Computer Methods in Applied Mechanics and Engineering 437 (2025) 117772

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

Data-driven multifidelity topology design with multi-channel variational auto-encoder for concurrent optimization of multiple design variable fields

Hiroki Kawabe [a,*], Kentaro Yaji [a], Yuichiro Aoki [b]

[a] Department of Mechanical Engineering, Graduate School of Engineering, Osaka University, 2-1, Yamadaoka, Suita, Osaka 565-0871, Japan
[b] Aviation Technology Directorate, Japan Aerospace Exploration Agency, 6-13-1, Mitaka, Tokyo, 181-0015, Japan

ARTICLE INFO

Keywords:
Topology optimization
Data-driven approach
Evolutionary algorithm
Mini-max optimization problem
Multi-channel variational auto-encoder

ABSTRACT

Topology optimization can generate high-performance structures with a high degree of freedom. Regardless, it generally confronts entrapment in undesirable local optima especially in problems characterized by strong non-linearity. This study aims to establish a gradient-free topology optimization framework that facilitates more global solution searches to avoid the entrapment. The framework utilizes a data-driven multifidelity topology design (MFTD), where solution candidates initially generated by solving low-fidelity (LF) optimization problems are iteratively updated by a variational auto-encoder (VAE) and high-fidelity (HF) evaluation. A key procedure of the solution update is to construct HF models by extruding material distributions obtained by the VAE to thickness distribution, which is spcatially constant across all solution candidates in the conventional data-driven MFTD. This constant assignment leads to no exploration of the thickness space, which necessitates extensive parametric studies outside the optimization loop. To enable a more comprehensive optimization in a single run, we propose a multi-channel image data architecture that stores material distributions in the first channel and other design variable fields like thickness distribution in the second or subsequent channels. This significant shift enables a thorough exploration of the additional design variable fields space with no necessity of parametric studies afterwards, by simultaneously optimizing both material distributions and those variable fields. We apply the framework to a maximum stress minimization problem, where the LF optimization problem is formulated with approximation techniques, whereas the HF evaluation is conducted by accurately analyzing the stress field, bypassing any approximation techniques. We first validate that the framework can successfully identify high-performance solutions superior to the reference solutions by effectively exploring both material and thickness distributions in a fundamental stiffness maximization. Then we demonstrate the framework can identify promising solutions for the original maximum stress minimization problems.

1. Introduction

In structural design, there is an increasing demand for designing lightweight structures that are optimized for performances such as stiffness and strength. One of the most successful tools to meet the demand is topology optimization, the concept of which is based on a formulation that shifts the focus from optimizing structural configurations to optimizing material distribution within the design

* Corresponding author.
E-mail address: kawabe@syd.mech.eng.osaka-u.ac.jp (H. Kawabe).

https://doi.org/10.1016/j.cma.2025.117772
Received 9 September 2024; Received in revised form 15 December 2024; Accepted 15 January 2025
Available online 21 January 2025

domain, as introduced by Bendsøe and Kikuchi [1]. This elegantly simple yet effective formulation renders topology optimization sufficiently versatile for addressing a wide range of design problems, such as structural design and thermal-fluid design [2,3].

In practical use, topology optimization has been predominantly utilized for stiffness maximization problems that are one of the most fundamental optimization problems in structural design [4]. Despite its widespread application in such contexts, topology optimization has been less frequently applied to problems characterized by strong non-linearity, such as maximum stress minimization and buckling load maximization, compared to its use in stiffness maximization. Notably, maximum stress minimization is crucial for designing structures within industrial applications, where failure criteria are primarily stress-based. Consequently, numerous researchers are actively tackling the challenges associated with maximum stress minimization problems [5–7].

Maximum stress minimization has a notable challenge called "singularity phenomenon". This challenge arises when material removal leads to rapid increases in stress, causing numerical instabilities. To mitigate this, traditional strategies have adopted stress-relaxation techniques. Moreover, because discrete maximum stress values are impractical for optimization, these methods often resort to approximation techniques, substituting them with more continuous functions. These approximations, however, introduce strong non-linearities, and in the prevalent approach of gradient-based topology optimization, there is a tendency to converge to local minima under these conditions. Consequently, traditional approaches, through their reliance on relaxation and approximation techniques, can be viewed as addressing a modified version of the original optimization problem, rather than directly solving itself.

To avoid entrapment in local optima and promote a more global search, gradient-free topology optimization methods, such as Evolutionary Algorithms (EAs) [8], could offer a viable approach for addressing the challenges posed by strong non-linearities, which motivated a number of researchers to explore the application of EAs to topology optimization [9–11]. Although EAs enable a broader exploration of the solution space, the scalability of design variables in EAs is constrained to the order of 10^2 to 10^3, which is markedly less than the capacity of conventional gradient-based topology optimization methods, typically extending to the order of 10^5 to 10^6 [12]. This constraint arises because two primary procedures of EAs, i.e., crossover and mutation, conventionally handle design variables as individual entities with no consideration of their spatial relationships. Handling such a large number of design variables individually leads to the *curse of dimensionality*, where the EA operations require an exponentially increasing computational cost as the number of design variables increases.

As a solution to the scalability issue, Estimation of Distribution Algorithm (EDA) [13] has been proposed to reduce the dimension of design variables by estimating the joint probability distribution of design variables. However, the initial design candidates are randomly generated in general EDAs, which often makes it difficult to identify high-performance designs when the number of design variables is large.

To address the scalability issue of design variables in topology optimization, deep generative models [14] can be employed to reduce the dimension of the design space with a degree of freedom maintained. In the realm of engineering design, data-driven designs based on those deep generative models has been actively investigated recently [15,16]. Guo et al. pioneered a data-driven topology optimization method that employs a Variational Auto-Encoder (VAE) [17] to encode and decode material distributions in the latent space, demonstrating the effectiveness of deep generative models in representing the topologically optimized designs [18]. Oh et al. introduced a Generative Adversarial Network (GAN) [19] to generate diverse material distributions, which successfully identified high-performance structures [20]. Wang et al. utilized a Latent Variable Gaussian Process (LVGP) to embed a mixed-variable inputs consisting of both qualitative microstructure concepts and quantitative microstructure design variables into a continuous and differentiable design space, enabling the multiscale topology optimization [21].

To bypass the *curse of dimensionality* in EA-based topology optimization, a series of researches have focused on a approach to replace the conventional crossover and mutation procedures with EA-like ones. These EA-like operations are designed to inherit their essential concept and be specifically tuned for topology optimization requiring numerous design variables. Yamasaki et al. proposed a data-driven topology design based on a crossover-like operation using a VAE that can compress high-dimensional input data into a low-dimensional latent space and then reconstruct high-dimensional output from this latent space. They demonstrated that deep generative model-based crossover is effective to efficiently generate offspring material distributions that inherit the topological features from the parent ones [22]. Yaji et al. introduced a mutation-like operation that deduces a mutant material distribution by solving topology optimization additionally constrained with referential material distribution in the dataset under optimization [23]. They confirmed that the mutation-like operation is effective to prevent premature convergence in the iterative process through several numerical examples. These methods can be regarded as a type of EDAs because they construct the probabilistic models with elite solutions and generate new solutions based on these models, and this generative process is iteratively conducted. However, the data driven topology design can be distinguished from the conventional EDAs in that the initial design candidates are not randomly generated but are the material distributions derived from the topology optimization.

Moreover, Yaji et al. integrated this EA-based topology optimization method with Multi-Fidelity Topology Design (MFTD), based on the concept of multifidelity method known for its success in indirectly solving the original problem by addressing both low-fidelity (LF) and high-fidelity (HF) problems [24]. In MFTD, the original topology optimization problem is solved through two primary steps: LF optimization and HF evaluation. The beauty of this multifidelity formulation is that a topology optimization problem, analytically or numerically difficult to directly solve, can be indirectly solved by replacing itself with a simplified optimization problem, then ensuring mechanical performance accurately evaluated through the optimization process. The MFTD's effectiveness has been evidently confirmed by a vast array of mechanical applications, such as heat exchangers and redox flow batteries [25–27].

To tackle the strong non-linearity and the inaccurate stress evaluation by approximation techniques in maximum stress minimization problems, Kato et al. applied the data-driven MFTD to a maximum stress minimization problem [28]. They confirmed the accurate stress evaluation and more global solution search of the data-driven MFTD resulted in material distributions superior to those derived from the conventional gradient-based topology optimization. Additionally, Kii et al. proposed a crossover-like

operation called *latent crossover* [29] that enhances the convergence of the data-driven MFTD by generating offspring material distributions with latent vectors sampled using the simplex crossover [30] to inherit the parental characteristics more efficiently than the conventional random latent vectors uniformly sampled from the latent space.

Despite the significant progress in data-driven MFTD for maximum stress minimization, the conventional data-driven MFTD has a limitation in exploring parameters for constructing the HF model. In previous data-driven MFTD approaches, to construct the HF model, the thickness of the material distribution is typically assigned as a constant value across all solution candidates, with no exploration of the thickness value during the optimization process. However, the thickness of the material distribution, we refer to as HF modeling parameters, has a significant impact on structural performance metrics, such as stiffness and stress, for the entire structure. Therefore, exploring the HF parameter space in addition to the material distribution space is crucial for identifying high-performance structures. To explore the HF parameter space, the conventional data-driven MFTD requires extensive parametric studies, which often highly depend on user intuition.

In this study, to expand the optimization search to the HF modeling space, we focus on an architecture of image data used in the data-driven MFTD. The data-driven MFTD conventionally adopts a single-channel image to store material distributions, where HF modeling parameters are exclusively assigned as a constant value across each material distribution. Instead of representing material distribution as a single-channel image, we introduce a multi-channel image where design variables for constructing the HF model are stored in the second or subsequent channels of multi-channel images, with the material distributions maintained in the first channel. This significant shift enables a thorough exploration of the HF modeling parameter space with no necessity of parametric studies afterwards, by simultaneously optimizing both material distributions and HF modeling parameters.

To transition the data architecture in the data-driven MFTD from a single-channel to multi-channel image, we primarily consider two aspects for adaptation: the crossover operation facilitated by VAE and the mutation operation achieved through solving constrained LF optimization. Unlike the original data-driven MFTD, where the VAE's machine learning model architecture consisted simply of dense layers, this study incorporates a convolutional neural network (CNN) to efficiently learn the relationship between material distributions in the first channel and the HF parameters in subsequent channels [31]. For the mutation operation, material distributions are generated in a manner similar to the traditional data-driven MFTD and then stored in the first channel of the multi-channel images. After preparing the first channel, subsequent channels are designated to globally represent the HF parameter space.

The primary goal of this study is to develop an enhanced data-driven MFTD capable of identifying high-performance structures within the HF modeling parameter space in addition to traditional material distribution space, especially for optimization problems characterized by significant non-linearity, such as maximum stress minimization. Initially, we apply this approach to stiffness maximization for a simple reinforced skin structure, where the thickness of the reinforcement plays a crucial role in determining structural performance metrics, such as stiffness and stress. Subsequently, we demonstrate the effectiveness of the proposed framework in addressing maximum stress minimization using the same structural model for the stiffness maximization.

The rest of the paper is organized as follows. In Section 2, we introduce a concept of data-driven MFTD for the simultaneous optimization of material distributions and HF modeling parameters using the multi-channel image, focusing on the significant shift of image data architecture. In Section 3, we describe the brief formulation of the optimization problem discussed in this study, including the LF optimization and HF evaluation. In Section 4, we elaborate the detailed implementation of the proposed framework, including the architecture of the multi-channel VAE and the EA-based optimization process. In Section 5, we examine the effectivenss of the proposed framework through numerical examples of stiffness maximization and maximum stress minimization. Finally, we conclude the paper in Section 6.

2. Framework

To facilitate the simultaneous optimization of the material distributions and HF modeling parameters, we first introduce a concept of MFTD specifically defined for the HF modeling parameter search. A multi-objective topology optimization problem is formulated in general, as follows:

$$\begin{aligned} \underset{\gamma}{\text{minimize}} \quad & [J_1(\gamma), J_2(\gamma), \dots, J_{r_o}(\gamma)] \\ \text{subject to} \quad & G_j(\gamma) \le 0, \quad (j = 1, 2, \dots, r_c), \\ & \gamma(\boldsymbol{x}) = 0 \text{ or } 1, \quad \forall \boldsymbol{x} \in D, \end{aligned} \tag{1}$$

where $J_i(\gamma)$ and $G_j(\gamma)$ represent objective and constraint functionals, respectively. γ denotes the design variable field, which is a binary material density field determined by the spatial position $\boldsymbol{x}$ in the design domain D, where D denotes a predefined two- or three-dimensional design domain. In MFTD, solving this topology optimization problem directly is assumed to be difficult or even impossible in some cases, as it is formulated as a non-linear optimization problem with an enormous number of design variables.

To indirectly tackle the original intricate topology optimization problem, MFTD divides the original optimization problem into two components: a LF topology optimization and a HF evaluation. The LF topology optimization problem is formulated, as follows:

$$\begin{aligned} \underset{\gamma}{\text{minimize}} \quad & [\tilde{J}_1(\gamma^{(k)}, l^{(k)}), \tilde{J}_2(\gamma^{(k)}, l^{(k)}), \dots, \tilde{J}_{r_o}(\gamma^{(k)}, l^{(k)})] \\ \text{subject to} \quad & \tilde{G}_j(\gamma^{(k)}, l^{(k)}) \le 0, \quad (j = 1, 2, \dots, r_c), \\ & 0 \le \gamma^{(k)}(\boldsymbol{x}) \le 1, \quad \forall \boldsymbol{x} \in D, \\ \text{for given} \quad & l^{(k)}, \end{aligned} \tag{2}$$

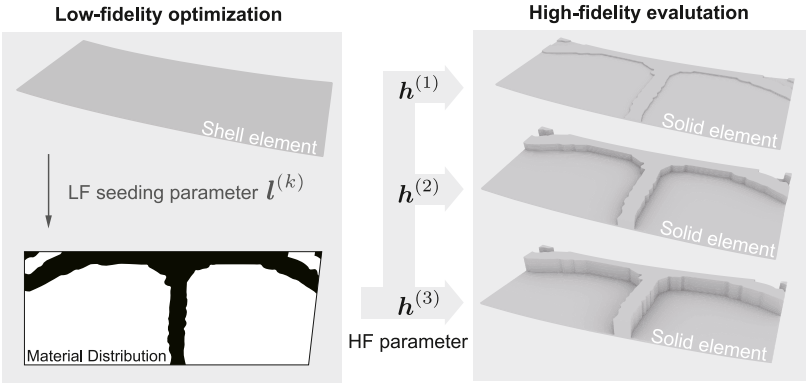

Fig. 1. Illustration of LF and HF parameter concept.

where $\tilde{J}_i$ and $\tilde{G}_j$ represent the objective and constraint functionals for the LF optimization problem, respectively. The $l^{(k)}$ is a scalar value called LF seeding parameter, which is used to generate a variety of design candidates by varying the LF topology optimization problem. The superscript k with $k = 1, 2, \ldots, N_{\text{lf}}^{\text{sd}}$ corresponds to different topology optimizations with respect to each $l^{(k)}$ such as volume fraction. In other words, the optimization problem (2) is solved a total of $N_{\text{lf}}^{\text{sd}}$ times for each time using a different sample point from $l^{(k)}$, resulting in $N_{\text{lf}}^{\text{sd}}$ distinct topologically optimized material distributions $\gamma_{\text{to}}^{(k)}$. Once all candidates from the LF topology optimization problems are collected, the HF evaluation is performed to identify the best solution, denoted as $\gamma_{\text{to}}^{(k^*)}$, where:

$$k^* \in \arg\min_{k} \{[J_1(\gamma_{\text{to}}^{(k)}, h^{(k)}), J_2(\gamma_{\text{to}}^{(k)}, h^{(k)}), \ldots, J_{r_o}(\gamma_{\text{to}}^{(k)}, h^{(k)})] \mid G_j(\gamma_{\text{to}}^{(k)}, h^{(k)}) \leq 0\}. \tag{3}$$

Herein, HF seeding parameters are introduced as an extension to the conventional MFTD. These parameters are represented as a spatially distributed scalar field $h^{(k)}$ for each $\gamma_{\text{to}}^{(k)}$. Each scalar field h_m contains parameter distributions used to configure the Finite Element (FE) models for HF evaluation. The HF seeding parameters can include various configurations, such as thickness distributions for creating three-dimensional FE models by extruding the material distribution in the horizontal plane to a specified height in the vertical direction, as illustrated in Fig. 1. The original objective and constraint functionals, J_i and G_j, are utilized to find the best solutions in the HF evaluation. When $i \geq 2$, J_i exhibits trade-offs with each other, resulting in the best solutions being Pareto optima represented by the vector k^* that contains multiple sample point indices.

