ICAS 2021 SHANGHAI
International Council of the Aeronautical Sciences
32nd Congress of the International Council of the Aeronautical Sciences
September 6-10, 2021 Pudong Shangri-La, Shanghai, China

MICROPHONE ARRAY TECHNIQUES FOR TRAILING EDGE NOISE BASED ON CONVOLUTIONAL NEUTRAL NETWORK

Zhangchen Song[1], Peiqing Liu[1, #], Hao Guo[1]

[1]Key Laboratory of Aero-acoustics(Beihang University), Ministry of Industry and Information technology, Beijing 100191, People's Republic of China
[#] e-mail: lpq@buaa.edu.cn

Abstract

Trailing edge noise is a major contribution to the airframe noise and has been studied in various processing methodologies such as wavelet analysis and velocity field comparison. To further understand the noise mechanism and improve the result quality in wind tunnel, microphone array technique has been applied to obtain the sound source distribution by different traditional processing algorithms such as beamforming algorithms and deconvolution algorithms. Rather than solving the sound transfer function, a convolutional neutral network (CNN) method has been proposed based on machine learning for complex objects, and proven efficient in limited number of incoherent monopole sources. However, trailing edge noise is an emission of distributed sound sources and most phased array methods including the CNN method assume the monopole sources. In this study, CNN method is investigated from monopole sources to simple line sources in order to find out an applicable condition for trailing edge noise. By using different definitions of accuracy and loss, the network is trained, tested, and compared with other researchers' results to validate on monopole sources. Then a new test strategy has been proposed and proven helpful in understanding the wrong prediction of the CNN model. The confidence level of the applied region and the method to use the prediction results are proposed to standardize and normalize the applicable range of the phased array. Then, a new training strategy is proposed by adding statistical parameters of the line sources as training data to get a direct source strength distribution of line sources. An experiment of the trailing edge noise of NACA0012 airfoil is conducted in Beihang D5 aeroacoutics wind tunnel to validate the methods. These results indicate that CNN method might be able to detect a more general feature of the sound source distributions, which might be helpful in constructing a network with a statistical output.

Keywords: microphone array, trailing edge noise, CNN, deep learning, conventional beamforming

1. Introduction

Trailing edge noise is a remarkable problem of the airframe noise reduction concerning about the next generation quiet aircraft [1]. To understand the location and strength distribution of the trailing edge noise, microphone array [2,3] has been used to study the noise mechanism and noise reduction methods such as serrations [4,5]. However, a robust and high spatial-resolution map of the source strength distribution requires sophisticated methods and high computational costs [3,6]. In order to increase the calculating efficiency while maintain a high spatial resolution, machine learning methods [7] have been introduced in dealing with source localization. Convolution neural networks [8] have been widely used in computer vision [9] and shown potential to solve the problem in recognizing the direction of arrival (DOA) estimation [10,11] and locations of incoherent monopole sound sources [12,13]. Still, for large number of monopole sources or the line sources, there lacks further research in the training and testing parts.

The raw data acquired by multiple microphones are processed based on the phase and amplitude difference between each microphone and the assumptions of the noise characteristics. A single microphone or pressure receiver is affected by the background noise and other unconcerned or interference noise such as the facility noise. More than one microphone is needed in trailing edge noise experiments to exclude the facility noise influence on the experiment result [14]. A classical microphone technique called beamforming [2] has been a standard method for aeroacoustics measurement due to high efficiency and robustness [15]. Using Delay-and-Sum method, conventional frequency domain beamforming (CFDBF) [16] has a high signal noise ratio (SNR) in high frequency but has poor spatial resolution in low frequency and poor dynamic range due to high side lobes in the source strength distribution map. Aimed at SNR improvement in intensive background noise, high spatial resolution and strength precision and low-to-moderate frequency range application, various processing methods have been developed. Functional beamforming (FB) [17] and orthogonal beamforming (OB) [18] are aimed at reducing side lobes. Adaptive beamforming for coherent signal [19], and blind beamforming for non-gaussian signals [20] are developed for sophisticated noise condition. In the past twenty years, deconvolutional methods such as DAMAS [21,22], DAMAS2 [23,24], CLEAN [25], CLEAN-SC [26], HR-CLEAN-SC [27] and NNLS [28] are applied in phased array to increase the source accuracy in spatial resolution especially in high frequency. But deconvolutional methods have a sacrifice in calculation efficiency.

Furthermore, due to the Rayleigh limit [29], the microphone array techniques are commonly used in high frequency condition and treated as an indication of the noise source location validation for the noise data acquired by the microphones. To guarantee a reliable result, a balance between efficiency and accuracy should be achieved. Due to the rapid development of computer science in deep learning (DL), the correlation between the input microphone signals and the output source distribution has attracted researchers in using deep neural network (DNN). In 2017, Chakrabarty et al. [11] published a breakthrough finding in using CNN to process microphone array data in speech recognition field on NIPS (Conference and Workshop on Neural Information Processing Systems). The inputs are M × K matrix as the amplitudes of each frequency band of each microphone in a given time frame(M is the number of microphones; K is the number of Fourier transform blocks). And the outputs are the vectors as the amplitudes in different angles. With limited number of microphones, the learning result is better in spatial resolutions compared with SRP-PHAT (Steered Response Power-Phase Transform-based) [30]. The idea of using neural networks attracted attentions from researchers in phased array. In 2018, Ma Wei et al. [12] proposed a CNN algorithm to processed multi-monopole equal-strength independent sound sources. It proved that under the condition of a certain frequency and grid density, the results of the distribution of up to 6 sound sources showed that the spatial resolution was not worse than the DAMAS algorithm, and the calculation speed is much faster than the DAMAS algorithm. Vera-Diaz J, Pizarro D, Macias-Guarasa J [13] has applied CNN to the three-dimensional sound source localization according to a similar idea. Because deep learning methods are trained, tested, and validated on tremendous number of samples to prove its efficiency, researchers need to use statistical parameters to defend their models. And because these machine learning methods treat the spatial position relationship between the microphones and the grid density of the scanning map, and the physical model of the sound source as hidden variables,

researchers also need to validate their results in a form of normalization description. Rather than using the descending order to judge accuracy like Ma Wei [12], Xu Pengwei et al. [31] used the mean square error between the true value and the predicted value to investigate the effectiveness of a deep neural network (DNN) algorithm in normalizing the frequency with distance, array diameter and scanning grid density. However, the pattern of these statistical parameters is unconvincing for quantifying the reliability of the predictions, and new distinct statistical parameters are needed to strictly refine the appliance range of the deep learning models. Therefore, it is necessary to see the results of monopole sources in details.

For trailing edge noise, the present results have shown potential in recognizing the correct locations of sound sources with fewer number of sources for training parts than the testing parts [31]. However, due to the flatten layers in the CNN architecture which fail to catch the topology relationship between each grid, the predicted results are unconvenicing. Based on the idea from recent result by Castellomo etc in Applied Acoutics [32] who paid attention to grid-less noise and proved the ability of CNN in acquiring detailed information of the sound sources, new statistical parameter is proposed to strengthen the topology information in the training process.

In this study, core issues are validation of the result correctness, patterns of these significant parameters and the new training strategies. Firstly, to validate the methods, the basic CNN model would be almost the same as Ma Wei [12]. Secondly, the accuracy and the loss functions are discussed in detailed with the normalization of the frequency, which relates to the grid length, array bandwidth or Rayleigh limit, distance between the array and scanning planar and the opening angle. And the wrong predictions are examined by different test sets and focusing area. Then the results are discussed about the reasons of failure. At last, an adjustment of input data according to the accuracy definition is proposed and applied to predict simulations of line sources. For trailing edge noise as line sources, the spatial resolution in the flow-direction resolution is focused on and experiment data of NACA0012 is used to justify the results. The predictions are illustrated and judged by the new criterion to offer relatively reliable description of the sound source distributions.

Section 2 is composed of the methods of traditional phased array, CNN structures, phased array configurations and array ratio as a normalization factor of frequency, general simulation configurations of training and testing sets, the definitions of accuracy and loss and a general result of line sources. Section 3 consists of the validation of the CNN method on monopole sources, a thorough investigations of array ratio, a detailed illustration of accuracy and loss with respects to the frequency, and the CNN procedure into simulations of line sources and predicted results of experiment data.

2. Methods

2.1 Beamforming and Deconvolutional Algorithms

By using equation (1), Leclère [33] classified frequency domain methods to solve the source strength distribution map problem into two categories, namely beamforming-related methods and inverse methods.

$$p = Gq \tag{1}$$

where p stands for the microphone pressure signal, G is the sound transfer function or matrix which depends on the noise characteristics and the environment parameters, and q is the demanding unknown noise source strength. The basic setup for an array would be like figure 1. The scanning planar is assumed to be a square. D is the diameter of the microphone array, $\boldsymbol{r_m}$ is the vector towards the number m microphone, $\boldsymbol{r}$ is the vector towards a point on the scanning planar, z is the distance between the array scanning planar, Nx and Ny are the number of grids on each line, and α is the opening angle. In this study, the scanning surface is square, so $Ngrid=Nx^2=Ny^2$ and the length L would be $L=2z\tan(\alpha/2)$. The distance between two nearest points in vertical or horizontal line is identical as Δx, which also means that $L=Nx\Delta x$.

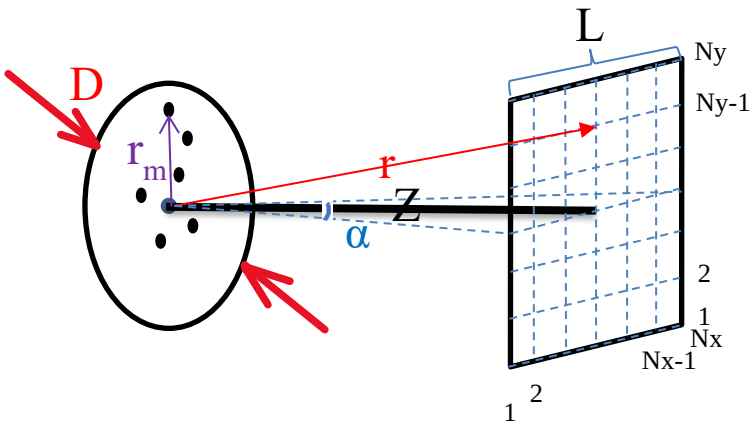


Figure 1 – Array and the scanning grids.

For frequency domain methods, microphone vector *p* is

$$p = [p_1(f), \ldots, p_M(f)]^T \tag{2}$$

where *M* is number of all microphones and $p_j(f)$ is a frequency domain pressure result of number *j* microphone by Fast Fourier Transformation (FFT) at frequency *f*.

Sound source distribution vector *q* on the grids is

$$q = [q_1(f), \ldots, q_{Ngrid}(f)]^T \tag{3}$$

where *Ngrid* is number of grids on the scanning surface and $q_j(f)$ is source amplitude in terms of the pressure produced at grid *j*.

The cross-spectrum matrix (CSM) is defined as *C* by

$$C = < pp^* > \tag{4}$$

where $(\bullet)^*$ denotes the complex conjugate transpose, and $<\bullet>$ denotes time average of the several time snaps.

The conventional beamforming is

$$b(\boldsymbol{r}) = \frac{e(\boldsymbol{r})^* Ce(\boldsymbol{r})}{\|e(\boldsymbol{r})\|^4} \tag{5}$$

where $\|\bullet\|$ is the 2-norm, vector $\boldsymbol{r}$ is the center of the array or the reference point towards the focus point and the steering vector $e(\boldsymbol{r})$ at vector *r* is

$$e(\boldsymbol{r}) = [e_1(\boldsymbol{r}), \ldots, e_M(\boldsymbol{r})]^T \tag{6}$$

Although there are many different formulations for steering vector [35], under monopole source assumption and for the standard strength of noise source to the center of array is

$$e_m(\boldsymbol{r}) = \frac{\|\boldsymbol{r} - \boldsymbol{r_m}\|}{\|\boldsymbol{r}\|} \exp\{-j2\pi f / c_0 \|\boldsymbol{r} - \boldsymbol{r_m}\|\} \tag{7}$$

where c_0 is the sound speed and vector $\boldsymbol{r_m}$ is from the center of the center of the array or the reference point towards the location of microphone number m.

