Proceedings of the 1st international Symposium on AI and Fluid Mechanics
Paper No S2 P2

Multi-source Heterogeneous Aerodynamic Data Fusion Neural Network Embedding Reduced-dimension Features

[1]Chenjia Ning , [1]Weiwei Zhang*

[1] *School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China*

Abstract

Constructing a high-fidelity aerodynamic database at low cost is of great significance to solve the contradiction between the aerodynamic analysis accuracy and development cost of modern aircraft. Data fusion technology opens up fresh perspectives for the efficiency improvement and cost reduction of aerodynamic acquisition. The majority of the current studies aim at homogeneous data fusion with various fidelities, while the heterogeneous data fusion methods are rarely involved in aerodynamics. In many research fields, just a few concentrated loads are used throughout most aerodynamic modeling. As the aerodynamic is most closely related to the concentrated force, the distributed force is frequently underutilized because of its mismatched data type. This research focuses on multi-source heterogeneous aerodynamic data fusion and reduced-dimension features embedding, and a multi-source heterogeneous aerodynamic fusion method (MHA-Net) based on autoencoders is proposed. The chief idea of this method is to fully use the neglected aerodynamic distributed load for extracting the physical reduced-dimension features and embedding the features into the prediction model by imposing constraints, which significantly increases the accuracy and reliability of the few-shot modeling. The MHA-Net is an indirect fusion technique that effectively implements multi-source heterogeneous aerodynamic data fusion. In this study, the CAS350 wind turbine airfoil serves as the basic shape with apparent flow separation. There are 116 sample pairs of experimental data consisting of the whole data set, covering different flow conditions. Compared with the single-source concentrated force model, this method can significantly reduce the prediction error and model dispersion. Statistically, under the few-shot learning region, the MHA-Net can reduce the aerodynamic modeling MAE by more than 20 % on average, as well as the model dispersion by more than 45 % on average.

Keywords: Multi-source modeling, Heterogeneous model, Aerodynamic data fusion, Reduced-dimension extraction, Feature embedding

Introduction

In aircraft design, using high-fidelity (HF) aerodynamic data can significantly reduce the need for iterative cycle and correction while also enhancing the capabilities of model design. Nonetheless, the acquisition of HF aerodynamic data slows down the design iteration cycle and process seriously in terms of cost and time consumption. How to achieve "less investment and high profits" and obtain HF data at low cost has become one of the key issues in engineering design. Data fusion technology duly echoes this demand and can be considered as an effective manner to balance speed and accuracy at a given cost. In recent years, emerging approaches such as deep learning and artificial intelligence has drawn substantial research attention with the development of data science, offering a good opportunity for breakthroughs in aerodynamic data fusion.

Data fusion technology commonly refers to a group of prediction methods that combine multi-fidelity data to mine potential correlations and effectively exploit them. One of the main methods to implement data fusion is to construct surrogate models. This form of surrogate model is defined as multi-fidelity model (MFM)[1]. The majority of multi-fidelity models are constructed from some fundamental data-driven models with broad applicability, such as RBF[2], Kriging[3], deep neural networks (DNN), and deep Gaussian processes[4]. MFM has been successfully implemented in several fields, including aerospace design[5], electromagnetic simulation[6], and image recognition and classification[7].

As the name implies, multi-source refers to several data sources, whereas heterogeneous refers to a diversity of data formats or data types. The majority of the aforementioned studies are aimed at homogeneous data fusion with various fidelities, while the heterogeneous data fusion methods are rarely involved in aerodynamics. Most of the current research focuses on multimodal deep learning and big data analysis[8]. The massive aerodynamic data collected by numerical simulation (CFD) or experimental measurements often have complicated data formats and sources. From a different classification perspective, it can be separated into two categories: concentrated aerodynamic force and distributed aerodynamic force, depending on the type of aerodynamic force. Aerodynamic data with different types should be utilized independently. As the most common data type in aerodynamics, concentrated aerodynamic force is widely used in steady/unsteady aerodynamic modeling[8-10], aerodynamic optimization design and other fields. Unfortunately, just a few concentrated loads are used throughout most aerodynamic modeling, while distributed loads

* Corresponding author, Weiwei Zhang aeroelastic@nwpu.edu.cn

are frequently underutilized. Although data of various heterogeneous types cannot be utilized directly in aerodynamic physical systems, there must be relationships between them. It is a meaningful and ingenious research on how to mine the potential correlations between aerodynamic heterogeneous data, extract the effective information as compensating physical features, and incorporate these features into aerodynamic models.