The effectiveness of achieving a satisfactory solution set depends on the diversity of candidates collected, considering a range of joint seeding parameters (l, h). Notably, both LF and HF seeding parameters have an impact on structural performance and should be optimized to eventually obtain high-performance structures. The conventional data-driven MFTD focuses on searching for satisfactory solutions in the material distribution space, that is initially determined by LF seeding parameters and automatically updated by the EA-based optimization process, as HF models are evaluated with a consistent configuration applied across all candidates. On the contrary, our proposed framework updates the initial candidates derived from various sets of LF and HF seeding parameters utilizing the EA-based optimization process, similar to the conventional one but specifically adaptive to the data architecture of a multi-channel image.

The overview of our proposed framework is illustrated in Fig. 2. For ease of understanding, the flowchart illustrates a case with three-channel images and specific numbers of solution candidates, which can be adjusted according to the optimization problem. In Fig. 2, soludtion candidates of three-channel images are placed on the left column, with the number of images noted on the top and status of objective functionals on the bottom. The right column represents the primary steps of the optimization process, including the LF topology optimization, HF evaluation, selection, crossover, and mutation.

In this case, the LF topology optimization (2) is performed with 100 different seeding parameters, resulting in 100 different material distributions. Those material distributions are then mapped from the design domain to the two-dimensional image space and stored in the first channel of the images, using a technique called Design Domain Mapping (DDM) [32]. DDM, as proposed by Yamasaki et al. serves to map material distributions from the original domain to another domain while preserving the boundary conditions inherited from the original domain. In the previous data-driven MFTD, DDM was employed to resize the original data to fit within a unit square domain denoted as $\bar{D}$, which was discretized with square elements, i.e., structured quad mesh, for VAE's convenience. While the previous research focused primarily on resizing the mesh to a unit square, we extend its focus to mapping a 3D mesh onto a 2D plane mesh, in addition to mesh resizing. The detailed methodologies behind this dimensional extension is elaborated in Appendix. We assume that the additional design variable fields, such as thickness distributions, are closely related to the material distribution or physical fields of the material distribution, such as displacement and stress distributions. Therefore, the initial design candidates store the material distributions in the first channel and those physical fields in the subsequent channels, in this figure. Note that this initialization of the subsequent channels can vary depending on the optimization problem, but their initial distributions need to closely relate to the corresponding design variables. Fig. 2 on the left column shows the 100 empty images are initialized by LF topology optimization process. At this point, the objective functionals are not evaluated with the status being question marks.

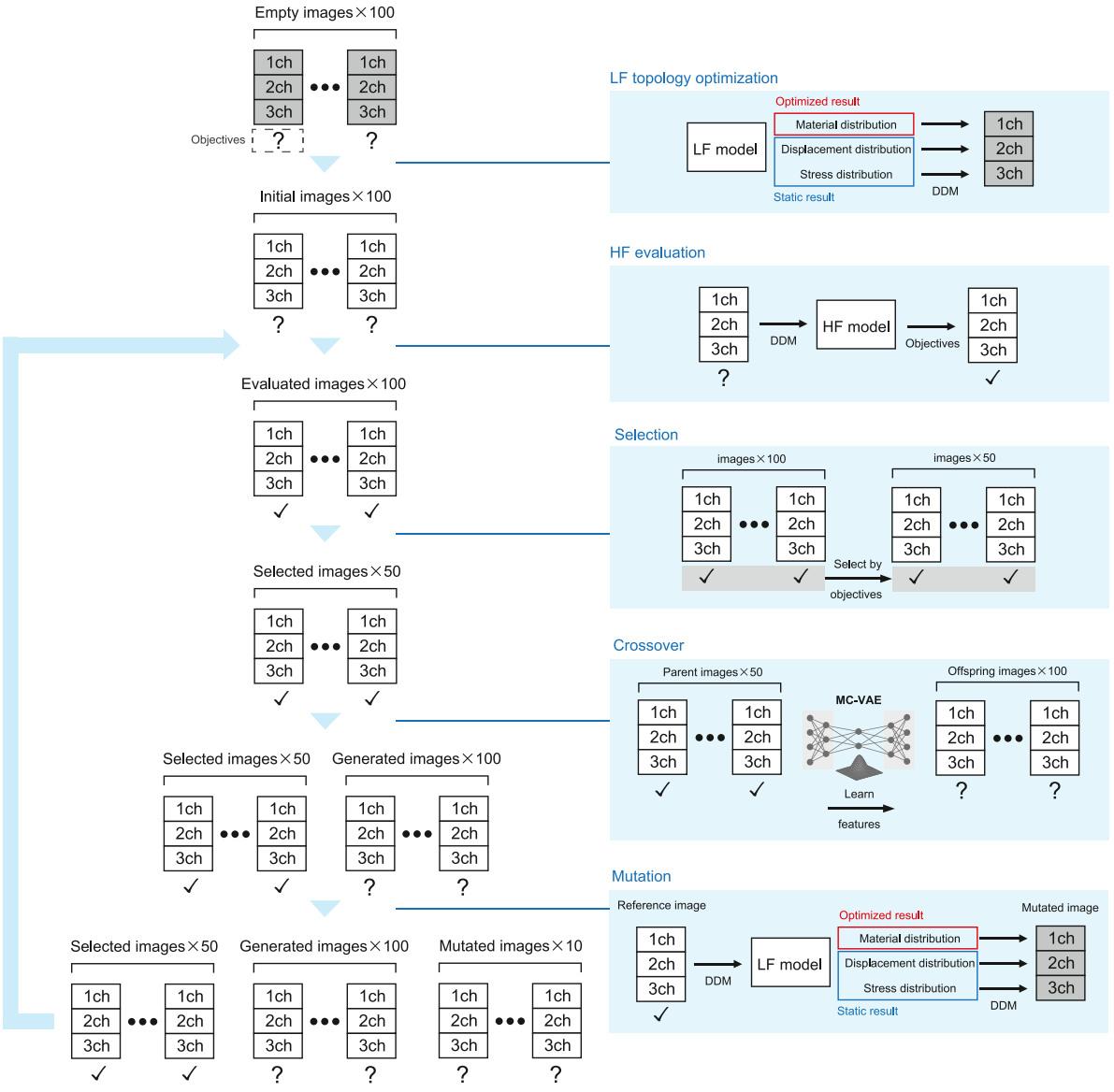

Fig. 2. Schematic flowchart of the proposed data-driven multifidelity topology design.

Following the initial setup, the HF evaluation process is conducted to evaluate the objective functionals of the solution candidates. The HF evaluation process initiates by mapping each channel of multi-channel images to a three-dimensional design domain using the inverse DDM. Then the HF model is constructed based on the mapped design variables, such as reinforcement regions and thickness distributions. The objective functionals are evaluated through forward analysis (e.g., static analysis in this context) to derive precise objective functionals on the constructed HF models. The derived objective functionals are then assigned to the corresponding images, with the status of the objective functionals updated in the left column of Fig. 2.

After accurately evaluating the performance of each solution candidate, the selection process is conducted to select only promising solutions from the solution candidates with non-elite solutions pruned. This process sorts the solution candidates based on the dominance relationship of the corresponding objective functionals, and select a certain number of elite solutions. In the case of Fig. 2, the selection process retains 50 elite solutions, with the number of the images reduced to 50.

For the elite solutions selected in the previous step, the crossover operation is performed to generate offspring solutions. This operation is facilitated by the VAE, which compresses the high-dimensional input data into a low-dimensional latent space and then reconstructs the high-dimensional output from this latent space. Through the data compression, the VAE is known to generate the offspring solutions that mix the characteristics of the parent solutions. Whereas the conventional data-driven MFTD utilized a single-channel VAE (SC-VAE) for candidate generation, our framework employs a multi-channel VAE (MC-VAE) [33] to enable the

neural network of VAE to learn the interactive relationship between multiple design variable fields. To generate offspring solutions that inherit the characteristics of the most promising solutions at each iteration, the MC-VAE is re-trained with the updated elite solutions at each iteration. Then the random latent vectors are sampled using the gaussian distribution, and the offspring solutions are generated by decoding the latent vectors. In the right column of Fig. 2, the crossover operation is performed on the 50 elite solutions, resulting in 100 offspring solutions with the objective functionals remaining unknown. The left column shows the 100 offspring solutions with unknown objective functionals are added to the solution candidates.

The mutation operation is conducted every several iterations to introduce mutants that significantly diverge from the existing dataset yet likely survive the selection process. Several solution candidates are selected as reference solutions, and the material distribution in the first channel of the corresponding multi-channel images is mapped to the design domain. The LF optimization is then performed with the constraints of how divergent the mutants should be from the material distribution of the reference solutions. Similar to the LF optimization, the subsequent channels are initialized with the physical fields of the material distribution, such as displacement and stress distributions. In this study, the initialization of the mutated solutions is kept consistent with the initial setup of the solution candidates. The left column of Fig. 2 shows the 10 mutated solutions are added to the solution candidates, with the objective functionals remaining unknown. After the mutation operation, or after the crossover operation if the mutation operation is not conducted, the HF evaluation is performed to evaluate the objective functionals of the solution candidates. Fig. 2 shows the objective functionals of the 160 solution candidates needs to be updated after the mutation operation. The number of solution candidates can differ depending on the iteration, and the process is repeated until the convergence criterion is met.

The optimization process converges when a hypervolume indicator [34] falls below a predefined threshold or the maximum number of iterations is reached. The hypervolume indicator is calculated for the selected solution set $\gamma_{\mathrm{to}}^{(k^*)}$ using a reference point $r_{\mathrm{hv}} \in \mathbb{R}^{r_o}$, defined as follows :

$$\mathrm{HV}(k^*, r_{\mathrm{hv}}) = \mathcal{L}\left(\bigcup_{k_p \in k^*} \{ J(\gamma_{\mathrm{to}}^{(k)}) \mid J(\gamma_{\mathrm{to}}^{(k_p)}) \leq J(\gamma_{\mathrm{to}}^{(k)}) \leq r_{\mathrm{hv}} \} \right) \tag{4}$$

where the Lebesgue measure, denoted by $\mathcal{L}(\cdot)$, is utilized to quantify the hypervolume indicator. Specifically to bi-objective problems, this indicator measures the area spanned by the non-dominated solutions, identified by k_p, and a reference point r_{hv}.

2.1. Selection

As introduced in the previous data-driven MFTD research [22], the selection of promising solutions is performed using a concept of elitism through the application of the non-dominated sorting genetic algorithm II (NSGA-II) [35], which is a well-established algorithm in the realm of multi-objective EAs. NSGA-II focuses on generating a set of non-dominated solutions, satisfying the following relationship, especially in a multi-objective minimization problem:

$$\forall i : J_i(\gamma_{\mathrm{to}}^{(k_1)}) \leq J_i(\gamma_{\mathrm{to}}^{(k_2)}) \quad \wedge \quad \exists j : J_j(\gamma_{\mathrm{to}}^{(k_1)}) < J_j(\gamma_{\mathrm{to}}^{(k_2)}). \tag{5}$$

This relationship signifies that a solution $\gamma_{\mathrm{to}}^{(k_1)}$ is not dominated by another solution $\gamma_{\mathrm{to}}^{(k_2)}$, meaning the k_1-th solution is considered superior to the k_2-th one.

NSGA-II selects solutions using non-dominated sorting and crowding distance sorting. In the non-dominated sorting phase, a set of solutions is categorized into different fronts based on the dominance relationship (5). The most dominant solutions, i.e., those not dominated by any other solution, are assigned to the first front called *Pareto front*. Successive fronts contain solutions that are dominated only by those in the previous fronts. Once solutions are sorted into fronts, the crowding distance for each solution is calculated. This metric measures the density of solutions surrounding a particular solution in the objective space. Within each front, solutions with larger crowding distances are given higher priority, as they are more isolated from other solutions and represent more diverse parts of the objective space. Finally, when selecting solutions for the next generation, NSGA-II starts with those in the lowest rank (first front) and proceeds to higher ranks. Within a rank, solutions with higher crowding distances are preferred. This process continues until the desired number of solutions for the next generation is achieved.

2.2. Crossover: multi-channel variational auto-encoder

Before employing the MC-VAE to generate a new candidate solution set, the input data undergo smoothing via the Helmholtz PDE [36], as formulated below:

$$-r^2 \nabla^2 \bar{\gamma} + \bar{\gamma} = \gamma. \tag{6}$$

Here, γ represents the mapped material distribution, $\bar{\gamma}$ is the smoothed material distribution, and r denotes the filtering radius.

Building upon the fundamental principles of the standard VAE, a MC-VAE adds an extra layer of complexity by simultaneously processing multiple channels of input data. The primary advantage of the MC-VAE lies in its capacity to capture the underlying relationships between different data channels, which would be challenging for separate SC-VAEs to accomplish.

Consider a set of samples $\boldsymbol{X} = \{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(N_{\mathrm{in}})}\}$ where $\boldsymbol{X}^{(i)} \in \mathbb{R}^{N_{\mathrm{mc}}}$ denotes normalized pixel data with the total number of pixels N_{mc}, and multiple channels of input data represented as $\boldsymbol{X}^{(i)} = \{\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(i)}, \ldots, \boldsymbol{X}_C^{(i)}\}$, where C is the total number of channels, and each $\boldsymbol{X}_c^{(i)} \in \mathbb{R}^{N_{\mathrm{sc}}}$ corresponds to the data from the cth channel with each channel's number of pixels N_{sc}. The total number of pixels N_{mc} accordingly results in CN_{sc}. The first channel $\boldsymbol{X}_1^{(i)}$ contains the material distribution for the sample $\boldsymbol{X}^{(i)}$, while the

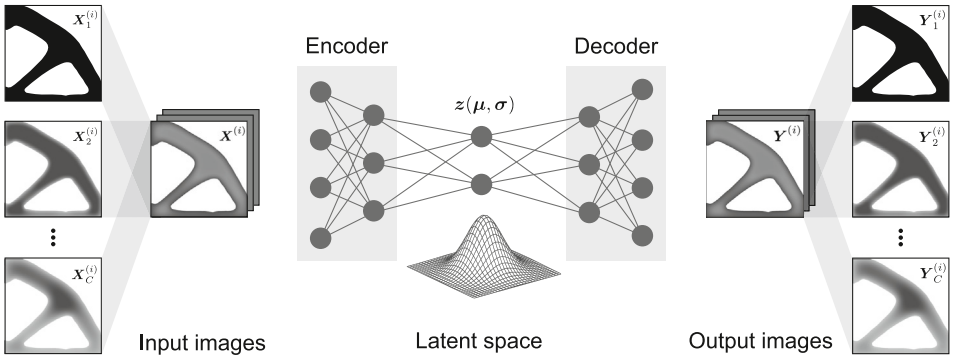

Fig. 3. Illustration of the data structure of MC-VAE.

second and subsequent channels $\boldsymbol{X}_{c+1}^{(i)}$ contain each HF seeding parameters' vector h_c. Fig. 3 depicts the concept of the MC-VAE's data structure.