In equation (1), the deconvolution method based on beamforming is to assume that *p* can be decomposed into *S* sources:

$$p = \sum_{i=1}^{S} e(\boldsymbol{r}_i) q_i \tag{8}$$

And for incoherent sources, CSM would be:

$$\mathrm{C} = \sum_{i=1}^{S} |q_i|^2 \, e(\boldsymbol{r_i}) e(\boldsymbol{r_i})^* \tag{9}$$

And the CFDBF result would be:

$$b(\boldsymbol{r}) = \sum_{i=1}^{S} |q_i|^2 \frac{e(\boldsymbol{r})^* e(\boldsymbol{r_i}) e(\boldsymbol{r_i})^* e(\boldsymbol{r})}{\|e(\boldsymbol{r})\|^4} = \sum_{i=1}^{S} |q_i|^2 \frac{|e(\boldsymbol{r}) e(\boldsymbol{r_i})|^2}{\|e(\boldsymbol{r})\|^4} \tag{10}$$

For a unit-power monopole source at location $\boldsymbol{r}$, equation (10) is known as point-spread function (PSF) of the array to evaluate the spatial resolution performance and is defined as:

$$PSF(\boldsymbol{r} \mid \boldsymbol{r_s}) = \frac{|e(\boldsymbol{r}) e(\boldsymbol{r_s})|^2}{\|e(\boldsymbol{r})\|^4} \tag{11}$$

And the deconvolution method like DAMAS is to solve equation (1) by using equation (9) in constructing linear equations within the whole scanning grids or some preferable grids

$$[b(\boldsymbol{r_1}), \ldots, b(\boldsymbol{r_{Ngrid}})]^T = A \cdot [|q_i|^2, \ldots, |q_{Ngrid}|^2]^T \tag{12}$$

where A is a $Ngrid \times Ngrid$ matrix as:

$$A = [A_{ij}] = [\frac{|e(\boldsymbol{r_i}) e(\boldsymbol{r_j})|^2}{\|e(\boldsymbol{r_i})\|^4}] \tag{13}$$

In this study, to ensure that the input CSM and the output x would be normalized of the same unit respectively, the strength q would be either 1 Pa or 0 Pa to represent existence and non-existence of the source, and the output x would be:

$$x = [|q_1|^2, \ldots, |q_{Ngrid}|^2]^T \tag{14}$$

2.2 The Structures of CNN

The basic network structure applied in this work is illustrated in table 1, which is almost the same as the structure by Ma [12]. However, the microphone number (M) is 28 rather than 30 and number of grids (N) is 121 rather than 100. The backend is a Tensorflow backend on Keras framework. The input is CSM (an M×M matrix) divided into real and imaginary as a tensor of M×M×2. The output is a N×1 tensor of the square noise strength on each grid. Conv2D is a two-dimensional convolutional layer, MaxPooling2D is a two-dimensional pooling layer to acquire the maximum values over each kernel size sub-region, Flatten is a flatten layer for the input data only, and Dense is a regular densely connected neural networks layer. The padding is valid, if the layer changes the output size and it will be same, if not. A rectified linear unit (ReLU) filter is applied as the activation for the output of convolutional layer. In this study, the number of trainable parameters is 1,018,617, which is a large number on the same level with the maximum number of training samples.

The noise sources to train are incoherent monopole sources. A CSM and a corresponding N-tuple vector compose a sample to test and train, whilst the values in a training sample should be either 0.0 or 1.0 Pa^2 for non-source or source grid with the input CSM in accordance with. The input data are randomly divided into two parts: the training part and the validation or testing part. The model is assessed on the testing part after training a sufficient large number of epochs of samples on the training part with a frequency ranging from 400Hz to 5000Hz and a maximum total samples of 100,000, which still occupies less than 0.0025% of all possible samples. During the training process, the training part comprises 80% of the total samples and the testing part comprises 20%. The evaluation and validation of the model are based on the difference between the output or predicted vectors by the CNN model and the vector corresponding to the input CSM in the testing part. The vectors corresponding to the input CSM are N-tuple vectors consisting of true values T_i, and the output vectors or the predicted vectors are N-tuple vectors consisting of predicted values P_i.

In the calculations of the model, the default configurations for Keras are applied. The number of epochs is 20 with a batch size of 32 samples and the Adam stochastic optimization algorithm is used with a learning rate of 0.001. The network training takes several hours on a desktop computer with a processor of 3.2 GHz Intel Core i7 8700.

Table 1 – CNN structure

Layer no.	Layer type	Kernel number	Kernel size	Stride	Activation	Padding	Output size
1	Conv2D	64	3×3	1×1	ReLU	Yes	M×M×64
2	Conv2D	64	3×3	1×1	ReLU	Yes	M×M×64
3	MaxPooling2D	-	2×2	2×2	-	No	$\frac{M}{2}\times\frac{M}{2}\times64$
4	Conv2D	128	3×3	1×1	ReLU	Yes	$\frac{M}{2}\times\frac{M}{2}\times128$
5	Conv2D	128	3×3	1×1	ReLU	Yes	$\frac{M}{2}\times\frac{M}{2}\times128$
6	MaxPooling2D	-	2*2	2×2	-	No	$\frac{M}{4}\times\frac{M}{4}\times128$
7	Flatten	-	-	-	-	-	$(\frac{M}{4}*\frac{M}{4}*128)\times1$
8	Dense	1	Ngrid×$(\frac{M}{4}*\frac{M}{4}*128)$	-	-	-	121×1

2.3 Phased Array Configurations and Array Ratio

The microphone array used for source localization contains 28 channels within a 0.45m diameter circle region. 0.45m is the diameter of the array (*D*). The scanning planar is parallel to the array surface at 1m and is a square area as is shown in Figure 1. The scanning grids are uniformly distributed in the vertical and horizontal lines. Number of grids on a horizontal line *Nx* is 11 and the length of a line *L* is 1m. To quantify the performance of the array and normalize the relative parameters, researchers have used bandwidth and Rayleigh limit as the spatial resolutions of phased array. Defined by the width of the PSF in -3dB, Bandwidth (Bwd) is a function of the microphone array spatial distribution and the distance between the array and the scanning surface.

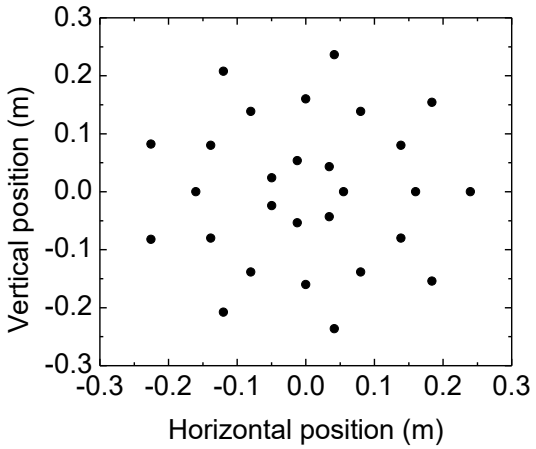


Figure 2 – Scheme of the microphone array

Rayleigh limit (RL, has been defined in [29] but different in the assumption of opening angle α) is about the distance *z*, sound speed *c*, array diameter *D*, frequency *f* and the opening angle α:

$$RL = \frac{1.22zc}{Df\cos^3(\alpha/2)} \tag{15}$$

Figure 1 shows that length *L* of the square evenly distributed scanning grid is

$$L = N_x\Delta x = z\tan(\alpha/2) \tag{16}$$

where *Nx* is the number of points on x-line of scanning grids and Δ*x* is the distance between two nearest points in vertical or horizontal line. Then it can be known that:

$$RL = \frac{1.22cN_x\Delta x}{Df\sin(\alpha/2)\cos^2(\alpha/2)} \tag{17}$$

Three variables could be used for normalizing the frequency by the performance of the array. They are the bandwidth (Bwd), Rayleigh limit (RL) and Rayleigh limit simply (RLS, assume in equation (17) that $\alpha\rightarrow 0$). Figure 3 shows the three variables with respects to frequency. To normalize these variables, a dimensionless coefficient, namely array ratio (AR), is defined by using one of these variables divided by Δ*x*. Table 2 shows the different array ratios as normalizations of frequency.

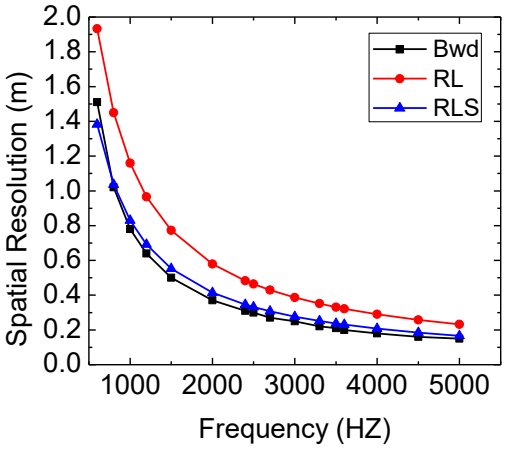


Figure 3 – Array spatial resolutions at 1m.

Table 2 – Frequency normalization

Name	A	B	AR=B/A
AR_Bwd	Bandwidth	Δx	$\frac{\Delta x}{Bwd}$
AR_RL	RL	Δx	$\frac{Df\sin(\alpha/2)\cos^2(\alpha/2)}{1.22cN_x}$
AR_RLS	RL(α→0)	Δx	$\frac{Df\Delta x}{1.22zc}$

These array ratios have been proposed in previous studies in different forms to define the application of deep learning methods in phased array and proven effective in limit condition. Ma [12] suggested that AR_RL should be larger than 0.3 to ensure the accuracy. Xu [31] concluded a trend for the loss in convergence when AR_RLS was larger than 0.1. However, they are not thoroughly investigated in comparison in using deep learning methods. Nevertheless, considering that the array and the shape of scanning grids are constant, diameter of the array D and number of grids on a line N_x should be unchangeable. α could be obtained by $\alpha=2\arctan(z/(2N_x\Delta x))$, so Δ*x*, *z*, *f* are the main variables to change array ratio. In this study, different array ratios were compared in simulation results to evaluate in a limited condition.

2.4 Accuracy and Loss

The general definitions of accuracy and loss parameters used in this study are shown in Table 3. Accuracy and loss are both evaluations of the average difference between the predicted samples and the true samples. The predicted samples and the true samples are both *Ngrid*-tuple vectors consisted of true values and predicted values respectively.

The loss is an average of samples in the mean variance between the predict values and the true values in training and testing samples. Loss_total is the mean variance of the subtract of predicted value and the true value. Loss_source is the fluctuation of the prediction near the true value in the

source existence location. Loss_zero is the fluctuation of the prediction near the true value in the source non-existence location. Loss_sratio is a ratio of the contribution to the total variance by the source existence location. The four loss functions are aimed at quantifying the predictions' correctness ratio in grids with or without sources.

Accuracy is the correctness ratio to predict the values on the right locations. Acc_5 is the accuracy criterion applied by Ma [12], which is sorting the output vector in a descending order and comparing locations of the maximum values in the predicted vectors. Unlike Acc_5, for accuracy criteria named Acc_1, Acc_2, Acc_3 and Acc_4, threshold is applied as a value to determine whether the predicted value is large enough to be recognized as a source. If the prediction is larger than the threshold, it would be selected to be compared with the source existence area with true values, otherwise it would be ignored and compared with non-existence area with zero values. After filtering, the accuracy is the about predicting the values with the right locations. Acc_1 needs the prediction to be right on the correct location of source and non-source, while Acc_2 just needs to offer the source location correctly. Acc_3 and Acc_4 compare only 6 locations in this study and these locations are selected by the maximum values in the descending order respectively. Acc_3 sort the predicted values in a descending order, and the locations of the filtered predicted values should be correct to the locations of the source and non-source. Acc_4 is a weak version of Acc_3 like Acc_2 as a weakened version of Acc_1. Acc_4 only needs the predicted values larger than threshold are in the right positions of the source.

