Article

Real-Time Structure Generation Based on Data-Driven Using Machine Learning

Ying Wang 1,*, Feifei Shi 1 and Bingbing Chen 2

1 Department of Mechanical and Electrical Engineering, Suzhou Institute of Industrial Technology, Suzhou 215100, China
2 Department of Energy Science and Engineering, Nanjing Tech University, Nanjing 210009, China
* Correspondence: wangying@siit.edu.cn

Abstract: Topology optimization results are highly dependent on the given design constraints and boundary conditions. Moreover, small changes in initial design conditions can result in different topological configurations, which makes topology optimization time-consuming in a given design constraint domain and inefficient in structural design. To address this problem, a data-driven real-time topology optimization framework and method coupled with machine learning by using a principal component analysis algorithm combined with a feedforward neural network are developed in this paper. Meanwhile, through the offline training, the mapping relationship between initial design conditions and topology optimization results is obtained. From this mapping, we estimate the optimal topologies for novel loading configurations. Numerical examples display that the online prediction results are consistent with the results of the topology optimization method. Furthermore, the network parameters are calibrated, and accurate structure prediction is achieved based on the algorithm. In addition, this method ensures the accuracy of high-resolution structural prediction on the premise of small samples.

Keywords: structure optimization; topology design; machine learning; data dimension reduction

1. Introduction

With the increasing demand for structural functions, topology optimization has been introduced into various fields for structure design, such as aerospace, engineering applications, electronic devices, and biomedicine [1–5]. So far, topology optimization has been developed to be one of the most effective ways to acquire new structures satisfied various design requirements [6–10]. With the continuous improvement of the theoretical system of topology optimization, varieties of methods have been developed to achieve the optimization of structures. At present, the solid isotropic material with penalization (SIMP) method [11], level set method [12], and moving morphable components (MMC) method [13] are widely used and implemented in topology optimization. Additionally, the evolutionary structural optimization (ESO) method [14], genetic algorithm (GA) method [15], and phase field method [16] are also efficient methods for structure optimization. Multiple constraints such as stress constraints [17], multi-material constraints [18], and other constraints are habitually considered in topology optimization design. Moreover, for the structural design of additive manufacturing, the thickness constraint of the minimum printable thickness [19] and the printing angle are as well as considered during the molding process [20]. Therefore, topology optimization provides a reference for structural design, in which multiple constraints make the design results much closer to engineering practice. Compared to the traditional structure design, topology optimization under multiple constraints provides more ideas for structure design. However, the iterative process of optimizing calculation is extremely time-consuming unless the time of iteration is greatly reduced or replaced by another more efficient method.
The neural networks, providing convenience for people to reduce the time for calculation of topology optimization, have been proven to be powerful functions in data learning and generalization, which can be used in classification, clustering, and regression [21–23]. Many researchers have aimed to reduce the time of structural design by combining neural networks with the topology optimization method. Taking the topological optimization of the heat conduction problem as an example, Lin et al. [24] proposed a deep learning approach combining the traditional solid isotropic material with the penalization method to reduce the time consumption of the optimization process. Ulu et al. [25] established a mapping between the load and the optimization result based on the SIMP method combined with the neural network algorithm. The results indicate that the proposed method can successfully predict the optimal topologies in novel loading configurations and also can be used as effective initial conditions for conventional topology optimization routines, resulting in substantial performance gains. Lei et al. [26] established a mapping between the load and optimization results on the basis of MMC combined with the supported vector regression (SVR)/K-nearest neighbors (KNN) algorithm. The training data and dimension of parameter space are substantially reduced. Yu et al. [27] solved the problem of low sample resolution by using generative adversarial networks (GAN) in view of the establishment of a predictive network with convolutional neural networks (CNN). The proposed method can determine a near-optimal structure in terms of pixel values and compliance with negligible computational cost. Cang et al. [28] proposed a theory-driven mechanism for a neural network model that performs generative topology design in one shot given a problem setting, circumventing the conventional iterative process. Thus, it can be seen, the combination of topology optimization and neural network learning is an effective way to reduce the cost of computation. However, the aforementioned methods still have some problems. For example, some methods only accelerate a portion of the optimization process, some methods have a low dataset and small sample size that is not enough to explain various problems, and some methods need a huge sample size and expensive calculation cost. Based on these observations, a data-driven real-time topology optimization framework and method coupled with machine learning are developed. The relationship between initial design conditions and topology optimization is obtained. The results ensure the accuracy of high-resolution structural prediction under the premise of small samples.