VAEs are generally structured around two principal neural networks: an encoder that transforms high-dimensional input data $\boldsymbol{X}^{(i)}$ into latent variables $\boldsymbol{z} \in \mathbb{R}^{N_{\text{lt}}}$, where N_{lt} is significantly smaller than N_{mc}, and a decoder that reconstructs high-dimensional output data $\boldsymbol{Y}^{(i)} \in \mathbb{R}^{N_{\text{mc}}}$ from these latent variables. Within the MC-VAE framework, data from each channel is processed by its dedicated encoder network, yielding several sets of latent variables, namely $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots \boldsymbol{z}_C$. These sets are then concatenated to form a unified latent representation $\boldsymbol{z}$. The MC-VAE's decoder network is engineered to utilize this unified latent representation to reconstruct the data for each channel, producing $\boldsymbol{Y}^{(i)} = \{\boldsymbol{Y}_1^{(i)}, \boldsymbol{Y}_2^{(i)}, \ldots, \boldsymbol{Y}_C^{(i)}\}$. Following encoding and decoding, the first channel of the decoded image is remapped back onto the original domain to represent the material distribution. The VAE generates new data through its decoder, facilitated by ensuring that the compressed data adheres to a Gaussian distribution within the latent space.

The fundamental idea of the VAE lies in finding a probability distribution $P(\boldsymbol{X})$ from an approximated model $P_{\boldsymbol{\theta}}(\boldsymbol{X}) = P(\boldsymbol{X}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the parameters of the decoder to be optimized through learning process so that the likelihood of $P_{\boldsymbol{\theta}}(\boldsymbol{X})$ is maximized ultimately. In other words, the likelihood to generate data points similar to the input data $\boldsymbol{X}$ is desired to be maximized during the iterations. However, it is hard to calculate the likelihood $P_{\boldsymbol{\theta}}(\boldsymbol{X})$ due to computational efficiency, so VAEs employ an approach known as variational inference, specifically leveraging the 'variational lower bound'. This approach lets us optimize the expression on the right side of the following inequality instead of maximizing the $P_{\boldsymbol{\theta}}(\boldsymbol{X})$ directly:

$$\log P_{\boldsymbol{\theta}}(\boldsymbol{X}) \geq -D_{\text{KL}}(Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}) \parallel P_{\boldsymbol{\theta}}(\boldsymbol{z})) + \mathbb{E}_{Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X})}\left[\log P_{\boldsymbol{\theta}}(\boldsymbol{X}|\boldsymbol{z})\right], \tag{7}$$

where $Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}^{(i)})$ and $P_{\boldsymbol{\theta}}(\boldsymbol{X}^{(i)}|\boldsymbol{z})$ respectively correspond to the encoder and decoder model where $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are the parameters of the encoder and decoder, respectively. The latent variable, denoted by $\boldsymbol{z}$, is given by:

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \tag{8}$$

where the mean and standard deviation of the approximate posterior, denoted as $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, respectively, represent the encoder's outputs for each individual sample. The symbol $\odot$ signifies element-wise multiplication, and $\boldsymbol{\epsilon}$ denotes a random vector drawn from the standard normal distribution. Eq. (8) is commonly referred to as the "reparameterization trick", a technique that facilitates gradient computation in probabilistic sampling within a neural network. It accomplishes this by breaking down the stochastic variable into two components: a deterministic element (such as the output of a neural network) and a sample originating from an independent noise source. It is important to note that the latent space captures features relevant to all channels while still preserving the characteristics of the traditional VAE.

The first term, $D_{\text{KL}}(Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}^{(i)}) \parallel P_{\boldsymbol{\theta}}(\boldsymbol{z}))$, on the right-hand side of Eq. (7), represents the Kullback–Leibler (KL) divergence, used to quantify the disparity between the encoder's distribution and a standard Gaussian distribution. Minimizing this term leads to an approximation of the encoder distributions that closely resemble Gaussian distributions. The KL divergence is calculated according to its definition, as follows:

$$D_{\text{KL}}(Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}) \parallel P_{\boldsymbol{\theta}}(\boldsymbol{z})) = -\frac{1}{2}\sum_{j=1}^{N_{\text{lt}}}\left(1 + \log((\sigma_j)^2) - (\mu_j)^2 - (\sigma_j)^2\right). \tag{9}$$

The second term in Eq. (7) represents the reconstruction error between the original input material distributions and the decoded probability distributions, computed as binary cross-entropy for each channel, as follows:

$$\begin{aligned}\mathbb{E}_{Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}_c)}\left[\log P_{\boldsymbol{\theta}}(\boldsymbol{X}_c|\boldsymbol{z})\right] &= \frac{1}{N_{\text{in}}}\sum_{i=1}^{N_{\text{in}}}\log P_{\boldsymbol{\theta}}(\boldsymbol{X}_c^{(i)}|\boldsymbol{z}_j) \\ &= \frac{1}{N_{\text{in}}}\sum_{i=1}^{N_{\text{in}}}\left(\boldsymbol{X}_c^{(i)}\log(\boldsymbol{Y}_c^{(i)}) + (\mathrm{I} - \boldsymbol{X}_c^{(i)})\log(\mathrm{I} - \boldsymbol{Y}_c^{(i)})\right),\end{aligned} \tag{10}$$

where I denotes the identity matrix. Finally, a loss function L_{VAE} is computed by introducing a parameter to adjust the weight of the KL divergence, as shown below:

$$L_{\mathrm{VAE}}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{X}) = w_{\mathrm{KL}} \cdot D_{\mathrm{KL}}(Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}) \parallel P_{\boldsymbol{\theta}}(\boldsymbol{z})) - \sum_{c=1}^{C} \mathbb{E}_{Q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{X}_c)} \left[\log P_{\boldsymbol{\theta}}(\boldsymbol{X}_c|\boldsymbol{z})\right], \tag{11}$$

where $w_{\mathrm{KL}} \in \mathbb{R}^+$ represents the weight parameter known as KL weight, which governs the influence of the KL divergence, encouraging the distribution in the latent space to approximate a standard normal distribution. A higher KL weight urges the model to prioritize aligning the latent space distribution closer to the standard normal distribution. The value of the KL weight impacts the trade-off between generative diversity and reconstruction capability of the model. As shown in Eq. (11), the reconstruction error is computed for each channel first, and then the errors from all the channels are summed. After training the CNN-VAE, the latent variables are randomly obtained using the gaussian sampling method, and the latent variables are passed to the decoder of the CNN-VAE to generate new solutions.

2.3. Mutation

To prevent premature convergence and maintain dataset diversity throughout the entire design process, we adopt a mutation-like operation that Yaji et al. originally proposed for the single-channel D2MFTD [23]. We extend this operation to the multi-channel image dataset, focusing on the first channel that contains the material distributions. The concept of this multi-channel mutation-like operation adopted in this study is depicted in Fig. 4.

In the single-channel D2MFTD, the mutation operation is designed to generate new material distributions that differ significantly from the existing dataset, yet are likely to survive the selection process. In order to achieve this, the mutation operation solves a LF topology optimization problem with an additional constraint that references the existing dataset. This operation is achieved by solving the LF topology optimization problem defined in Eq. (2), incorporating an additional constraint function:

$$\tilde{G}_{\mathrm{mut}}(\gamma^{(m)}) = \int_D (1 - |\gamma^{(m)} - \gamma_{\mathrm{ref}}^{(m)}|)\, d\Omega \;\leq\; \tilde{G}_{\mathrm{mut}}^{\max} \int_D d\Omega, \tag{12}$$

where $m = 1, 2, \ldots, N_{\mathrm{mut}}$, and $\tilde{G}_{\mathrm{mut}}^{\max}$ is a parameter controlling the discrepancy between a selected reference material distribution $\gamma_{\mathrm{ref}}^{(m)}$ and the optimized material distribution for mutation. This equation ensures that the material distribution $\gamma^{(m)}$ is sufficiently different from the reference material distribution $\gamma_{\mathrm{ref}}^{(m)}$ by penalizing the difference between the two distributions. The concept behind this mutation process is detailed in the Ref. [23]. Notably, a smaller value of $\tilde{G}_{\mathrm{mut}}^{\max}$ results in a more divergent material distribution, which inherits fewer topological features from the original material distribution. Careful consideration is required when determining the value of this parameter, as a highly divergent material distribution faces greater challenges in surviving during the selection process. The mutated material distribution $\gamma^{(m)}$ is then mapped onto the unit square domain using the DDM and stored in the first channel of the multi-channel images.

In the multi-channel framework, this mutation operation is applied to the first channel of the multi-channel images, which contains the material distributions. For the second and subsequent channels, one can directly assign random distributions or phsicallly meaningful distributions, such as displacement and stress fields. When each element is assigned a random value, the random distributions should be instantly eliminated by the selection process, as they are likely to obtain poor performances. Therefore, the random distributions can be computed by multiplying a random scalar value by the material distribution in the first channel. On the other hand, the physical fields can be more effective in generating new candidates with higher performances than the random distributions, if the physical fields are properly designed.

3. Topology optimization

3.1. Preliminaries

This section briefly explores the formulation of topology optimization, specific to the solid isotropic material with penalty (SIMP) method, one of the most widely used methods in topology optimization. To solve the optimization problem such as (2), the design variable field γ is discretized into a set of design variables $\gamma_e \in [0, 1]$ for each element e, where $D \subseteq \mathbb{R}^d$ with $d \in 2, 3$ denotes a predefined design domain, using Finite Element Method (FEM) in this study. The governing equations are formulated based on the SIMP method, detailed as follows:

$$\begin{aligned} \boldsymbol{K}\boldsymbol{U} &= \boldsymbol{F}, \\ \boldsymbol{K} &= \sum_{e}^{N_{\mathrm{elm}}} E_{\mathrm{SIMP}}(\hat{\tilde{\gamma}}_e)\boldsymbol{K}_e, \\ E_{\mathrm{SIMP}}(\hat{\tilde{\gamma}}_e) &= E_{\min} + \hat{\tilde{\gamma}}_e^{p}(E_0 - E_{\min}), \end{aligned} \tag{13}$$

where the stiffness matrix $\boldsymbol{K}$ represents a system of stiffness equations for N_{elm} elements, with a displacement vector, denoted by $\boldsymbol{U}$, containing the displacements at each node or element, subject to applied loads represented by a force vector $\boldsymbol{F}$. Here, $\boldsymbol{K}_e$ denotes an element stiffness matrix, and E_{SIMP} represents the modified Young's modulus. E_0 and $E_{\min}$ correspond to the Young's modulus of solid and void materials, respectively. The parameter p serves as a penalty factor that encourages the binarization of design variables. Furthermore, $\hat{\tilde{\gamma}}_e$ denotes a filtered and projected design variable, as elaborated below.

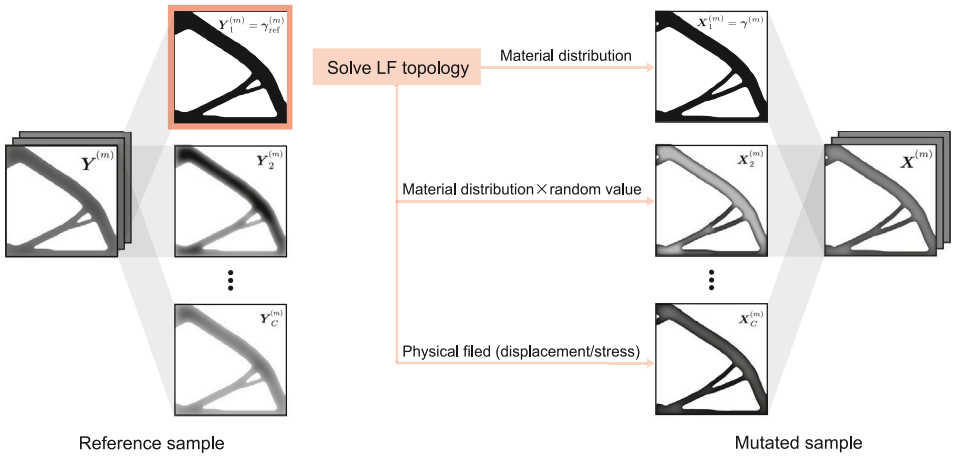

Fig. 4. Illustration of the mutation process.

To mitigate the emergence of overly complex topologies during optimization, a filtering process is typically incorporated into topology optimization. The design variables are subjected to filtering using a method proposed by Bourdin [37] to smooth the design variables by averaging each one with its neighbors within a filtering radius r_{f}, as shown below:

$$\begin{aligned} \tilde{\gamma}_e &= \frac{\sum_{j\in\mathbb{N}_{d,e}} H(\boldsymbol{x}_j)\gamma_j}{\sum_{j\in\mathbb{N}_{d,e}} H(\boldsymbol{x}_j)}, \\ H(\boldsymbol{x}_j) &= \max(0, r_{\mathrm{f}} - |\boldsymbol{x}_j - \boldsymbol{x}_e|), \end{aligned} \tag{14}$$

where $\mathbb{N}_{d,e}$ represents a set of design variables located within a distance less than the filter radius r_{f} from the position of each element $\boldsymbol{x}_e$. For 2D or 2D-like optimization problems solved in the LF optimization step, a Heaviside projection is incorporated to remove the grayscale results and achieve simpler yet sharper topologies [38], as illustrated below:

$$\hat{\tilde{\gamma}}_e = \frac{\tanh(\beta(\tilde{\gamma}_e - \eta)) + \tanh(\beta\eta)}{\tanh(\beta(1-\eta)) + \tanh(\beta\eta)}, \tag{15}$$

where β represents a scaling parameter that determines the sharpness of the projection, and η is a threshold value to determine whether the projected variable should be adjusted closer to zero or one. For 3D optimization problems under casting constraints, 2D design variable field is projected onto 3D design variable field using the following equation [39]:

$$\hat{\tilde{\gamma}}_e = \frac{\tanh(\beta\tilde{\gamma}_e) + \tanh(\beta(\gamma_{\mathrm{intrp}} - \tilde{\gamma}_e))}{\tanh(\beta\tilde{\gamma}_e) + \tanh(\beta(1-\tilde{\gamma}_e))}. \tag{16}$$

Here, γ_{intrp} is a temporary design variable field that interpolates the 2D design variable field in xy-plane to the 3D design variable field along the z-axis.

The design variable field γ is iteratively updated using the Method of Moving Asymptotes (MMA), a well-established technique of sequential convex programming method [40], based on the sensitivity of the objective functional against the design variable field, calculated using the chain rule.

3.2. Stiffness maximization

A bi-objective topology optimization to maximize stiffness and minimize volume is written as follows:

$$\begin{aligned} \text{find} \quad & \gamma_e \quad (e = 1, 2, \ldots, N_{\mathrm{elm}}), \\ \text{that minimize} \quad & W = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{U}, \\ & V = \sum_{e=1}^{N_{\mathrm{elm}}} v_e\gamma_e, \\ \text{subject to} \quad & \gamma_e \in \{0, 1\}, \end{aligned} \tag{17}$$

where v_e represents the volume of each element. The objective functional W is defined as the compliance of the structure, or strain energy, calculated by $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{U}$, which can be considered as the inverse of stiffness. Notably, the material distribution derived from this formulation is entirely discrete by its nature, allowing for a more precise evaluation of objective functionals than methods involving relaxation.

The bi-objective optimization problem (17) is converted to a mono-objective optimization problem with a constraint of maximum volume $V_{\max}$, as follows:

$$\begin{aligned} \text{find} \quad & \gamma_e \quad (e = 1, 2, \dots, N_{\text{elm}}), \\ \text{that minimize} \quad & W = \boldsymbol{U}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{U}, \\ \text{subject to} \quad & V = \sum_{e=1}^{N_{\text{elm}}} v_e \gamma_e \ \leq \ V_{\max}, \\ & \gamma_e \in [0, 1], \end{aligned} \tag{18}$$

where γ_e is transitioned from a discrete to a continuous variable. Solving this optimization problem (18) with various values of the maximum volume constraint $V_{\max}$ allows us to generate a variety of material distribution candidates to be stored in the first channel of multi-channel images as the initial designs, combined with randomly assigned HF seeding parameters.

The material distributions derived from LF topology optimization or the first channel of the MC-VAE output undergo smoothign process and projected onto the voxel mesh, which are subsequently isosurfaced using a density threshold $\gamma_e = 0.5$. This isosurfacing process ensures that the material distribution is clearly defined as either solid or void, with only the elements designated as solid being extruded to the height of the HF seeding parameters.