Table 3 – Definition of accuracy and loss

Category	Name	Equation	Definition
Loss	Loss_total	$\frac{1}{Ngrid}\sum_{i=1}^{Ngrid}(x_{i,predict}-x_{i,true})^2$	The mean square of the predict and the true
	Loss_source	$\frac{1}{Nsource}\sum_{i=1}^{Nsource}(x_{i,predict}-x_{i,true})^2$	The mean square of the predict and the true in source
	Loss_zero	$\frac{1}{Nzero}\sum_{i=1}^{Nzero}(x_{i,predict}-x_{i,true})^2$	The mean square of the predict and the true in non-source
	Loss_sratio	$\frac{\sum_{i=1}^{Nsource}(x_{i,predict}-x_{i,true})^2}{\sum_{i=1}^{Ngrid}(x_{i,predict}-x_{i,true})^2}$	The source contribution ratio to the total
Accuracy	Acc_1	$\frac{Ncorrect}{Ngrid}$	Under the threshold setting, all sound source points and zero source points are recognized
	Acc_2		Under the threshold setting, all sound source points are recognized
	Acc_3		Under the threshold setting, sorted from large to small results within a finite number, both sound source points and zero source points are recognized
	Acc_4		Under the threshold setting, sorted from large to small results within a finite number, and the sound source point is recognized
	Acc_5		Sorted from large to small results within a finite number, both sound

source points and zero source points are recognized

2.5 Simulations of Trailing Edge Noise

Trailing edge noise can be simulated as an emission of line sources. And the pattern of line sources could be assumed as a set of equal-strength sources distributed in the vertical line. Therefore, a CNN model of training a few monopole sources is applied to predict on the samples with vertical line sources distributed in the flow direction. Figure 4 shows some samples of the vertical line pattern and the prediction results. The CNN predictions are based on the previous network about limited number of incoherent monopole sources. In this study, Section 3.4 would check the results of line source simulation in details.

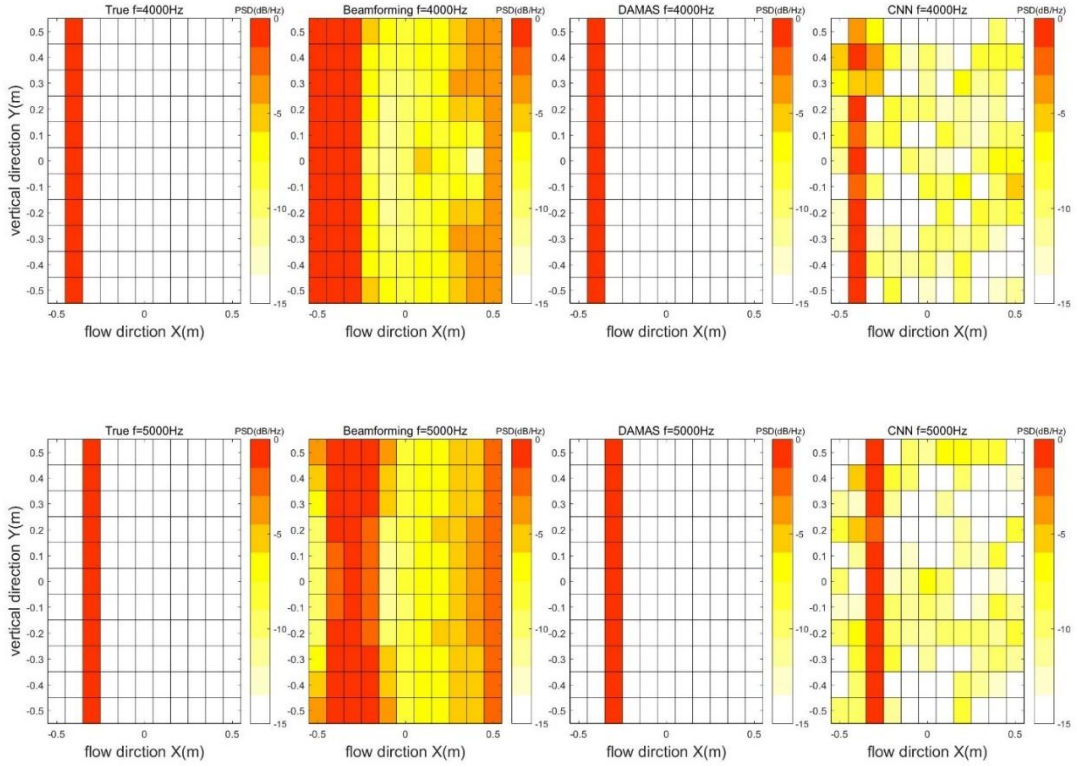


Figure 4 – The source distribution map in 4000Hz and 5000Hz without normalizing the input CSM (From left to right are the true monopole sources, beamforming result, DAMAS result and CNN predictions).

3. Result and Discussion

3.1 The validation of the methods

The training result is compared with the result of Ma Wei [12]. The network structures are almost the same and the accuracy criteria are both Acc_5. As is shown in figure 5, the accuracy trends are not close enough. This pattern is due to the different conditions of the array in the microphone spatial distributions and the distance between the array and the scanning planar. To evaluate the network in different array configurations, the results should be normalized.

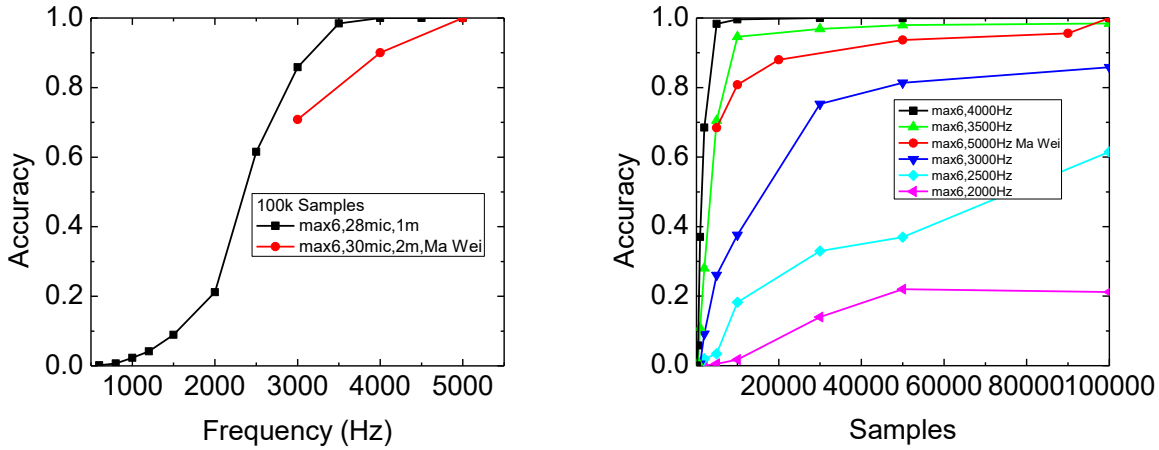


Figure 5 – A curve of accuracy with respects to frequency and samples.

As is discussed in Section 2.3 about the array ratio (AR), the frequency should be normalized by the characteristics of the array spatial resolution and the grid characteristics. For a total of 100k samples to train, the result is shown in figure 6, which is also a normalization of figure 5. *D* is 0.45m and *z* is 1m.

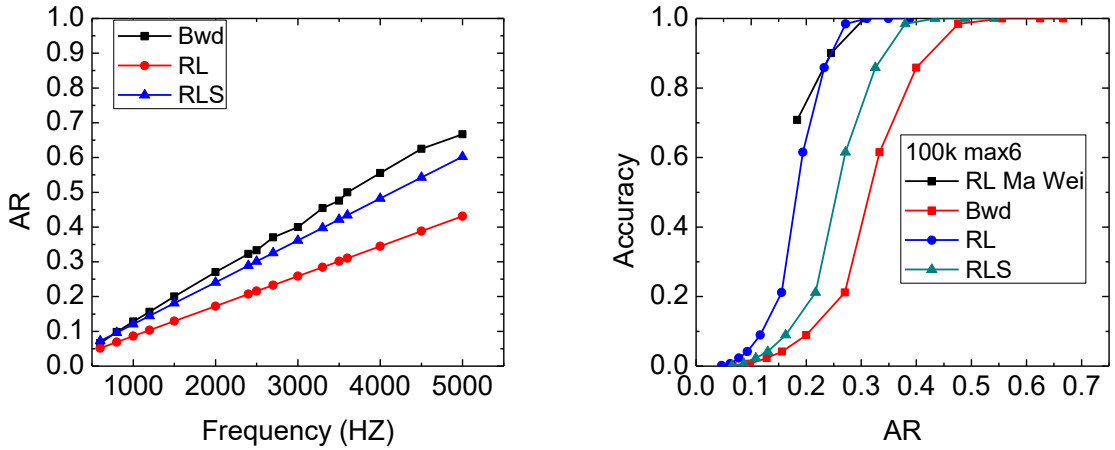


Figure 6 – A curve of AR with respects to frequency and a curve of accuracy with respects to AR.

As is shown in Section 2.4 and Table 2, *z*,*f* and Δ*x* are what matter to change the result. And what of much interest is Δ*x* which decides the lower limit of the spatial resolution of the methods. To give the users of the CNN model a prudent prediction, at least Δ*x* should be given in reference to a relatively lower value. Assuming that accuracy should be larger than 0.8, the minimum AR on each criterion would be that *AR_RL* should be 0.23, *AR_Bwd* should be 0.39 and *AR_RLS* should be 0.32. Assume Bwd linear to f. It can be inferred from equation (17) and the definition of opening angle α that $AR\_RL = \frac{Df(N_x\Delta x)^2}{0.61c((N_x\Delta x)^2 + 4z^2)}$ and $AR\_RLS = \frac{Df\Delta x}{1.22zc}$. To normalized Δx by f and c, some equations can be introduced that:

$$\Delta x = \frac{Bwd(z, f)}{AR\_Bwd} \tag{18}$$

$$\Delta x = \frac{2z}{N_x}\sqrt{\frac{0.61c * AR\_RL}{Df - 0.61c * AR\_RL}} \tag{19}$$

$$\Delta x = \frac{1.22zcAR\_RLS}{Df} \tag{20}$$

Figure 7 shows different criteria of Δx with respects to the lower limits of different array ratios respectively. In comparison, in order to be small enough and easy to do in the experiments, RLS is used as criteria to normalize the frequency by AR_RLS.

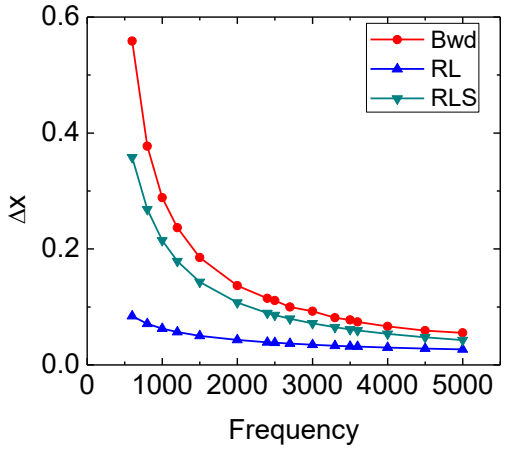


Figure 7 – A curve of Δx with respects to different frequency according to the lower limit of different AR.

3.2 New definition of accuracy and loss

As is defined in Section 2.4, Acc_3 is a further accuracy definition from Acc_5, because Acc_5 is a relatively unstrict evaluation criterion. In above Section 3.1, a comparison between different array ratio is applying Acc_5 as an accuracy criterion. Acc_3 differs in the threshold, which is a value as a filter to quantify the correctness ratio of the prediction to the sources. A high threshold would mean that the result is correct only the value is high enough compared with the true value. Figure 8 shows the source strength map in a dynamic range of 10dB. The map which is correct for Acc_5 criteria is wrong for Acc_3 criteria if the threshold is too high to filter too many or too low to filter too few sources.

