To cite this article: JI Y Y, WANG D Y. Near-field acoustic reconstruction method based on three-dimensional N-shaped convolution neural network and frequency focal-KH regularization [J/OL]. Chinese Journal of Ship Research, 2023, 18(6). http://www.ship-research.com/en/article/doi/10.19693/j.issn.1673-3185.03127 (in both Chinese and English).

DOI: 10.19693/j.issn.1673-3185. 03127

Near−field acoustic reconstruction method based on three−dimensional N−shaped convolution neural network and frequency focal−KH regularization

JI Yuyang[1,2], *WANG Deyu*[*1,2]

1 State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

2 Institute of Marine Equipment, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: [**Objectives**] Low sampling rates on near-field acoustic reconstruction cause high reconstruction error in near-field acoustic holography. Therefore, a deep learning-based approach which is applicable to planar sound sources and high-precision reconstruction with low sampling rates is put forward. [**Methods**] A three-dimensional N-shaped convolution neural network for near-field acoustic reconstruction is established to extract features in the frequency dimension in order to make up for sparse sampling in the spatial dimension. A frequency focal mechanism, namely an adaptive frequency weight focus mechanism, is put forward to improve reconstruction precision in the natural frequency and high frequency. Moreover, this paper also raises frequency-scaled focal loss and frequency-scaled focal Kirchhoff-Helmholtz (KH) loss, which are considered regularization. To validate the proposed methods, datasets are created with COMSOL Multiphysics and Matlab. [**Results**] The mean error range of 100−2 000 Hz of the algorithm proposed in this paper is only 4.96%, higher than those of SRCNN and PV-NN. [**Conclusions**] The proposed method is verified as having the potential to reconstruct the accurate velocity fields of sound sources under low sampling rates.

Key words: near-field acoustic reconstruction; sound source recognition; 3D convolution; Kirchhoff−Helmholtz (KH) regularization

CLC number: TB52; U661.44

0 Introduction

Accurate identification and localization of noise sources are essential prerequisites for downstream tasks such as noise reduction analysis and fault detection in the mechanical structure and equipment of ships[1-3]. Near-field acoustic reconstruction provides a non-contact method for source identification and acoustic visualization. By sampling holographic acoustic quantities of the near-field source, the surface vibration velocity of the source can be reconstructed. This method finds widespread application in the field of ship engineering, including underwater noise source localization and identification[4-6], analysis and control of noises in ship compartments [7], and analysis of noise sources in lightweight materials [8].

According to Shannon's sampling theorem, the holographic spatial distance between two points in the wavenumber domain during the reconstruction of the surface vibration velocity of a sound source should satisfy $|k_x| < \pi/\Delta_x$ and $|k_y| < \pi/\Delta_y$, where k_x and k_y, and Δ_x and Δ_y are the wavenumbers and sampling intervals in the x and y directions

Received: 2022 − 10 − 12 **Accepted**: 2022 − 12 − 22

Authors: JI Yuyang, male, born in 1998, master's degree candidate. Research interest: ship noise source identification. E-mail: jiyuyang@sjtu.edu.cn

WANG Deyu, male, born in 1963, Ph.D., professor, doctoral supervisor. Research interests: structural mechanics of ship and ocean engineering, structural optimization design and reliability analysis, structural ultimate strength and test technology research. E-mail: dywang@sjtu.edu.cn

***Corresponding author**: WANG Deyu

respectively. When the spatial sampling interval does not meet this condition, overlap errors, known as wrap-around errors[9], may occur in the wavenumber domain reconstruction, leading to significant reconstruction errors. Expanding the sampling range and increasing the sampling interval will significantly increase the measurement cost for sound source identification and localization in ship machinery. Additionally, the confined space in some compartments of the ship makes measurements inconvenient. Therefore, studying near-field acoustic reconstruction problems under sparse sampling conditions is highly practical for ship engineering. The near-field acoustic holography based on compressed sensing theory[10-12] provides a solution for high-resolution physical field reconstruction under sparse sampling conditions[13-17]. Chen et al.[7], combining compressed sensing theory with the plane equivalent source algorithm, reconstructed the surface sound pressure distribution of high-frequency weak sound sources in ship compartments. However, the effectiveness of compressed sensing methods depends on the selection of sparse bases[16], and different types of sound sources correspond to different sparse bases. Additionally, the reconstruction errors in the high-frequency region are also relatively high.

In recent years, scholars have begun to incorporate deep learning theory into the near-field acoustic reconstruction problems. Deep learning, known for its powerful feature extraction capabilities, has been widely applied in the field of ship sound and vibration analysis [18-19]. Olivieri et al. [20], utilizing a two-dimensional convolutional neural network, introduced a super-resolution convolutional neural network for near-field acoustic holography (SRCNN-NAH) and achieved favorable results on a violin-shaped plate. Subsequently, Olivieri et al. [21] proposed a loss function based on the Kirchhoff-Helmholtz (KH) forward sound radiation equation and released an open dataset. Wang et al. [8], leveraging an autoencoder, presented a pressure-velocity neural network model (PV−NN) and validated a data normalization method suitable for near-field acoustic reconstruction problems.

This paper focuses on the acoustic field generated by the stimulated vibration of a rectangular thin plate in an air medium, intending to propose a three-dimensional N-shape convolutional neural network framework for near-field acoustic holography (3D NCNN-NAH). This framework aims to enhance the accuracy of near-field acoustic reconstruction under conditions of hologram surface and low sampling rate, particularly in practical ship applications. Additionally, the paper introduces a loss function comprising frequency focal-normalization for reconstruction mean square error, and KH regularization terms to mitigate the challenge of lower accuracy in the reconstruction of high-frequency and certain source-specific frequency intervals. Finally, the effectiveness of the proposed method will be validated through the batch construction of datasets using COMSOL Multiphysics software.

1 Near-field acoustic identification

As shown in Fig. 1, the sound pressure, denoted as $p(\boldsymbol{r}, \omega)$, at a point in space, $\boldsymbol{r} = [x, y, z]^{\mathrm{T}}$, can be expressed using the KH integral equation.

$$p(\boldsymbol{r},\omega) = \int_S p(\boldsymbol{s},\omega)\frac{\partial}{\partial \boldsymbol{n}} g_\omega(\boldsymbol{r},\boldsymbol{s})\mathrm{d}\boldsymbol{s} - \mathrm{i}\omega\rho_0 \int_S v_n(\boldsymbol{s},\omega) g_\omega(\boldsymbol{r},\boldsymbol{s})\mathrm{d}\boldsymbol{s} \tag{1}$$

where ω represents the circular frequency; S denotes the surface of the sound source, $\boldsymbol{s} = [x', y', z']^{\mathrm{T}}$ is the vector of a point on the sound source surface S; $\boldsymbol{n}$ is the normal direction of the sound source; $g_\omega(\boldsymbol{r}, \boldsymbol{s})$ is the Green's function; i is the imaginary unit; ρ_0 is the density of the medium of the acoustic field; v_n is the normal vibration velocity on the sound source surface. The free-field Green's formula satisfying the second kind of boundary condition (Neumann boundary condition) is given by

$$g_\omega(\boldsymbol{r},\boldsymbol{s}) = \frac{1}{4\pi}\frac{\mathrm{e}^{-\mathrm{i}\frac{\omega}{c}\|\boldsymbol{r}-\boldsymbol{s}\|}}{\|\boldsymbol{r}-\boldsymbol{s}\|} \tag{2}$$

where c is the sound propagation velocity.