3.3. Maximum stress minimization

A bi-objective topology optimization to minimize maximum stress and volume is written as follows:

$$\begin{aligned} \text{find} \quad & \gamma_e \quad (e = 1, 2, \dots, N_{\text{elm}}), \\ \text{that minimize} \quad & \sigma_{\max} = \max(\sigma_{\text{vm},e}), \\ & V = \sum_{e=1}^{N_{\text{elm}}} v_e \gamma_e, \\ \text{subject to} \quad & \gamma_e \in \{0, 1\}, \end{aligned} \tag{19}$$

where $\sigma_{\text{vm},e}$ is von Mises stress of each element.

This bi-objective optimization problem (19) is so-called a mini-max problem, which is not applicable to gradient-based optimization methods due to its non-differentiability. To allow for gradient-based optimization, a p-norm function [41,42] is employed to approximate the maximum stress within the design domain, transforming the original bi-objective optimization problem (19) into a mono-objective optimization problem, as follows:

$$\begin{aligned} \text{find} \quad & \gamma_e \quad (e = 1, 2, \dots, N_{\text{elm}}), \\ \text{that minimize} \quad & \sigma_{\text{PN}} = \left(\frac{1}{N_{\text{elm}}} \sum_{e=1}^{N_{\text{elm}}} \hat{\sigma}_{\text{vm},e}^{P}(\gamma_e) \right)^{\frac{1}{P}}, \\ \text{subject to} \quad & V = \sum_{e=1}^{N_{\text{elm}}} v_e \gamma_e \ \leq \ V_{\max}, \\ & \gamma_e \in [0, 1], \end{aligned} \tag{20}$$

where the stress field $\hat{\sigma}_{\text{vm},e}(\gamma_e) = \gamma_e^q \sigma_{\text{vm},e}$ is modified to facilitate the binarization of design variables and mitigate numerical challenges encountered when optimizing design variables near material boundaries, where stress can change abruptly. This modification ensures that design variables transition smoothly, avoiding sharp increases as their values approach 0. The approximated stress, denoted by σ_{PN}, converges to the actual maximum stress $\sigma_{\max}$ as P approaches infinity. However, as P increases, so does the computational complexity. Consequently, the value of P should be carefully determined compromising the computational stability and evaluation accuracy.

Similar to the approach for stiffness maximization, material distributions are isosurfaced, and HF models are constructed using the defined material boundaries and HF parameters. Subsequently, von Mises stress $\sigma_{\max}$ and volume V are evaluated within these FE models. It is important to note that the maximum stress is determined directly from the von Mises stress distribution across the design domain, bypassing any relaxation or approximation strategies.

4. Numerical implementation

This section outlines the detailed implementation of the proposed framework. Building upon the core concepts previously discussed, a pseudo-code of the proposed framework is presented in Algorithm (1), tailored for bi-objective optimization problems such as those defined by Eqs. (17) and (19). The parameters governing the entire process are detailed in Table 1. The LF seeding parameter l and the HF seeding parameter h are determined through Latin Hypercube Sampling (LHS) from the intervals $[l^{\min}, l^{\max}]$ and $[h^{\min}, h^{\max}]$, respectively. LHS is a statistical technique designed to generate a diverse set of plausible parameter combinations from a multidimensional distribution, commonly employed in computer simulations to efficiently explore varied design spaces. The design iteration continues until the relative error of the hypervolume indicator falls below ε_{HV}, or the iteration count reaches a

Table 1
Parameters for the overall procedures.

Parameter	Symbol	Value
Number of LF seeding parameters	$N_{\mathrm{lf}}^{\mathrm{sd}}$	100
Number of channels	C	2
Maximum iterations	$N^{\max}$	100
Number of seeding parameters in mutation	$N_{\mathrm{mut}}^{\mathrm{sd}}$	5
Number of mutant generation for each seeding parameter	N_{mut}	10
Interval of mutation	$N_{\mathrm{mut}}^{\mathrm{int}}$	5
Threshold value of hypervolume indicator	ϵ_{HF}	1.0×10^{-5}
Number of samples generated by MC-VAE	N_{VAE}	256

Table 2
Parameters for LF optimization problem.

Parameter	Symbol	Value
Filter radius	r_{f}	0.03
Scaling factor of projection	β	4
Threshold value of projection	η	0.5
Young's modules of solid material	E_0	1.0
Young's modules of void material	$E_{\min}$	1.0×10^{-9}
Number of elements	N_{elm}	64×64
Penalty factor	p	3

predefined maximum $N_{\max}$. The reference point for hypervolume calculations, as per Eq. (4), is set to values marginally exceeding the highest initial objective values found on the Pareto front. The mutation process occurs every $N_{\mathrm{mut}}^{\mathrm{int}}$ iterations. In the mutation process, N_{mut} solutions are selected as reference solutions, and the material distribution in the first channel of the corresponding multi-channel images is mapped to the design domain. The LF optimization (12) is then performed with the constraints of how divergent the mutants should be from the material distribution of the reference solutions for $N_{\mathrm{mut}}^{\mathrm{sd}}$ times with different LF parameters, i.e., volume fraction in this case. The subsequent channels are initialized with the beforementioned random fields or physical fields. The reference material distributions $\gamma_{\mathrm{ref}}^{(m)}(m = 1, 2, \ldots, N_{\mathrm{mut}})$ are uniformly selected from the Pareto front solutions denoted as $\Theta^{(i)}$. Among all the processes, HF evaluation demands the most computational resources, as it involves running numerical simulations for each sample in the temporary dataset Θ_{tmp} at every iteration i. To optimize computation time, HF evaluations (lines 9–11) were performed in parallel, leveraging the independence of each simulation from the others.

In the selection process, the number of offspring should be carefully determined due to its influence on the efficiency of both the entire optimization and MC-VAE's learning performance. The MC-VAE's learning performance is significantly influenced by both the network architecture and the volume of the input dataset, particularly when this dataset is relatively limited. A consistent offspring count for the MC-VAE's input dataset promotes more stable learning outcomes because the MC-VAE's network architecture remains unchanged throughout the optimization iterations. On the contrary, the more offspring count allow the MC-VAE to learn the characteristics of the input dataset more effectively. Therefore, we opt to determine the number of offspring by setting a minimum threshold and selecting all Pareto front points as offspring when their total number surpasses this minimum.

The framework outlined in Algorithm (1) was developed using Python (version 3.9.13) as the primary programming environment, incorporating a variety of tools for its implementation. The LF topology optimization and HF evaluation, along with the solution of Helmholtz PDEs, were executed in COMSOL Multiphysics (version 6.1), a commercial FEM software. In this study, COMSOL Multiphysics was controlled via scripts written in MATLAB (version 2023a). The MC-VAE was constructed using TensorFlow (version 2.12.0), an open-source machine learning framework renowned for its versatility in developing and deploying sophisticated machine learning models. DDM was accomplished with the aid of a mapping function from the IGL Python library (version 2.4.1). The traditional EA operations, such as selection and hypervolume calculations, were facilitated by DEAP (version 1.3.3), an open-source Python library designed for crafting and executing genetic algorithms and other evolutionary computing strategies.

4.1. Low-fidelity topology optimization

For both the stiffness maximization problem (18) and the maximum stress minimization problem (20), the raw design variable field γ_e, filtered design variable field $\tilde{\gamma}_e$ and the displacement field $\boldsymbol{U}$ are represented using shell elements that feature five intergal points: one at each corner of the quadrilateral and one at the center. In contrast, the projected design variable field $\hat{\tilde{\gamma}}_e$ are represented using $\mathbb{P}^0$ Lagrange quadrilateral finite elements, which are suited for element-wise material distribution. The parameters for the LF topology optimization are detailed in Table 2.

The design variables are iteratively updated using the MMA, with a move limit set to 0.05 for all LF optimization simulations. The sensitivity analysis of the objectives relative to the design variables employs a discrete adjoint method installed in COMSOL Multiphysics. The maximum number of iterations was set to 50 for stiffness maximization and increased to 100 for maximum stress minimization to accommodate the latter's greater non-linearity and complexity. The P in Eq. (20) was initially set to 8, and increased as $16, 32, 64$ every 30 iterations based on the continuation method [43]. The penalty parameter for stress relaxation was set as $q = 0.5$.

Algorithm 1 Pseudo-code of the expanded data-driven MFTD for a bi-objective optimization problem

for $k = 1$ to $N_{\mathrm{lf}}^{\mathrm{sd}}$ **do**
  Solve the LF optimization problem (18) or (20) to find $\gamma_e^{(k)}(e = 1, 2, ...,)$ on $l^{(k)}$
end for
Contain the material distributions in the 1ch of a temporary dataset, $\Theta_{\mathrm{tmp},1} \leftarrow \{\gamma_e^{(1)}, \gamma_e^{(2)}, ..., \gamma_e^{(N_{\mathrm{lf}}^{\mathrm{sd}})}\}$
for $c = 2$ to C **do**
  Define HF parameters for the c-ch of the temporary dataset, $\Theta_{\mathrm{tmp},c} \leftarrow \{h^{(1)}, h^{(2)}, ..., h^{(N_{\mathrm{lf}}^{\mathrm{sd}})}\}$
end for
for $i = 0$ to $N^{\max}$ **do**
  for $k = 1$ to $\mathrm{len}(\Theta_{\mathrm{tmp}})$ **do**
    Calculate HF evaluation functionals, $\{J_1^{(i)}(\gamma_e^{(k)}), J_2^{(i)}(\gamma_e^{(k)}), G^{(i)}(\gamma_e^{(k)})\}$
  end for
  Squeeze Θ_{tmp} so that all the material distributions satisfy $G^{(i)} \leq 0$
  if $i = 0$ **then**
    Set the current dataset, $\Theta^{(i)} \leftarrow \Theta_{\mathrm{tmp}}$
  else
    Augment the dataset, $\Theta^{(i)} \leftarrow \Theta^{(i)} \cup \Theta_{\mathrm{tmp}}$
  end if
  Squeeze $\Theta^{(i)}$ by the selection algorithm based on $\mathcal{J}(\Theta^{(i)}) \leftarrow \bigcup_{k=1}^{\mathrm{len}(\Theta^{(i)})}\{J_1^{(i)}(\gamma_e^{(k)}), J_2^{(i)}(\gamma_e^{(k)})\}$
  if the hypervolume indicator HV $\leq \varepsilon_{\mathrm{HV}}$ regarding $\mathcal{J}(\Theta^{(i)})$ **then**
    Return the obtained dataset $\Theta^{(i)}$ with $\mathcal{J}(\Theta^{(i)})$
  end if
  Convert $\Theta^{(i)}$ to normalized dataset by design domain mapping, $\boldsymbol{X} \leftarrow \{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, ..., \boldsymbol{X}^{(\mathrm{len}(\Theta^{(i)}))}\}$
  Acquire a new dataset by MC-VAE using $\boldsymbol{X}$ as the input, $\boldsymbol{X}_{\mathrm{VAE}} \leftarrow \{\boldsymbol{X}_{\mathrm{VAE}}^{(1)}, \boldsymbol{X}_{\mathrm{VAE}}^{(2)}, ..., \boldsymbol{X}_{\mathrm{VAE}}^{(N_{\mathrm{VAE}})}\}$
  Convert $\boldsymbol{X}_{\mathrm{VAE}}$ to Θ_{tmp} by the inverse design domain mapping
  if $\mathrm{mod}(i, N_{\mathrm{mut}}^{\mathrm{int}})$ **then**
    Sample reference material distributions for mutation, $\{\gamma_{e,\mathrm{ref}}^{(1)}, \gamma_{e,\mathrm{ref}}^{(2)}, ..., \gamma_{e,\mathrm{ref}}^{(N_{\mathrm{mut}})}\} \in \Theta^{(i)}$
    for $n = 1$ to $N_{\mathrm{mut}}^{\mathrm{sd}}$ **do**
      for $m = 1$ to N_{mut} **do**
        Solve the LF optimization problem (18) or (20) for $\gamma_e^{(m)}$ on $l^{(n)}$ with the constraint (12)
      end for
      Assemble a dataset of the optimized material distributions, $\theta_{\mathrm{mut},1}^{(n)} \leftarrow \{\gamma_e^{(1)}, \gamma_e^{(2)}, ..., \gamma_e^{(N_{\mathrm{mut}})}\}$
      for $c = 2$ to C **do**
        Define HF parameters for c-ch of the mutation dataset, $\theta_{\mathrm{mut},c}^{(n)} \leftarrow \{h^{(1)}, h^{(2)}, ..., h^{(N_{\mathrm{mut}})}\}$
      end for
    end for
    Augment the temporary dataset with all the mutation dataset, $\Theta_{\mathrm{tmp}} \leftarrow \Theta_{\mathrm{tmp}} \cup \bigcup_{n=1}^{N_{\mathrm{mut}}^{\mathrm{sd}}} \theta_{\mathrm{mut}}^{(n)}$
  end if
end for
Return the obtained dataset $\Theta^{(i)}$ with $\mathcal{J}(\Theta^{(i)})$

4.2. Multi-channel variational auto-encoder

Here, we explore the detailed implementation of MC-VAE, including its network architecture. Unlike the approach taken in previous research [23], which utilized a basic VAE architecture comprising a multilayer perceptron with two dense layers, we adopt CNNs, which is engineered specifically for processing and understanding visual data. Whereas dense layers flatten the input image's pixels and assign independent weights to each pixel, convolutional layers apply a series of localized dot products across small sections of the input image using a filter. Furthermore, while the number of weights in a dense layer simply corresponds to the total pixel count of multi-channel images, CNNs significantly reduce the number of weights by utilizing a filter that matches the input image's channel count and sliding this filter across all image pixels. Consequently, CNNs excel in capturing the spatial relationships within multi-channel data through their convolutional layers. This ability to extract spatial features enables VAEs to generate more meaningful and compact representations within the latent space, leading to enhanced image reconstruction during decoding. The powerful feature extraction capabilities of CNNs motivated us to implement MC-VAE using CNNs to more efficiently manage the growing complexity of input data, which now includes multiple channels.

In this study, two distinct MC-VAE architectures were developed for constant thickness cases (Cases 1 and 2) and spatially varying thickness cases (Cases 3 and 4). The network architecture for Cases 1 and 2 is illustrated in Fig. 5. The encoder of MC-VAE is composed of an input layer accepting 64 × 64 pixels and two channels, the first convolutional layer with Relu function, the second

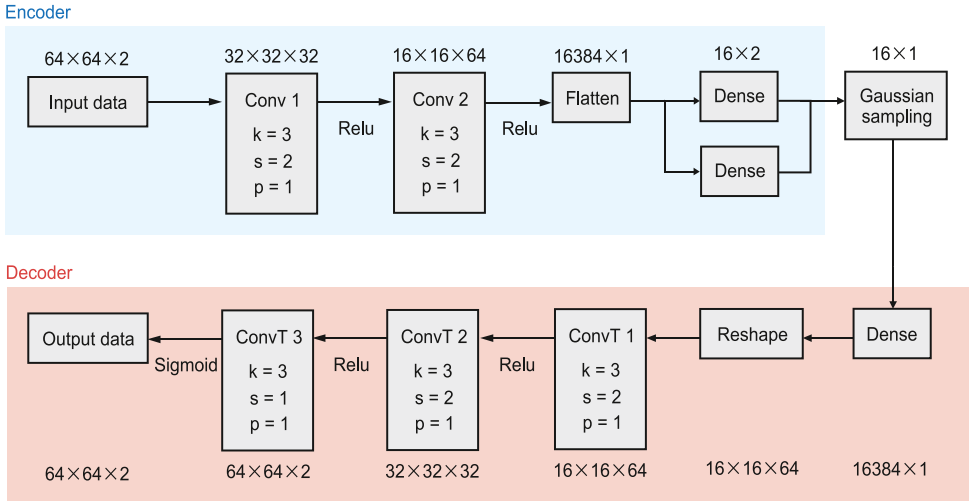

Fig. 5. Network architecture of the MC-VAE used for Case 1 and 2.