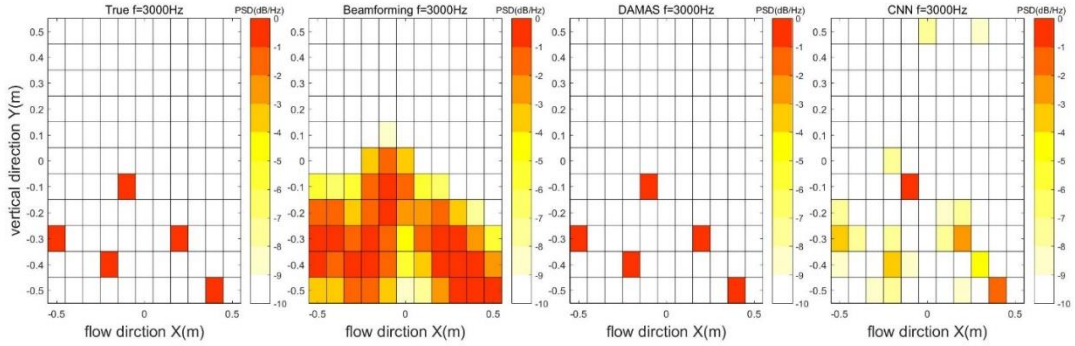


Figure 8 – The source distribution map in 3000Hz (From left to right are the true monopole sources, beamforming result, DAMAS result and CNN predictions).

Using different threshold would mean different dynamic range to be selected as the potential source locations for CNN prediction, which means that there would be different results for Acc_1, Acc_2, Acc_3 and Acc_4. The test set used in comparing different accuracy criteria contains 50k samples of 6 sources. Figure 9 shows the evaluation results in different accuracy definitions.

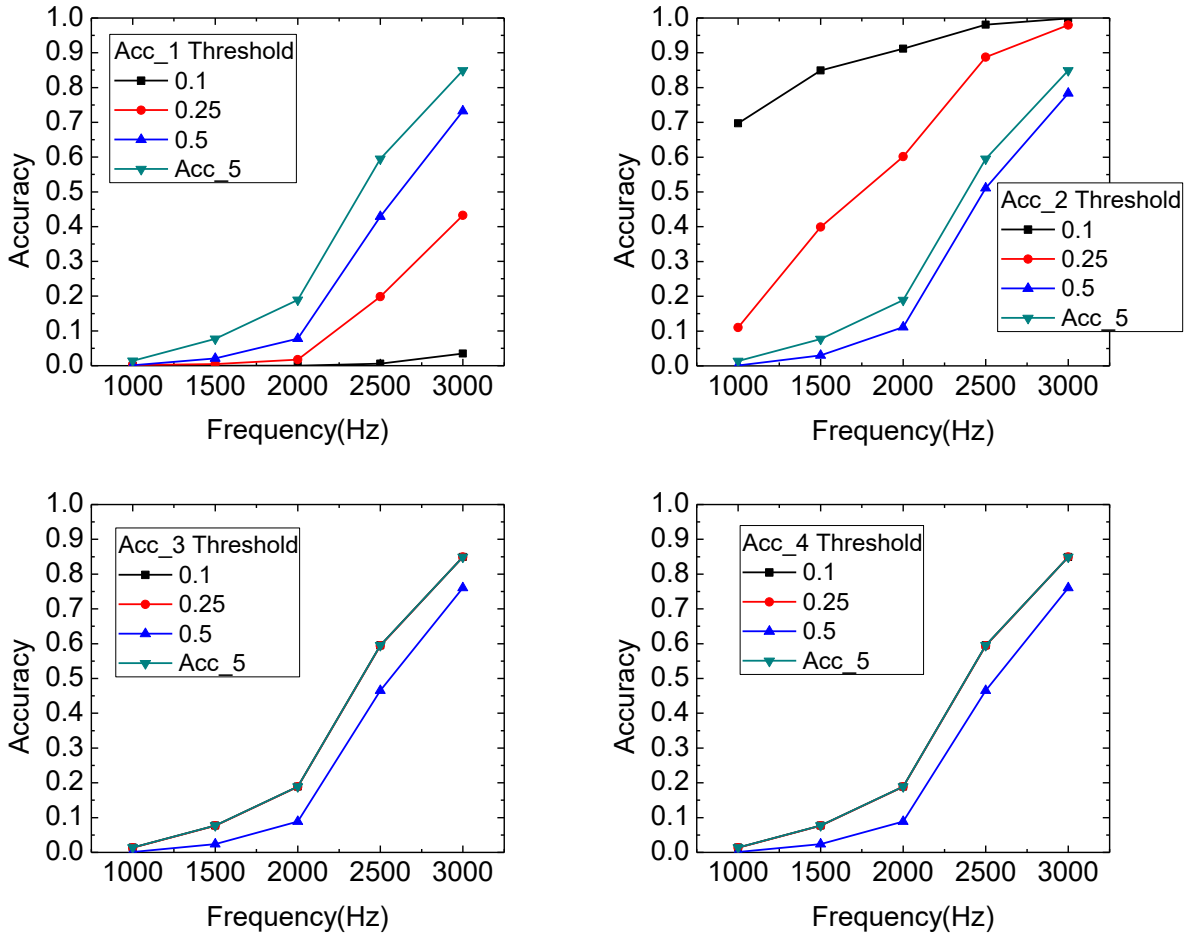
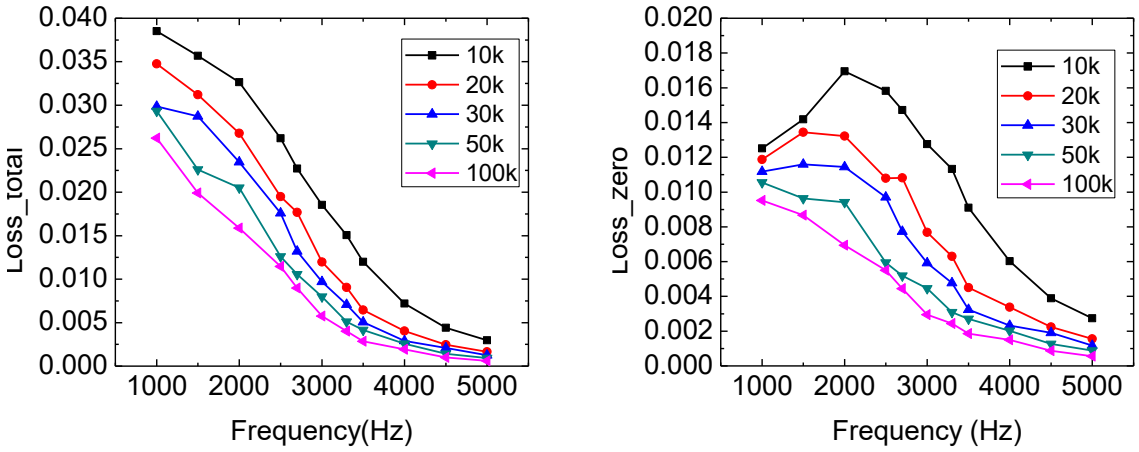
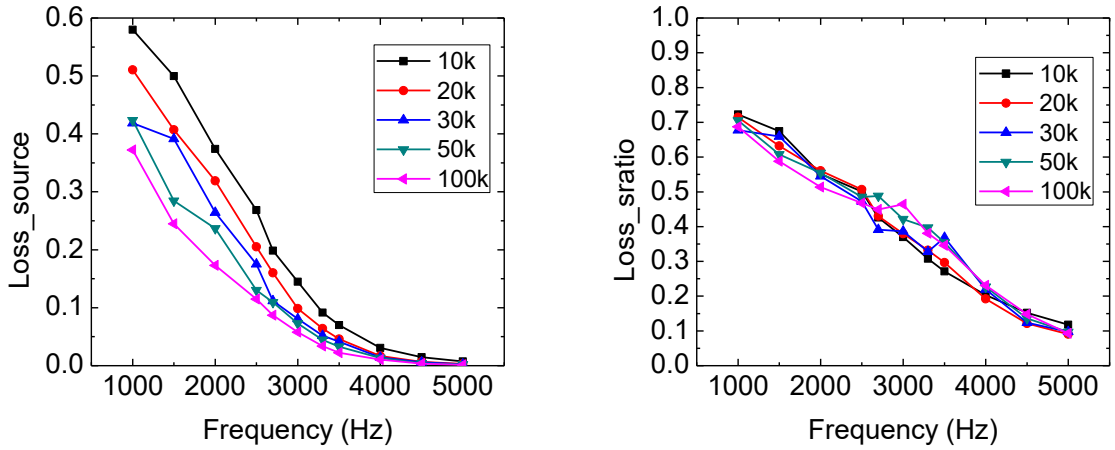


Figure 9 – A curve of accuracy with respects to frequency in different threshold.

The loss is not enough for researchers to have a general view of how close the result is to the true sources, as is shown in Section 2.4 about the loss_source, loss_zero and loss_ratio. They would help scientists understand the result better than loss_total, namely the mean square error. Figure 10 shows the result of new and old accuracy and loss definition.



(a) Loss_total (b) Loss_zero



(c) Loss_sorce (d) Loss_sratio

Figure 10 – A curve of different loss functions with respects to frequency and training samples.

Loss_source is the average of samples in mean square of the difference between the predicted values and the true values of 1 Pa^2. Figure 10 shows that if the frequency is larger than 1000Hz, loss_source should be below 0.4 Pa^4. Assuming the prediction following Gaussian distribution, achieving a probability of over 0.9 would mean that the threshold should be smaller than $1-\phi^{-1}(0.9)*\sqrt{0.4} \approx 0.1588$ for frequency around 1000Hz. However, in this study, the usable frequency should be larger than 2500Hz. As a result, the mean square is smaller than 0.15 Pa^4 and the threshold should be $1-\phi^{-1}(0.9)*\sqrt{0.15} \approx 0.4849$, which means that 0.5 Pa^2 or -3dB is preferred as the threshold. -3dB is the value compared with 1 Pa^2 as 0dB in sound power.

Due to above analysis by accuracy and loss, the threshold should be decided as 0.5 Pa^2. Figure 11 shows the result of 100k training samples for different accuracy criteria. Array ratio is AR_RLS. Acc_3 is chosen as the accuracy criteria in this study. In order to obtain an accuracy with a lower limit of around 0.8, the frequency should be not less than 3000Hz and array ratio should be not less than 0.374.

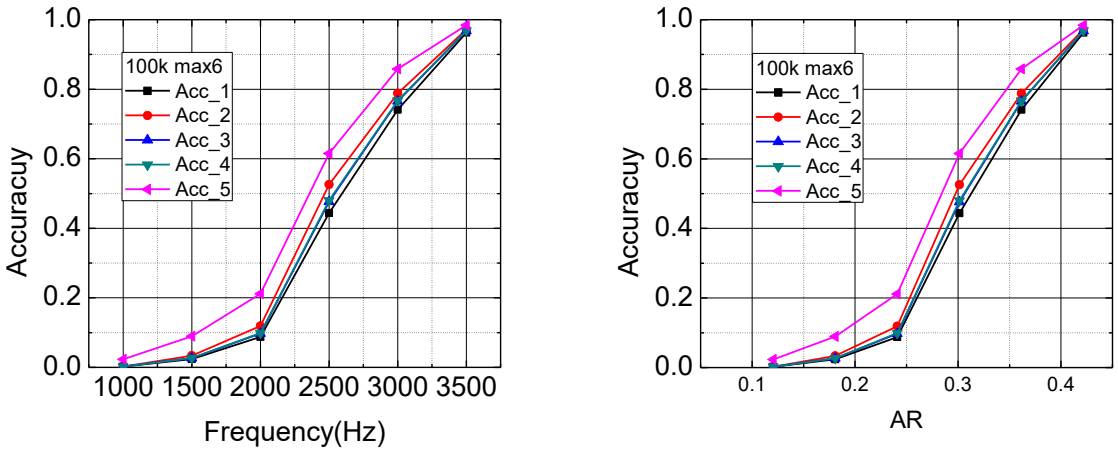


Figure 11 – A curve of accuracy with respects to frequency and AR.

3.3 Testing strategy and result

Another factor to influence accuracy is the random selection of test samples. The above analysis uses the testing data occupying 20% of the total samples. But this selection is influenced by the regions of the grids and number of sources. To evaluate the predictions in an overall way, the test samples are selected according to the strategy in Section 2.4.

For number of sources, the model is tested on a series of testing sets in Table 5. Figure 12 shows

that the accuracies in different number of sources test sets are different. The more training samples, the higher accuracy for 5,6 sources and the less accuracy for 1,2,3 sources would happen.

