

# Speeding up Computational Morphogenesis with Online Neural Synthetic Gradients

Yuyu Zhang<sup>\*§</sup>, Heng Chi<sup>†§</sup>, Binghong Chen<sup>\*</sup>,  
Tsz Ling Elaine Tang<sup>†</sup> Lucia Mirabella<sup>†</sup>, Le Song<sup>\*</sup> and Glaucio H. Paulino<sup>\*</sup>

<sup>\*</sup>Georgia Institute of Technology

<sup>†</sup>Siemens Corporate Technology

<sup>§</sup>Email: yuyu@gatech.edu, heng.chi@siemens.com

**Abstract**—A wide range of modern science and engineering applications are formulated as optimization problems with a system of partial differential equations (PDEs) as constraints. These PDE-constrained optimization problems are typically solved in a standard discretize-then-optimize approach. In many industry applications that require high-resolution solutions, the discretized constraints can easily have millions or even billions of variables, making it very slow for the standard iterative optimizer to solve the exact gradients. In this work, we propose a general framework to speed up PDE-constrained optimization using online neural synthetic gradients (ONSG) with a novel two-scale optimization scheme. We successfully apply our ONSG framework to computational morphogenesis, a representative and challenging class of PDE-constrained optimization problems. Extensive experiments have demonstrated that our method can significantly speed up computational morphogenesis (also known as topology optimization), and meanwhile maintain the quality of final solution compared to the standard optimizer. On a large-scale 3D optimal design problem with around 1,400,000 design variables, our method achieves up to 7.5x speedup while producing optimized designs with comparable objectives.

**Index Terms**—computational morphogenesis, PDE-constrained optimization, neural synthetic gradients, deep learning

## I. INTRODUCTION

Many science and engineering applications in various domains are addressed by turning them into optimization problems constrained by physical laws in forms of a system of partial differential equations (PDEs). The resulting mathematical problems are commonly known as the PDE-constrained optimization [1], [2]. In practice, PDE-constrained optimization is widely used in optimal design [3], [4], optimal control [5], [6], and inverse estimation [7], [8] problems. More concretely, PDE-constrained optimization has been applied in many scientific problems such as optical tomography in biomedical imaging [9], radiation treatment planning in radiotherapy [10], magnetic drug delivery in nanoscience [11] and architected materials design in material science [12]. It has also been used in a variety of engineering problems, such as flow control in fluid mechanics [13], structure design in aerospace, civil and architectural engineering [14]–[16], and optimal design in automobile industry and robotics [17]–[19].

PDE-constrained optimization problems are typically solved in a standard discretize-then-optimize approach, which first discretizes the governing PDE into finite-dimensional residual

form, and then optimizes the design variables using iterative gradient-based method. In many applications that require high-resolution solution, such as optimal design, the discretized constraints are high-dimensional, which can easily have millions or even billions of variables. When solving such problems, each iteration can take hours or even days to complete, and solving the entire problem typically takes hundreds of iterations [20]. It is of important practical value to speed up large-scale PDE-constrained optimization.

Recent works [21]–[24] have attempted to accelerate the optimal structure design by directly predicting the final design using deep learning models. These iteration-free approaches require to collect plenty of training data beforehand, and each training sample is obtained by completely solving a problem with the standard optimizer. These previous works have severe limitations in four aspects. First, collecting sufficient optimization results as training samples is computationally prohibitive, especially for large-scale 3D design problems which can take days to generate a single training sample. Second, due to the fixed dimensionality of training data, their trained models are not able to be generalized to solve new design problems with different design domains. Third, these approaches are not scalable due to the capacity limit of deep learning models. These works only consider 2D or tiny-scale 3D problems with less than 10,000 design variables. Lastly, these approaches are not accurate enough and the predicted final designs tend to have flaws like hanging or disconnected members. This happens because the above-mentioned approaches fail to embed underlying physical laws during the offline learning. Due to these limitations, a general framework for speeding up PDE-constrained optimization that can be accurate, scalable, efficient and generalizable at the same time is much needed.

To address the above challenges, we propose an online neural synthetic gradients (ONSG) framework. Our method is involved in the optimization process, and employs the optimization history to train a deep learning model in online fashion. The deep learning model can generate synthetic gradients to drive the optimization process. Synthetic gradients are much cheaper to obtain, and can replace exact gradients to skip the expensive solve. During the optimization, ONSG switches between applying exact and synthetic gradients. Since the model is trained online, our method is generalizable and

can be directly applied to new problems without any pre-collected training data. Moreover, in ONSG, we also propose a novel two-scale optimization scheme, which decomposes the design variables into local subsets but still incorporates global information. With the two-scale scheme, the ONSG can deal with large-scale optimization problems with millions of design variables. We experiment on challenging 3D optimal design problems with more than one million design variables. Our method shows superior performance in terms of accuracy and scalability, which achieves up to 7.5x speedup over the standard optimizer with comparable or even better final objective achieved.

By designing ONSG, we make three key contributions:

- We propose a general online learning framework named ONSG for speeding up PDE-constrained optimization. By interacting with the optimization process, our method can be directly applied to new problems without requiring to pre-collect the training data.
- We propose a two-scale optimization scheme to decompose the training data into local subsets and incorporate global information. With this technique, the dimensionality of training data can be constant, no matter how large-scale the optimization problem is.
- We successfully apply ONSG to computational morphogenesis, which is a challenging and representative application in large-scale PDE-constrained optimization. Our method achieves significant speedup over the standard optimizer with comparable or even better final objective.

## II. RELATED WORK

Solving large-scale PDE-constrained optimization is a computationally intensive task as the computation of exact gradient at each iteration is computationally expensive and, meanwhile, the complete optimization process typically takes hundreds of iterations. Traditionally, to accelerate such large-scale problems, one has to resort to parallel computing [20], [25], [26], advanced iterative solvers [27], [28], or multi-scale and multi-resolution approaches [29]–[31].

More recently, attempts have been made to apply machine learning to accelerate PDE-constrained optimization problems. The surrogate-assisted optimization approaches [32], which use various machine learning models such as Gaussian process [33] and deep neural networks [34] to replace expensive function evaluations, are popular to speed up gradient-free optimization problems. However, those approaches are not suitable for topology optimization problems which typically require gradient-based optimizers [35]. In topology optimization, the dominant idea of the existing attempts roots in an iteration-free and offline approach, that is, using machine learning to replace the optimizer either partially or completely so that, once the models are trained, one could directly employ them to the map initial guesses or partially converged solutions to the final one. This idea has been widely explored in optimal structure design problems. For example, Ulu et al. [36] developed a data-driven approach for predicting optimized topologies under various loading cases.