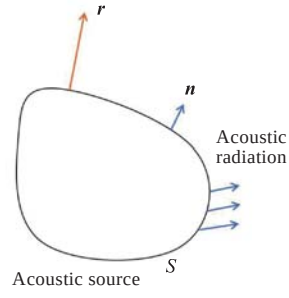

Fig. 1 Schematic diagram of acoustic radiation from vibrational sound source

Taking a plate vibrating under external loads and generating a radiating acoustic field as the example, from Eq. (1) and Eq. (2), we can observe that if the

acoustic pressure on the plane where the sound source is located is known, the acoustic pressure at any point in the radiating field can be derived. The Euler equation in Eq. (3) is employed to establish a connection between the acoustic pressure at a specific point in space and the particle velocity. In other words, the sound pressure at any point within the sound field can be inferred from the known velocity at the sound source surface.

$$\mathrm{i}\omega\rho_0\boldsymbol{v} = \nabla p \tag{3}$$

where $\boldsymbol{v}$ represents the particle velocity at a given point in space.

When the acoustic pressure on any plane of the radiating sound field is known, it is not possible to retroactively deduce the distribution of sound pressure and vibrational velocity on the sound source surface using the KH integral equation. However, when the distance between the holographic surface and the sound source is much smaller than the wavelength of the sound wave, it is possible to capture the evanescent waves in the high wavenumber domain. In this case, the inverse problem of acoustic field radiation becomes the near-field acoustic reconstruction. The equation for calculating the normal vibrational velocity on the sound source surface, inverted from the acoustic pressure in the radiating sound field, is given by

$$v_n(x,y,z_\mathrm{s}) = F_x^{-1}F_y^{-1}\left[F_xF_y\left[p(x,y,z_\mathrm{h})\right]G(k_x,k_y,z_\mathrm{s}-z_\mathrm{h})\right] \tag{4}$$

where $v_n(x, y, z_\mathrm{s})$ represents the normal vibration velocity on the sound source surface at a height z_s; $p(x, y, z_\mathrm{h})$ is the sound pressure in the radiated sound field at a height z_h; F_x and F_y are velocity transfer functions, indicating Fourier transforms in spatial domain with respect to x and y to convert to the wavenumber domains k_x and k_y, where k_x and k_y are the components of wavenumber in the x and y directions, respectively; F_x^{-1} and F_y^{-1} are functions representing the inverse Fourier transform in the wavenumber domain to convert back to the spatial domains x and y; $G(k_x, k_y, z_\mathrm{s}-z_\mathrm{h})$ is a positional parameter function.

Due to the ill-posed nature of the inversion process in Eq. (4), direct solutions are not feasible. Consequently, various approximate methods have been developed for near-field acoustic reconstruction, primarily involving fitting the inverse of the nonlinear equation *Function* in Eq. (5).

$$v_n(x,y,z_\mathrm{s}) = Function^{-1}(p(x,y,z_\mathrm{h})) \tag{5}$$

Based on the powerful feature extraction capabilities of convolutional neural networks (CNNs) based on deep learning, which can approximate complex nonlinear processes. In light of the challenges posed by the super-resolution problem in near-field acoustic reconstruction, this paper aims to develop a 3D N-type convolutional neural network framework (3D NCNN-NAH) and a novel loss function tailored for near-field acoustic reconstruction. The objective is to effectively fit the inverse of the nonlinear equation *Function*, thereby reducing the number of sampling points on holographic surfaces while ensuring a high level of reconstruction accuracy.

2 Super-resolution near-field acoustic reconstruction based on 3D NCNN-NAH

2.1 Problem description

In the Cartesian coordinate system, a thin rectangular plate undergoes harmonic excitation vibration and radiates a sound field outward. This paper aims to reconstruct high-resolution velocity information of the source surface by sampling low-resolution sound pressure signals on the holographic surface. As illustrated in Fig. 2, surface H represents the holographic surface located in the near field of the sound source, while surface S represents the plane of the thin plate (sound source). The sampling resolution of the holographic surface is $N_\mathrm{h} \times M_\mathrm{h}$, where N_h and M_h denote the numbers of sampling points along the x and y directions, respectively. The reconstruction resolution of the reconstruction surface is $N_\mathrm{s} \times M_\mathrm{s}$, where N_s and M_s represent the numbers of reconstruction points along the x and y directions, respectively. In this study, N_s is set to $4N_\mathrm{h}$, and M_s is set to $4M_\mathrm{h}$.

When the same model is subjected to harmonic excitations of different frequencies at the same point, it will induce vibrations of the sound source in different modes. Although the mode forms are diverse, they are inherently correlated, all reflecting the intrinsic characteristics of the excited object. Therefore, rational utilization of frequency domain information can compensate for the sparsity of spatial domain features caused by sparse sampling. Similar to 2D convolution that computes and extracts features in the spatial domain, this paper will employ 3D convolution as a feature extractor. By sliding the convolutional kernel in both spatial

and frequency domains and using convolution calculations and nonlinear activation functions, we extract high-dimensional features in this study.

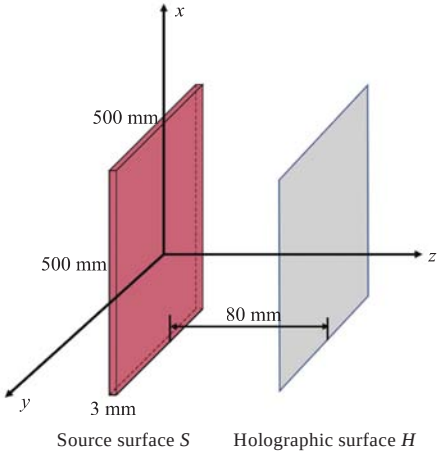

Fig. 2 Illustration of source surface and holography surface

To facilitate the use of 3D convolution for feature extraction in both frequency and spatial domains under the pytorch deep learning framework, this study adopts a five-dimensional tensor for input and output. The first dimension represents the batch dimension, facilitating batch input to ensure a smoother gradient descent curve. Batch input helps better represent the overall distribution compared to individual sample inputs, and it is advantageous for leveraging the compute unified device architecture (CUDA) designed for GPU parallel computing, thereby accelerating training speed. The second dimension represents the frequency domain, with harmonic excitations selected at an interval of 10 Hz within the range of 100−2 000 Hz. Thus, the number of elements in the second dimension is 191. The third dimension is the feature channel dimension, assuming there are m extracting features of convolution kernels in each step of gradient descent. This results in a total of m different features. The fourth and fifth dimensions represent the spatial domain, with input dimensions being the sampling point resolution $N_h \times M_h$ of the holographic surface, and the output dimensions being the resolution $N_s \times M_s$ of the reconstruction surface.

In the deep learning algorithm, a neural network framework with parameters to be optimized is constructed. Through continuous iterations based on a substantial amount of data and the back-propagation algorithm, the mapping from the input data distribution to the output data distribution can be obtained. Fig. 3 illustrates the training procedure of the near-field acoustic reconstruction problem based on the deep learning method.

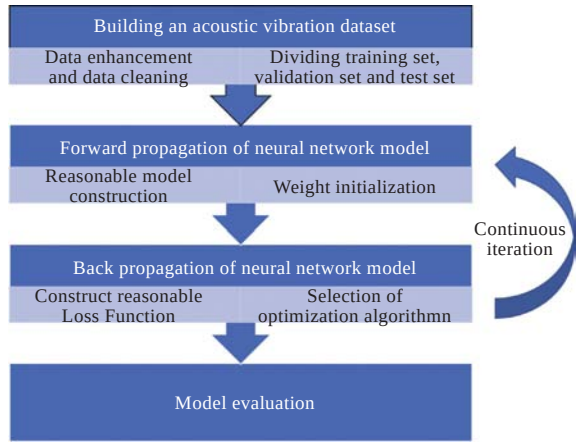

Fig. 3 Procedure of deep learning training

2.2 Network framework

For reconstruction-type problems, the encoder-decoder structure is widely utilized. In the field of near-field acoustic reconstruction, Olivieri et al.[20] employed deconvolution in the UNet decoder to upsample to a resolution equal to the input resolution. They further extracted features to obtain a super-resolved reconstructed image. However, due to the lack of semantic information fusion in the lower layers of the super-resolution part, the reconstruction errors are big. Wang et al. [8], in their designed PV-NN, also used an encoder-decoder structure, incorporating an autoencoder in the reconstruction process of the vibrating field on the source surface.