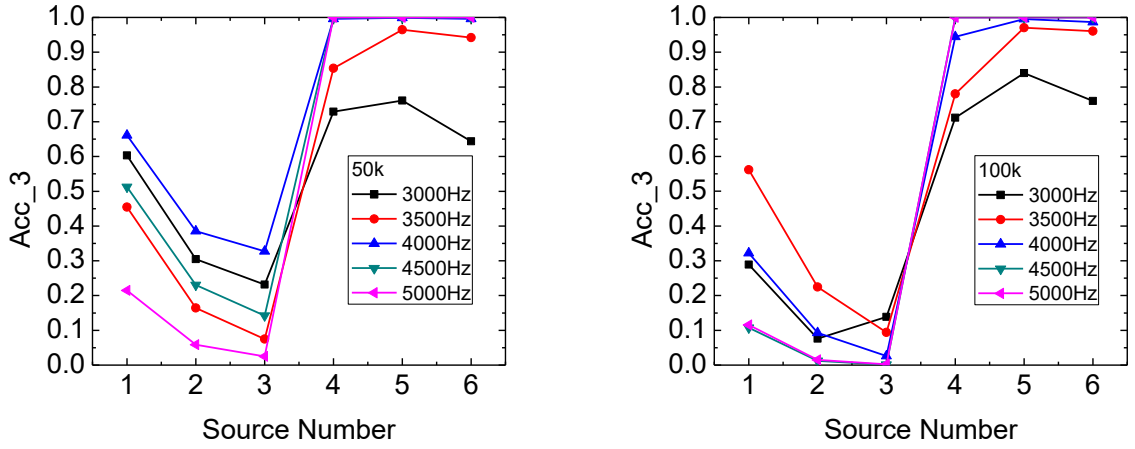


Figure 12 – A curve of accuracy with respects to different number of sources for test sets in different frequency.

For different regions, the scanning planar can be divided into regions from outer to inner. Figure13 shows the probability ratio of these regions. The grid on scanning planar has are colored with 0,1,2,3,4, or 5. Region 'all' contains 0,1,2,3,4,5, 'inner1' contains 1,2,3,4,5, 'inner2' contains 2,3,4,5, 'inner3' contains 3,4,5 and 'inner4' contains 4,5. The composition of samples in different regions are not selected evenly. To select different number of sources, the probability ratios are obtained by the combinations in different regions over the whole area of 121 grids. Figure 14 shows the accuracy of different regions in Acc_3 with respects to frequency and AR_RLS. The test sets are 50k samples on regions in Figure 13 with 6 sources. From high frequency to low frequency or high AR_RLS to low AR_RLS, the accuracy is not changed evenly for all regions. The inner regions are more sensitive to the frequency change than the outer regions and the inner regions contains more wrong prediction ratios.

There are two theories to explain what wrong predictions occur after large enough number of training samples for the model:

1.A not fully learning of the sources at the edge.

2.The increasing AR due to the frequency which is especially for the sources close enough.

The first theory is related to small number of individual sources. The second theory is concerned with the probability ratio due to the area of the region for the samples with large number of sources aggregating together. Because the number of sources influences a lot on the theories, the accuracy for different number of sources should be investigated.

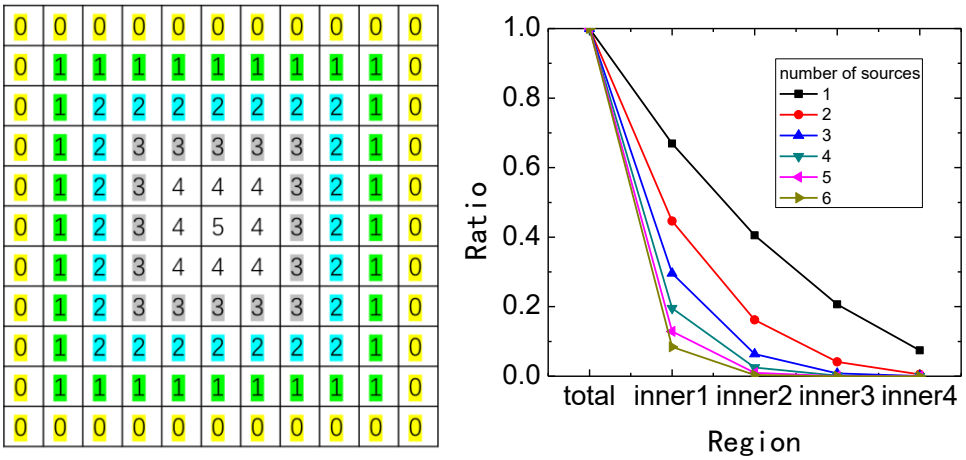


Figure 13 – The different regions of the scanning planar.

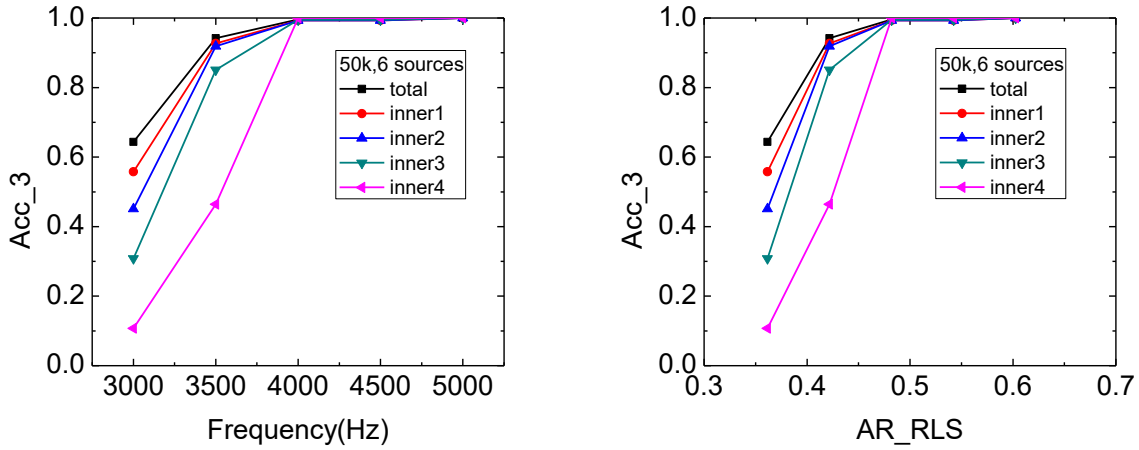


Figure 14 – The accuracy of test sets in different regions with respects to frequency and AR_RLS.

Due to the high slope in Figure 5, 3000Hz model is chosen to test different number of sources in different regions. Figure 15 shows that when in 3dB threshold, the less noise detected in the inner region, the more reliable the prediction is. And the more sources detected in the outer edge, the less convincing the result is.

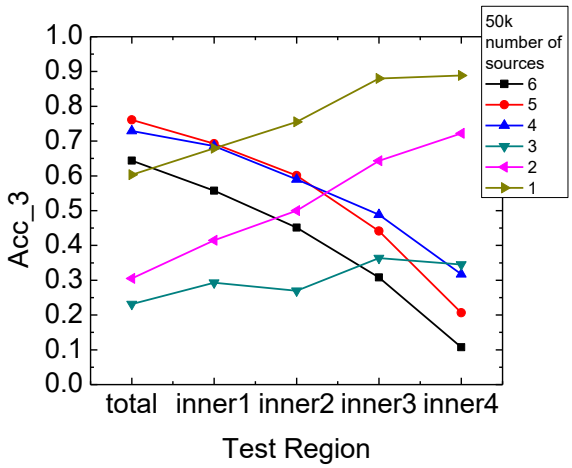


Figure 15 – The accuracy of test sets in different regions with respects to number of sources.

To further understand the result, a loss evaluation is done for the 6 sources test sets. Figure 16 shows that the loss_total (as a mean square error of all grids) becomes larger with test samples centralizing. It indicates that failure to predict sources close together contributes significantly to the wrong predictions. And the predictions with sources near the edge don't contribute greatly to the mean square error. The loss_zero (as a mean square error of only grids without sources) remains a relatively low value which indicates that the CNN model predicts the non-source girds close to zero. But the loss_sratio(as a ratio of the square error of all grids with sources over the square error of all grids) becomes small when the frequency is around 3000Hz, which indicates that although the non-source grids are detected at a stable level, more source grids are not detected. The wrong predictions appear first as a form of ignoring true source on the grids especially in the central regions. However, due to the nonrandom selection of test sets in Figure 13, wrong predictions of the edge region might be blamed. Figure 17 shows the loss_sratio multiply the probability ratio in Figure 13. Ma [12] indicated that the wrong predictions should be in the edge region. But according to Figure 17, it was due to the non-random selection of the test sets.

In conclusion, the wrong predictions of the sources near the edge of the scanning grids are always present but the wrong predictions due to the change of frequency or array ratio appears first and influence strongly in the inner regions, which indicated that the inner regions need to be handled

carefully for more clusters of sources to train and test. The new proposed statistical parameters in Section 3.4 about the average sound pressure square across the line is inspired by this pattern.

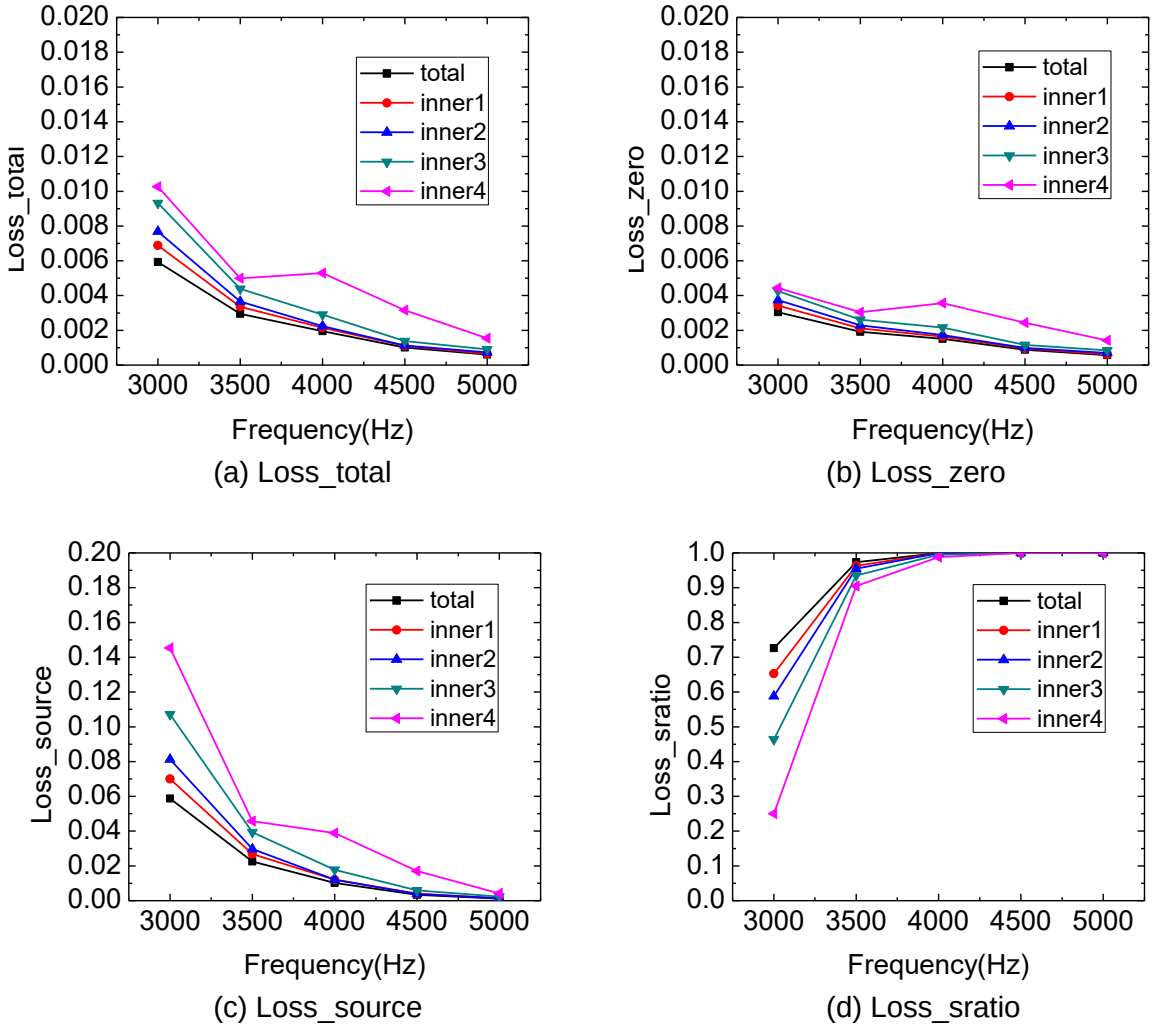


Figure 16 – Different loss functions of test sets in different regions with respects to frequency.