In this approach, the optimized topologies obtained under a wide range of loading cases are treated as training samples and the a feed-forward neural network is adopted together with the Principal Component Analysis (PCA) to establish a mapping for predicting the optimized topology under a given loading scenario. Sosnovik and Oseledets [21] introduced a deep learning-based framework which uses CNN to predict the final optimal design based on two input images: a partially converged design and its change with respect to the previous step. By doing so, the total number of optimization steps are reduced. Later, Banga et al. [23] took a similar idea and extended the work of Sosnovik and Oseledets to 3D design problems and to incorporate additional inputs such as external loads and boundary conditions. More recently, Yu et al. [22] proposed a two-stage prediction procedure to produce nearly-optimal structural design without iterations. When given a design domain, load and boundary condition, the procedure first uses a trained CNN-based encoder and decoder to predict a low resolution structural design and, then, the predicted low-resolution design is mapped to a high resolution one using a trained Generative Adversarial Network (GAN). All the above-mentioned frameworks are proposed in the context of density-based topology optimization approach. With the same goal, Lei et al. [37] developed a machine learning-based framework for the Moving Morphable Component (MMC) approach. The framework first uses the MMC approach to generate a set of training samples with different external loads locations and subsequently apply either Supported Vector Regression (SVR) or K-Nearest Neighbors (KNN) to establish an instantaneous mapping between the design parameters and optimized topology. Although achieving partial successes on small- to medium-scale problems, these iteration-free and offline approaches have severe limitations which prevent their success in large-scale problems: 1) they are not *efficient* due to the requirement of collecting a lot of optimization results as training samples; 2) they are not *generalizable* to solve new problems; 3) they are not *scalable* due to the capacity limit of deep learning models; 4) they are not *accurate* and the predicted solution tend to have flaws.

## III. METHODOLOGY

In this section, we first introduce the background of PDE-constrained optimization problems. We then present the online neural synthetic gradients (ONSG) framework to speed up PDE-constrained optimization, which includes an online gradients learning method and a two-scale optimization scheme. As an algorithmic overview, the procedure of speeding up PDE-constrained optimization with ONSG is summarized in Algorithm 1.

### A. Problem Background

The general form of PDE-constrained optimization can be expressed as

$$\begin{aligned} & \min_{\mathbf{z}, \mathbf{u}} \mathcal{J}(\mathbf{z}, \mathbf{u}) \\ & \text{subject to } \mathbf{r}(\mathbf{z}, \mathbf{u}) = \mathbf{0}, \end{aligned} \quad (1)$$

where  $\mathbf{z} \in \mathbb{R}^{N_d}$  and  $\mathbf{u} \in \mathbb{R}^{N_s}$  are the design variables and the state variables, respectively;  $\mathcal{J}(\cdot)$  is the objective function;  $\mathbf{r}(\mathbf{z}, \mathbf{u}) = \mathbf{0}$  are the discretized PDE constraints. Due to the infinite-dimensional nature of the governing PDE, the standard discretize-then-optimize approach first discretizes it into finite-dimensional residual form using the finite element method [38], [39]. The discretization results in a system of equations denoted as  $\mathbf{r}(\mathbf{z}, \mathbf{u}) = \mathbf{0}$ , which can be either linear or nonlinear in terms of the state variables  $\mathbf{u}$  depending on the problem.

To solve the optimization problem above, the Nested Analysis and Design (NAND) approach is typically used [2], [3], which treats the state variables  $\mathbf{u}$  as an implicit function of the design variables  $\mathbf{z}$ , namely  $\mathbf{u}(\mathbf{z})$ . Thus the original optimization problem in Eq. (1) is recast as

$$\min_{\mathbf{z}} \mathcal{J}(\mathbf{z}, \mathbf{u}(\mathbf{z})). \quad (2)$$

The NAND approach requires to solve the discretized PDE constraints  $\mathbf{r}(\mathbf{z}, \mathbf{u}) = \mathbf{0}$  for the design variables  $\mathbf{z}$  at every iteration in the optimization process. In many industry-scale applications, high-resolution solutions are required, which means that the state variables  $\mathbf{u}$  are usually of high dimensionality, e.g., in the order of million or even higher. If we denote the design and state variables at optimization iteration  $k$  as  $\mathbf{z}^{(k)}$  and  $\mathbf{u}^{(k)}$ , respectively, the gradients of the objective function  $\mathcal{J}(\cdot)$  at that iteration is given by

$$\begin{aligned} \mathbf{g}^{(k)} &= \left. \frac{\partial \mathcal{J}}{\partial \mathbf{z}}(\mathbf{z}, \mathbf{u}(\mathbf{z})) \right|_{\mathbf{z}=\mathbf{z}^{(k)}} = \left. \frac{\partial \mathcal{J}}{\partial \mathbf{z}}(\mathbf{z}, \mathbf{u}) \right|_{\substack{\mathbf{z}=\mathbf{z}^{(k)} \\ \mathbf{u}=\mathbf{u}^{(k)}}} \\ &- \left[ \left( \left. \frac{\partial \mathbf{r}}{\partial \mathbf{u}}(\mathbf{z}, \mathbf{u}) \right|_{\substack{\mathbf{z}=\mathbf{z}^{(k)} \\ \mathbf{u}=\mathbf{u}^{(k)}}} \right)^{-\top} \left. \frac{\partial \mathcal{J}}{\partial \mathbf{u}}(\mathbf{z}, \mathbf{u}) \right|_{\substack{\mathbf{z}=\mathbf{z}^{(k)} \\ \mathbf{u}=\mathbf{u}^{(k)}}} \right] \left. \frac{\partial \mathbf{r}}{\partial \mathbf{z}}(\mathbf{z}, \mathbf{u}) \right|_{\substack{\mathbf{z}=\mathbf{z}^{(k)} \\ \mathbf{u}=\mathbf{u}^{(k)}}}. \end{aligned} \quad (3)$$

As shown in the equation above, to compute the gradients  $\mathbf{g}^{(k)}$ , one needs to solve  $\mathbf{r}(\mathbf{z}^{(k)}, \mathbf{u}) = \mathbf{0}$  for  $\mathbf{u}^{(k)}$ , and compute  $(\partial \mathbf{r} / \partial \mathbf{u})^{-\top} (\partial \mathcal{J} / \partial \mathbf{u})$ , both of which are very computationally expensive in large-scale optimization problems.