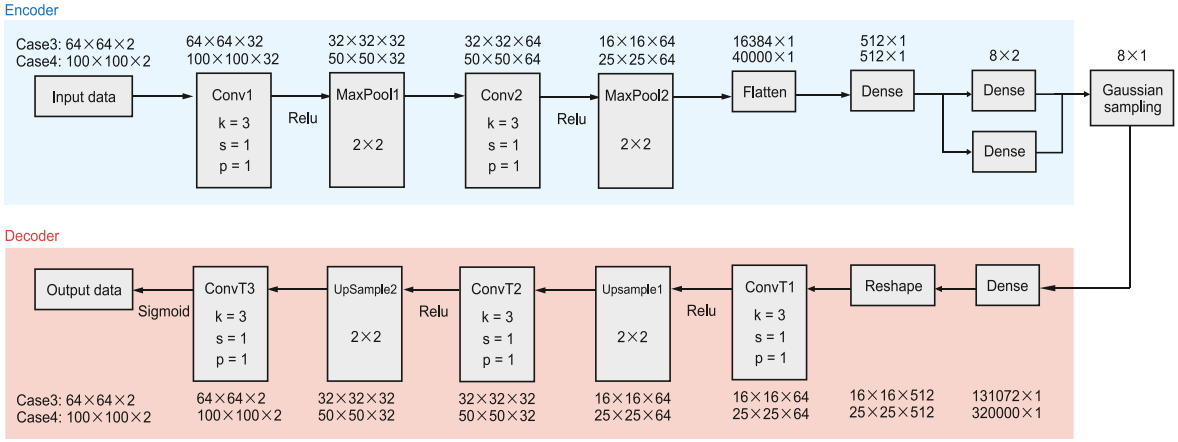

Fig. 6. Network architecture of the MC-VAE used for Case 3 and 4.

convolutional layer with Relu function, a flattening layer, and a dense layer. The decoder consists of an input layer accepting a latent vector of size 16, a dense layer, the first deconvolutional layer with Relu function, the second deconvolutional layer with Relu function, and the third deconvolutional layer with sigmoid function. The use of the sigmoid function in the final layer ensures that the output values are within the range $[0, 1]$. The convolutional and the first two deconvolutional layers have 32 and 64 filters, respectively, with a kernel size of 3×3, a stride of 2×2, and a padding of 1×1. The third deconvolutional layer has 2 filters, a kernel size of 3×3, a stride of 1×1, and a padding of 1×1, to match the input image size.

The network architecture for Cases 3 and 4 is illustrated in Fig. 6. The differences between the two architectures lie in the number of stride of the covolutional filters and the first two deconvolutional filters; the maximum pooling layer inserted after each convolutional layer; and a dense layer additionally inserted in the encoder. These differences are implemented to accommodate the spatially varying thickness cases, where the learning of the spatial relationship between the two channels is more critical.

It is important that the output values for the first channel can be directly used as material distributions $\gamma \in [0, 1]$, but for the second channel, the values needs denormalization that inversely normalizes the output values from $[0, 1]$ to $[h^{\min}, h^{\max}]$. As the optimizer in training neural networks, we used Adam [44], an extension of the stochastic gradient descent method, which is widely used as the optimizer of neural networks including VAEs. The maximum epochs, the learning rate, and the mini-batch size are set as 300, 1.0×10^{-4}, and 16, respectively. The KL weight w_{KL} for controlling the influence of the KL divergence is set as $w_{\mathrm{KL}} = 1.0 \times 10^{-3}$. The neural networks are learned until the epoch is reached to the maximum number of epochs or the loss function L_{VAE} is not improved for 40 epochs in total.

To validate the implementation of the MC-VAE's network architecture, the MC-VAE was trained on two distinct example datasets. In both datasets, the first channel contained identical material distributions, while the second channel held a scalar HF parameter for each sample. For one dataset, this scalar HF parameter was directly proportional to the volume fraction derived from the first

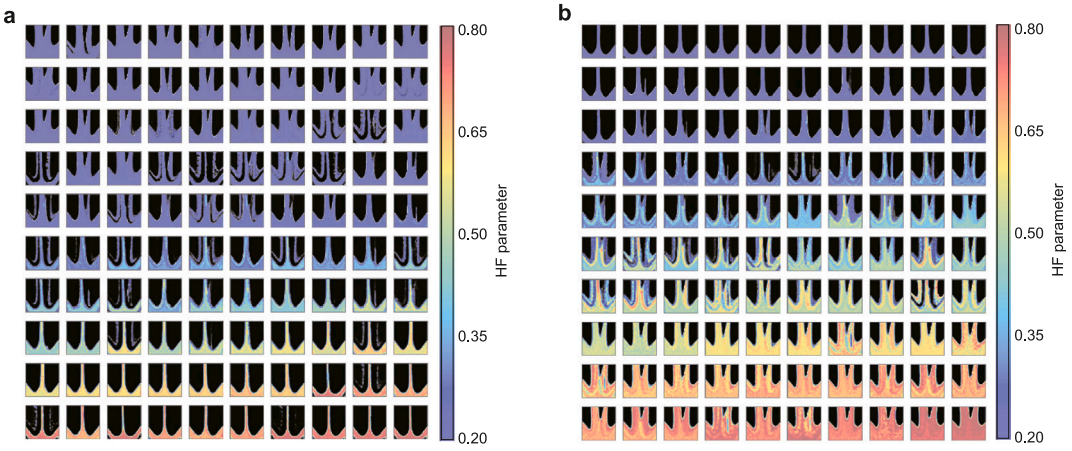

Fig. 7. Generated images comparison between (a) the proportional and (b) inversely proportional dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

channel; for the other, it was inversely proportional. The images generated from these datasets are displayed in Fig. 7. Here, solid materials in the first channel are color-coded to represent the HF parameter in the second channel, with black indicating void areas. Despite the relatively small size of each dataset, consisting of only 100 samples – significantly less than the typical dataset sizes for VAEs, which are on the order of 10^4 – the results, as shown in Fig. 7, demonstrate that the MC-VAE successfully captured both the topological features of the material distributions and the relationship between these features and the HF parameter. This is evident as samples with a higher volume in the proportional dataset correspondingly stored higher HF parameters in the second channel, and the inverse was true for the inversely proportional dataset. This effective learning from a small dataset can be attributed to the simplicity of the features in the input images and the straightforward relationship between the material distribution and the HF parameter. Essentially, the MC-VAE needed to grasp either a proportional or inversely proportional relationship between the two channels. As the number of channels increases, there may be a need to enlarge the input dataset or adjust the network architecture for more efficient learning. Furthermore, Fig. 7 suggests that the HF parameters in the second channel are represented not as scalar values but as distributions, even when the input dataset's second channel contained scalar values. Therefore, when calculating HF models, if a scalar value for the HF parameter is required, it must be extracted from the decoded images as a scalar value through a mathematical computation like averaging. As a preliminary study, this research utilized scalar values for the HF parameter in each sample to assess the framework's capability to identify promising samples within the HF modeling space.

The size of decoding samples is fixed at 256, while the size of input samples often exceeds 100, due to the implementation of a minimum offspring number and the inclusion of all Pareto front samples in the selection process. This excess raises the potential for overfitting within the MC-VAE, as the input sample size surpasses the initial dataset size used for network tuning. Nonetheless, minor overfitting does not critically impact the framework's effectiveness, as MC-VAE's primary role is to facilitate crossover by producing a multitude of samples that inherit target dataset's features. These decoded samples, while bearing resemblance to the input dataset in terms of topological features and HF parameters, exhibit slight variations. It is important to note that the generation of samples with significantly different features is achieved through the mutation-like operation, as detailed in Section 2.3.

A particular challenge for MC-VAE is maintaining sample diversity, especially when the input dataset is significantly imbalanced. The selection of solutions based on LF and HF parameters, coupled with the optimization problem's nature, often results in a disproportionate representation of samples with extreme values for one objective and the inverse for the other. This imbalance is particularly pronounced for samples with medium values of both objectives, becoming more acute with each iteration and leading to an increasingly biased input dataset for MC-VAE.

To mitigate the iterative bias in the MC-VAE dataset, we employed a sampling strategy called *oversampling* [45], applied at the input stage of MC-VAE. Oversampling is utilized to rectify class imbalances within datasets by duplicating under-represented samples until a balance is achieved. This encourages MC-VAE within the data-driven MFTD framework to develop a more comprehensive understanding of less frequent data, thereby enhancing its capability to learn the input dataset's characteristics more effectively maintaining the diversity.

4.3. High-fidelity evaluation

In the HF evaluation process, the design variables stored in the multi-channel images are utilized to construct HF models with adaptive meshes, which facilitate more accurate evaluations of the displacement and stress fields. An example of this modeling process is depicted in Fig. 8, where the design domain is not curved but flat for simplicity. As shown in Fig. 8(a), the multi-channel images contain two channels: the first channel represents the reinforcement region mask $X_1^{(k)}$, while the second channel represents the thickness distribution $X_2^{(k)}$. The reiforcement region mask contains pixel values ranging from 0 to 1, with 0 indicating void areas

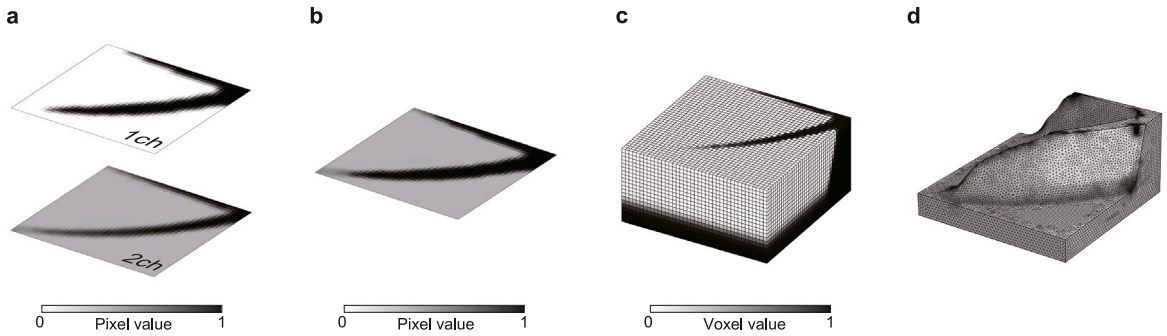

Fig. 8. An example of HF modeling: (a) mask field in the first channel and thickness distribution in the second channel; (b) filtered design variable field; (c) projected design variable field; (d) adaptive mesh constructed by isosurface.

and 1 indicating solid areas. On the other hand, the thickness distribution is represented by pixel values ranging from normalized skin thickness γ_{skin} to 1, with γ_{skin} indicating the minimum thickness and 1 indicating the maximum thickness. Each channel image is mapped to the design domain, converting the pixel values to design variables $\gamma_{e,1}$ and $\gamma_{e,2}$ by DDM. Then the mapped design variables are utilized to calculate the reinforced thickness distribution γ_e, as follows:

$$\gamma_e = \gamma_{\text{skin}} + \gamma_{e,1} \cdot (\gamma_{e,2} - \gamma_{\text{skin}}). \tag{21}$$

The reinforced thickness distribution γ_e is then filtered by Eq. (14) to obtain $\tilde{\gamma}_e$ and projected to the design domain by Eq. (16) to obtain $\hat{\tilde{\gamma}}_e$. Fig. 8(b) shows the reinforced thickness distribution, which is projected to the design domain using the casting method, as depicted in Fig. 8(c). The adaptive mesh is constructed by extracting isosurface at $\gamma_e = 0.5$, as shown in Fig. 8(d). The displacement and stress fields are discretized using $\mathbb{P}^1$ Lagrange tetrahedral finite elements.

Note that the beforementioned 3D topology optimization with casting constraints utilizes the same projection strategy as the HF evaluation process, ensuring consistency between the conventional 3D topology optimization and the proposed method. However, the casting-constrained 3D topology optimization has difficulty in using the adaptive mesh during the optimization process, while the proposed method can easily utilize the adaptive mesh for HF evaluation due to the decoupling of LF and HF optimization. Significantly, the mesh configuration can differ between LF and HF models without issue, leveraging the multifidelity approach where LF optimization and HF evaluation are independently processed. Consequently, this study employs solid meshes for HF models to ensure detailed evaluations, and 2D or shell meshes for LF models to save computational resources.

4.4. Mutation

A subset of N_{mut} multi-channel images $\boldsymbol{Y}^{(m)}$ is earmarked for mutation by uniformly selecting reference samples from solution candidates in the existing dataset. Then the first channel of the reference images $Y_1^{(m)}$ are mapped onto the design domain as $\gamma_{e,\text{ref}}^{(m)}$ using the DDM. The mapped design variables are utilized as constraints in the LF topology optimization for mutation of material distribution, as follows:

$$\Sigma_e^{N_{\text{elm}}} \left(1 - |\gamma_e^{(m)} - \gamma_{e,\text{ref}}^{(m)}|\right) \leq \Sigma_e^{N_{\text{elm}}} \tilde{G}_{\text{mut}}^{\max}. \tag{22}$$

This equation is a discrete form of Eq. (12), which is used to ensure that the material distribution of the mutated samples is similar to or different from that of the reference samples to a specified degree. Eq. (22) is added to the constraint in the Eq. (18) or Eq. (20) for the LF optimization of mutation, and the LF optimization is solved to find the material distribution $\gamma_e^{(m)}$. An example of the LF optimization for mutation is shown in Fig. 9, where the reference design variable $\gamma_{e,\text{ref}}^{(m)}$ and the mutated design variable $\gamma_e^{(m)}$ are compared. In this example, the discrepancy parameter $\tilde{G}_{\text{mut}}^{\max}$ is set to 0.8, roughly indicating that the material distribution of the mutated sample is constrained to resemble the reference sample within 80% of the design domain. Fig. 9 demonstrates that the material distribution of the mutated sample is similar to that of the reference sample yet exhibits slight differences around the right bottom corner.

After the LF optimization for mutation, the resultant material distribution $\gamma_e^{(m)}$ is mapped onto the design domain and stored in the first channel of the mutated sample image as $\boldsymbol{X}_1^{(m)}$. For subsequent channels in the mutated samples, the multiplication assignment of a random scalar and the material distribution $\gamma_e^{(m)}$ is adopted in aforementioned case 1 and 2, while the displacement and stress filed of the LF optimization result are directly inherited in the subsequent channels in case 3 and 4, respectively. This assignment is consistent with the initital dataset, where the material distribution is derived from the LF optimization result and the random scalar is multiplied to the material distribution or the displacement and stress fields are inherited in the subsequent channels.

5. Numerical examples

The efficacy of the proposed framework is illustrated through four numerical examples. To validate the capability of the framework to navigate the HF parameter space, the first two examples (Case 1 and 2) concurrently optimize reinforcement layout and its scalar thickness for each sample. The subsequent two examples (Case 3 and 4) concurrently optimize reinforcement layout

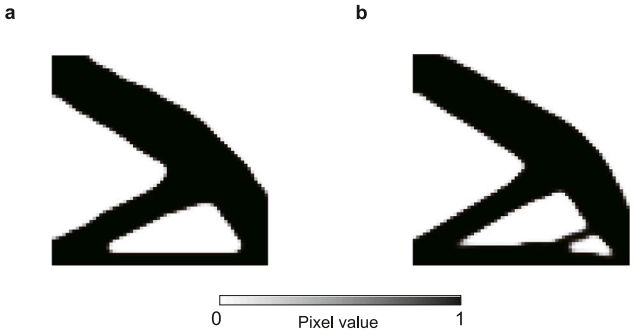

Fig. 9. Example results of the LF topology optimization for mutation with $\tilde{G}_{\mathrm{mut}}^{\max} = 0.8$: (a) reference design variable $\gamma_{e,\mathrm{ref}}^{(m)}$; (b) mutated design variable $\gamma_{e}^{(m)}$.

Table 3
Condition of numerical examples.