The remaining sections are organized as follows. Section 2 presents the structural optimization theory and numerical examples. Section 3 describes the construction of neural networks and the theory of dimension reduction algorithms in detail. The effects of various network parameters are also discussed. Finally, the advantages and disadvantages of this method are systematically described, and the corresponding solutions are proposed.

2. Structural Optimization Theory and Data Sample

The purpose of topology optimization is to provide the best material distribution in the designated design area according to the given load and boundary conditions. In other words, a certain mechanical performance index of the optimized configuration can reach the maximum value. Messerschmitt–Blkow–Blohm (MBB) beam problem is used as an example of real-time topology optimization implementation in this study. In the MBB model, the length and width of the beam are represented by $N_x$ and $N_y$, respectively. The design domain is discretized into $N_x \times N_y$ rectangular grids for finite element calculations, and the density $\rho$ of each element controls the presence or absence of the material. $F(x, y)$ describes the load at the $x, y$ points in the Cartesian coordinate system as seen in Figure 1. To reduce the amount of calculation, the beam is processed symmetrically and the DOF (degree of freedom) of $x$ in the middle part is constrained.
The goal of topological optimization is to seek the topological configuration of the MBB beam with maximum stiffness at a specific volume fraction. The objective function, $C$, represents the work performed by various configurations under load. The optimal solution is obtained by seeking the minimum value of the objective function. In the present research, the goal of topology optimization is maximum stiffness, and the objective function is equivalent to minimum compliance. The constraints can be described as volume constraints, equilibrium equations, and density constraints, so the problem description can be expressed as Equation (1):

$$
\min_{x} : c(x) = U^TGU = \sum_{j=1}^{N} E_j(x_j)u_j^Tu_j
$$

subject to:

$$
\begin{align*}
\sum_{j=1}^{N} V_j(x_j) & = f \\
GU & = F \\
0 < x_{\min} & \leq x_j \leq 1
\end{align*}
$$

where $U$ and $F$ are the displacements and global force, respectively. $G$ is the global stiffness matrix. $g_j$ and $u_j$ are the stiffness matrix and displacement vector of the element. $x_j$ is the design variable. $V$ and $V_0$ are the design material volume and design area volume.

The density interpolation function is chosen to give the modulus of the element as Equation (2):

$$
E_j = E_{\min} + x_j^p (E_0 - E_{\min}) \quad x_j \in [0, 1]
$$

where $E_0$ and $E_{\min}$ are the material modulus and minimum modulus, respectively. $p$ is the penalty coefficient of the modulus, which is set to 3 here.

A batch of random loads is used as input data for the topology optimization program, and a series of optimization results (1500) is obtained as a database. The input of the topology optimization program TOP88 is set to (100, 50, 0.25, 3, 1.5, 1, $F(x,y)$), which corresponds to (nelx, nely, volfrac, penalization coefficient, filter radius, filter function, load coordinates). The topology optimization program in this paper is adapted from the program of the group of Andreassen et al. [29], and part of the optimization results are exhibited in Figure 2. The red arrow in Figure 2 represents load, and the value of $C$ indicates compliance.
The MATLAB toolbox is used to build a neural network for predicting topology-optimized structural designs. The network is divided into two stages: the offline training stage and the online prediction stage. Before the offline training, the sample should be collected by the topology optimization program as the training set and contrast set. Then, the network is built by using a principal component analysis (PCA) algorithm combined with a feedforward neural network. Firstly, the training samples are reduced to the feature image by the PCA algorithm. Secondly, the load–weight relationship is established by the feedforward neural network. Finally, the relationship between the load and the optimization result is obtained. Based on the above stages, new load information is selected and input into the neural network to return a new weight W by online prediction, and real-time topology optimization is realized.

3.1. Data Reduction

The PCA algorithm is a dimensionality reduction method for preprocessing high-dimensional feature data, which improves the data processing speed by removing noise [30]. The optimization results need to be reduced to gain a certain number of feature images via the PCA algorithm before training. The PCA algorithm flow is shown in Figure 3. If the design domain size is \( N_x \times N_y \) matrix, the single optimization result is represented by a matrix of rows. Then, the result of optimization \( O \) is a \( K \times L \) matrix, and the centralized image of the result can be expressed as \( Y_j = O_j - \bar{O} \), where \( \bar{O} \) is the mean value of the optimization result. In order to reduce the dimension of data, the covariance matrix \( C = YY^T \) is constructed. The \( C \) is a K-order square matrix with \( K \) eigenvectors, and the eigenvectors with smaller eigenvalues are unimportant features. According to the size of eigenvalues, the first \( M \)-order eigenvectors are selected as the new covariance matrices \( J \), which is the result of PCA dimensionality reduction (\( M \ll K \)).