### B. Online Gradients Learning

To address all limitations of the iteration-free and offline approaches dominated in the literature, we take a completely different path from previous methods. Instead of collecting data beforehand and training model offline, we employ the optimization history as data, and train the deep learning model online. The deep learning model can quickly generate synthetic gradients on-the-fly, and drive the optimization forward. As illustrated in Algorithm 1, ONSG switches between applying exact and synthetic gradients. For iterations using exact gradients, ONSG collects data as training samples and updates the deep learning model. For iterations using synthetic gradients as a shortcut, ONSG skips the expensive solve of exact gradients. There are various ways to design  $\mathcal{I}_s$ , i.e., which iterations are to apply synthetic gradients. We adopt a relatively simple strategy: use exact gradients for the initial  $N_I$

---

### Algorithm 1: Online Neural Synthetic Gradients

---

**input** : Initialized design variables  $\mathbf{z}^{(0)}$ ; original discretized PDE constraints  $\mathbf{r}(\mathbf{z}, \mathbf{u}) = \mathbf{0}$ ; coarse-scale discretized PDE constraints  $\mathbf{r}_C(\mathbf{z}_C, \mathbf{u}_C) = \mathbf{0}$ ; objective function  $\mathcal{J}(\mathbf{z}, \mathbf{u})$ ; total number of iterations  $K$ ; iterations applying synthetic gradients  $\mathcal{I}_s$ ; deep learning model  $\mathcal{M}$

**output**: Final solution of design variables  $\mathbf{z}^{(K)}$

```

1  $\mathcal{D}_{train} \leftarrow \emptyset$ 
2 for  $k = 0$  to  $K$  do
3    $\mathbf{z}_C^{(k)} \leftarrow$  coarse-grain design variables  $\mathbf{z}^{(k)}$ 
4    $\mathbf{u}_C^{(k)} \leftarrow$  solve  $\mathbf{r}_C(\mathbf{z}_C^{(k)}, \mathbf{u}_C) = \mathbf{0}$ 
5   if  $k \in \mathcal{I}_s$  then /* apply synthetic gradients */
6      $\tilde{\mathbf{g}}^{(k)} \leftarrow$  get synthetic gradients predicted by  $\mathcal{M}$ 
7   else /* apply exact gradients */
8      $\mathbf{u}^{(k)} \leftarrow$  solve  $\mathbf{r}(\mathbf{z}^{(k)}, \mathbf{u}) = \mathbf{0}$ 
9      $\mathbf{g}^{(k)} \leftarrow$  get exact gradients with  $\mathbf{u}^{(k)}$  using (3)
10     $\mathcal{D}_{train} \leftarrow \mathcal{D}_{train} \cup \{(\mathbf{z}^{(k)}, \mathbf{u}_C^{(k)}, \mathbf{g}^{(k)})\}$ 
11    update model  $\mathcal{M}$  with training data  $\mathcal{D}_{train}$ 
12   $\mathbf{z}^{(k+1)} \leftarrow$  update  $\mathbf{z}^{(k)}$  with  $\mathbf{g}^{(k)}$  or  $\tilde{\mathbf{g}}^{(k)}$ 

```

---

iterations to train a decent deep learning model, and afterwards use exact gradients every  $N_F$  iterations. The exact gradients serve as the supervision signal to train the deep learning model. Since the model is trained online, the inaccurate prediction at the beginning can be corrected in later iterations. Also, our framework can be directly plugged into new optimization problems, and has no need of any pre-collected training data. Therefore, ONSG is both *efficient* and *generalizable*. We will explain why ONSG is *scalable* in the next section, and show that ONSG is *accurate* in the section of experiments.

### C. Two-scale Optimization

As discussed before, large-scale PDE-constrained optimization problems can easily have millions or even billions of variables in many industry applications. It is very challenging to build a scalable online deep learning model for such problems. The reasons are three folds: 1) as the number of variables increases, the dimensionality of the input and output grows, so that learning the mapping from input to output becomes much more difficult due to the limited model capacity; 2) such increase of learning difficulty may be alleviated in small-scale problems by providing more supervision, i.e., collecting exact gradients in more iterations for model training. However, for large-scale problems, evaluating exact gradients is very time-consuming, and it can significantly hurt the speedup performance if more supervision is provided; 3) the size of deep learning model grows together with the dimensionality of input and output, which costs more GPU memory for training and predicting. Therefore, the maximum GPU memory available bounds the scale of the optimization problem that can be handled.

To address the challenges above in the single-scale deep learning model, we propose a novel two-scale optimization scheme, which decomposes the training data into local subsets but still incorporates global information. More specifically, we segment the design variables  $\mathbf{z}$  into small subsets by grouping them based on their physical locations in the design space, and each subset serves as one training sample. Based on the segmentation, we also introduce the corresponding coarse-grained design variable vector  $\mathbf{z}_C$  so that each of its component is the averaged value of  $\mathbf{z}$  in each subset. Since we are optimizing all design variables together, it is not feasible to build a deep learning model using only local information within each subset. To incorporate global information, we introduce coarse-scale discretized PDE constraints  $\mathbf{r}_C(\mathbf{z}_C, \mathbf{u}_C) = \mathbf{0}$  and solve it for  $\mathbf{u}_C$ , which is used as additional features for each subset of  $\mathbf{z}$ . In this way, we maintain a constant dimensionality of the input and output spaces, thus a constant number of model parameters, no matter how large-scale the optimization problem is. In our experiments, we quantitatively show the scalability of the two-scale optimization scheme.