Addressing the near-field acoustic reconstruction problem with super-resolution, in this study we propose a three-dimensional convolutional neural network consisting of a pre-encoder, encoder, and decoder. Given its framework shape resembling an uppercase "N" (Fig. 4), it is abbreviated as 3D NCNN-NAH. The 3D NCNN-NAH presented in Fig. 4 introduces a pre-encoder module to overcome the deficiency in semantic information for the super-resolution part as identified by Olivieri et al. [20] in SRCNN. In this study we replace 2D convolution with 3D convolution to address the sparsity of spatial sampling on the holographic surface by extracting serialized features in the frequency domain. Simultaneously, the adoption of the deep layer aggregation (DLA) structure achieves feature map fusion of multi-layer size and hierarchy. The parameter settings in Fig. 4 are as follows: the batch length B of the input size is 16; the frequency dimension length F is 191; the channel dimension length C is 1; $N_h \times M_h$ varies based on the shape of the rectangular thin plate,

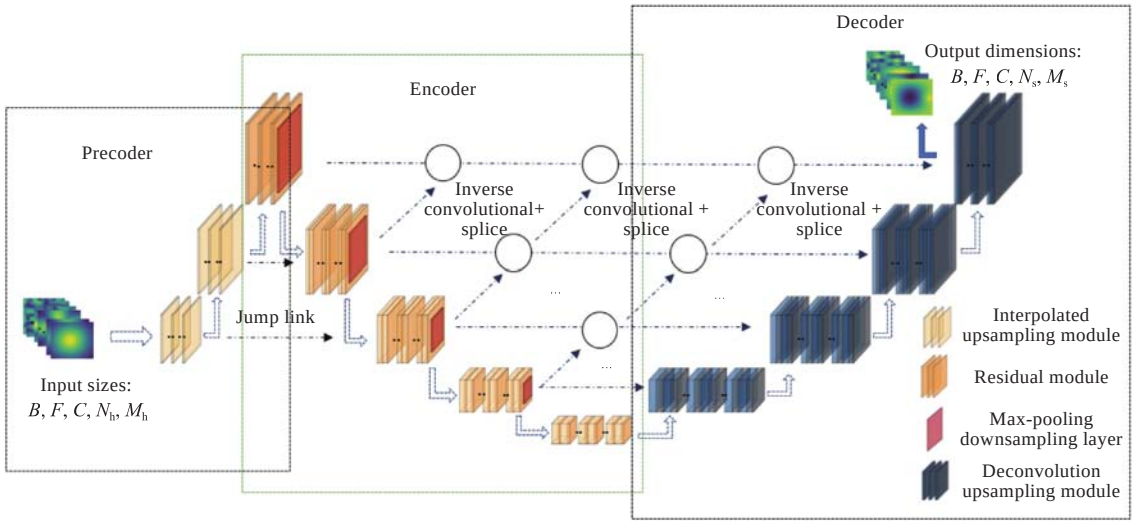

Fig. 4 Framework of 3D NCNN-NAH

with 8 × 8 used for validation in this study. The dimensions of the output size are the same as those of input, with $N_s \times M_s = 32 \times 32$.

2.2.1 Interpolated upsampling module

Interpolated upsampling involves enhancing the resolution of low-resolution images through interpolation. In the 3D NCNN-NAH, the interpolated upsampling module is positioned within the precoder module. To supply semantic information with the original resolution for the super-resolution features in the encoder and decoder based on the acoustic pressure features of the initial holographic surface, the precoder includes two consecutive layers of interpolated upsampling. These layers exclusively interpolate the spatial dimensions, namely dimensions 4 and 5. In the first layer, bilinear interpolation is employed, as depicted in Fig. 5 and Eq. (6). In the second layer, nearest-neighbor interpolation is utilized. In Fig. 5, point 1, point 2, point 3, point 4 represent the contribution units of the interpolated data, the interpolated point is the position of the target data, and x_1, x_2, y_1, y_2 are the data parameters of the contribution units.

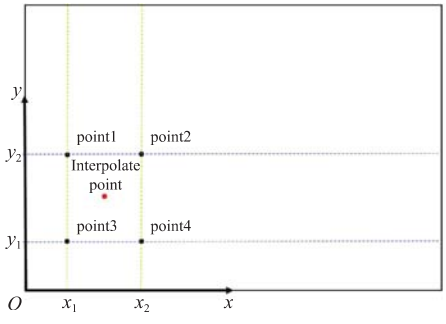

Fig. 5 Bilinear interpolation block in precoder

$$p(x,y)=\frac{y_2-y}{y_2-y_1}\left(\frac{x_2-x}{x_2-x_1}p(x_1,y_1)+\frac{x-x_1}{x_2-x_1}p(x_2,y_2)\right)+\frac{y-y_1}{y_2-y_1}\left(\frac{x_2-x}{x_2-x_1}p(x_1,y_1)+\frac{x-x_1}{x_2-x_1}p(x_2,y_2)\right) \quad (6)$$

where $p(x, y)$ is the sound pressure value at coordinate (x, y), and $\frac{y_2-y}{y_2-y_1}=\frac{y-y_1}{y_2-y_1}=\frac{x_2-x}{x_2-x_1}=\frac{x-x_1}{x_2-x_1}=\frac{1}{2}$ are the sound pressure values at points (x_1, y_1) and (x_2, y_2) respectively.

2.2.2 Residual module

For deep neural networks, to address issues such as gradient vanishing caused by the compounding effect during backpropagation and error dispersion due to deep-layer structures, in this paper we adopt residual modules as the fundamental structure of the encoder. Each residual structure comprises two layers of 3D convolution-batch normalization (BN)-ReLu activation function stacked together, with residual edge connection on the outer layer, as depicted in Fig. 6. The 3D convolution kernel size for feature extraction in the spatial and frequency domains is (5, 3, 3), while the 3D convolution kernel size for the neck structure connecting the encoder and decoder is (5, 1, 1). BN layers are employed to alleviate data shift issues during the forward transfer in the network, ensuring data distribution falls within the non-linear region of the activation function and providing a certain regularization effect.

Following the residual module, a maximum pooling down-sampling layer will be employed to

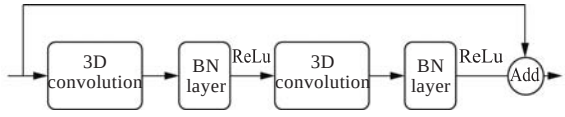

Fig. 6 Illustration of residual block

decrease the resolution of the feature map. This aims to enhance the receptive field of the subsequent convolutional kernel and decrease the computational load of the network. In this study, the kernel size for max-pooling downsampling is (1, 2, 2). Considering that the correlation between high-frequency and low-frequency domains in the frequency domain is lower than that in the spatial domain, in this paper we only extend the receptive field range for the spatial domain.