This research focuses on multi-source heterogeneous aerodynamic data fusion and reduced-dimension features embedding, and a multi-source heterogeneous aerodynamic fusion method based on autoencoder (AE) is proposed. The chief idea of this method is to fully use the neglected aerodynamic distributed load for extracting the physical reduced-dimension features, and embed the physical features into the prediction model by imposing constraints, which significantly increases the accuracy and reliability of the few-shot modeling. In contrast to most homogeneous aerodynamic data fusion approaches, this method offers an efficient means for the collaborative use of heterogeneous data, and fills the gap of aerodynamic heterogeneous modeling in scientific research and engineering applications. This method has important engineering application value in high-efficiency and low-cost construction of high-fidelity aerodynamic database, correlation of aerodynamic force and pressure data, etc., and also offers new insights for the efficiency improvement and cost reduction of aerodynamic data.

Methods

Figure 1 illustrates the modeling flow chart of the MHA-Net, which is divided into three portions. The generation of sample pairs in the data set required for heterogeneous aerodynamic fusion modeling is depicted in Figure 1(a). The common single-source concentrated force modeling takes the concentrated force calculated by integrating the wall pressure distribution as the model output, and distributed force data are not taken into account throughout the modeling process. In MHA-Net, we divide each sample pair into three parts: input variables, distributed aerodynamic forces derived directly from wall pressure distribution, and concentrated aerodynamic force via integrating, which are represented by purple, blue, and green cylinders, respectively. Figure 1(b) describes the compression and extraction of dimensionality reduction features using AE for distributed loads in detail. The locations of pressure points on the wall surface are used to extract the pressure distribution map, which is then vectorized to produce the s-dimensional pressure coefficient vector x_{cp} arranged in a fixed sequence. x_{cp} is employed as the input of the encoder, and $\hat{x}_{cp}$ is served as the supervision information of AE for self-supervised learning, where $\hat{x}_{cp} = x_{cp}$. The nonlinear reduced-dimension features $\boldsymbol{F_r}$ extracted from the distributed load is obtained from the output of the encoder or the input of the decoder. The mathematical expression of the autoencoder is:

$$\begin{aligned} \boldsymbol{F}_r &= \mathrm{G}_e(\boldsymbol{x}_{cp};\boldsymbol{\phi}) \\ \hat{\boldsymbol{x}}_{cp} &= \mathrm{G}_d(\boldsymbol{F}_r;\boldsymbol{\varphi}) = \boldsymbol{x}_{cp} \end{aligned} \tag{1}$$