#### D. ONSG for Computational Morphogenesis

In this section, we showcase the application of ONSG on a challenging and representative class of PDE-constrained optimization problems: the optimal structure design [20], [40]. The goal of the optimal structural design is to find the best structure that can sustain a given load. The corresponding PDE-constrained optimization formulation is stated as [40]:

$$\begin{aligned} \min_{\mathbf{z}} \quad & \mathcal{J}(\mathbf{z}, \mathbf{u}(\mathbf{z})) = \mathbf{f}^\top \mathbf{u}(\mathbf{z}) \\ \text{subject to} \quad & \mathcal{V}(\mathbf{z}) = \mathbf{v}^\top \mathbf{P}\mathbf{z} - V_{\max} \leq 0 \\ & 0 \leq z_i \leq 1 \quad \forall i \in \{1, \dots, N_d\}, \end{aligned} \quad (4)$$

where  $\mathbf{v}$  is a constant vector collecting element volumes,  $\mathbf{P}$  is the density filter matrix [40] and  $V_{\max}$  is the maximum volume imposed on the design. In the about NAND formulation, the implicit function  $\mathbf{u}(\mathbf{z})$  is defined through the discretized PDE constraints  $\mathbf{r}(\mathbf{z}, \mathbf{u}) = \mathbf{K}(\mathbf{z})\mathbf{u} - \mathbf{f} = \mathbf{0}$ , where  $\mathbf{K}$  and  $\mathbf{f}$  are the stiffness matrix and the global force vector, respectively. According to Eq. (3), the gradients of the objective function  $\mathcal{J}(\cdot)$  at iteration  $k$  is computed as

$$g_i^{(k)} = -(\mathbf{u}^{(k)})^\top \frac{\partial \mathbf{K}}{\partial z_i}(\mathbf{z}^{(k)}) \mathbf{u}^{(k)}, \quad (5)$$

where  $g_i^{(k)}$  is the  $i$ -th component of  $\mathbf{g}^{(k)}$ . We emphasize that, to compute the gradient  $\mathbf{g}^{(k)}$ , one needs to first solve  $\mathbf{u}^{(k)}$  through  $\mathbf{K}(\mathbf{z}^{(k)})\mathbf{u} - \mathbf{f} = \mathbf{0}$ , which is very computationally expensive for large-scale optimization problems. Iterative solvers such as preconditioned conjugate gradient (PCG) and multi-grid methods are typically used. On the other hand, the gradient of the volume constraint function  $\mathcal{V}$  is simply given by  $\partial \mathcal{V} / \partial \mathbf{z} = \mathbf{P}^\top \mathbf{v}$ , which does not require expensive computation. Once the gradients are obtained, we use the optimality criteria (OC) method [40], which the standard optimizer for (4), to update design variables  $\mathbf{z}$ .

To apply the two-scale optimization scheme to computational morphogenesis, we segment the design variables  $\mathbf{z}$  into local subsets (3D blocks) of uniform dimensions  $N_B \times N_B \times N_B$ , where  $N_B$  is the block size, as illustrated in Figure 1. The resulting coarse-scale discretized PDE constraints become  $\mathbf{r}_C(\mathbf{z}_C, \mathbf{u}_C) = \mathbf{K}_C(\mathbf{z}_C)\mathbf{u}_C - \mathbf{f}_C = \mathbf{0}$ , where  $\mathbf{K}_C$  and  $\mathbf{f}_C$  are the coarse-scale stiffness matrix and the force vector, respectively.

#### E. Model Architecture

We explored several different architectures for the deep learning model, including fully-connected neural networks, convolutional neural networks and its variants residual networks [41]. We finally choose fully-connected DenseNet [42] since it performs the best on various design domains. Our model connects all layers in a fully-connected neural network and preserve the feed-forward nature. To be more specific, each layer obtains additional inputs from all preceding layers and passes on its own output to all subsequent layers. We concatenate the incoming features to serve as input of each layer. Our model is different from DenseNet: we use fully-connected layers while DenseNet uses convolutional layers. The hyperparameters of our model, such as the number of layers, learning rate and hidden layer size, are provided in the experiment section.

## IV. EXPERIMENTS

In this section, we conduct extensive experiments to demonstrate the effectiveness of ONSG in speeding up large-scale PDE-constrained optimization problems. We first show the scalability of the proposed ONSG framework by increasing the scale of the optimization problem. Then, we study the impact of an important ONSG hyperparameter, the block size  $N_B$ , which can further speedup the optimization process. We evaluate ONSG by tracking the objective achieved along with the wall-clock time. Also, we plot the final solution for each benchmark problem. For all the experiments, we fairly compare ONSG with the standard optimizer, running with exactly the same computational resources.

**Benchmark problems.** We aim to construct a challenging benchmark, and choose two large-scale PDE-constrained optimization problems in optimal structure design: 1) a cantilever beam design problem, whose dimensions and boundary conditions are illustrated in Figure 2(a). The design domain is fixed on its face  $x = 0$  and subjected to a distributed load  $\tau = 1$  in the negative  $z$  direction at the lower edge of the face  $x = 2$ . We construct three fine-scale meshes (Mesh 1–3) from lowest to highest level of refinement. The volume fraction is  $V_{\max} = 12\%$ , and the density filter radius is  $R = 0.08$ ; 2) a Messerschmitt–Bölkow–Blohm (MBB) beam design problem, whose dimensions and boundary conditions are illustrated in Figure 2(b). A downward force  $F = 1$  is applied on the center of the top surface of the design domain. We construct a highly refined mesh (Mesh 4) for this problem. The volume fraction and the density filter radius are set to  $V_{\max} = 12\%$  and  $R = 0.08$ , respectively. Both are large-scale optimal design