2.2.3 Deconvolution upsampling module

For the deconvolutional up-sampling, zero-padding is applied to the feature maps. Other steps are similar to the regular convolutional modules, involving the sliding of convolution kernels on the feature map to extract features and simultaneously increase the feature map resolution. In this paper we will utilize deconvolution upsampling module as the fundamental structure of the decoder and employ a residual structure to enhance the decoder's performance. The structure of each deconvolution up-sampling module is illustrated in Fig. 7. In comparison with the residual structure, the first 3D convolution is replaced by 3D deconvolution. Additionally, low-level semantic information from previous layers is fused through feature concatenation, i.e., stacked in the second dimension (feature dimension).

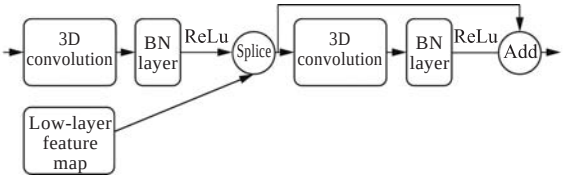

Fig. 7 Illustration of deconvolution module

In this study, when the deep feature fusion method is employed, a tree-like structure is utilized to fuse feature maps of different sizes from the pre-encoder, encoder, and decoder. This choice is motivated by the need for fine-grained, point-wise super-resolution reconstruction in the near-field sound source. Consequently, both local fine-grained features and global holistic features significantly influence the final reconstruction results. The deep feature fusion structure maximally facilitates the interaction of features with different granularities at both low and high levels.

2.3 Loss function

The essence of deep neural networks lies in an optimization problem. The optimization parameters in 3D NCNN-NAH consist of the convolutional kernel weight parameters of 3D convolution and 3D deconvolution, as well as the scaling and bias parameters of the BN layer. The objective function is the loss function comprising frequency scaled focal loss and the KH regularization term. The optimization process will be carried out using the Adam gradient descent method.

2.3.1 Frequency scaled focal loss function (FSF-Loss)

Mean squared error loss (MSE-Loss) function is commonly employed for deep learning reconstruction problems, as shown in Eq. (7). The performance of the network is evaluated by calculating the mean squared difference between the reconstructed velocity values of the sound source surface through the network and the supervised values.

$$L_{\mathrm{MSE}} = \frac{1}{B\times F\times H\times W}\cdot \sum_{\alpha}\sum_{\beta}\sum_{\gamma}\sum_{\varepsilon} E(\|v_{\mathrm{GT}_{\alpha,\beta,\gamma,\varepsilon}} - v_{\mathrm{PRED}_{\alpha,\beta,\gamma,\varepsilon}}\|_2) \tag{7}$$

where L_{MSE} represents the MSE-Loss function; B, F, H, W denote the data lengths of a specific dimension in each dataset within the neural network; E signifies the mathematical expectation; v represents the normal velocity of the sound source surface, where the subscript GT denotes the true value of the normal velocity, and PRED denotes the reconstructed value of normal velocity; α, β, γ, ε refer to the batch dimensions, frequency domain dimensions, and the width and height directions of the thin plate, respectively.

However, when the MSE-Loss function is directly applied to the near-field acoustic reconstruction problems, the following issues arise:

1) The MSE-Loss function measures absolute errors. However, in different frequency domains, especially in the vicinity of the frequency domain with the inherent features of the sound source, there is a significant difference in the normal velocity when compared to other frequency bands. Fig. 8 shows the phase shift (proportional to velocity) of the normal velocity on the sound source surface of various points. If the MSE-Loss function is directly applied, the majority of the loss is composed of reconstruction loss values near the inherent features, thereby neglecting the reconstruction accuracy in non-inherent frequency bands.

2) The accuracy on near-field acoustic reconstruction in the high-frequency range is generally low.

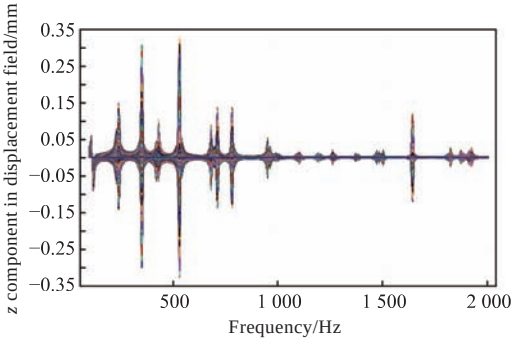

Fig. 8 Displacement of the reconstructed points on sound source varies in frequency domain in training dataset

Samples with low reconstruction accuracy directly using the MSE-Loss function become hard-to-train in the frequency domain. This is because, compared to the low-frequency range, the distribution of the reconstructed normal velocity field in the high-frequency range is more complex, and the acoustic field details are richer. The MSE-Loss function lacks the ability to distinguish between easy and difficult-to-learn samples in frequency domain intervals.

To address the above issues, this paper proposes a FSFLoss function, as shown in Eq. (8) and Eq. (9). Eq. (8) represents the error at frequency β and normalizes across the batch and spatial dimensions, ensuring that the difference between the true and predicted values is at the same order of magnitude. This helps resolve the issue that the majority of the loss values come from the inherent frequency domain of the sound source in the MSE-Loss function.

Due to the different details contained in the acoustic field, when the features are complex, the difficulty of feature extraction between different frequency domains will be greater, leading to a decrease in reconstruction accuracy. Eq. (9) adaptively measures the difficulty of training in different frequency bands by considering the Frobenius (F) norm of the difference between the reconstructed velocity field and the true velocity field for each frequency domain. As the F norm increases, indicating a larger normalized error in the reconstructed velocity field, the loss function weight should be increased. Conversely, as the F norm decreases, indicating a smaller normalized error in the reconstructed velocity field, the loss function weight can be reduced. This adaptive mechanism allows the loss function to focus more on learning hard-to-train samples in different frequency domains, effectively implementing a frequency-scaled attention mechanism.

$$L_{\mathrm{FS}\beta}=\frac{E(\|v_{\mathrm{GT}_{\alpha,\gamma,\varepsilon}}-v_{\mathrm{PRED}_{\alpha,\gamma,\varepsilon}}\|_2)}{E(\|v_{\mathrm{GT}_{\alpha,\gamma,\varepsilon}}\|_2)} \tag{8}$$

$$L_{\mathrm{FSF}\beta}=\frac{1}{F}\sum\nolimits_{\beta}E(\|v_{\mathrm{GT}_{\alpha,\gamma,\varepsilon}}-v_{\mathrm{PRED}_{\alpha,\gamma,\varepsilon}}\|_F)^{\mu}\cdot L_{\mathrm{FS}\beta} \tag{9}$$

where $L_{\mathrm{FS\beta}}$ is the normalized reconstructed MSE loss function; $L_{\mathrm{FSF\beta}}$ is the FSFLoss function; μ is a hyperparameter utilized to regulate the focus level and is set to 0.5 in this paper.

2.3.2 Frequency scaled focal-KH regularization term

The near-field acoustic reconstruction problem delineates the inverse process of the radiated sound field, whereas the KH equation describes the positive process of the radiated sound field. The near-field acoustic holography problem, utilizing the equivalent source method, has shown that physical prior information aids in the regularization of near-field acoustic reconstruction. In light of this, Olivieri et al.[21] proposed incorporating the discrete form of the forward-propagating KH equation as part of the loss function to exploit its regularization effect. Eq. (10) represents the discrete form of the KH equation in the free field.

$$p_{\mathrm{KH_{PRED}}}(x_1,y_1,f_{\mathrm{a}})=-\mathrm{i}\rho_0 f_{\mathrm{a}}\sum\nolimits_{\beta}\sum\nolimits_{\alpha}\dot{w}(x_\alpha,y_\beta,f_{\mathrm{a}})\cdot\frac{\mathrm{e}^{\mathrm{i}\frac{2\pi f_{\mathrm{a}}}{c_0}\sqrt{(x_1-x_\alpha)^2+(y_1-y_\beta)^2+\Delta z^2}}}{4\pi\sqrt{(x_1-x_\alpha)^2+(y_1-y_\beta)^2+\Delta z^2}}\Delta S \tag{10}$$

where $p(x_1, y_1, f_{\mathrm{a}})$ represents the sound pressure value at coordinates (x_1, y_1) in the radiation sound field when the frequency is f_{a}; $\dot{w}(x_\alpha,y_\beta,f_{\mathrm{a}})$ denotes the normal vibration velocity at f_{a} at coordinates (x_α, y_β) on the reconstructed surface; ΔS stands for the unit area of the reconstructed surface; Δz is the distance from the holographic surface to the reconstructed surface; c_0 is the sound speed in the propagating medium of sound field.