Case	2ch design variable	Objective function	Element type for LF	Element type for HF	B.C. for LF	B.C. for HF
1	Scalar thickness	Stiffness, volume	Structured shell	Adaptive tetrahedra	2D-like CB	3D CB
2	Scalar thickness	Max-stress, volume	Structured shell	Adaptive tetrahedra	2D-like CB	3D CB
3	Thickness distribution	Stiffness, volume	Structured 2D	Adaptive tetrahedra	2D CB	3D CB
4	Thickness distribution	Max-stress, volume	Structured 2D	Adaptive tetrahedra	2D L-bracket	3D L-bracket

and its varying thickness distribution for each sample to fully utilize the potential of the framework. The first two exmples can be achieved by a combination of a white-and-black density distribution (obtained via a single-channel VAE) and a scalar thickness value (obtained through regression), but that combination is not suitable for the latter two examples. Additionally, in these examples, the reinforced thickness is calculated by multiplying the reinforcement region stored in the first channel and the spatially varying thickness distribution stored in the second channel. This reinforced thickness is used to express the thickness extrusion from the bottom plane to the top surface of the HF model. In this context, it is true that a single continuous field with a projection method could achieve the same results as the two-channel approach in theory. However, we chose to use two distinct channels because we intended to confirm that the proposed framework can effectively optimize multiple individual design variable fields in a single optimization process. More specifically, if the framework fails to optimize the two design variable fields, it can easily obtain the reinforcement region completely unrelated to the thickness distribution. In this case, the HF models constructed by the combined reinforced thickness distribution should result in poor performance. Therefore, the two-channel approach is adopted to verify the capability of the framework to optimize multiple design variable fields simultaneously.

Case 1 and 3 adress stiffness maximization problems that can be directly solved by conventional topology optimization, so that the proposed framework can be validated by comparing the results with the conventional topology optimization. Case 2 and 4 tackle maximum stress minimization problems, which inherently require relaxation and approximation methods for stress to be solvable. Prior research has shown that the mini-max stress problem could be indirectly addressed through the data-driven MFTD, yet without specifically aiming to explore the HF parameter space for superior solutions [28]. We aim to demonstrate that the proposed framework can successfully navigate the multi-channel design space to identify high-performance solutions for these problems without relaxation and approximation methods. To achieve this, the stress optimization problem was solved using relaxation and approximation methods, followed by an evaluation of the objective values with HF models, bypassing relaxation and approximation. For case 2, as an alternative LF optimization within the data-driven MFTD, the stiffness maximization problem was also solved, subsequently assessing the stress and volume objectives to gauge the impact of the LF optimization problem on the final solution's performance.

Additionally, the first two examples target a curved design domain to demonstrate the capability of the framework to handle complex geometries, while the subsequent two examples address non-curved design domains for simplicity. Case 1–3 focus on a cantilever beam (CB) problem, one of the most representative benchmarks, while Case 4 addresses a L-bracket problem, another common benchmark especially for maximum stress minimization. The conditions of each numerical example are summarized in Table 3.

5.1. Case 1: stiffness maximization optimizing constant thickness

5.1.1. Problem setup

In practical design scenarios, the target surface is often three-dimensional thin-walled structures with curvature, necessitating the capability of the framework to handle such structures. To facilitate this, a part of a quadric surface was selected for the target surface, as expressed:

$$f(x, y) = \alpha\sqrt{x^2 + y^2}, \tag{23}$$

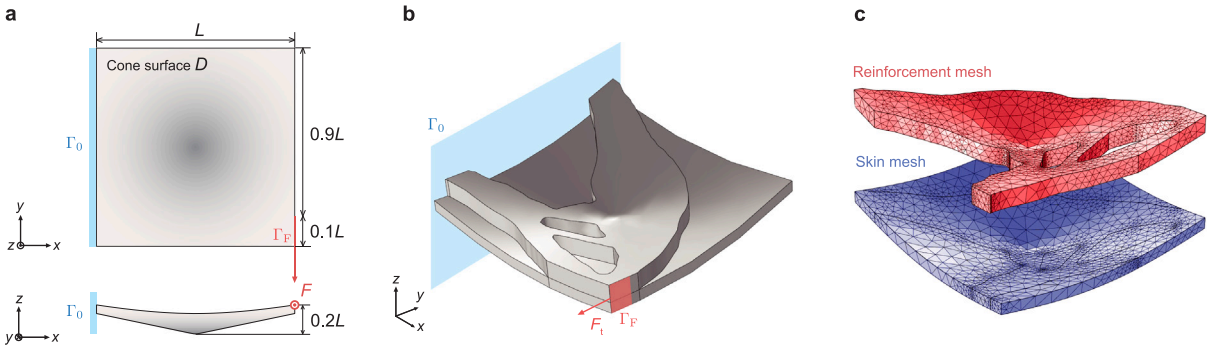

Fig. 10. Setup for the LF optimization and HF evaluation for stiffness maximization: (a) dimensions and boundary conditions for the LF optimization; (b) boundary conditions for the HF evaluation; (c) adaptive mesh for the HF evaluation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where α was chosen to be 0.3. The target surface was modeled by the mesh grids projected from x–y plane to the $f(x, y)$ surface using Eq. (23). For the LF optimization, the design domain was modeled as a two-dimensional shell structure with a thickness of 0.01. For the HF evaluation, the skin thickness was set to 0.01 and the maximum reinforcement thickness was set to 0.1.

The setup for LF optimization, including dimensions, loading conditions, and boundary conditions, is depicted in Fig. 10(a). The design domain D is modeled as a cone surface with square-trimmed edges, where the width is defined by the reference length L. A uniform edge load of $\boldsymbol{F} = 1$ is applied at the lower right corner Γ_{F} in the negative y direction. The Dirichlet boundary condition $\boldsymbol{U} = 0$ is applied to the left wall Γ_0 in the design domain D. This setup is widely recognized as the cantilever beam problem, a prevalent benchmark for topology optimization, albeit with the design domain D being a three-dimensional bending surface.

The HF configuration, particularly regarding loading and boundary conditions, is depicted in Fig. 10(b). These configurations aim to extend the LF optimization settings for thin-walled shell models, as shown in Fig. 10(a), to those suitable for extruded solid models. Consequently, the edge load in the LF configuration is broadened to an area load in the HF setup. Tetrahedral meshes for the skin and reinforcement structures are shown in Fig. 10(c), colored in blue and red, respectively. The mesh density is adjusted based on proximity to the boundary surfaces, enhancing the accuracy of stress evaluation in areas where stress is likely to concentrates.

5.1.2. Results & discussion

The material distributions derived from the LF optimization are depicted in Fig. 11(a), organized by the maximum volume constraint $V_{\max}$ ranging from 0.2 to 0.8. Samples with higher $V_{\max}$ exhibit a thicker topology of material distribution. The central ribs, connecting the arch-shaped outer rib to the bottom rib, are notably thinner compared to the connecting ribs. These configurations are characteristic outcomes of the cantilever beam problem addressed through topology optimization, specifically using the SIMP method, albeit with slightly thicker and curved ribs at the cone's apex. This curvature is attributed to the top of the cone being more prone to deformation without significant support, leading to a potential surge in the strain energy, represented by $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{U}$. Consequently, the results from the LF stiffness maximization align qualitatively with the anticipated outcomes.

Since the thickness is constant in this numerical example, the thickness distribution of the HF model is obtained by multiplying the reinforcement region mask and a random scalar value. To prepare more comprehensive initial designs for the HF evaluation, the random scalar value and LF parameter $V_{\max}$ were randomly sampled using LHF, as illustrated in Fig. 11(b). The samples are sorted based on the assigned random scalar values for the thickness distribution. Fig. 11(b) demonstrates that the initial dataset comprehensively represents all categories of HF parameters, thanks to the even distribution achieved by LHS. This dataset was then prepared for input into the MC-VAE, excluding samples that did not meet the $V_{\max}$ constraint, as material boundaries might shift during the smoothing phase.

To validate the proposed framework, 100 reference solutions were obtained by solving the same stiffness maximization problem, and then extruded to heights uniformly sampled from the same range as the proposed framework's HF parameters. The reference solutions were then evaluated using the HF model, and the objective values were compared with those from the proposed framework. Objective values derived from the final solutions of the proposed framework and those from the reference solutions are juxtaposed in Fig. 12. This figure indicates the effectiveness of the proposed framework in identifying high-performance solutions across broader objective spaces, as the proposed framework's solutions with lower volume exhibit considerably superior performance compared to the reference solutions, with higher volume's performance being equivalent. This implies that such a parametric study by uniformly sampled HF parameters can yield suboptimal performance without a comprehensive search for more global solutions through the exploration of enormous HF parameters.

The structural performance of the optimized model is compared with the reference model in Fig. 13 on the right and left, respectively. Although the volume of both models is same, as shown in Fig. 13(a), the optimized model's reinforcement has a taller and thinner sectional shape without a middle bridge that the reference model has, leading to a 24% lower strain energy than that of

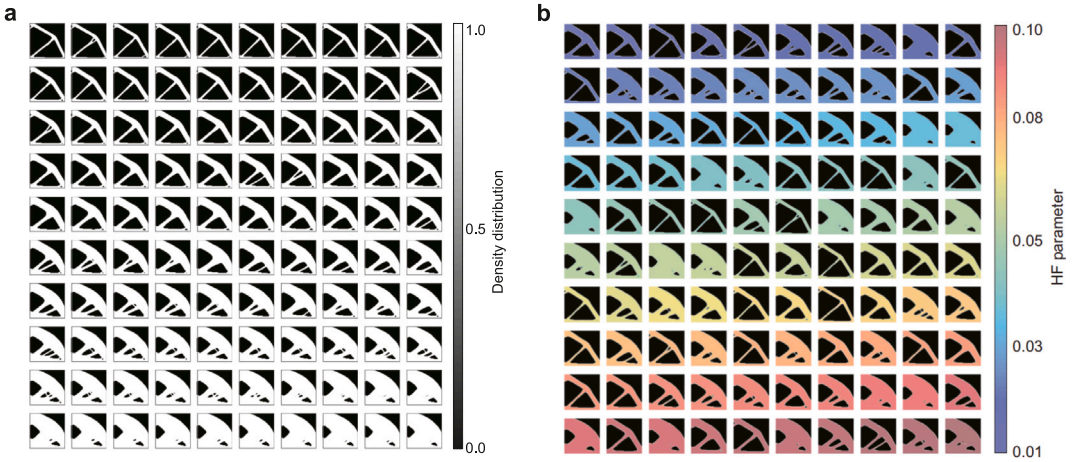

Fig. 11. Material distributions used as initial designs: (a) material distributions obtained from LF stiffness maximization; (b) initial designs constructed by assinging HF parameters to corresponding material distributions.

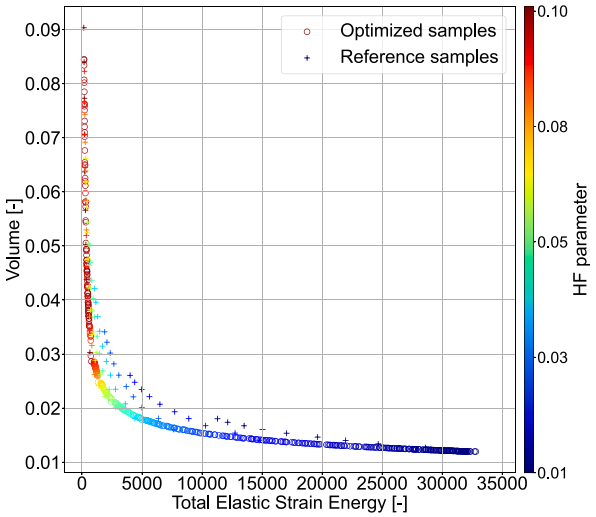

Fig. 12. Comparison of objective functionals with color of HF parameters between the optimized and reference samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the reference model. This is evident in the stress and displacement fields, as shown in Fig. 13(b) and (c), respectively. Both models exhibit similar displacement and stress distributions, where the displacement is significant at the top and bottom right corner and the stress concentrates on the top and bottom corners of design and the center of the design domain with the most curvature. This is because their topological features are nearly identical, however, the optimized model's reinforcement is more efficiently supporting the structure with thinner sectional shape and larger height, resulting in the entirely lower stress and displacement fields than the reference model.

The evolution of the hypervolume indicator from the stiffness maximization is depicted in Fig. 14. Hypervolume indicators were computed using $r_{\mathrm{hv}} = [J_1, J_2] = [35000, 0.1]$, where the objective functions J_1 and J_2 represent strain energy and volume, respectively. The optimization process converged at the 24th iteration. This occurred when the relative error of the hypervolume indicator dropped below the threshold. The hypervolume indicator improved over iterations, indicating an expansion of the Pareto front towards the utopia point in the objective space. The improvements in the hypervolume indicator were 2.0%. Notably, the convergence of these cases occurred significantly sooner, and the enhancements in the hypervolume indicator were more modest compared to the outcomes for the data-driven MFTD in previous studies [23], which tackled thermal-fluid problems characterized by greater non-linearity than the stiffness maximization problem. The earlier convergence and lesser improvement in the hypervolume indicator for stiffness maximization should be attributed to the relative simplicity of its optimization problem.

Fig. 15 displays the denormalized images decoded from the MC-VAE at the (a) initial and (b) last iteration, where the images are sorted by the corresponding total volume values. At the last iteration, the simpler topologies with center ribs and no holes are

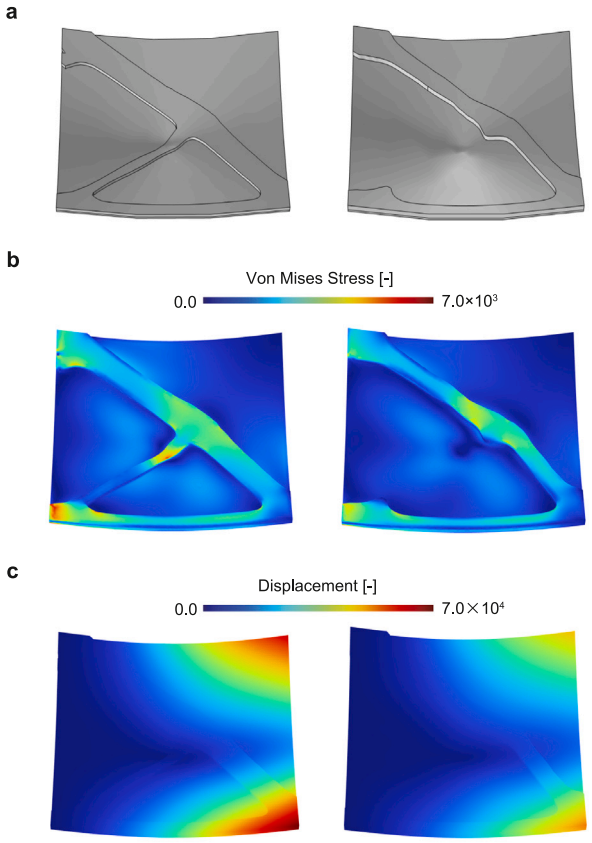

Fig. 13. Comparison of structural performance between reference (left) and optimized (right) model: (a) HF model; (b) stress field; (c) displacement field. The volume of both models is 0.015, whereas the strain energy is 1.3×10^4 (reference); 9.9×10^3 (optimized).

