algorithm, determining labeling information based on formula (8). f_{min} and f_{max} represent the starting frequency and the cutoff frequency, f_i is the ith identified natural frequency, ξ_i is lab model's ith damping ratio, κ is bandwidth coefficient, Δf represents the frequency resolution of the data involved in the calculation.

The problem of inconsistent SV dimensions with different resolution ratios is solved by linear interpolation. The interpolated SV will form a 4×n dimensional matrix(n representing the number of SV data points at each frequency accuracy). Both SVs with high-frequency accuracy and SVs with low-frequency accuracy are referenced when selecting band parameters automatically. The whole 'data' contains $18 \times 3 = 54$ sets of data, each 600 second long and sampled at 256Hz. We choose first 36 sets of data for training and the left 18 sets of data for testing. The detailed description for data could be found in references. Initialy the size of sampled data is 16×600, we extends the sample to size of 16×1024 by filling zeros to fit the network. The samples for the peak prediction network are the SV information within the selected bands, whose size is extended to 16×256 by filling zeros.

Fig.5 Lab shear frame model

$$\begin{aligned} f_{min} &= min\,\{f_i(1-\kappa\xi_i), (f_i - 50 * \Delta f)\} \\ f_{max} &= max\,\{f_i(1+\kappa\xi_i), (f_i + 50 * \Delta f)\} \end{aligned} \quad (8)$$

4.2 Experiment Result

An overview of the modes is shown in figure 6, from which it can be seen that there are three bands to select. Besides, there are close modes within the first and third bands. The comparison between the automatic identification results and artificial labels of frequencies is shown in figure 7. The first two figures are results for band 1 and band 3 that are close modes; the last figure shows the well-separated mode. It's worth mentioning that the modified 1D network functions well under the circumstance of closely-spaced mode. Figure 8 illustrates the comparison between the automatic identification with artificial labels of damping ratios. Similarly, the network has a good performance for both well-separated and closely-spaced modes.

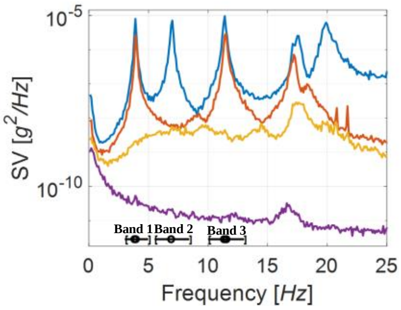
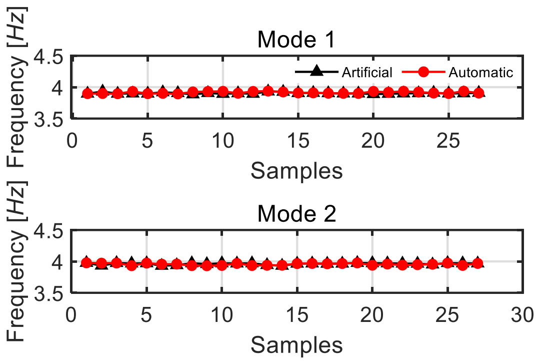
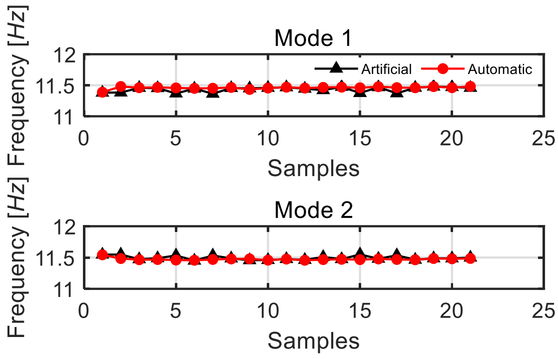