\[
\begin{align*}
\text{Centralization} & \quad Y = O - \text{Ave}(Y) \\
\text{Covariance calculation} & \quad C = YY^T \\
\end{align*}
\]

The eigen-images shown in Figure 3 are calculated as \( V = Y^T J \), where \( V \) equals to \( K \times M \). It indicates that the optimization results can be linearly combined by these eigen-images. An equation of a linear mapping relationship between optimization results and eigen-images \( O_{K \times L} \) established as Equation (3):
The eigen-images shown in Figure 3 are calculated as $TV Y J = [W_1, W_2, \ldots, W_k]^T$ (3)

where $W$ is the mapping weight.

The original $K$ optimization results are integrated into $M$ eigen-images by the PCA method for data dimensionality reduction. The weight $W$ corresponding to the new design condition is the key to realizing real-time topology optimization. Subsequently, establishing the mapping relationship between load $F$ and weight $W$ is necessary.

3.2. Feedforward Neural Network

It is well known that a feedforward neural network is the most widely used neural network, which can realize the functions of classification, regression, and clustering. The regression function can be used to express the mapping relationship between load and eigen-image. Generally, the network can be divided into the input layer, the hidden layer, and the output layer. A four-layer feedforward neural network is shown in Figure 4. The information obtained from the input layer is transmitted to the hidden layer for training, and then the trained information goes into the output layer and returns the output value. This process of offline training is to train a feedforward neural network and establish the nonlinear mapping relationship between load $F_i$ and weight $W_i$. The load information $F_i$ is an input and the weight $W$ is set as a label. Trainlm is selected as the training function, the learning function is leaern, and the mean square error is used to calculate the error of samples. The trained network can realize online prediction, and real-time topology optimization can be achieved by combining the PCA algorithm.

Figure 4. The online prediction realization using the neural network.

3.3. The Implementation of Real-Time Topology Optimization Algorithm

The real-time topology optimization is generally divided into four stages, as shown in Figure 5. The first stage is sample collection, in which the training set and contrast set are obtained through the topology optimization program. The second stage is the data dimension reduction stage. In this stage, the training set is transformed into a featured image by the PCA algorithm, and the linear mapping relationship is established to obtain the weight $W$. The third stage is the offline training stage, in which the load information corresponding to the training set is input and labeled as weight $W$ to train the neural network. The fourth stage is the final online prediction stage. New load information is selected and input into the neural network to return the new weight $W$, and new optimization results are obtained by combining the eigen-images. The online prediction phase is completed instantaneously, which represents the realization of real-time topology optimization.
3.4. Discussion on Network Parameters

The training effect involves many parameters, such as the number of samples \( D \), the size of design domain \( L \), the number of feature images \( M \), the number of layers \( H \) of the neural network, and the amount of neurons \( N \). In order to determine the optimal value of the parameters, the control variable method is used to analyze the influence of the parameters. The TRAINLM function is selected as the training function, the LEARNNGDM function as the adaptation learning function and the TANSIG function is selected as the regularization function. Meanwhile, the function of subsequent network training settings remains the same. The prediction accuracy is expressed by the mean square error (MSE) as Equation (4):

\[
MSE = \frac{1}{D} \sum_{d=1}^{D} (y_d - \bar{y}_d)^2
\]  

where \( D \) is the number of samples, and the \( y_d \) is the density value of the predicted structure.

It is very important to determine the influence of eigen-images \( M \) on network training. Considering the computational efficiency, 500 samples are used to train the network. The layer of networks is set to 2, the neurons are set to 100, and the eigen-images \( M \) is taken from 10 to 200. One of the optimization results is selected to compare with the predicted results as shown in Figure 6.

The number of neurons \( N \) is one of the parameters of neural network. It is necessary to analyze the influence of the number of neurons on network training. Considering the computational efficiency, 200 samples are used to train the network. The layers of network are set to 2, the eigen-images are set to 20, and the numbers of neurons are taken from 10 to 150 to produce 15 sets of samples for training. The prediction images are shown in Figure 7.