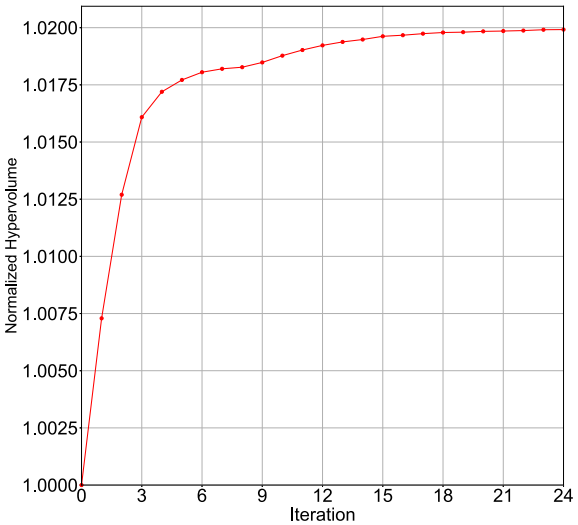

Fig. 14. Iteration history of hypervolume indicator for stiffness maximization.

more prevalent, however, as the total volume increases, the topologies feature more complex reinforcements with holes and ribs to support the entire structure more effectively. This trend is consistent with the height of reinforcement as well; as the total volume increases, the height of reinforcement increases, and the strain energy decreases. These trends validate that the MC-VAE effectively

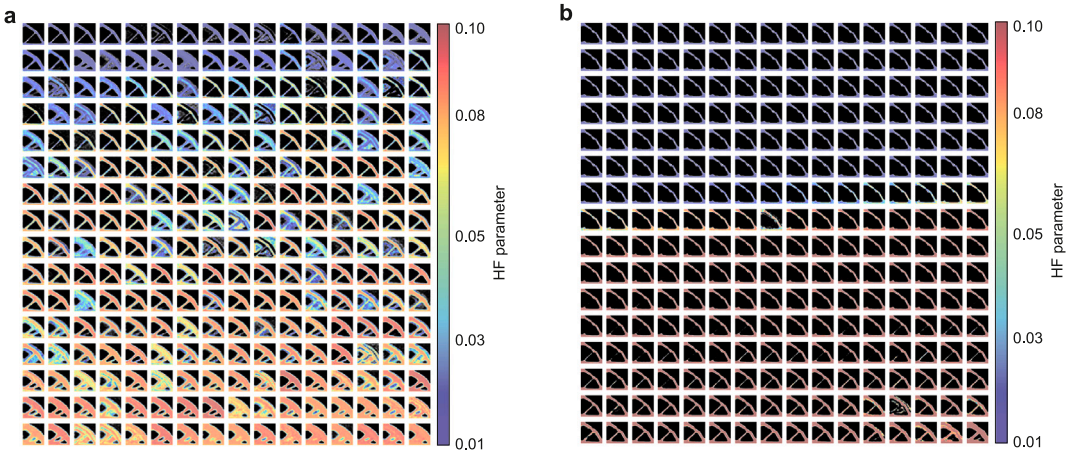

Fig. 15. Denormalized images decoded by MC-VAE: (a) at the initial iteration and (b) at the last iteration.

learns the physical relationship between the sectional shape and height of reinforcements, stored in the first and second channels, respectively.

Additionally, the material boundaries of the material distributions decoded at the last iteration are more distinct than those decoded at the initial iteration. This discrepancy arises because the dataset for the MC-VAE has no physical relationship between the material distribution and the HF parameter at the initial iteration, increasing the difficulty in effectively learning the relationship between the two channels.

5.2. Case 2: maximum stress minimization optimizing constant thickness

Now we apply the proposed framework, validated through its application to a fundamental topology optimization problem, to the maximum stress minimization that involves strong non-linearity. The problem setup is identical to that of Case 1, as detailed in Section 5.1.

5.2.1. Comparison with conventional framework

To assess the capability of the proposed data-driven MFTD in identifying high-performance samples throughout the HF parameter space, the conventional data-driven MFTD approach was employed for maximum stress minimization using the same problem settings and HF parameters $h = [0.01, 0.04, 0.07, 0.1]$. The objective functionals derived from both the conventional data-driven MFTD with SC-VAE and the proposed data-driven MFTD with MC-VAE are illustrated in Fig. 16. The conventional data-driven MFTD with SC-VAE yielded samples that are partly positioned on the Pareto front for the MC-VAE case, whereas samples with smaller volumes and higher maximum stress are often positioned away from the Pareto front, indicating they are less effective. Predicting the placement of optimized samples within the objective space prior to applying the conventional data-driven MFTD with a singular HF parameter is challenging. Therefore, to achieve a Pareto front as comprehensive as that obtained with the proposed data-driven MFTD, extensive parametric studies involving various HF parameters would be necessary if relying solely on the conventional data-driven MFTD. This underscores the significant advantage of the proposed data-driven MFTD in globally searching for optima across material distributions and HF parameters. It is noteworthy that the conventional data-driven MFTD with $h = 0.1$ yielded some samples outperforming those from the enhanced data-driven MFTD, suggesting that the latter could further enhance its capability to discover more promising solutions through a more global search across material distributions and HF parameters.

The structural performance of the optimized model obtained from LF maximum stress minimization and LF stiffness maximization, and the initial model from LF maximum stress minimization are juxtaposed in Fig. 17 on the right, middle and left, respectively. Although all models have the same volume of 0.015, as shown in Fig. 17(a), the initial model's reinforcement has a more complex topology with partially narrow ribs and many holes, causing a higher stress concentration as shown in Fig. 17(b). The optimized model from both LF maximum stress minimization and LF stiffness maximization exhibit simpler topologies with uniformly thick ribs and fewer holes, leading to lower stress concentrations as shown in Fig. 17(b). Interestingly, while the maximum stress increases in the order of the stress-based optimized model, the stiffness-based optimized model, and the stress-based initial model, the displacement follows a different trend, becoming larger in the order of the stiffness-based optimized model, the stress-based initial model, and the stress-based optimized model, as shown in Fig. 17(c). This discrepancy arises from the stress concentration depending on the topology of the reinforcement more strongly than the stiffness or deformation, clearly demonstrated by comparing the optimized models from LF stiffness maximization and LF maximum stress minimization, where the latter exhibits a lower maximum stress but a higher displacement. It is noteworthy that the stiffness-based model has the lower displacement than both stress-based initial and optimized models, indicating that the optimized model from LF stiffness maximization reduces the stress concentration by increasing the stiffness, as expected.

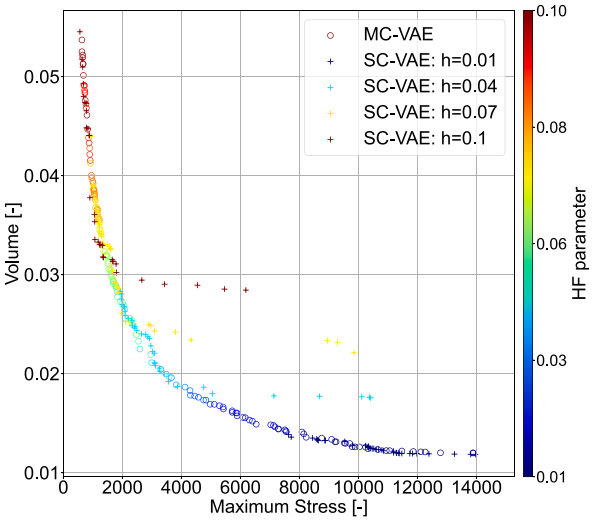

Fig. 16. Comparison of objective functionals with color of HF parameters between the conventional data-driven MFTD with SC-VAE and the expanded data-driven MFTD with MC-VAE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

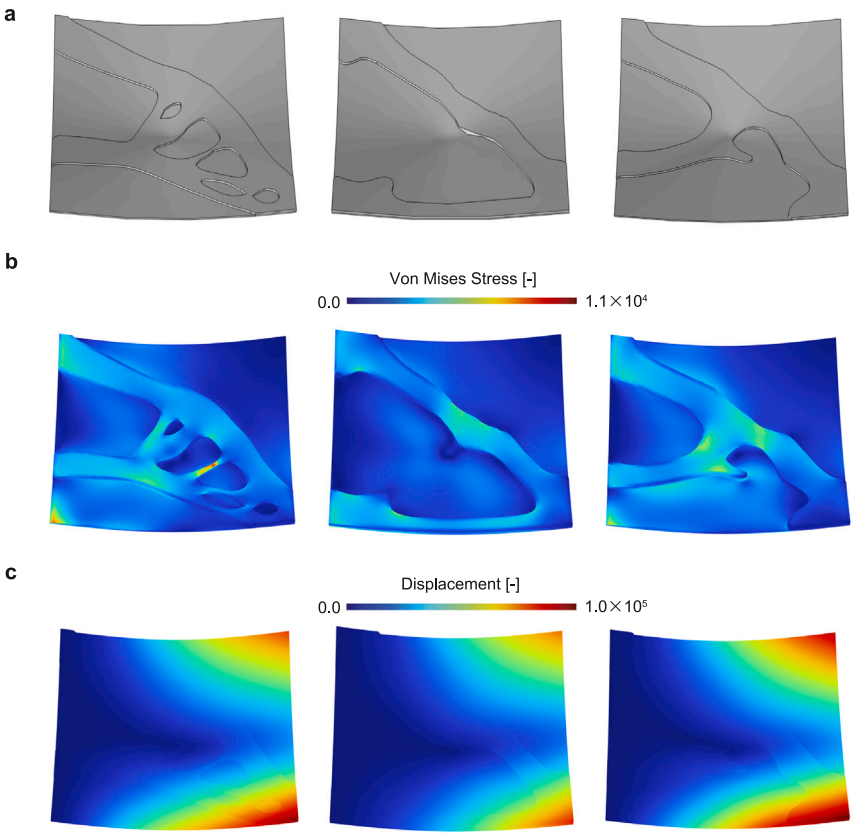

Fig. 17. Comparison of structural performance between stress-based initial (left), stiffness-based optimized (middle), and stress-based optimized (right) model: (a) HF model; (b) stress field; (c) displacement field. The volume of these models are 0.015, whereas the maximum stress is 1.1×10^4 (stress-based initial); 7.3×10^3 (stiffness-based optmized); 7.2×10^3 (stress-based optimized).

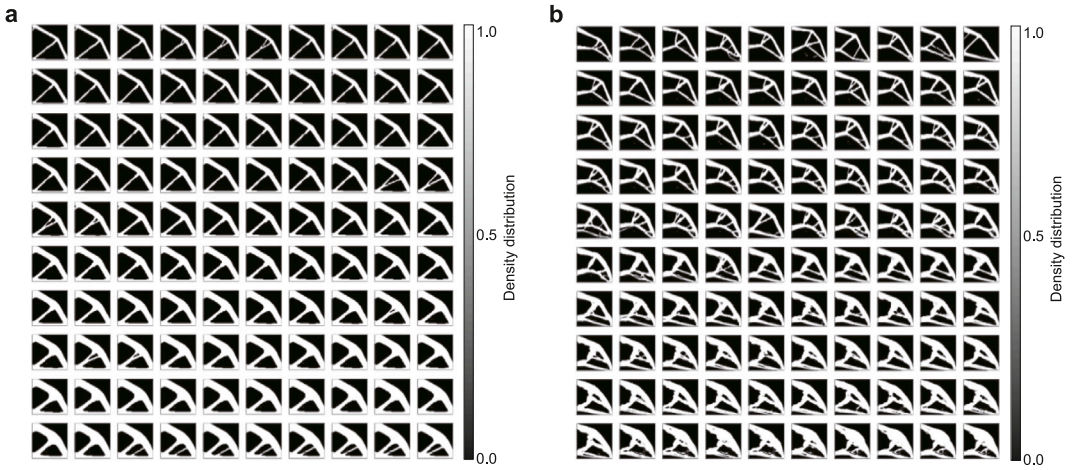

Fig. 18. Comparison of material distribution with $V_{\max} \in [0.2, 0.5]$ from (a) LF stiffness maximization and (b) LF maximum stress minimization.

5.2.2. *Effect of low-fidelity optimization formulation*

In addition to the comparison with the conventional data-driven MFTD, the proposed data-driven MFTD was further investigated by comparing the final optimized results from two different couplings of LF optimization and HF evaluation: LF stiffness maximization and HF stress minimization, and LF maximum stress minimization and HF stress evaluation. Initially, maximum stress minimization was approached through LF stiffness maximization coupled with HF stress evaluation. This strategy is anticipated to indirectly yield promising solutions characterized by low maximum stress and volume by identifying samples with reduced strain energy, indicative of higher stiffness. In a linearly elastic system, a structure with greater stiffness is expected to undergo less deformation under identical loading conditions, potentially leading to decreased stress levels within the structure, given constant applied forces. While LF optimization for stiffness maximization is an indirect method, it offers the benefit of generally lacking non-linearity and allows for the derivation of material distributions without the need for relaxation or approximation methods. In the second coupling, maximum stress minimization was tackled using LF approximated maximum stress minimization and HF stress evaluation, a more direct approach. This method is expected to achieve promising solutions less indirectly through the objective functional of stress approximated by a p-norm, which, while still indirect, is more straightforward than stiffness maximization, where the objective functional is strain energy. A more direct LF optimization could potentially enable the samples that inherit the LF results to attain higher performance after iterative processes.

The material distributions resulting from the LF stiffness maximization and LF maximum stress minimization are depicted in Fig. 18, organized according to the maximum volume constraint $V_{\max}$ ranging from 0.2 to 0.5. The topology of these outcomes with LF stiffness maximization closely resembles those derived from the stiffness maximization, as illustrated in Fig. 11. This resemblance is consistent with the slight distinction between two LF stiffness maximization problems, where the loading area is marginally broader for the maximum stress minimization problem.

Fig. 18(b) reveals a wide variety of topologies can be observed in the scenario with LF maximum stress minimization, even with results based on similar volume constraints showing significant differences. This diversity stems from the strong non-linearity inherent in the approximated maximum stress minimization problem, leading to a tendency for the optimization process to become ensnared in local optima. Furthermore, these local optima often feature small holes and pronounced steep curves, contributing to stress concentration. Such characteristics, arising from the approximated maximum stress minimization, present substantial opportunities to reduce the true objective functional, specifically, the original von Mises stress.

The evolution of the hypervolume indicator for the scenarios using LF stiffness maximization and LF maximum stress minimization is depicted in Fig. 19, following normalization against the initial hypervolume indicator from the LF maximum stress minimization scenario. The hypervolume indicators were computed using $r_{\text{hv}} = [J_1, J_2] = [14000, 0.06]$, with J_1 and J_2 representing maximum stress and volume, respectively. The optimization concluded at iteration $i = 30$ for LF stiffness maximization and at iteration $i = 34$ for LF maximum stress minimization, upon the relative error of the hypervolume indicator falling below the threshold $\varepsilon_{\text{HV}} = 1.0 \times 10^{-5}$. The improvement in the hypervolume indicator was noted as 4.2% for LF stiffness maximization and 5.2% for LF maximum stress minimization, relative to the initial value from LF maximum stress minimization. It is noteworthy that the initial hypervolume for the LF stiffness maximization scenario was higher than that for the LF maximum stress minimization. This discrepancy arises because LF maximum stress minimization yielded topologies that were qualitatively sound but included numerous stress-concentrating features due to its pronounced non-linearity. Once the iterative evolutionary process commenced, the hypervolume indicator for the LF maximum stress minimization scenario immediately exceeded that of the LF stiffness maximization scenario, improving the qualitatively sound topologies to push the Pareto front towards the utopia point in the objective space. Furthermore, MC-VAE facilitated this rapid advancement by generating samples characterized by fewer holes and smoother material boundaries, effectively blending the features of all samples in the dataset.

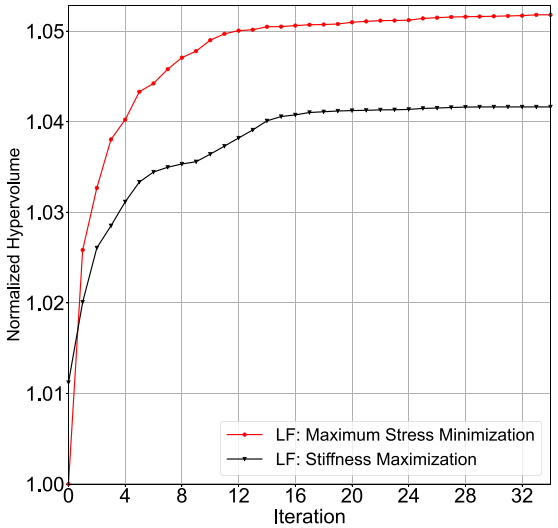

Fig. 19. Comparison of iteration history of hypervolume indicator between LF stiffness maximization and LF maximum stress minimization.