The reconstruction loss function based on the KH equation involves calculating the mean squared error between the velocity distribution obtained from the 3D NCNN-NAH reconstruction and the secondary reconstruction of the holographic surface pressure derived using the KH equation. Therefore, directly using their MSEs as the loss function would lead to the dominance of the MSE values at characteristic frequencies in the loss function. This would similarly result in neglecting the reconstruction effects in non-intrinsic frequency bands and an inability to distinguish between easy-to-train and difficult-to-train samples. Eq. (11)

represents two-norm of the reconstructed holographic surface pressure calculated based on the Kirchhoff-Helmholtz (KH) formula at frequency β, normalized by the two-norm of the true holographic surface sampling pressure values at that frequency. Eq. (12) adaptively measures the difficulty of training in different frequency bands based on the Frobenius norm of the difference between the reconstructed and true pressure fields. An increase in the Frobenius norm indicates a larger normalized error in the reconstructed pressure field, necessitating an increase in the loss function weight. Conversely, a decrease in the Frobenius norm indicates a smaller normalized error, allowing a reduction in the loss function weight, thereby focusing the loss function on the learning of hard-to-train samples in the frequency domain—termed the frequency scaled focal-KH Loss (FSF-KHLoss) function.

$$L_{\text{S-KH}-\beta}=\frac{E(\left\|p_{\text{KH}_{\text{GT}}}(f_{\text{a}}=\beta)-p_{\text{KH}_{\text{PRED}}}(f_{\text{a}}=\beta)\right\|_2)}{E(\left\|p_{\text{KH}_{\text{GT}}}(f_{\text{a}}=\beta)\right\|_2)} \tag{11}$$

$$L_{\text{FFS-KH}-\beta}=\frac{1}{F}\sum\nolimits_{\beta}E(\|p_{\text{KH}_{\text{GT}}}(f_{\text{a}}=\beta)-p_{\text{KH}_{\text{PRED}}}(f_{\text{a}}=\beta)\|_F)^{\mu}\cdot L_{\text{S-KH}-\beta} \tag{12}$$

where $L_{\text{S-KH-}\beta}$ is the normalized KH loss function; $L_{\text{FFS-KH-}\beta}$ is the FSF-KHLoss function; $p_{KH_{GT}}(f_a=\beta)$ represents the true pressure value at frequency β; $p_{KH_{PRED}}(f_a=\beta)$represents the reconstructed pressure value at frequency β.

In summary, the loss function L proposed in this study for near-field acoustic reconstruction under sparse sampling conditions within the super-resolution framework 3D NCNN-NAH comprises the FSFLoss function L_{FSF} and the loss function ($L_{\text{FFS-KH}}$) with KH regularization term.

$$L=L_{\text{FSF}}+AL_{\text{FFS-KH}} \tag{13}$$

where A is the weight parameter used to regulate the regularization effect of the physical prior information.

3 Example verification

3.1 Building an acoustic vibration dataset

This study utilized the COMSOL Multiphysics and Matlab to obtain the vibroacoustic responses of thin panels under harmonic excitations at various positions. These data are then employed as the dataset for training and validating the performance of the 3D NCNN-NAH method. The parameters of the vibroacoustic model come from Reference [8]. After the model is established, the normal vibrational velocity is sampled for the thin panel (source) with a resolution of 32 × 32. A plane located 80 mm away from the thin panel is selected as the holographic surface, and the corresponding sound pressure values are sampled with the same resolution of 32×32. Given that the proposed method is applicable to super-resolution problems, the sound pressure obtained from the holographic surface is downsampled by a factor of 4 using uniform sampling, resulting in a final holographic surface resolution of 8 × 8.

For each sampling point for vibrational velocity on the sound-source thin plate, in this study harmonic excitation is applied with an amplitude of 5 N. Each sample includes the vibroacoustic velocity distribution of 191 × 1 × N_{s} × M_{s} and sound pressure distribution of 191 × 1 × N_{h} × M_{h} in the holographic surface, where 191 represents the number of samples in the frequency domain dimension with range (100, 2000, 10); 1 is the PyTorch framework's channel dimension, generally denoting the number of features. To construct the dataset for the vibroacoustic model, an 8:2 ratio is employed to split it into training and testing datasets.

Due to the non-uniformity in the distribution of normal vibrational velocity-sound pressure within different frequency domains, especially with large values in the intrinsic feature region and significant differences between peaks and valleys in the same frequency domain, it is essential to normalize the input within each frequency domain. Wang et al. [8] proposed a novel normalization and regularization method, performed normalization within each frequency domain and subsequently applied regularization across the entire frequency domain range. Since this study uses the relative error form of the L_{FSF} and $L_{\text{FFS-KH}}$ function, i.e., considering the non-uniformity between frequency domains within the loss function, normalization is only conducted within each frequency domain, as shown in Eq. (14). Additionally, this normalization approach contributes to a more regularized solution space, facilitating faster convergence of the Adam gradient descent algorithm.

$$x^*=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{14}$$

where x and x^* represent the physical quantities before and after normalization, respectively; $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of

the data, respectively. In addition, considering the varying levels of noises in the actual measurement process, to enhance the model's robustness to noises, this paper also randomly added noise with signal-to-noise ratios ranging from 0 to 50 dB during the training process of sound pressure of sampling points on holographic surface. This was done to simulate measurement noise, and this method can be considered as a means of data augmentation.

3.2 Model performance verification

To validate the performance of the acoustic vibration model proposed in this paper, in this section we employ the PyTorch framework to train the dataset established in Section 3.1. The input for the 3D NCNN-NAH comprises holographic-domain and spatial-domain sampled signals of acoustic pressure, while the output corresponds to the super-resolved normal vibrational velocity of the reconstructed source surface. For detailed configurations, please refer to Section 2.1.

The relative error δ of the source reconstruction is

$$\delta = \frac{\|v_{\mathrm{PRED}} - v_{\mathrm{GT}}\|_2}{\|v_{\mathrm{GT}}\|_2} \tag{15}$$

where v_{PRED} is the normal vibrational velocity obtained by the 3D NCNN-NAH reconstruction; v_{GT} is the true value of the normal vibrational velocity obtained by the simulation.

Fig. 9 depicts the reconstruction results of 3D NCNN-NAH in various frequency domains, with the excitation source located at the center of the length direction of the plate and one-fourth of the plate width below in the width direction. As shown in Fig. 9, 3D NCNN-NAH can accurately reconstruct the high-resolution surface velocity distribution of the sound source under low sampling rate conditions, with reconstruction accuracy of 97.45%, 96.68%, 95.20%, 95.98%, 96.12%, 94.88% at frequencies of 300, 600, 900, 1 200, 1 500, 1 800 Hz, respectively. As the frequency increases, the number of peaks and troughs in the reconstructed surface velocity distribution also increases, indicating higher complexity in its features. Consequently, the reconstruction accuracy slightly decreases with the increase in frequency domain. However, at 1 800 Hz, the reconstruction accuracy still reaches 94.8%, confirming the effectiveness of the proposed 3D NCNN-NAH framework and the loss function comprising L_{FSF} and $L_{\mathrm{FFS\text{-}KH}}$ in this study.