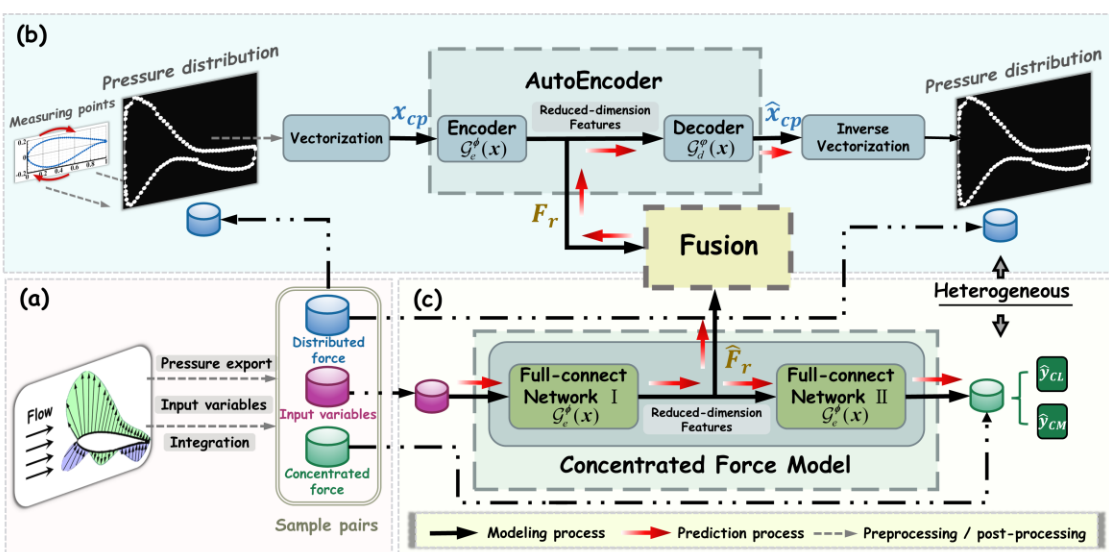


Figure 1 The flow chart of MHA-Net

* Corresponding author, Weiwei Zhang aeroelastic@nwpu.edu.cn

Figure 1(c) demonstrates the concentrated force modeling module in heterogeneous aerodynamic fusion modeling. The expression is displayed in Eq.(2). In Figure 1, the dash-dot-dot-lines indicates the application position of each sample part in the model training. The black arrow represents the training process of the model， while the red arrow denotes the prediction process after the model training is completed. In contrast, the red arrow is used to the prediction process after training. In MHA-Net, the extracted reduced-dimension features are used as additional physical constraint, so that the dimensionality reduction information of the distributed load is added into aerodynamic model realized the multi-source heterogeneous data fusion.

$$\begin{aligned} \hat{\boldsymbol{F}}_r &= \mathrm{P}_1(\boldsymbol{x};\boldsymbol{\xi}) \\ \hat{\boldsymbol{y}} &= \mathrm{P}_2(\hat{\boldsymbol{F}}_r;\boldsymbol{\vartheta}) \end{aligned} \tag{2}$$

Results

In this case, the CAS350 wind turbine airfoil is served as the basic shape with relatively thick airfoil thickness and obvious flow separation. Accordingly, aerodynamics exhibit strong nonlinear characteristics. There are 116 sample pairs of experimental data consisted in the whole data set, covering different Reynolds numbers and angles of attack with $Re \in [1.0 \times 10^6, 1.5 \times 10^6, 3.0 \times 10^6, 4.0 \times 10^6]$ and $\alpha = -10°\sim 18°$, and the angle of attack interval of the test samples under fixed Re is 1°. The Reynolds number in the low-speed wind tunnel is adjusted by the incoming wind speed, so the Mach number is not taken into account. Each sample pairs of test states contains multi-source heterogeneous aerodynamic data, including distributed load data with 97 pressure points on the airfoil surface, lift coefficient C_L and moment coefficient C_M. Based on the actual wind turbine experimental data, this example constructs a multi-source heterogeneous aerodynamic model with the flow parameters as input to verify the applicability and application value of MHA-Net in practical engineering problems.

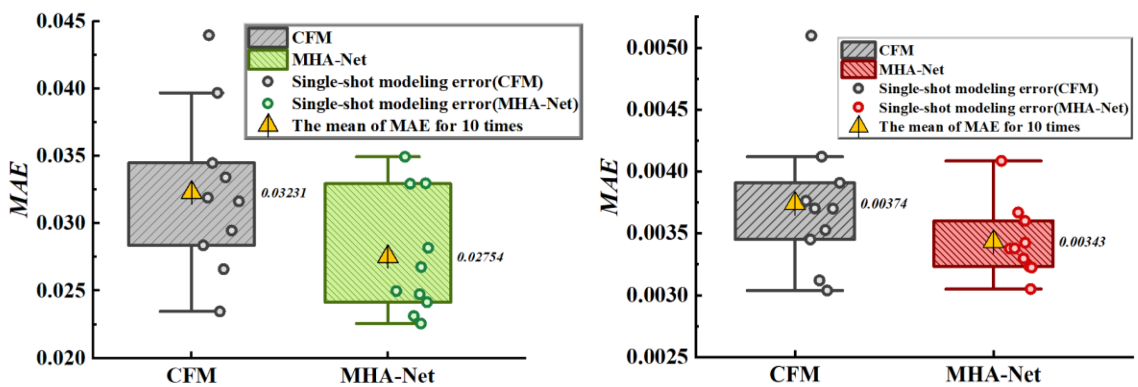


Figure 2 The MAE of test set under repetitive modeling when training set accounted for 60%