Figure 8 displays the error analysis in view of the \( M \) and the \( N \). When \( M \) is in the range of 15–25% of the sample size, the error is stable at the range of 2.65–2.7%, and the minimum error is 2.663% when \( M \) is 22% of the sample size \( D \). The same trend is also reflected in the prediction image. When \( M \) is 110, the best prediction effect is obtained. When \( M \) is greater than 130, the image begins to appear over-fitting. This is because the excessive value of \( M \) leads to unnecessary features and too small a value of \( M \) leads to the phenomenon of under-fitting as well. The error analysis indicates that when \( N \) is in the range of 15–30% of the sample size, the error is stable at 3.6–3.8%, and the minimum error is 3.658% when \( M \) is 20% of the sample size. This is also reflected in the prediction image. When \( N \) is 40, the best prediction effect is obtained. When \( N \) is greater than 100,
the image begins to appear over-fitting. This is also because the excessive value of N leads to unnecessary features, and too small a value of N leads to the phenomenon of under-fitting. It is obvious that the training effect is optimal when \( N \approx M \). According to the results obtained by analyzing the errors of the control variables, when the values of the M and the N are both at 20% of the training samples, the minimum prediction error is obtained. It is not difficult to understand from the principle of neural networks, each neuron corresponds to an intrinsic image. In this way, when the weights are mapped as labels, a better training effect can be achieved.

As for depth H of the network, the most basic structure is the two-layer structure with one hidden layer and one output layer. When the number of hidden layers is over two, it is called a deep neural network. In order to determine the influence of \( H \), 200 samples are used to train the network. The number of neurons N is set to 20, the number of eigen-images M is set to 20, and the number of hidden layers is taken from 2 to 8 to produce 7 sets of samples for training. The prediction images and the errors are shown in Figures 9 and 10, separately. The error analysis indicates that when the value of \( H \) goes into 2–5, the error is less than 3.8%. When \( H \) is greater than 5, the image begins to appear over-fitting. The error of the three-layer network and four-layer network is 3.742% and 3.764%, respectively. Considering the computational cost, the number of hidden layers is set to 2.
The prediction result of training 3 with different H.

Figure 7. The prediction result of training 2 with different N.

Figure 8. The influence of number of feature images and number of neurons on prediction accuracy.

Figure 9. The prediction result of training 3 with different H.
In order to analyze the influence of the number of samples $D$, the layer of the network is set to 2, the eigen-images $M$ are set to 20 and the number of neurons $N$ is set to 20. The number of samples is taken from 100 to 1400. The prediction images are shown in Figure 11. It indicates that when the number of samples exceeds 500, the detail loss and prediction errors begin to appear on the prediction images, which is not consistent with common sense. This is because the values of $N$ and $M$ do not change with the number of samples, which leads to the inadequacy of network fitting ability. When the number of samples ranges from 100 to 400, the proportion of $M$ is in 5–20%, which further validates the previous conclusion of the $M$ value.

![Figure 10](image-url)  
**Figure 10.** The influence of the number of network layer on prediction accuracy.

![Figure 11](image-url)  
**Figure 11.** Initial prediction result of training 4 with different $D$.  

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Predict</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D=100$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=200$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=300$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=400$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=500$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=600$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=700$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=800$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=900$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=1000$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=1100$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=1200$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=1300$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td>$D=1400$</td>
<td><img src="image-url" alt="Image" /></td>
</tr>
</tbody>
</table>
3.5. Realization of Real-Time Topology Optimization

An efficient real-time topology optimization network is established through parameter impact analysis, which gives an accurate calibration of the network parameters. Based on the research, the prediction effect of real-time topology optimization is first determined by the sample size, and a sufficient number of samples can compensate for the problem of insufficient network fitting ability to some extent. When the number of neurons is about 20% of the sample size, the network fitting ability is the strongest. Under the premise of comprehensively considering the calculation and calculation consumption, a proper increase in the number of layers of the neural network can obtain a better prediction effect.

After the parameter calibration, the value or value range of the network parameters are determined. The number of neurons in the input layer is equal to the first dimension K of the weight w, and the number of neurons in each hidden layer is coincident with the input layer. The number of neurons in the output layer is concurrent with the dimension of the label (load), which is set constant to 2. The number of hidden layers is set to 2. As shown in Table 1, the value of eigen-images M and the number of neurons N are adjusted according to the value of sample size D. The prediction images are shown in Figure 12, and the errors are shown in Figure 13, respectively. As can be seen from the two figures, in the beginning, the prediction effect is very poor due to the lack of sample sizes. As the number of samples increases, the accuracy of image prediction increases synchronously. When the D reaches 500, the results are predicted as acceptably accurate. When D is 500, 600, 700, the network training time is 121.43 S, 178.54 S, and 252.76 S, respectively. The new load conditions are printed into the network, and the online prediction time is only 10-2 S, which is realized in real-time topology optimization. The prediction accuracy can be improved by segmenting and filtering the results of the preliminary prediction.