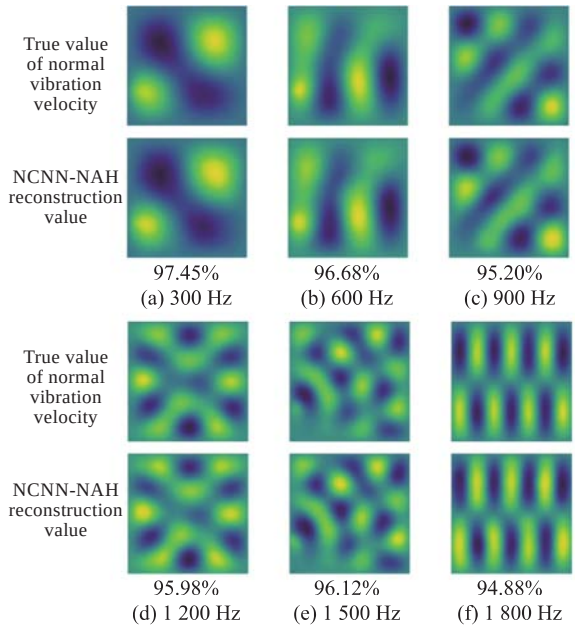

Fig. 9 Comparison of theoretical velocity and reconstruction effect from 3D NCNN-NAH in different frequencies

In this section, to further validate the reconstruction efficacy of 3D NCNN-NAH, the PV-NN model proposed by Wang et al. [8] is compared with the SRCNN model proposed by Olivieri et al. [20]. Specifically, the training parameters for the 3D NCNN-NAH model and PV-NN model are adopted from Reference [8], while those for the SRCNN model are derived from Reference [20]. The PV-NN model consists of two components: the first part comprises an autoencoder for source velocity, and the second part comprises a holography feature extraction neural network (HFENN) obtained from 3D convolution stacking [8]. On the other hand, SRCNN utilizes 2D convolution for feature extraction in the spatial domain, followed by super-resolution extension using U-NET[20]. To demonstrate model performance across different frequency domains, in this study we plot the reconstruction errors of the validation set at various frequency ranges, as illustrated in Fig. 10.

Fig. 10 shows that in the low-frequency range (100−600 Hz) and mid-frequency range (600−1200 Hz), the reconstruction errors of 3D NCNN-NAH are lower than those of the other two algorithms, with average reconstruction errors of 0.039 and 0.046, respectively. In the high-frequency region, except for the region around 1 900 Hz, the reconstruction errors of 3D NCNN-NAH are relatively low, resulting in an average reconstruction error of 0.059.

As shown in Fig. 10, the reconstruction error of SRCNN, which uses 2D convolution as the feature

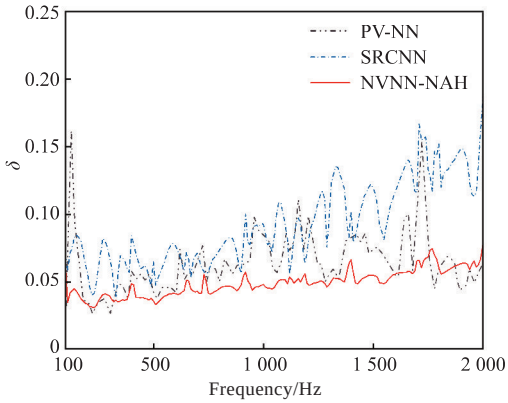

Fig. 10 Comparison of the reconstruction errors of 3D NCNN, SR-CNN, PV-NN in frequency domains

extractor, is significantly higher than those of 3DNCNN-NAH and PV-NN, both utilizing 3D convolution. This difference is more pronounced in the middle and high-frequency intervals. This observation confirms that 3D convolution is more effective in extracting features in the frequency domain. In other words, leveraging features in the frequency domain near a specific frequency can compensate for the feature loss caused by sparse spatial sampling.

The three algorithms exhibit a decline in reconstruction accuracy as the frequency increases, except for PV-NN, which demonstrates a peak error below 150 Hz. Notably, the 3D NCNN-NAH framework, introduced in this paper, stands out with the smallest reconstruction error. Moreover, its error increment with rising frequency is the smallest. As analyzed in Section 3.3.1 below, this superior performance of the framework is attributed to the pre-encoder. Furthermore, the curve of 3D NCNN-NAH in Fig. 10 appears smoother, validating the efficacy of the frequency scaled focal mechanism in focusing learning on the frequency bands with lower reconstruction accuracy, as discussed in Section 3.3.2 below.

3.3 Discussion and analysis of application effects

3.3.1 Analysis on precoder effect

In addressing the super-resolution challenges in near-field acoustic reconstruction, this paper introduces a novel approach by incorporating a precoder into the conventional encoder-decoder structure. This results in a new architecture known as the precoder-encoder-decoder N-type structure. To validate the effectiveness of this structure, in this section we conduct comparative analysis on the reconstruction accuracy in the frequency domain between the 3D NCNN-NAH N-type network frame incorporating the precoder and the U-type frame without the precoder. The training process is conducted under the combined supervision of the proposed loss function L_{FSF} and the loss function $L_{\mathrm{FFS\text{-}KH}}$. All other parameters are kept constant. The comparison results are illustrated in Fig. 11.

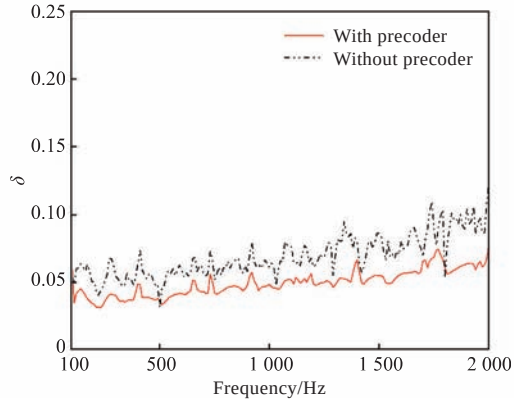

Fig. 11 Comparison of reconstruction errors with and without precoder in the frequency domain

As indicated by Fig. 11, the average reconstruction errors are 0.039 and 0.051 in the low-frequency range (100−600 Hz), 0.046 and 0.063 in the mid-frequency range (600−1 200 Hz), and 0.059 and 0.078 in the high-frequency range (1 200−2 000 Hz), respectively. Therefore, the N-type structure (with a precoder) exhibits higher reconstruction accuracy than the U-type structure (without a precoder) across various frequency domains. This is because that the pre-encoder, which performs upsampling of the original features provides more primitive fused features for subsequent encoders and decoders. In contrast, the U-type structure lacks a pre-encoder. The reconstruction accuracy is decreased when the upsampling region of the decoder exceeds the sampling resolution of the holographic surface, resulting in a lack of fusion of spatial and frequency domain local features. Moreover, as the frequency domain increases, the disadvantage of the U-type structure becomes more apparent. Since the sound-vibration model contains more features in spatial and adjacent frequency domain ranges, the problem of reduced feature fusion information due to the absence of the pre-encoder will be more prominent in high-frequency ranges.

3.3.2 Analysis on effectiveness of loss function

To validate the effectiveness of the loss function proposed in this paper, in this section we compare the reconstruction accuracy of the 3D NCNN-NAH

model trained under the combined supervision of the loss function FSFLoss and the loss function FSF-KHLoss, as detailed in Section 2.3, with that of the model supervised by the mean square error loss function (MSELoss).