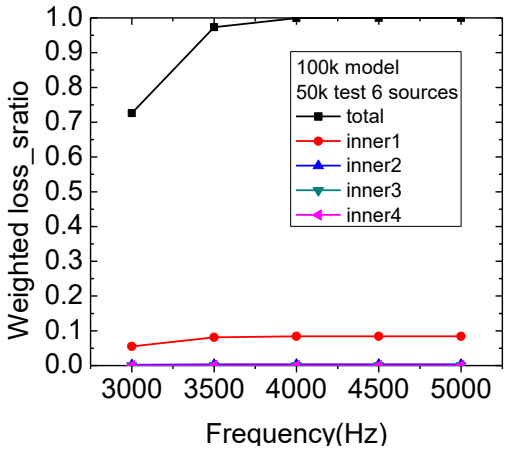
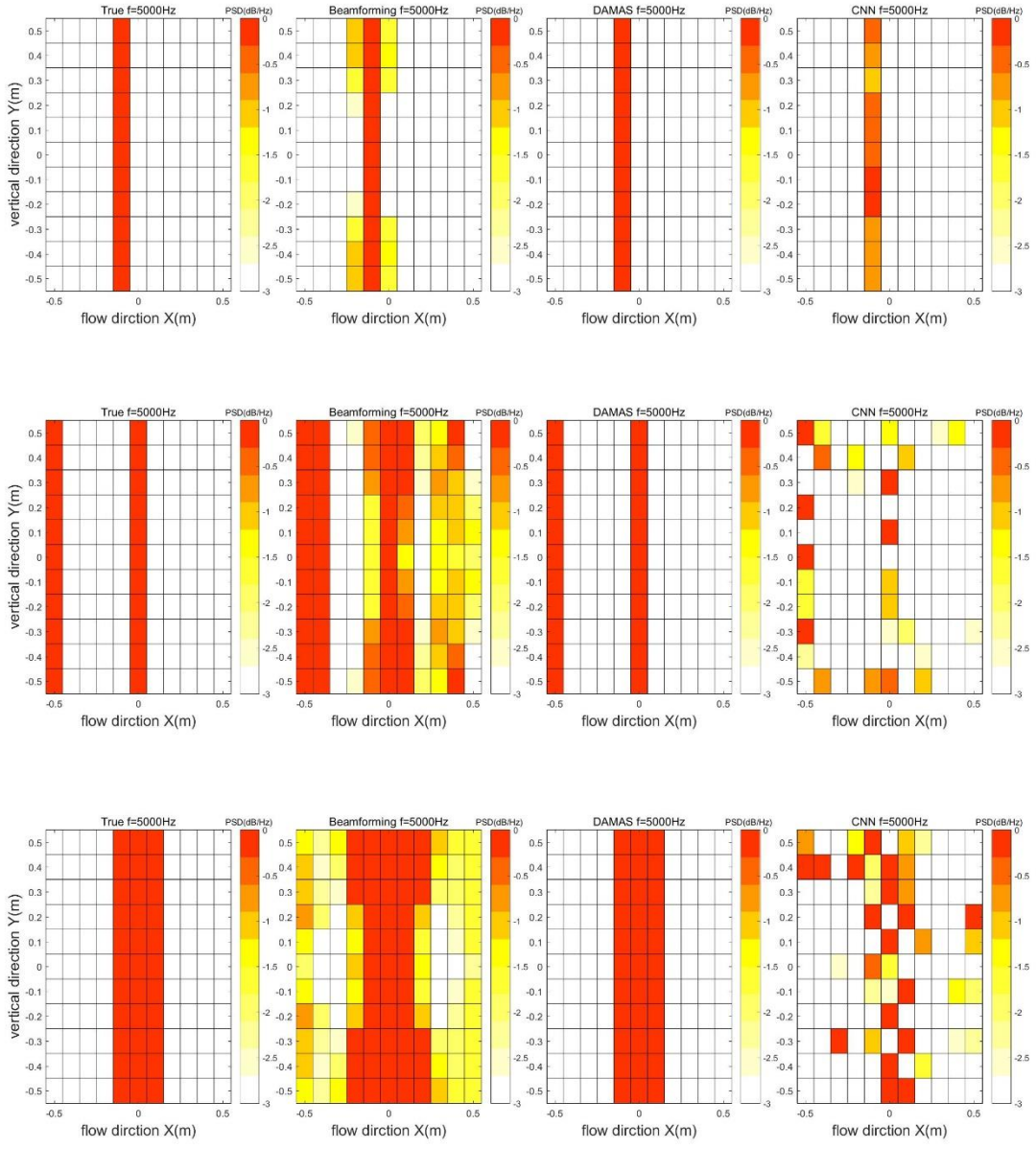


Figure 17 – Weighted loss_sratio of test sets in different regions with respects to frequency.

3.4 New training strategy and simulations of trailing edge noise

The CNN algorithm should be at least convincing for a limited condition. The limited condition is a large enough array ratio, a model of up to 11 monopole sources with over 100k samples to train and test, and -3dB threshold as a dynamic range. Array ratio is controlled by phased array spatial

distribution, scanning planar setting and frequency. Trailing edge noise could be assumed as line sources. Figure 4 has shown a few sound source distribution maps of lines sources. In addition, Figure 18 shows the map of source distributions in different dynamic ranges. By filtering out the relatively small result, a -3dB map performs better in the flow-direction spatial resolution. The input CSM has not been normalized, which means that the model trained on 100k samples of maximum 6 sources must predict with more sources and larger CSM. The simulation results indicate that if the line sources are larger than a number and result is not acceptable. By directly using CNN model to predict, the maximum number of line sources should be 3.



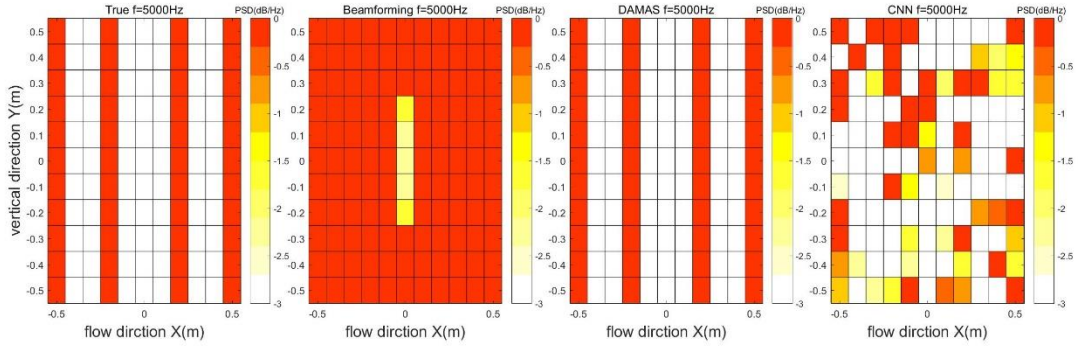


Figure 18 – The source distribution map (From left to right are the true results, beamforming result, DAMAS result and CNN predictions. All maps have a dynamic range of 3dB).

Due to the input and output value of the same unit, a CNN procedure is proposed to normalize the input to process microphone data in wind tunnel test. There are two alternative ways to normalize R. One is to use the maximum value in the diagonal of R, and the other is to use beamforming at first to get a value of the whole scanning area and find the maximum values of all the grids on the source strength map. The repeated simulations have showed that using beamforming at first performs better. Figure 19 shows the process for the methods to deal with the microphone data.

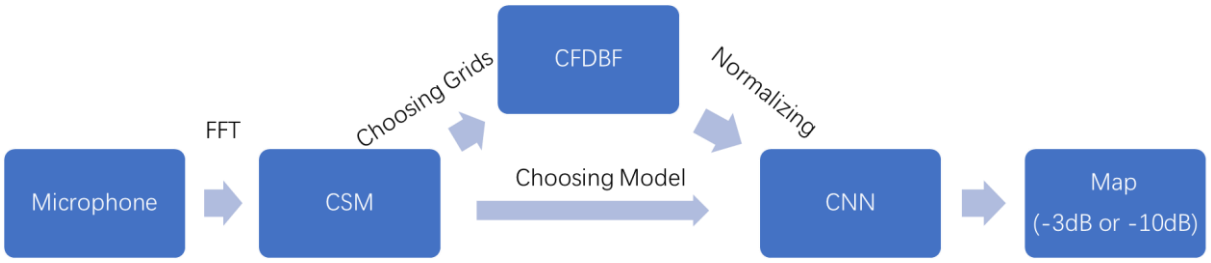
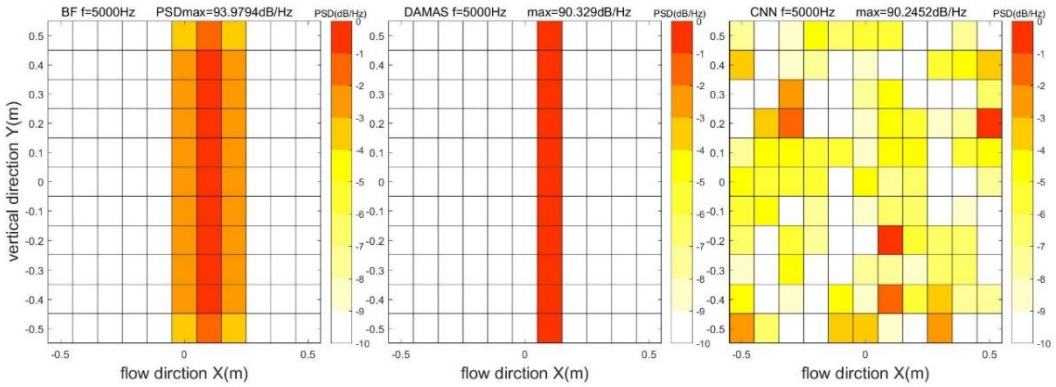
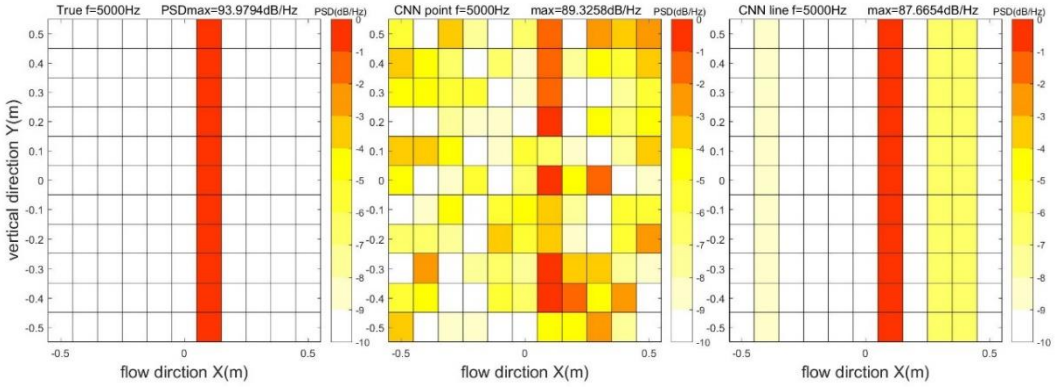


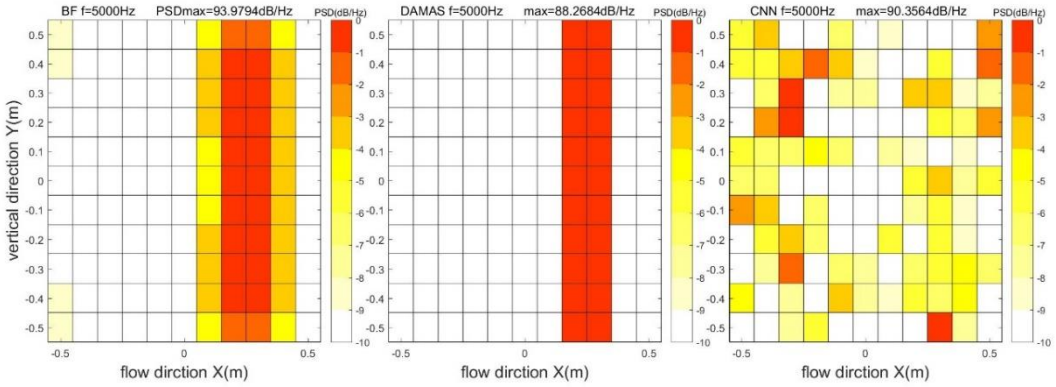
Figure 19 – CNN procedure in experiment measurement.