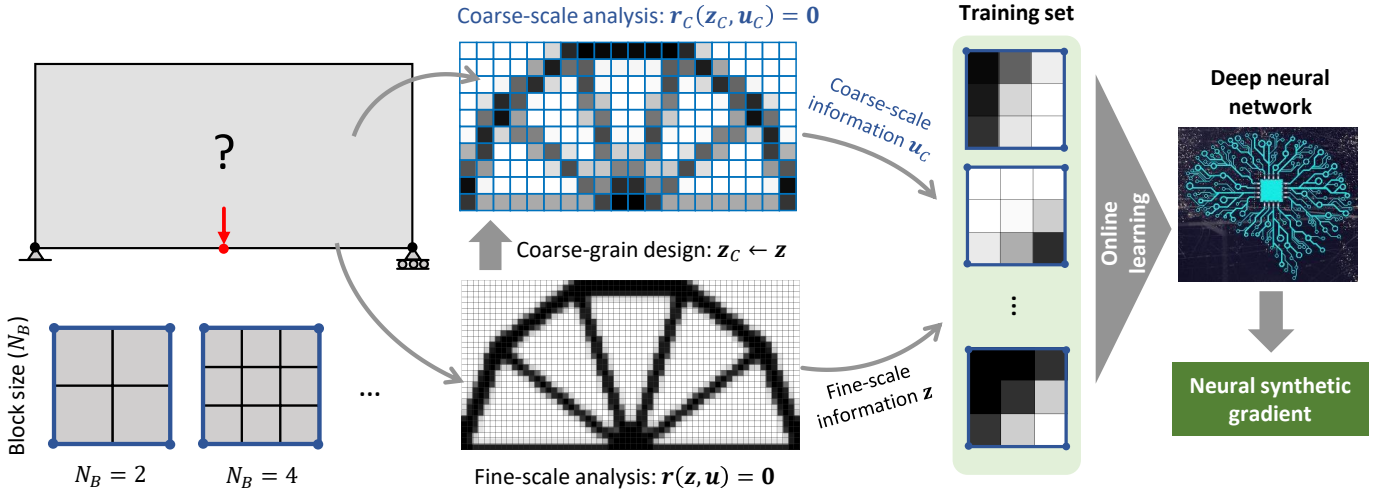


Fig. 1. Illustration of our two-scale optimization scheme, which groups the design variables into a coarse-scale mesh based on their physical locations in the design space. For the computational morphogenesis problem, we incorporate both global and local information from the coarse-scale and fine-scale mesh to provide input for training the deep learning (DL) model to generate synthetic gradients. As illustrated in the bottom left corner, the block size  $N_B$  determines the ratio between the side of an element of the fine-scale mesh to the side of an element of the coarse-scale mesh.

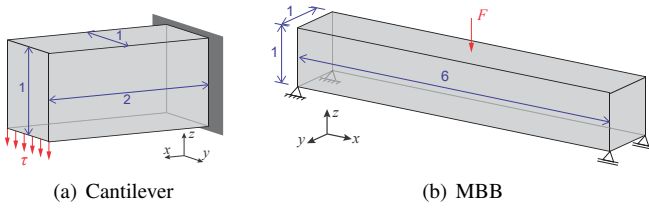


Fig. 2. Illustration of the design domains of two benchmark problems.

TABLE I  
STATISTICS OF THE BENCHMARK PROBLEMS.

|                    | Cantilever |        |        | MBB       |           |           |
|--------------------|------------|--------|--------|-----------|-----------|-----------|
|                    | Mesh 1     | Mesh 2 | Mesh 3 | Mesh 4    |           |           |
|                    | $N_B = 5$  |        |        | $N_B = 2$ | $N_B = 4$ | $N_B = 8$ |
| # DVs              | 86K        | 250K   | 1.5M   | 1.4M      | 1.4M      | 1.4M      |
| $\mathbf{K}$ dim   | 276K       | 788K   | 4.5M   | 4.1M      | 4.1M      | 4.1M      |
| $\mathbf{K}_C$ dim | 3K         | 8K     | 40K    | 71K       | 10K       | 3K        |

problems in 3D, and their refined meshes have more than one million design variables. One can refer to Table I for the exact number of design variables (# DVs) and the dimension of the stiffness matrix  $\mathbf{K}$  of each mesh.

**Competitor method.** We compare our ONSG method with the standard optimizer in the literature of large-scale structural design [20], [26]. Note that the standard optimizer also employs GPUs to solve the discretized PDE constraints, which is the most computationally intensive part in the optimization process. The standard optimizer uses the PCG method with the Jacobi preconditioner [43], which is shown to be the state of the art in terms of efficiency [44]. We use the official GPU parallel computing implementation of the PCG method in Matlab.

**Experimental settings.** All the experiments are conducted on

a GPU-enabled (Nvidia Titan Xp) Linux machine powered by Intel Xeon Silver 4116 processors at 2.10GHz with 256GB RAM. The GPU memory is 12GB. To ensure fair comparison, we allocate exactly the same computational resources (CPU, GPU and memory) for all the experiments. For ONSG experiments, we fix the block size  $N_B = 5$  on three Cantilever meshes, and have different block sizes  $N_B = 2, 4, 8$  on the MBB mesh. For the deep learning model in ONSG, we use a fully-connected neural network with 4 hidden layers. The size of each hidden layer is set to 1000 for the Cantilever experiments, and set to 200, 500, 1000 for the MBB experiments with  $N_B = 2, 4, 8$  respectively. The learning rate is set to 0.0005 initially, and decays by half every 500 iterations. We train the neural network for 2000 iterations for the Cantilever experiments, and 3000 iterations for the MBB experiments. We use  $N_I = 10$  initial iterations with exact gradients. We fix the online update interval  $N_F = 10$  for the Cantilever problems, and set  $N_F = 50, 45, 10$  for the MBB experiments with  $N_B = 2, 4, 8$  respectively. We set the total number of iterations  $K = 200$  for all the experiments.

**Evaluation metrics.** We use the optimization objective to evaluate the model accuracy. Since we are minimizing the compliance, the objective is the lower the better. To evaluate the model efficiency, we time each experiment and report the wall-clock time in seconds. We also track the intermediate objective during the optimization process, and plot the objective curve along with the wall-clock time. With the plot, we compare the objective achieved by different methods given certain amount of time. Also, we compare the time cost of different methods to achieve a certain objective value.

#### A. Model Accuracy

As stated in Eq. (4), the goal of the design problems is to minimize the compliance of the structure, which is the objective of the optimization process. Thus, the objective is the

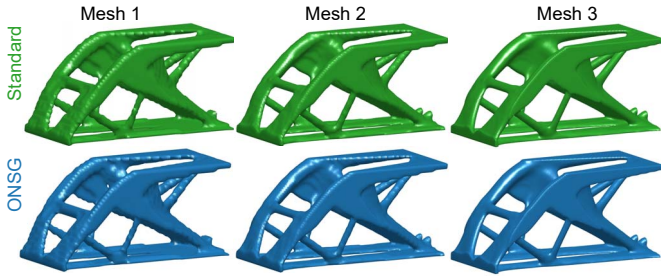


Fig. 3. Final Cantilever designs obtained by the standard optimizer and our method.

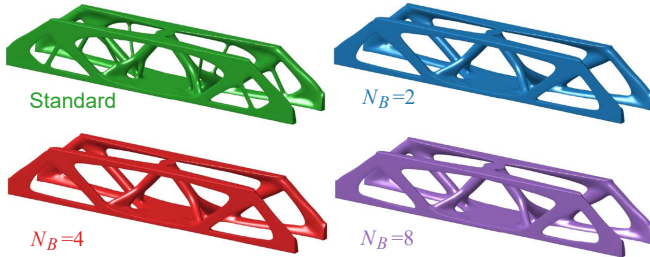


Fig. 4. Final MBB designs obtained by the standard optimizer and our method.

lower the better. Our method uses the online neural synthetic gradients to replace the standard gradients, which will guide the optimization to a different path. So we first check the accuracy of our method, i.e., how close it is to the standard optimizer in terms of the final objective achieved. We report the experimental results in Table II. For all the cases, ONSG achieves comparable or even better objective, compared to the standard optimizer. Since the optimization problem is highly non-convex, ONSG is able to drive the optimization to a better local minima and beat the standard optimizer to get negative difference in many cases.