As depicted in Fig. 12, in the low-frequency range (100−600 Hz), the reconstruction errors under the combined supervision of FSFLoss and FFS-KHLoss (denoted as $L_{\text{FFS-KH}}$) and under the supervision of MSELoss are 0.039 and 0.041, respectively. In the mid-frequency range (600−1 200 Hz), the reconstruction errors are 0.046 and 0.049 respectively. In the high-frequency range (1 200−2 000 Hz), the reconstruction errors are 0.059 and 0.088. Thus, it can be observed that the accuracy is relatively close in the low- and mid-frequency ranges, but in the high-frequency range, the reconstruction error supervised by MSELoss is higher. Namely $E(\|v_{\text{GT}_{\alpha,\gamma,\varepsilon}} - v_{\text{PRED}_{\alpha,\gamma,\varepsilon}}\|_F)^{\mu}$ in Eq. (9) and $\sum_{\beta} E(\|p_{\text{KH}_{\text{GT}}}(f_{\text{a}}=\beta) - p_{\text{KH}_{\text{PRED}}}(f_{\text{a}}=\beta)\|_F)^{\mu}$ in Eq. (12) are larger in the high-frequency interval. Therefore, "attention" should be paid to difficult-to-learn samples in this range; while loss functions based on attention mechanisms can learn them by focus, increasing their weight in gradient descent. Hence, the proposed loss function in this paper achieves higher reconstruction accuracy in the high-frequency range.

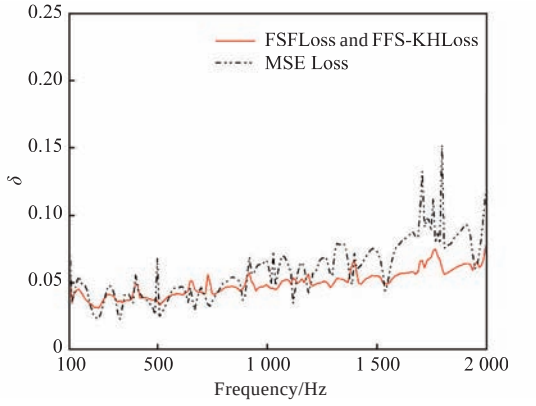

Fig. 12 Comparison of reconstruction errors of NCNN-NAH in frequency domain under combined supervision of FSFLoss and FFS-KHLoss, and MSELoss supervision

Furthermore, this paper employs a loss function incorporating normalized mean square error and the Helmholtz regular term, enabling a more uniform gradient descent on the reconstructed models within the frequency band. Consequently, smoother frequency-domain accuracy curves are obtained, facilitating further reduction of reconstruction errors in both intrinsic source feature intervals and high-frequency regions.

It is worth noting that the Helmholtz regularization term engages in a two-stage acoustic radiation process from a physical perspective, namely the inverse process of the near-field source reconstruction. This regularization mechanism contributes to the regularization of deep neural network models, not only aiding in error reduction but also enhancing model stability and robustness.

3.3.3 Holographic surface sampling rate analysis

The holographic sampling rate stands as a crucial factor affecting the cost of near-field source reconstruction, and a decrease in the sampling rate inevitably results in reduced accuracy in source reconstruction. In this section, we analyze the impact of hologram sampling rate on the reconstruction accuracy of 3D NCNN-NAH. Initially, hologram sampling points are set to 256, 144, 100, 64, 49, 36, corresponding to hologram sampling rates of 16 × 16, 12 × 12, 10 × 10, 8 × 8, 7 × 7, 6 × 6. Subsequently, the sampling number on holography surface is individually trained and tested while maintaining consistent settings for other network parameters. Within the low- , medium- , and high-frequency ranges, the source reconstruction errors of 3D NCNN-NAH model on the validation set under different hologram sampling rates are illustrated in Fig. 13.

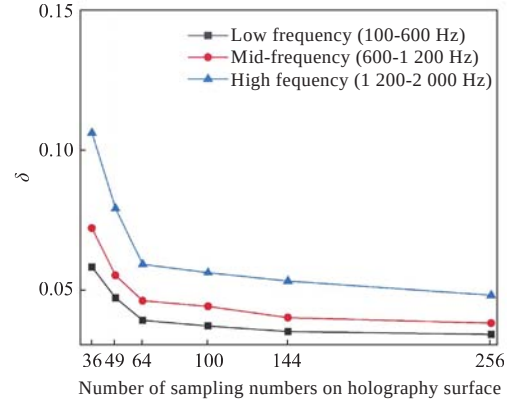

Fig. 13 The relationship between different sampling numbers on holography surface and reconstruction error

Fig. 13 reveals the following observations: 1) With an increase in sampling number on holography surface, the reconstruction error decreases. This is attributed to the augmented hologram sampling rate providing the network with more information for feature extraction, thereby reducing source reconstruction error. 2) As the sampling number increases, the decreasing trend of reconstruction error gradually diminishes, indicating that the feature extraction capability of

the network approaches saturation. 3) Compared to the medium- and low-frequency ranges, the influence of hologram sampling rate on the high-frequency range is more pronounced. This is because source modalities in the high-frequency region are more complex, leading to a more intricate distribution of stimulated source velocities.

3.3.4 Effect of signal-to-noise ratio on reconstruction accuracy

In practical engineering sampling, background noise and measurement noise are often present. To simulate real-world engineering scenarios, in this section we introduce noises with signal-to-noise ratio (SNR) ranging from 0 to 50 dB into the 3D NCNN-NAH model for training and testing purposes. Within the low- , medium- , high-frequency ranges, the source reconstruction errors of 3D NCNN-NAH under different SNR levels are illustrated in Fig. 14.

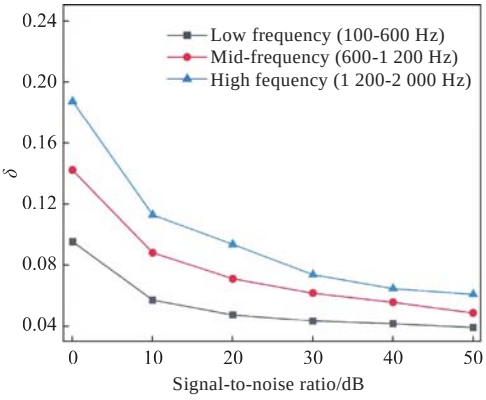

Fig. 14 The relationship between different SNRs and reconstruction errors

From Fig. 14, it can be observed that: 1) As the signal-to-noise ratio decreases, the proportion of background noise increases, leading to a reduction in the network's reconstruction accuracy. When the signal-to-noise ratio exceeds 10 dB, the mean reconstruction error in the 100−2 000 Hz range is less than 0.1, indicating a certain level of robustness in the 3D NCNN-NAH model. 2) Higher frequency domains are more affected by noise, with a reconstruction error as high as 0.187 in the high-frequency range when the signal-to-noise ratio is 0 dB. 3) With decreasing signal-to-noise ratio, the increasing trend of source reconstruction error becomes more pronounced, indicating that the robustness of 3D NCNN-NAH under low signal-to-noise ratios still needs improvement. This may be attributed to two factors: the low sampling rate on holography surface limits the performance of the network under high-noise conditions, necessitating more sampling information to provide features; under high-noise conditions, the feature extraction capabilities of network are relatively limited.