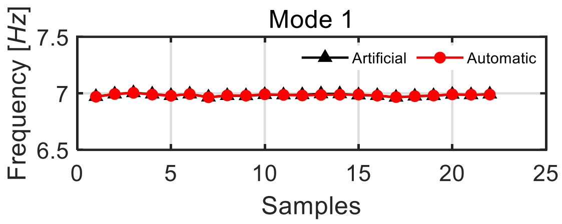
Fig.6 Bands to be selected



a. 3.5-4.5Hz

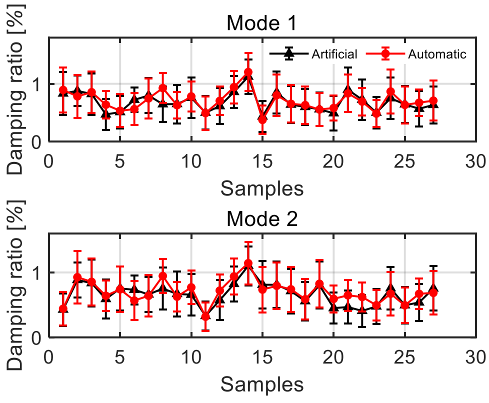


b. 11-12Hz

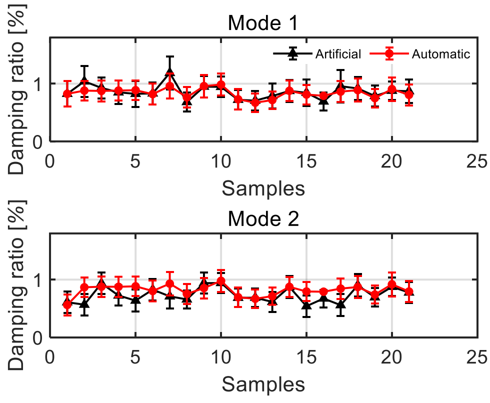


c. 6.5-7.5Hz

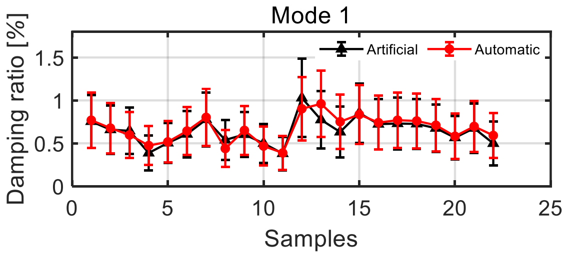
Fig.7 Identified modal frequencies



a. 3.5-4.5Hz



b. 11-12Hz



c. 6.5-7.5Hz

Fig.8 Identified damping ratios

5. Conclusion

A band selection network based on the RetinaNet and a peak prediction network is proposed to select the frequency bands to automate the Bayesian FFT algorithm for operational modal analysis. The proposed method utilizes SV information at multiple frequency resolutions and is capable of extracting modal information for both well-separated and closely-spaced modes. The examples show it can achieve human accuracy with only milliseconds and thus can significantly accelerate the modal analysis. More training data set is expected to feed in the proposed model from various types of civil engineering structures, e.g., bridges, buildings, wind turbines, to have more general applicability and better generalization ability.

6. Acknowledgements

This study is supported by the National Natural Science Foundation of China under grant No. 51908494. The funding support is greatly appreciated.

References

[1] Wang Y.Y., Wang Chao, et al (2019) Automatic ship detection based on RetinaNet using muli-resolution Gaofen-3 imagery [J]. Remote Sensing, 11(5).

[2] Kim H.J., Sim S.H. (2019) Automated peak picking using region-based convolutional neural network for operational modal analysis [J]. Structural Control Health Monitoring, No 26(11).

[3] Chen J, Xu Y L, Zhang R C. (2004) Modal parameter identification of Tsing Ma suspension bridge under Typhoon Victor: EMD-HT method[J]. Journal of Wind Engineering and Industrial Aerodynamics, 92(10): 805-827

[4] Li B., Au S.-K. (2019) An expectation-maximization algorithm for Bayesian operational modal analysis with multiple (possibly close) modes [J]. Mechanical Systems and Signal Processing, 132:490-511.

[5] Au S.-K., (2017) Operational Modal Analysis [M]. Singapore: Springer., pp 328-330.

[6] T. Lin, P. Goyal, R. Girshick, K. He and P. (2017) Dollár, Focal Loss for Dense Object Detection, IEEE International Conference on Computer Vision (ICCV), , pp. 2999-3007.

[7] Ren S, He K, Girshick R, et al. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 28: 91-99