In Figure 3 and 4, we visualize the final designs obtained by the standard optimizer and our method. For the Cantilever design problem, there is no visual difference observed between our design and the standard one, and both get smoother and higher-quality designs on finer meshes. For the MBB design problem, there is minor difference in our designs compared to the standard one. The standard design is more complex, which has several small members. According to the final objectives in Table II, our method actually finds better designs with lower compliance values.

The experimental results above demonstrate that ONSG has comparable or even better performance compared to the standard optimizer.

### B. Model Scalability

ONSG employs the two-scale optimization scheme to decomposes the global structure into local ones, which is important for solving large-scale optimal design problems. To evaluate the scalability of ONSG, and quantitatively show the impact of the two-scale scheme, we record the number of model parameters and the corresponding GPU memory cost

TABLE II  
FINAL OBJECTIVE ACHIEVED BY THE STANDARD OPTIMIZER AND OUR METHOD ON TWO BENCHMARK PROBLEMS. THE OBJECTIVE IS THE LOWER THE BETTER.

|            | Cantilever |        |        | MBB       |           |           |
|------------|------------|--------|--------|-----------|-----------|-----------|
|            | Mesh 1     | Mesh 2 | Mesh 3 | Mesh 4    |           |           |
|            | $N_B = 5$  |        |        | $N_B = 2$ | $N_B = 4$ | $N_B = 8$ |
| Standard   | 370.11     | 379.54 | 383.51 | 221.30    | 221.30    | 221.30    |
| ONSG       | 368.61     | 379.37 | 383.75 | 220.28    | 220.52    | 220.92    |
| Difference | -0.40%     | -0.05% | 0.06%  | -0.46%    | -0.35%    | -0.17%    |

TABLE III  
NUMBER OF MODEL PARAMETERS AND THE CORRESPONDING GPU MEMORY COST OF THE CANTILEVER DESIGN PROBLEM. A HYPHEN INDICATES RUNNING OUT OF GPU MEMORY.

| Scheme       |            | Mesh 1 | Mesh 2 | Mesh 3 |
|--------------|------------|--------|--------|--------|
| Single-scale | # params   | 175M   | 867M   | 3B     |
|              | GPU memory | 3.9GB  | -      | -      |
| Two-scale    | # params   | 3.3M   | 3.3M   | 3.3M   |
|              | GPU memory | 0.7GB  | 0.7GB  | 0.9GB  |

of the Cantilever design problem in Table III. For the single-scale model, as we increase the scale of mesh from Mesh 1 to Mesh 3, the number of parameters in the deep learning model increases drastically and so does the GPU memory required for training the model. Actually the single-scale model runs out of GPU memory on Mesh 2 and Mesh 3. On the contrary, ONSG with the two-scale scheme maintains a constant number of parameters on all meshes. The GPU memory cost is much less compared to the single-scale model, and has only increased by a modest margin, which is caused by more training data being cached on GPU.

The two-scale scheme requires to solve the coarse-scale PDE constraints with additional computational cost. In fact, this extra cost is almost negligible since the dimensionality of the coarse-scale stiffness matrix  $\mathbf{K}_C$  is cubically smaller than the original stiffness matrix  $\mathbf{K}$ , as listed in Table I. Therefore, with this small extra cost, the two-scale scheme has much better scalability and can easily handle large-scale optimization problems on a single GPU.

### C. Model Efficiency

To evaluate the model efficiency, we time the optimization process for both the standard optimizer and our method, and compare their time cost. We first study the performance on the Cantilever design problem. For Mesh 1–3, the wall-clock time is reported in Figure 6(a), and the objective curves along with the wall-clock time are plotted in Figure 5(a)–5(c). From these experimental results, we can draw several conclusions: 1) ONSG is significantly faster than the standard optimizer in all cases; 2) ONSG achieves higher speedup on larger scale meshes. On Mesh 3 with over one million design variables, ONSG achieves 4.2x speedup, which is higher than the 2.4x speedup on Mesh 2 and the 1.5x speedup on Mesh 1. For

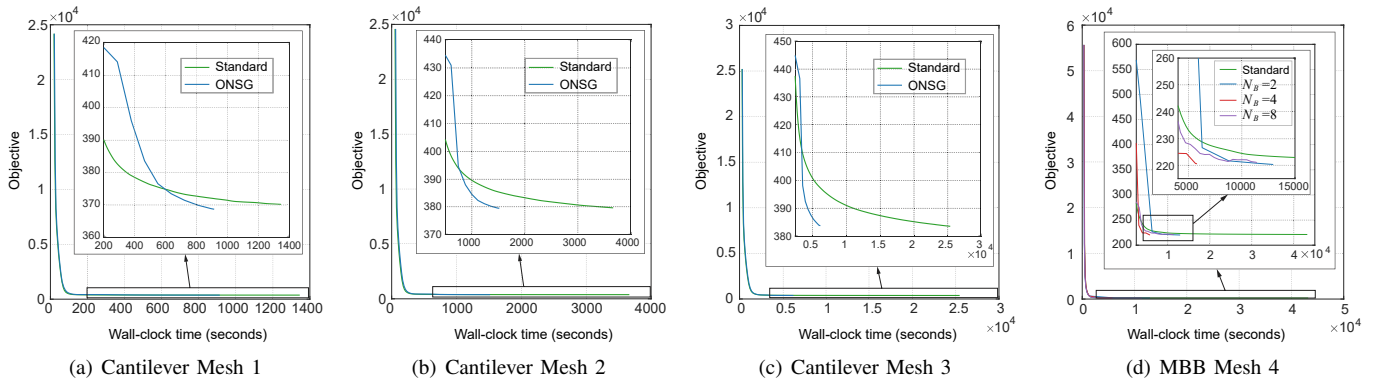


Fig. 5. Objective curves on the Cantilever and MBB design problems. The objective is the lower the better.