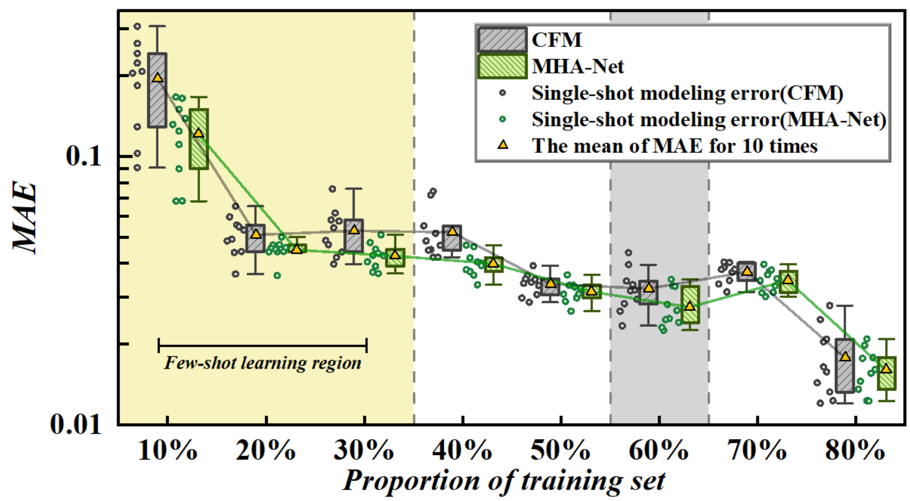
Aiming at the CAS350 airfoil variable state heterogeneous fusion case, this study carried out aerodynamic modeling based on a fixed number of training samples to verify the feasibility of MHA-Net. An extended analysis involves the modeling of variable training sample number to confirm the applicability of the heterogeneous fusion method in few-shot learning. The 116 sample pairs of experimental data were randomly disrupted, and were divided into two parts: 60 % of sample pairs were served as the training set, and the remaining 40 % were treated as the test set. Randomly and evenly extract 10 % from the training set as the validation set (6 %), which is used to optimize the hyperparameters of neural networks. For the purpose of setting the control group more reasonably, the Figure 1(c) is used as the baseline of the single-source concentrated force model, denoted as CFM. Besides, the random initialization of neural networks significantly affects the training results, leading to strong model uncertainty. In an effort to comprehensively assess the advantages of MHA-Net, this study randomly initializes the two methods and performs 10 repetitive modeling to derive statistical predictions. The modeling accuracy and dispersion of different methods are evaluated.

Figure 2 shows the MAE boxplot of MHA-Net and CFM on the test set when the training set accounts for 60%. The statistical MAE of CFM for 10 times of modeling is depicted in the grey boxplot. And the green and red boxplots are related to the statistical MAE by MHA-Net for C_L and C_M, respectively. Among them, the mean of MAE(MMAE) for

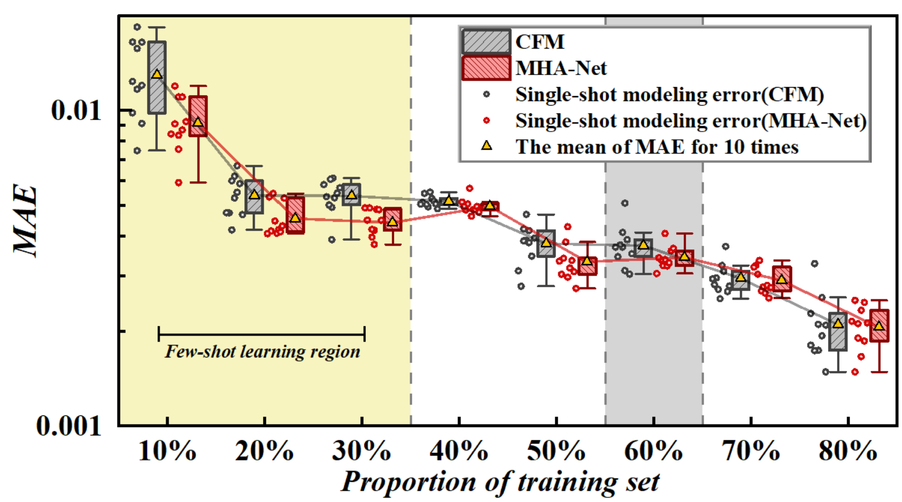
* Corresponding author, Weiwei Zhang aeroelastic@nwpu.edu.cn

repetitive modeling is represented as yellow triangle, which is noted in Figure 2 with the mean value. For C_L, the MMAE decreased by about 15%, and the dispersion of the aerodynamic model decreased by 23.7% through MHA-Net. The accuracy and robustness of C_M are also elevated, but the improvement is weaker than that of the C_L.