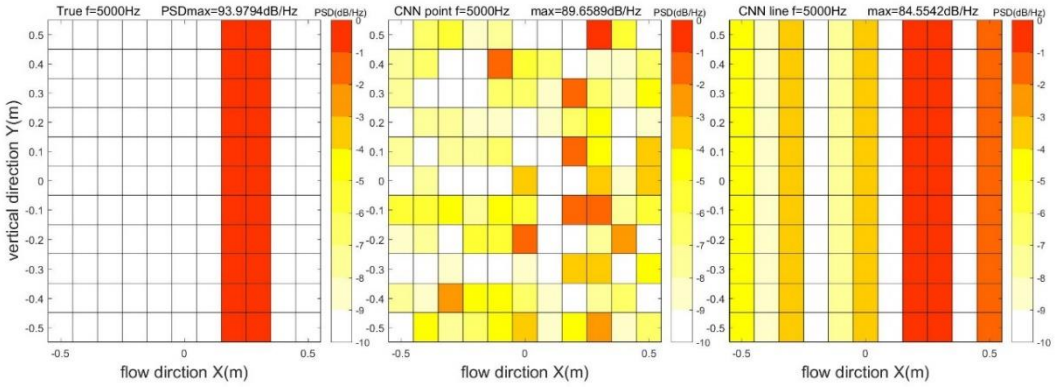
To strengthen the topology relationship between each grid, an average of the flow-direction or the x-direction is used as a statistical parameter to train and test. For present simulations, 11 parameters are added as an average of the 11 x-direction locations across the y-direction. As is shown in Figure 20, this method (a) decrease the influence of the over-normalization, (b) increase the dynamic range and spatial resolution (c) smooth the map and (d) increase the number of line sources to predict.



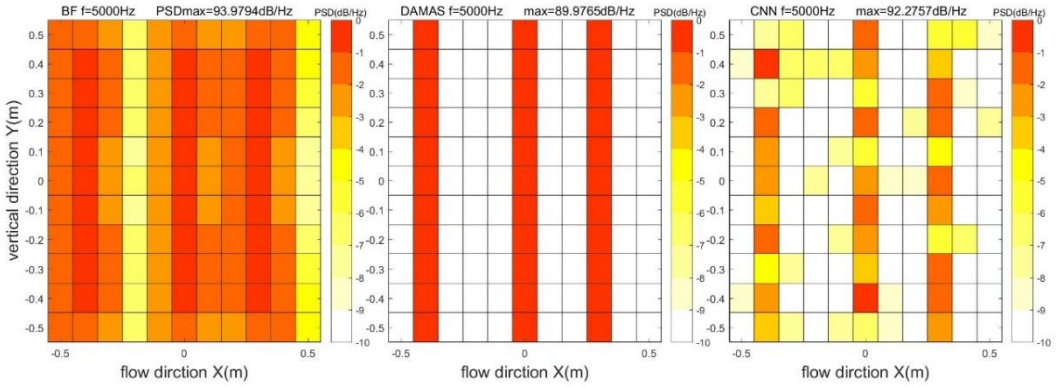


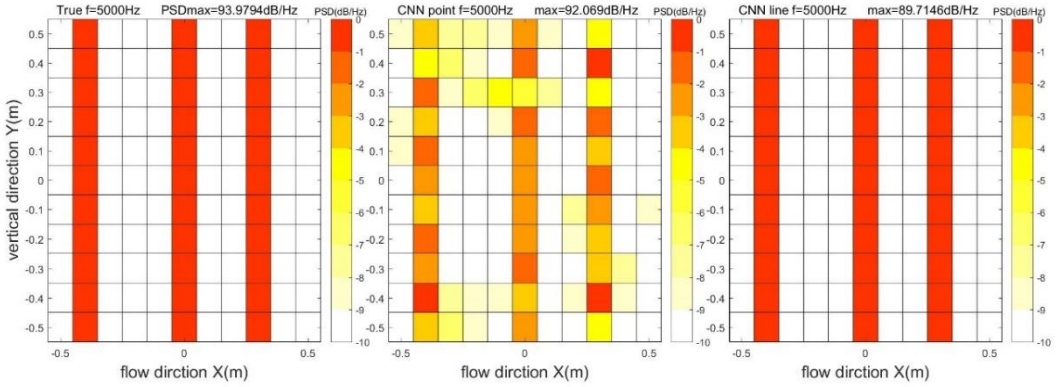
(a)CNN predictions of one line source.



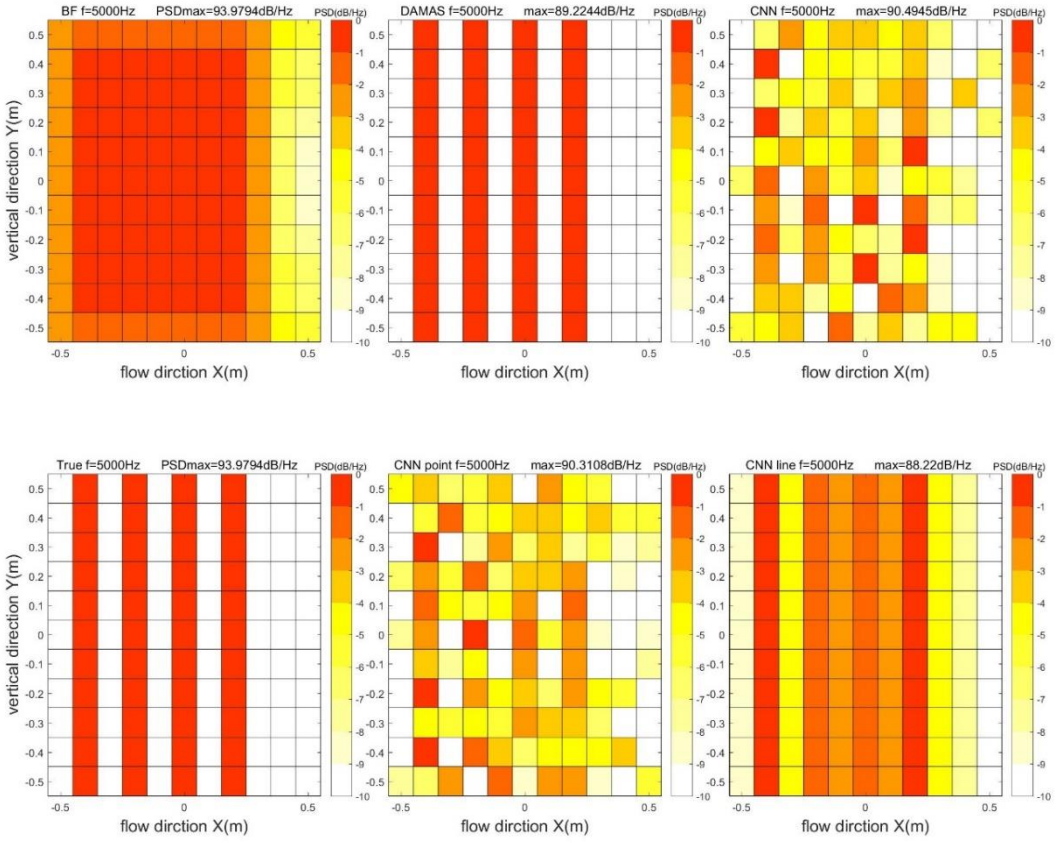


(b) CNN predictions of two line sources.





(c) CNN predictions of three line sources.



(d) CNN predictions of four line sources.

Figure 20 – The source distribution map (From left to right are beamforming result, DAMAS result original CNN predictions, true result, CNN point predictions and line predictions based on the new added parameters. All maps have a dynamic range of 10dB).

3.5 Validation of methods on wind tunnel experiment

An aeroacoustics experiment is conducted in Beihang D5 aeroacoustics wind tunnel. A 300mm-chordlength NACA0012 airfoil is tested in the Kevlar closed section. The incoming velocity ranges from 30m/s to 40m/s and the angle of attack is 2 degree and -2 degree. The phased array centers in the rotating axis of the airfoils. A general view of the experiment setup is shown in Figure 21. The scanning grids are 1.5 meters away from the array plane.

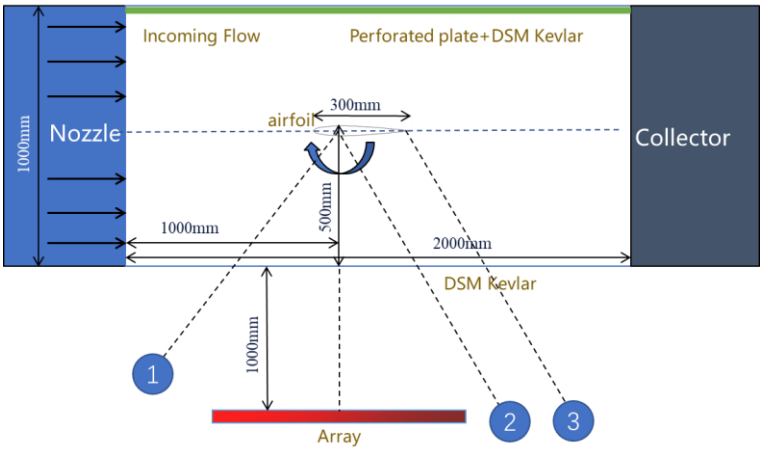
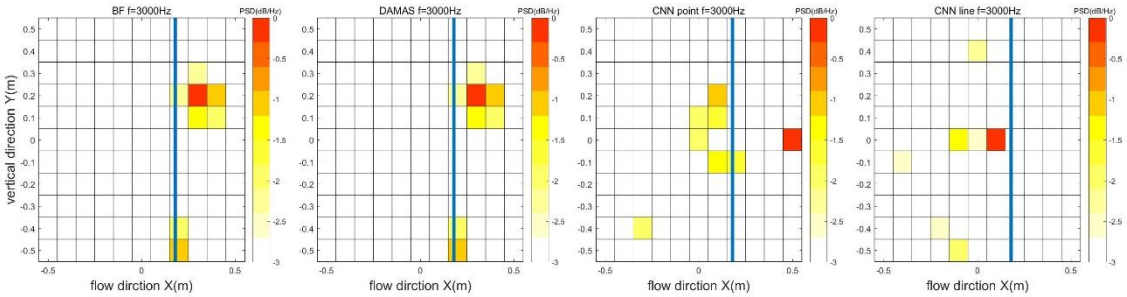
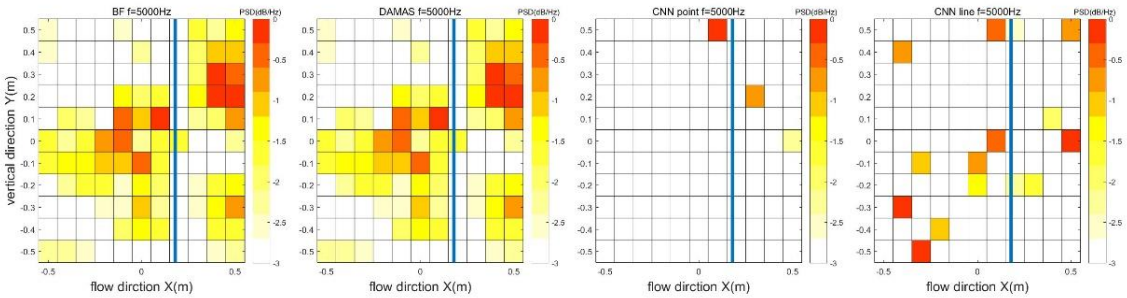


Figure 21 – Experiment setup in D5 wind tunnel.