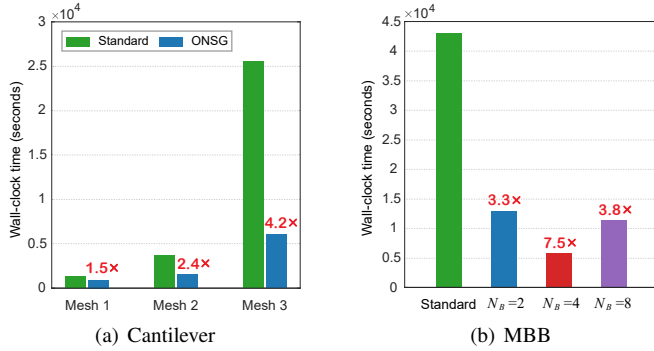


Fig. 6. Wall-clock time of the standard optimizer and our method. The corresponding speedup of our method is labeled in red above the bars. For Cantilever, Mesh 1–3 are from the smallest to the largest scale. For MBB, we experiment on a large-scale mesh (Mesh 4), and vary the block size  $N_B$ .

the standard optimizer, when the number of design variables increases, the dimensionality of the stiffness matrix in the PDE constraints  $\mathbf{K}$  grows drastically (as listed in Table I), so that on large-scale mesh it is very time-consuming to compute the gradients in every iteration. ONSG employs neural synthetic gradients as a shortcut to skip the expensive computation in most iterations, and it saves more time on larger scale meshes; 3) the objective curves in Figure 5(a)–5(c) show that at the beginning ONSG maintains the objective close to the standard one, and then overtakes the standard optimizer and quickly goes down. This is reasonable since the deep learning model has few training data at the beginning, and the synthetic gradients are not that accurate. As the optimization proceeds, more training data is collected to update the model online, so that the synthetic gradients become more accurate and guide the optimization back on track.

#### D. Further Speedup

In previous experiments on the Cantilever design problem, we fix the block size  $N_B = 5$  and our method ONSG achieves up to 4.2x speedup over the standard method. Now we discuss how the block size  $N_B$  can have impact on speedup, and explore whether our method can achieve further speedup by varying  $N_B$ .

In the extreme case, if we set  $N_B = 1$ , there is no difference between the coarse-scale and the original fine-scale PDE constraints, i.e.,  $\mathbf{K} = \mathbf{K}_C$ , which roughly reduces ONSG to the standard optimizer and there should be no speedup anymore. In the other extreme case, if we set  $N_B$  to be large enough to cover the entire mesh (suppose we have enough GPU memory), there is only one block to generate the training data. In this case, the learning becomes more difficult due to the high-dimensional input and the small number of training samples. Thus we have to increase the frequency of supervision for the deep learning model, and use larger deep neural networks for model training, which will definitely diminish the speedup of ONSG. Based on our analysis of the two extremes, there should be a trade-off of the block size  $N_B$  to achieve the best speedup.

We conduct experiment on the MBB design problem by setting  $N_B = 2, 4, 8$  respectively. As reported in Table II, the final objectives achieved are even better than the standard optimizer, showing that ONSG can maintain perfect accuracy with larger block size. The wall-clock time is reported in Figure 6(b). With  $N_B = 4$ , ONSG achieves 7.5x speedup over the standard optimizer, which almost doubles the speedup of other two cases. The corresponding objective curves are plotted in Figure 5(d). We can see that the  $N_B = 2$  curve takes long time to drop, since the coarse-scale PDE constraints are more refined and takes longer to solve. The  $N_B = 4$  curve is at the bottom, which is the fastest among all the curves. These experimental results demonstrate that a good trade-off of the block size  $N_B$  can lead to further speedup of ONSG.

## V. CONCLUSIONS AND FUTURE WORK

We propose a framework named ONSG to speed up computational morphogenesis. ONSG employs the optimization history to train a deep learning model, and obtains online synthetic gradients from the deep learning model to replace exact gradients as a shortcut to save computational time. To enable the framework to handle large-scale optimization problems, we propose a two-scale optimization scheme, which decomposes the design variables into local subsets but still incorporates global information. With the two-scale scheme,

ONSG can efficiently deal with computational morphogenesis problems with millions of design variables. In the future, we plan to formulate other PDE-constrained optimization applications into our framework, such as optimal control and inverse estimation problems.