For investigating the modeling accuracy of MHA-Net under various training samples, this research verifies the advantages of MHA-Net under few-shot learning by altering the ratio between training set and test set. Consistent with the previous instance, the statistical predictions of models are trained by 10 times of random initializations, and the modeling accuracy and dispersion are also evaluated accordingly. Figure 3 depicts the MAE boxplots of heterogeneous aerodynamic fusion modeling and single source modeling with various training samples by logarithmic axis. With the decrease of model training samples, the advantages of MHA-Net method gradually rise, indicating that this method has exceptional advantages and effects in few-shot modeling. Statistically, under the few-shot learning region in the CAS350 variable state aerodynamic case, the MHA-Net method can reduce the aerodynamic modeling MAE by more than 20 % on average, as well as the model dispersion by more than 45 % on average. The results fully reflect the rewards of heterogeneous aerodynamic data fusion for few-shot sparse modeling, effectively lessen the model's dependence on samples, and raise the accuracy and robustness of the few-shot aerodynamic modeling.



(a) C_L



(b) C_M

Figure 3 Concentrated force prediction error with the adjustment of training samples

* Corresponding author, Weiwei Zhang aeroelastic@nwpu.edu.cn

Conclusion

In a purpose of improving the reliability of aerodynamic modeling under limited samples and lessening the sample dependence of models, this study proposes a multi-source heterogeneous aerodynamic data fusion method (MHA-Net). In this research, the previously neglected distributed load is used as additional modeling information to extract the nonlinear reduced-dimension features through the autoencoder. By embedding reduced-dimension features into the concentrated aerodynamic model, the proposed method (MHA-Net) significantly enhance the accuracy and robustness of aerodynamic models, which fills the technical gap in heterogeneous aerodynamic data fusion. The single-source concentrated force method is treated as a benchmark, and comparative analysis and verification are carried out on the variable state experimental case of wind turbine airfoil. The results demonstrate that this embedded multi-source heterogeneous aerodynamic fusion method has substantial advantages in improving the accuracy and robustness of the model, especially in the aerodynamic sparse few-shot modeling. The MHA-Net has strong application in aerodynamic shape optimization and other fields because of its high accuracy and robustness under few-shot learning. Due to its exceptional strengths in robustness, the proposed method can be smoothly implemented in practical engineering problems.

References

[1]PEHERSTORFER B, WILLCOX K, GUNZBURGER M. Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization[J]. Siam Review, 2018, 60(3): 550-591.

[2]SOBESTER A, LEARY S J, KEANE A J. On the design of optimization strategies based on global response surface approximation models[J]. Journal of Global Optimization, 2005, 33(1): 31-59.

[3]KRIGE D G. A statistical approach to some basic mine valuation problems on the Witwatersrand[J]. Journal of the Southern African Institute of Mining and Metallurgy, 1951, Vol.52(No.6): 119-139.

[4]DAMIANOU A, LAWRENCE N D. Deep gaussian processes; proceedings of the Artificial intelligence and statistics, F, 2013 [C]. PMLR.

[5]ZHANG X S, XIE F F, JI T W, et al. Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization[J]. Computer Methods in Applied Mechanics and Engineering, 2021, 373.

[6]KOZIEL S, BEKASIEWICZ A. Rapid design optimization of antennas using variable-fidelity EM models and adjoint sensitivities[J]. Engineering Computations, 2016, 33(7): 2007-2018.

[7]LIU X, LI L L, LIU F, et al. GAFnet: Group Attention Fusion Network for PAN and MS Image High-Resolution Classification[J]. Ieee Transactions on Cybernetics, 2022, 52(10): 10556-10569.

[8]WANG X, KOU J Q, ZHANG W W, et al. Incorporating Physical Models for Dynamic Stall Prediction Based on Machine Learning[J]. Aiaa Journal, 2022, 60(7): 4428-4439.

[9]LI K, KOU J Q, ZHANG W W. Unsteady aerodynamic reduced-order modeling based on machine learning across multiple airfoils[J]. Aerospace Science and Technology, 2021, 119.

[10]WANG X, KOU J Q, ZHANG W W. A new dynamic stall prediction framework based on symbiosis of experimental and simulation data[J]. Physics of Fluids, 2021, 33(12).

* Corresponding author, Weiwei Zhang aeroelastic@nwpu.edu.cn