4 Conclusion

In the real-world application of super-resolution near-field acoustic reconstruction on ships, there exists a challenge of significant reconstruction errors under low sampling rate conditions. To address this, in this paper we first design a three-dimensional convolution-based N-type neural network framework (3D NCNN-NAH), incorporating a pre-encoder structure, and introduces residual structures, as well as DLA semantic information fusion structures. Subsequently, addressing the characteristics of near-field sound source reconstruction problems, a frequency domain attention mechanism is proposed, along with the design of a loss function comprising frequency scaled focal loss and the KH regularization term. Finally, based on Matlab for secondary development of COMSOL Multiphysics, a dataset is obtained, and the 3D NCNN-NAH structure based on frequency domain attention mechanism and Helmholtz regularization is compared and analyzed against SRCNN and PV-NN, validating the effectiveness of the algorithm, leading to the following conclusions:

1) The proposed 3D NCNN-NAH network, by incorporating a pre-encoder structure and employing a well-designed encoder and decoder structure, enhances the reconstruction accuracy of the normal vibration velocity under low holographic sampling rate conditions in planar sound source reconstruction problems, with a reconstruction error of only 4.96% over the validation set in the 100−2 000 Hz range.

2) The pre-encoder structure enables the feature map size in the decoder to exceed the input size of the feature layer, providing super-resolution point-by-point information (also known as semantic information) of the original sound field, thereby increasing the model's representational capacity and reducing reconstruction errors.

3) The attention mechanism effectively enhances the reconstruction accuracy of the intrinsic feature area and high-frequency region by adaptively increasing the weight of hard-to-train samples (generally the intrinsic feature area of sound sources and high-frequency regions) within the loss

function or regularization term. By normalizing mean squared error and Helmholtz regularization term in the loss function, the problem of large differences in absolute errors within the frequency domain is alleviated, simplifying the data processing. Therefore, the loss functions based on frequency domain attention-normalized reconstructed mean square error and Helmholtz regularization term effectively improve the performance of the 3D NCNN-NAH framework.

References

[1] ZHAO X Y, ZHU Y, MEI Z Y, et al. Experimental study on effect of different reinforcements on sound transmission performance of composite stiffened plates in water [J]. Chinese Journal of Ship Research, 2023, 18 (3): 197−204 (in Chinese).

[2] LI G S, CHEN M X, YUAN C H. Calculation method of underwater acoustic and vibration response of cabin segment based on onshore vibration test [J]. Chinese Journal of Ship Research, 2022, 17 (6): 252−260 (in Chinese).

[3] LIAO J, HE L, CHEN Z B, et al. Overview of submarine steering system noise [J]. Chinese Journal of Ship Research, 2022, 17 (5): 74−84 (in Chinese).

[4] ZUO X, CHEN H. Near-field and high-resolution cylindrical noise source location method based on vector sound pressure array [J]. Chinese Journal of Ship Research, 2017, 12 (4): 147−150 (in Chinese).

[5] HU Q Y, LI C C, GUO S X. Equivalent source nearfield acoustic holography algorithm based on virtual sound source localization [J]. Ship Science and Technology, 2022, 44 (11): 164−168 (in Chinese).

[6] LI B, LI X Y, WANG Z Q, et al. Research on parameter selection for the statistically optimal cylindrical nearfield acoustical holography [J]. Ship Science and Technology, 2018, 40 (3): 120−127 (in Chinese).

[7] CHEN H T, GUO W Y, HAN J G, et al. Near-field acoustic holography method for high frequency weak sound source in ship cabin [J]. Ship Science and Technology, 2019, 41 (11): 138−143, 147 (in Chinese).

[8] WANG J Z, ZHANG Z F, HUANG Y Z, et al. A 3D convolutional neural network based near-field acoustical holography method with sparse sampling rate on measuring surface [J]. Measurement, 2021, 177: 109297.

[9] WILLIAMS E G. Fourier acoustics: sound radiation and nearfield acoustical holography [M]. San Diego: Academic Press, 1999: 67−89.

[10] CHARDON G, DAUDET L, PEILLOT A, et al. Nearfield acoustic holography using sparse regularization and compressive sampling principles [J]. The Journal of the Acoustical Society of America, 2012, 132 (3): 1521−1534.

[11] FERNANDEZ-GRAND E, XENAKI A. Compressive sensing with a spherical microphone array [J]. The Journal of the Acoustical Society of America, 2016, 139 (2): EL45−EL49.

[12] HALD J. A comparison of iterative sparse equivalent source methods for near-field acoustical holography [J]. The Journal of the Acoustical Society of America, 2018, 143 (6): 3758−3769.

[13] HALD J. Fast wideband acoustical holography [J]. The Journal of the Acoustical Society of America, 2016, 139 (4): 1508−1517.

[14] FERNANDEZ-GRAND E, XENAKI A, GERSTOFT P. A sparse equivalent source method for near-field acoustic holography [J]. The Journal of the Acoustical Society of America, 2017, 141 (1): 532−542.

[15] BI C X, LIU Y, XU, L, et al. Sound field reconstruction using compressed modal equivalent point source method [J]. The Journal of the Acoustical Society of America, 2017, 141 (1): 73−79.

[16] BI C X, ZHANG F M, ZHANG X Z, et al. Sound field reconstruction using block sparse Bayesian learning equivalent source method [J]. The Journal of the Acoustical Society of America, 2022, 151 (4): 2378−2390.

[17] WU S, WEI S H, WU X L, Improvement of near-field reconstruction accuracy of plate using compressed sensing equivalent source method [J]. Mechanical Science and Technology for Aerospace Engineering, 2023, 42 (6): 870−877 (in Chinese).

[18] JANSSENS O, SLAVKOVIKJ V, VERVISCH B, et al. Convolutional neural network based fault detection for rotating machinery [J]. Journal of Sound and Vibration, 2016, 377: 331−345.

[19] ZHANG Y Y, LI X Y, GAO L, et al. Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method [J]. Measurement, 2020, 151: 107232.

[20] OLIVIERI M, PEZZOLI M, MALVERMI R, et al. Near-field acoustic holography analysis with convolutional neural networks [C] //INTER-NOISE and NOISECON Congress and Conference Proceedings. Seoul, Korea: Institute of Noise Control Engineering, 2020: 5607−5618.

[21] OLIVIERI M, PEZZOLI M, ANTONACCI F, et al. A physics-informed neural network approach for nearfield acoustic holography [J]. Sensors, 2021, 21 (23): 7834.

基于三维N型卷积神经网络和频域注意力–亥姆霍兹正则化的近场声源重建方法

籍宇阳[1,2],王德禹[*1,2]

1 上海交通大学 海洋工程国家重点实验室,上海 200240

2 上海交通大学 海洋装备研究院,上海 200240

摘　要:[**目的**]针对全息面、低采样率条件下近场声源重建误差较大的问题,提出一种高分辨率、低误差的平面声源表面法向振速重建的深度神经网络框架。[**方法**]首先,建立用于近场声源重建问题的三维N型卷积神经网络框架(包含预编码器),通过提取空间声场频域内的特征,以弥补空间信息的稀疏性;然后,提出频域注意力机制,设计包含频域注意力–归一化重建均方误差、亥姆霍兹正则项的损失函数,以自适应增加频域内难训练样本的损失权重,从而提升声源在高频和本征特征区间的重建精度;最后,通过Matlab对COMSOLMultiphysics软件进行二次开发,建立矩形薄板声振模型的训练集和测试集,开展对比验证。[**结果**]对比结果表明,该方法在验证集上100~2 000 Hz内的平均重建误差仅为4.96%,重建精度明显高于SRCNN和PV-NN。[**结论**]该研究成果可以降低近场声源重建实船应用中的全息面采样点数量,同时可保证较高的声源面法向振速重建精度。

关键词:近场声源重建;声源识别;三维卷积;亥姆霍兹正则化