## REFERENCES

- [1] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*. Springer Science & Business Media, 2008, vol. 23.
- [2] L. T. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, *Large-Scale PDE-Constrained Optimization*. Springer Science & Business Media, 2003, vol. 30.
- [3] M. J. Zahr and C. Farhat, "Progressive construction of a parametric reduced-order model for PDE-constrained optimization," *International Journal for Numerical Methods in Engineering*, vol. 102, no. 5, pp. 1111–1135, 2015.
- [4] M. J. Zahr, D. Amsallem, and C. Farhat, "Construction of parametrically-robust CFD-based reduced-order models for PDE-constrained optimization," in *21st AIAA Computational Fluid Dynamics Conference*, 2013, p. 2845.
- [5] W. Hu, "Approximation and control of the Boussinesq equations with application to control of energy efficient building systems," Ph.D. dissertation, Virginia Tech, 2012.
- [6] A. Carnarius, B. Gunther, F. Thiele, D. Wachsmuth, F. Tröltzsch, and J. C. de los Reyes, "Numerical study of the optimization of separation control," in *45th AIAA Aerospace Sciences Meeting and Exhibit*, 2007, p. 58.
- [7] T. van Leeuwen and F. J. Herrmann, "3D frequency-domain seismic inversion with controlled sloppiness," *SIAM Journal on Scientific Computing*, vol. 36, no. 5, pp. S192–S217, 2014.
- [8] K. Wang, T. Matthews, F. Anis, C. Li, N. Duric, and M. A. Anastasio, "Breast ultrasound computed tomography using waveform inversion with source encoding," in *Medical Imaging 2015: Ultrasonic Imaging and Tomography*, 2015.
- [9] G. S. Abdoulaev, K. Ren, and A. H. Hielscher, "Optical tomography as a PDE-constrained optimization problem," *Inverse Problems*, vol. 21, no. 5, p. 1507, 2005.
- [10] G. Leugering, P. Benner, S. Engell, A. Griewank, H. Harbrecht, M. Hinze, R. Rannacher, and S. Ulbrich, *Trends in PDE constrained optimization*. Springer, 2014, vol. 165.
- [11] H. Antil, R. H. Nochetto, and P. Venegas, "Optimizing the Kelvin force in a moving target subdomain," *Mathematical Models and Methods in Applied Sciences*, vol. 28, no. 01, pp. 95–130, 2018.
- [12] M. Osanov and J. K. Guest, "Topology optimization for architected materials design," *Annual Review of Materials Research*, vol. 46, pp. 211–233, 2016.
- [13] M. D. Gunzburger, *Perspectives in flow control and optimization*. Siam, 2003, vol. 5.
- [14] J.-H. Zhu, W.-H. Zhang, and L. Xia, "Topology optimization in aircraft and aerospace structures design," *Archives of Computational Methods in Engineering*, vol. 23, no. 4, pp. 595–622, 2016.
- [15] L. L. Stromberg, A. Beghini, W. F. Baker, and G. H. Paulino, "Application of layout and topology optimization using pattern gradation for the conceptual design of buildings," *Structural and Multidisciplinary Optimization*, vol. 43, no. 2, pp. 165–180, 2011.
- [16] L. L. Beghini, A. Beghini, N. Katz, W. F. Baker, and G. H. Paulino, "Connecting architecture and engineering through structural topology optimization," *Engineering Structures*, vol. 59, pp. 716–726, 2014.
- [17] L. Wang, P. K. Basu, and J. P. Leiva, "Automobile body reinforcement by finite element optimization," *Finite Elements in Analysis and Design*, vol. 40, no. 8, pp. 879–893, 2004.
- [18] J. Hiller and H. Lipson, "Automatic design and manufacture of soft robots," *IEEE Transactions on Robotics*, vol. 28, no. 2, pp. 457–466, 2012.
- [19] C.-H. Liu and C.-H. Chiu, "Optimal design of a soft robotic gripper with high mechanical advantage for grasping irregular objects," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017.
- [20] N. Aage, E. Andreassen, B. S. Lazarov, and O. Sigmund, "Giga-voxel computational morphogenesis for structural design," *Nature*, vol. 550, no. 7674, p. 84, 2017.
- [21] I. Sosnovik and I. Oseledets, "Neural networks for topology optimization," *arXiv preprint arXiv:1709.09578*, 2017.
- [22] Y. Yu, T. Hur, and J. Jung, "Deep learning for determining a near-optimal topological design without any iteration," *Structural and Multidisciplinary Optimization*, pp. 1–13, 2019.
- [23] S. Banga, H. Gehani, S. Bhilare, S. Patel, and L. Kara, "3D topology optimization using convolutional neural networks," *arXiv preprint arXiv:1808.07440*, 2018.
- [24] Y. Zhang, A. Chen, B. Peng, X. Zhou, and D. Wang, "A deep convolutional neural network for topology optimization with strong generalization ability," *CoRR*, vol. abs/1901.07761, 2019.
- [25] T. Borrvall and J. Petersson, "Large-scale topology optimization in 3D using parallel computing," *Computer Methods in Applied Mechanics and Engineering*, vol. 190, no. 46–47, pp. 6201–6229, 2001.
- [26] N. Aage, E. Andreassen, and B. S. Lazarov, "Topology optimization using PETS: An easy-to-use, fully parallel, open source topology optimization framework," *Structural and Multidisciplinary Optimization*, 2015.
- [27] S. Wang, E. d. Sturler, and G. H. Paulino, "Large-scale topology optimization using preconditioned Krylov subspace methods with recycling," *International Journal for Numerical Methods in Engineering*, vol. 69, no. 12, pp. 2441–2468, 2007.
- [28] O. Amir, N. Aage, and B. S. Lazarov, "On multigrid-CG for efficient topology optimization," *Structural and Multidisciplinary Optimization*, vol. 49, no. 5, pp. 815–829, 2014.
- [29] Y. Y. Kim and G. H. Yoon, "Multi-resolution multi-scale topology optimization—a new paradigm," *International Journal of Solids and Structures*, vol. 37, no. 39, pp. 5529–5559, 2000.
- [30] T. H. Nguyen, G. H. Paulino, J. Song, and C. H. Le, "A computational paradigm for multiresolution topology optimization (MTOP)," *Structural and Multidisciplinary Optimization*, vol. 41, no. 4, pp. 525–539, 2010.
- [31] H. Liu, Y. Wang, H. Zong, and M. Y. Wang, "Efficient structure topology optimization by using the multiscale finite element method," *Structural and Multidisciplinary Optimization*, pp. 1–20, 2018.
- [32] R. T. Haftka, D. Villanueva, and A. Chaudhuri, "Parallel surrogate-assisted global optimization with expensive functions—a survey," *Structural and Multidisciplinary Optimization*, vol. 54, no. 1, pp. 3–13, 2016.
- [33] J. A. Easum, J. Nagar, and D. H. Werner, "Multi-objective surrogate-assisted optimization applied to patch antenna design," in *2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*. IEEE, 2017, pp. 339–340.
- [34] G. Briffoteaux, M. Gobert, R. Ragonnet, J. Gmys, M. Mezmas, N. Melab, and D. Tuytens, "Parallel surrogate-assisted optimization: Batched bayesian neural network-assisted GA versus q-EGO," *Swarm and Evolutionary Computation*, vol. 57, p. 100717, 2020.
- [35] O. Sigmund, "On the usefulness of non-gradient approaches in topology optimization," *Structural and Multidisciplinary Optimization*, vol. 43, no. 5, pp. 589–596, 2011.
- [36] E. Ulu, R. Zhang, and L. B. Kara, "A data-driven investigation and estimation of optimal topologies under variable loading configurations," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 4, no. 2, pp. 61–72, 2016.
- [37] X. Lei, C. Liu, Z. Du, W. Zhang, and X. Guo, "Machine learning-driven real-time topology optimization under moving morphable component-based framework," *Journal of Applied Mechanics*, vol. 86, no. 1, p. 011004, 2019.
- [38] G. Strang and G. J. Fix, *An analysis of the finite element method*. Prentice-hall Englewood Cliffs, NJ, 1973, vol. 212.
- [39] T. J. Hughes, *The finite element method: linear static and dynamic finite element analysis*. Courier Corporation, 2012.
- [40] M. P. Bendsøe and O. Sigmund, *Topology optimization: theory, methods, and applications*. Springer Science & Business Media, 2013.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [42] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [43] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003, vol. 82.
- [44] L. S. Duarte, W. Celes, A. Pereira, I. F. Menezes, and G. H. Paulino, "Polytop++: an efficient alternative for serial and parallel topology optimization on CPUs & GPUs," *Structural and Multidisciplinary Optimization*, vol. 52, no. 5, pp. 845–859, 2015.