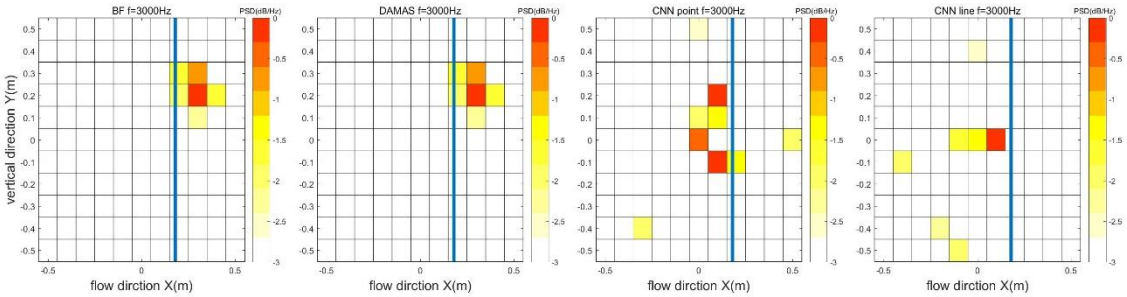
The CNN results are shown below in Figure 21. The trailing edge position is along the blue lines. Judging from the CNN point method (without the values along the x-direction of average square pressure across the y-direction to train) and CNN line method (with the values along the x-direction of average square pressure across the y-direction to train), the predicted source map is improved by adding new statistical parameter, which helps to locate the trailing edge that is indicated as the blue lines. As is shown in Figure 21, the CNN line methods decrease the possibility misidentifying the sound sources on the edge and strengthen the ability of detecting sound sources in the inner regions, which has been discussed and emphasized in Section 3.3.



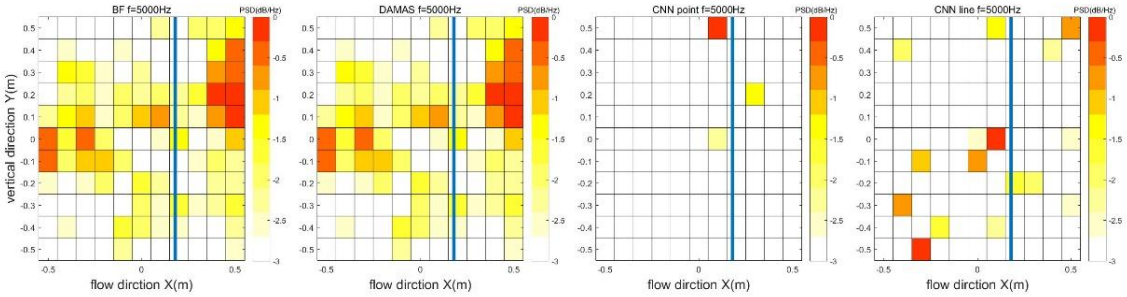
(a) 30m/s, 2 degree, 3000Hz, 1/3 octave



(b) 40m/s, 2degree, 5000Hz, single frequency



(c) 30m/s, -2 degree, 3000Hz, 1/3 octave



(d) 40m/s, -2 degree, 5000Hz, single frequency

Figure 21 – The results of the source distribution after normalization(From left to right are beamforming result, DAMAS result, CNN point predictions and CNN line predictions based on the new added parameters. All maps have a dynamic range of 3dB).

4. Conclusion

Deep learning methods like CNN algorithms have shown potential for producing the sound source distribution map efficiently. However, the usage of the CNN has been limited by the vague applicable range of finite number of monopole sources and ambiguous performance in frequency and scanning grids. In this study, to evaluate the prediction results on the simulations and experiments of trailing edge noise, a thorough investigation was done and the prediction result about trailing edge noise was evaluated.

To solve the range of application in quantifying the influence factors, a standardization of relevant parameters was achieved. In comparison with other researchers' results, different array ratios were defined. Among them, AR_RLS, namely Δx over Rayleigh limit simply (Rayleigh limit, while $\alpha \rightarrow 0$), were chosen to be the array ratio for normalizing frequency in this study. To filter out wrong predictions of the result in a strict way, new definitions of accuracy and loss to evaluate the prediction results of CNN algorithm were proposed. Acc_3, which used a threshold of -3dB to filter out the values in prediction results, was the accuracy criterion chosen in this study. By filtering out sources not larger enough in the predictions, the CNN model in this study had a relatively convincing performance when AR_RLS was larger than 0.374 under Acc_3, rather than 0.32 under Acc_5. To analyze the wrong predictions, models were tested on different test sets to conclude the trends of accuracy. According to Acc_3, the predictions had a reliable region dependent on the number of sources detected. Judging from loss_sratio, the wrong predictions should first appear in the inside central region than the outer region.

Because the topology information of different grids influences the correctness ratio of the predictions, new statistics parameters are proposed based on the flow-direction average of the sound strength on each grid. In comparison with the general CNN point methods of simulations and NACA0012 wind tunnel experiment, the new CNN line methods increase the ability of the CNN to some typical distribution of sound sources.

In the end, a procedure of using CNN in simulations of line sources and experiments were proposed.

Due to the network layers of non-linear operations, the values of output vectors might be a better representative of the source distribution maps when new parameters are trained in the dense layers. During future work, the network about the definition of the output should be further investigated in improving the spatial resolution for more complicated distributions of sound sources.

5. Contact Author Email Address

Corresponding author at: School of Aeronautical Science and Engineering, Beihang University, Beijing 100191, People's Republic of China

Peiqing Liu: lpq@buaa.edu.cn

6. Copyright Statement

7. Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (No. 11272034, 12072016).

References

[1] Dobrzynski, Werner. Almost 40 Years of Airframe Noise Research: What Did We Achieve?[J]. Journal of Aircraft, 2010, 47(2):353-367

[2] Mueller T J , Prasad M G . Aeroacoustic Measurements[J]. Applied Mechanics Reviews, 2003, 56(5):B66.

[3] Chiariotti P , Martarelli M , Castellini P . Acoustic beamforming for noise source localization – Reviews, methodology and applications[J]. Mechanical Systems and Signal Processing, 2019, 120(APR.1):422-448.

[4] Arce León, C., Merino-Martinez, R., Ragni, D., Avallone, F.,Snellen, M. Boundary layer characterization and acoustic measurements of flow-aligned trailing edge serrations. Exp.Fluids 57(182), 1–22 (2016).

[5] Arce León, C., Merino-Martinez, R., Ragni, D., Avallone, F., Scarano, F., Pröbsting, S., Snellen, M., Simons, D.G., Madsen,J. Effect of trailing edge serration-flow misalignment on airfoil noise emission. J. Sound Vib. 405, 19–33 (2017).

[6] Merino-Martinez R , Sijtsma P , Snellen M , et al. A Review of Acoustic Imaging Methods Using Phased Microphone Arrays[J]. CEAS Aeronautical Journal, 2019(1).

[7] Bianco, Michael & Gerstoft, Peter & Traer, James & Ozanich, Emma & Roch, Marie & Gannot, Sharon & Deledalle, Charles. (2019). Machine learning in acoustics: Theory and applications. The Journal of the Acoustical Society of America. 146. 3590-3628. 10.1121/1.5133944.

[8] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,”Bio. Cybern. 36(4), 193–202 (1980).

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems (2012), pp. 1097–1105.

[10] S. Chakrabarty and E. A. Habets, “Broadband DOA estimation using convolutional neural networks trained with noise signals,” in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE (2017), pp. 136–140.

[11] Chakrabarty S , Emanuël A. P. Habets. Multi-Speaker Localization Using Convolutional Neural Network Trained with Noise[J]. 2017

[12] Ma W, Liu X. Phased microphone array for sound source localization with deep learning[J]. arXiv preprint arXiv:1802.04479, 2018

[13] Vera-Diaz J, Pizarro D, Macias-Guarasa J. Towards End-to-End Acoustic Localization Using Deep Learning: From Audio Signals to Source Position Coordinates[J]. Sensors, 2018, 18(10): 3418

[14] Arcondoulis E , Liu Y , Xu P . An investigation of the facility effects on NACA 0012 airfoil tonal noise[C]// Aiaa/ceas Aeroacoustics Conference. 2019

[15] Michel U . History of acoustic beamforming[C]// 1st. Berlin Beamforming Conference. DLR, 2006

[16] Dougherty R P . Beamforming in Acoustic Testing[M] Aeroacoustic Measurements. Springer Berlin Heidelberg, 2002

[17] Dougherty, R.P.: Functional beamforming. 5th Berlin Beamforming Conference, February 19–20 2014, Berlin, Germany, GFaI,e.V., Berlin (2014)

[18] Sarradj, E., Schulze, C., Zeibig, A.: Identification of noise source mechanisms using orthogonal beamforming. Noise Vib Emerg Methods (2005)

[19] Shan T J , Kailath T . Adaptive beamforming for coherent signals and interference[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1985, 33(3):527-536

[20] Cardoso J F, Souloumiac A. Blind Beamforming for non Gaussian Signals[J]. IEE Proc.-F, 1993, 140(6)362-370

[21] Brooks, T.F., Humphreys, W.M.: A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. 10th AIAA/CEAS Aeroacoustics Conference, May 10–12, 2004, Manchester, UK, 2004, AIAA Paper 2004-2954 (2004)

[22] Dougherty R P . Extensions of DAMAS and Benefits and Limitations of Deconvolution in Beamforming[C]// 11th AIAA/CEAS Aeroacoustics Conference, Monterey, California, May 23-25, 2005. 2005

[23] Suzuki, T.: DAMAS2 using a point-spread function weakly varying in space. AIAA J. 48(9), 2165–2169 (2010)

[24] Haxter, S.: Extended version: improving the DAMAS 2 results for wavenumber-space beamforming. 6th Berlin Beamforming Conference, February 29–March 1 2016, Berlin,

Germany (2016)
[25]Högbom, J.A.: Aperture synthesis with a non-regular distribution of interferometer baselines. Astron. Astrophys. Suppl. Ser. 15,417–426 (1974)
[26]Sijtsma, P.: CLEAN based on spatial source coherence. Int. J.Aeroacoust. 6(4), 357–374 (2007)
[27]Sijtsma, P., Merino-Martinez, R., Malgoezar, A.M.N., Snellen, M.: High-resolution CLEAN-SC: theory and experimental validation.Int. J. Aeroacoust. 16(4–5), 274–298 (2017)
[28]Herold, G., Geyer, T.F., Sarradj, E.: Comparison of inverse deconvolution algorithms for high-resolution aeroacoustic source characterization. 23rd AIAA/CEAS Aeroacoustics Conference,June 5–9 2017, Denver, Colorado, USA, 2017, AIAA Paper 2017-4177 (2017)
[29]L. Rayleigh. “Xxxi. investigations in optics, with special reference to the spectroscope.” The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 8(49), 261–274, 1879
[30]M. S. Brandstein and H. F. Silverman, “A robust method for speech signal time-delay estimation in reverberant rooms,” in Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, Apr. 1997, pp. 375–378
[31]Xu Pengwei, Arcondoulis Elias, Liu Yu. DEEP NEURAL NETWORK MODELS FOR ACOUSTIC SOURCE LOCALIZATION. Conference: The 8th Berlin Beamforming Conference (BeBeC)At: Berlin, GermanyVolume: BeBeC-2020-D21
[32]Castellini P , Giulietti N , Falcionelli N , et al. A neural network based microphone array approach to grid-less noise source localization[J]. Applied Acoustics, 2021, 177:107947.
[33]Q. Leclère, Pereira A , Bailly C , et al. A unified formalism for acoustic imaging based on microphone array measurements[J]. International Journal of Aeroacoustics, 2017, 16(4-5)431-456
[34]Sarradj, E.: Three-dimensional acoustic source mapping with different beamforming steering vector formulations. Adv.Acoust. Vib. 2012(292695), 1–12 (2012)
[35]F. Chollet, Keras, GitHub Repository (2015)