Table 1. Network parameter setting.

<table>
<thead>
<tr>
<th>Design Domain Size L</th>
<th>Sample Size D</th>
<th>Eigen-Images Size M</th>
<th>Neurons Size N</th>
<th>Number of Hidden Layers H</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 × 50</td>
<td>100</td>
<td>20</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>100 × 50</td>
<td>200</td>
<td>40</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>100 × 50</td>
<td>300</td>
<td>60</td>
<td>60</td>
<td>2</td>
</tr>
<tr>
<td>100 × 50</td>
<td>400</td>
<td>80</td>
<td>80</td>
<td>2</td>
</tr>
<tr>
<td>100 × 50</td>
<td>500</td>
<td>100</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>100 × 50</td>
<td>600</td>
<td>120</td>
<td>120</td>
<td>2</td>
</tr>
<tr>
<td>100 × 50</td>
<td>700</td>
<td>140</td>
<td>140</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 12. The prediction result after adjusting parameters of training 4 with different D.
4. Results and Discussion

A topology optimization model based on artificial neural network learning is proposed. Through the PCA algorithm, the training data are screened in advance to eliminate interference or redundant data and the data dimension is reduced. Then, through the comparative analysis of different training network parameters, the best training framework is found, so that it can obtain more accurate prediction results with less sample data.

Compared with the work of Ulu et al. [25], our algorithm particularly studies the influence of dimensionality reduction parameters and neural network parameters on the prediction results. Selecting appropriate parameters can significantly reduce the offline training data samples required for network training, and the prediction structure is more accurate.

Compared with the convolutional decoding and coding network adopted by Yu et al. [27], our network adopts the architecture of data dimensionality reduction and a full connection network. Before network training, we preprocess the samples by means of data dimensionality reduction to remove redundant information. Therefore, compared with Yu et al. [27] using 10,000 training samples to achieve accurate prediction of the structure, our algorithm can achieve the same effect with only 700 samples, that is, our algorithm has less sample dependency and stronger learning ability.

5. Conclusions

This study proposes a data-driven real-time topology optimization method that combines data dimensionality reduction with machine learning to achieve a design structure at the instant of a given design condition. The impact of the network parameters on the prediction accuracy is discussed and then the optimal parameter configuration scheme is obtained. It indicates the network prediction effect is the best when the number of hidden layers H is two and the number of neurons N is at 20% of the number of samples. In addition, when the number of samples K is greater than 500 as a threshold, the network error converges, and it decreases with the increasing number of samples.

The predicted image obtained by this method has higher resolution and more computational efficiency, and the demand sample is less. However, this method has certain limitations. Although the generation of the load is random, the boundary conditions in the example are fixed. This limitation can be solved by a combination of boundary conditions.
and loads. This study demonstrates the advantages of combining machine learning with topology optimization and broad development prospects, providing more efficient and accurate optimization methods and tools for structural design.

**Author Contributions:** Conceptualization, Y.W. and B.C.; methodology, Y.W.; software, F.S.; validation, Y.W., F.S. and B.C.; formal analysis, Y.W.; investigation, B.C.; resources, B.C.; data curation, F.S.; writing—original draft preparation, Y.W.; writing—review and editing, B.C.; visualization, F.S.; supervision, Y.W.; project administration, Y.W.; funding acquisition, B.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the National Natural Science Foundation of China (1902144), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (19KJD460006).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** No potential competing interests were reported by the authors.

**References**


3. Han, Z.; Wang, Z.; Wei, K. Shape morphing structures inspired by multi-material topology optimized bi-functional metamaterials. *Compos. Struct.* 2022, 300, 116135. [CrossRef]


7. Sahimi, M.; Tahmasebi, P. Reconstruction, optimization, and design of heterogeneous materials and media: Basic principles, computational algorithms, and applications. *Phys. Rep.* 2021, 939, 1–82. [CrossRef]

8. da Silveira, T.; Baumgardt, G.; Rocha, L.; dos Santos, E.; Isoldi, L. Numerical simulation and constructal design applied to biaxial elastic buckling of plates of composite material used in naval structures. *Compos. Struct.* 2022, 290, 115503. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.