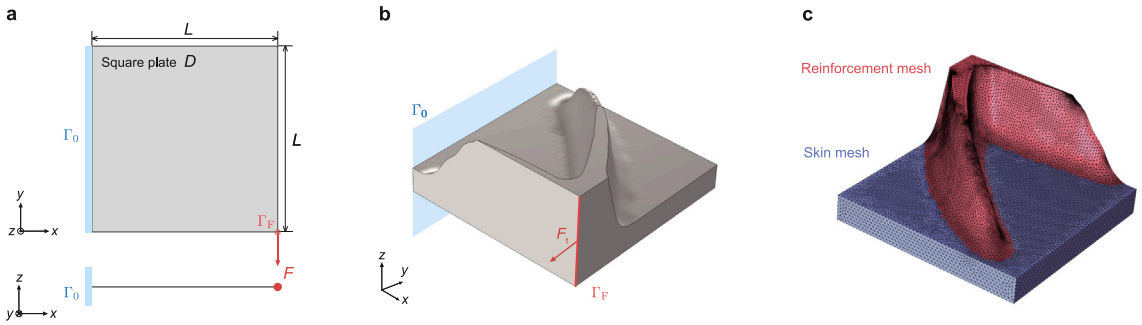

Fig. 20. Setup for the LF optimization and HF evaluation for stiffness maximization: (a) dimensions and boundary conditions for the LF optimization; (b) boundary conditions for the HF evaluation; (c) adaptive mesh for the HF evaluation.

5.3. Case 3: stiffness maximization optimizing spatially varying thickness

5.3.1. Problem setup

As shown in Fig. 20(a), the setup for LF optimization, including dimensions, loading conditions, and boundary conditions, is identical to that of Case 1 and 2, except the design domain is a two-dimensional square surface. Following this slight modification, the skin in the HF evaluation is modeled as a box with a square cross-section, as shown in Fig. 20(b). Fig. 20(c) illustrates the tetrahedral mesh for the skin and reinforcement structures, where the reinforcement mesh is smoothly connected to the skin mesh since its thickness is spatially varying.

For the initial dataset and mutated dataset, the reinforcement layout was initialized using material distribution directly obtained from the optimized results of LF stiffness maximization, whereas the thickness distribution was initialized using displacement fields subsequently obtained from the optimization. This initialization strategy was adopted to ensure that the initial dataset is composed of samples with qualitatively sound reinforcement layouts and thickness distributions, thereby easing the exploration of the design space.

5.3.2. Results & discussion

In order to validate the capability of the proposed framework to explore the multi-channel design space, we conducted a side-by-side comparison of the optimized results from our multi-channel D2MFTD and 3D topology optimization with casting constraints. Fig. 21(a) illustrates the comparison of the objective functional values resulting from MC-D2MFTD at the initial and final iterations and 3D topology optimization with casting constraints. This result clearly shows that the proposed framework can obtain solutions with equivalent or superior performance to those from 3D topology optimization with casting constraints. Additionally, it is interesting to note that the objective functional values at the last iteration moved largely towards the right bottom corner of the objective space, indicating that the robustness of the proposed framework against initial conditions. This can be quantitatively

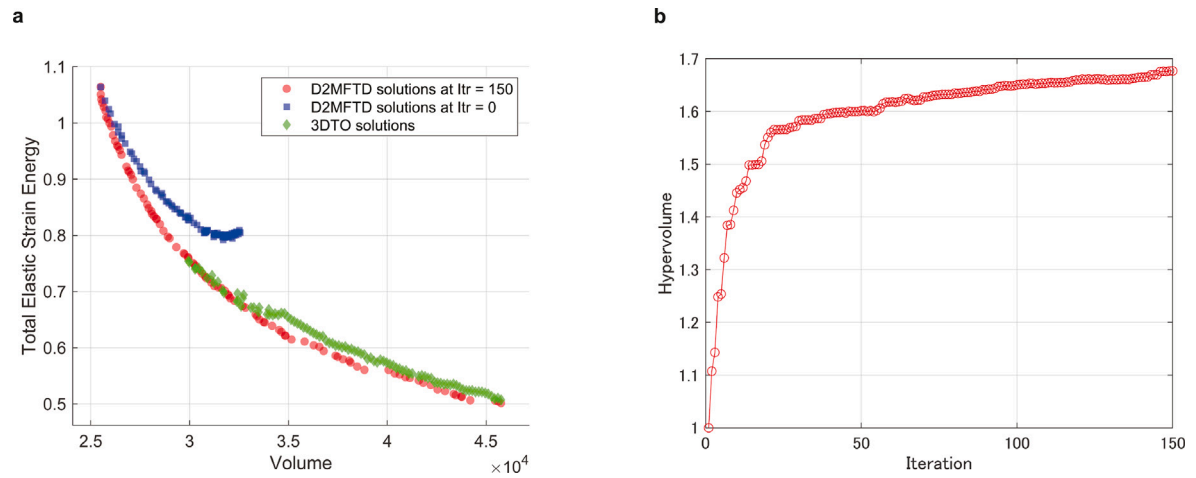

Fig. 21. Optimization results: (a) comparison of objective functional values at itr = 0 and itr = 150; (b) iteration history of hypervolume.

Table 4
Computational time of each method; unit is second.

Method	3DTO	2DTO (LF)	HF evaluation	Crossover	2DTO (mutation)	Selection	Total
MC-D2MFTD	–	32.4	14 946	36 229	152.8	1.3	51 329
3DTO with casting constraints	39 182	–	63.2	–	–	–	39 245

confirmed by the hypervolume indicator history depicted in Fig. 21(b), where the hypervolume indicator improved by 68% over the optimization process.

To evaluate the computational efficiency of the proposed framework, the computational time of each method was measured and summarized in Table 4. 3D topology optimization with casting constraints required 39 182 s to complete, whereas MC-D2MFTD took 51 329 s in total. 3D topology optimization, LF optimization, HF evaluation, and mutation process were conducted on 50 threads in parallel, whereas crossover and selection processes were conducted on 128 core single-threaded. In the MC-D2MFTD, the HF evaluation and crossover operations consumed the most time, accounting for 14 946 and 36 229 s, respectively. The HF evaluation process for each sample is not computationally intensive in this case, but this process is conducted on around 150 solutions for each iteration, leading to a significant computational time. The crossover process is also computationally intensive, as the MC-D2MFTD employes the learning process of the MC-VAE to generate new samples every iteration. Note that this process is expected to reduce the computational time significantly by utilizing a parallel computing environment such as a GPU.

Furthermore, it is critical to note that the proposed framework can optimize the design variables based on the accurate HF evaluation using the adaptive mesh, whereas the 3D topology optimization with casting constraints is conducted on a fixed mesh during the optimization process.

The structural performances of the optimized model obtained from the proposed framework and 3D topology optimization with casting constraints are compared in Fig. 22. Fig. 22(a) visualizes the HF model, stress field, and displacement field of specific samples with same volume of 3.5×10^4. The optimized model from the proposed framework exhibits a more efficient reinforcement structure with a thinner sectional shape and larger height, leading to a 5% lower strain energy than that of the 3D topology optimization with casting constraints. This is evident in the displacement fields, as shown in Fig. 22(b), where the optimized model exhibits lower displacement at the bottom right corner. This result confirms that the proposed framework can effectively explore the multi-channel design space when optimizing spatially varying thickness, leading to superior performance compared to 3D topology optimization with casting constraints.

5.4. Case 4: maximum stress minimization optimizing spatially varying thickness

5.4.1. Problem setup

Fig. 23(a) illustrates the L-shaped design domain for the LF optimization, where the top edge is fixed and the force is applied on the right tip. This setup is widely recognized as the L-bracket problem, a prevalent benchmark for maximum stress minimization. The HF configuration is depicted in Fig. 23(b), where the loading and boundary conditions are adjusted to accommodate the extruded L-bracket model. Similar to Case 3, the tetrahedral mesh of the reinforcement structures is smoothly connected to the skin mesh, as shown in Fig. 23(c).

For the initial dataset and mutated dataset, the reinforcement layout was initialized using material distribution directly obtained from the optimized results of LF maximum stress minimization, whereas the thickness distribution was initialized using stress fields subsequently obtained from the same optimization. The intention of using stress fields for thickness initialization lies in the

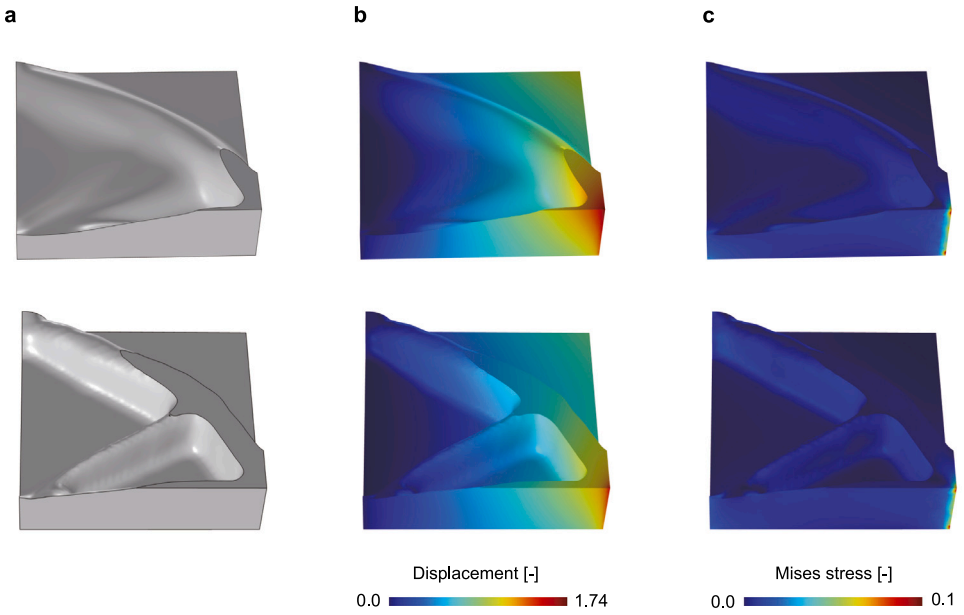

Fig. 22. Comparison of structural performance between 3D TO (top) and MC-D2MFTD (bottom) model: (a) HF model; (b) displacement field; (c) stress field. The volume of these models are 3.5×10^4, whereas the total elastic strain energy is 0.65 (3D TO); 0.62 (MC-D2MFTD).

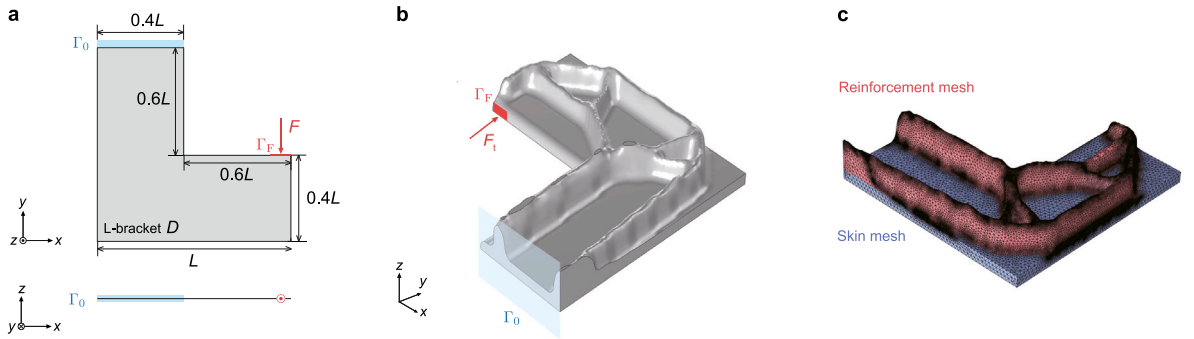

Fig. 23. Setup for the LF optimization and HF evaluation for stiffness maximization: (a) dimensions and boundary conditions for the LF optimization; (b) boundary conditions for the HF evaluation; (c) adaptive mesh for the HF evaluation.

assumption that the stress distribution is directly related to the thickness distribution, and the stress concentration can be effectively reduced by adjusting the thickness distribution.

5.4.2. Results & discussion

To ensure that the LF optimization process is conducted properly, the convergence history of the LF optimization is depicted in Fig. 24. In Fig. 24, the convergence history is visualized as a statistical plot of the objective functional values, where the mean and standard deviation are computed from the objective functional values of the population at each iteration. Since the LF optimization is conducted on $N_{\mathrm{LF}} = 100$ samples, the convergence history is visualized as a statistical plot of the objective functional values. The mean and standard deviation are computed from the objective functional values of the population at each iteration.

The standard deviation of the objective functional values are extremely high at the initial iteration, but decrease as the optimization progresses. This is reasonable because the volume fraction for the constraint is different for each sample, where the initial design variables are uniform in the design domain. Additionally, both of mean and standard deviation of the objective functional values decrease by stages, exactly when the continuation method is applied to the *beta* in the projection method (15). With all of these results, it is confirmed that the LF optimization process converged properly.

Fig. 25(a) illustrates the comparison of the objective functional values resulting from the proposed framework at the initial and final iterations. The objective functional at the final iteration is completely dominant over that at the initial iteration. The initial solutions have a wide range of maximum stress values, with the maximum stress minimization problem being inherently difficult

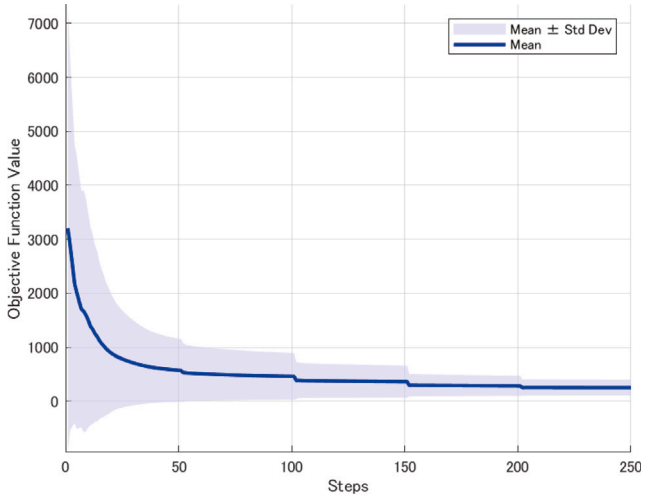

Fig. 24. Statistical plot of convergence history for LF optimization.

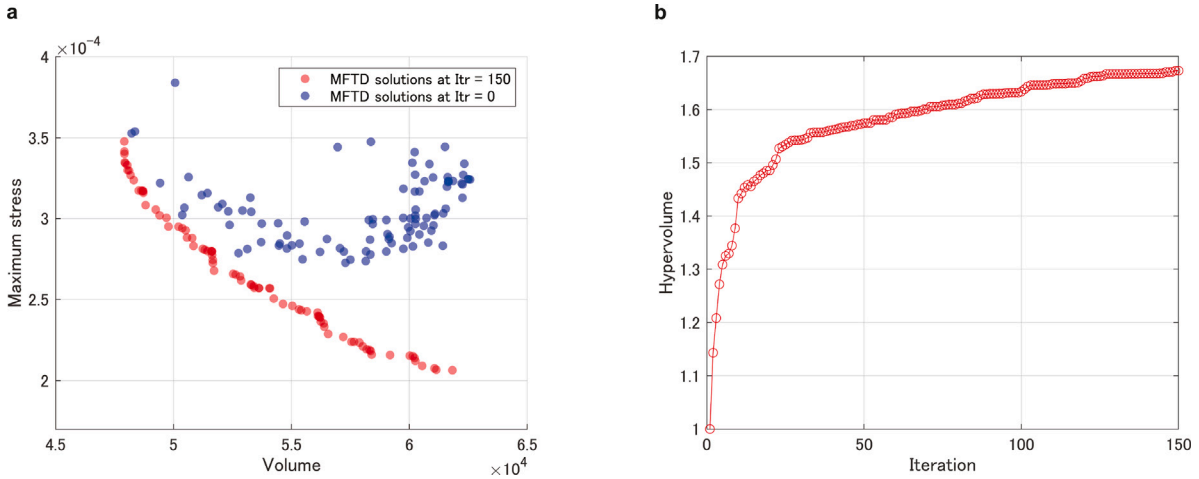

Fig. 25. Optimization results: (a) comparison of objective functional values at itr = 0 and itr = 150; (b) iteration history of hypervolume.

to solve directly due to its pronounced non-linearity. This difficulty is reflected in the initial dataset, where the maximum stress values are scattered across the objective space. The proposed framework successfully indentified high-performance solutions when optimizing spatially varying thickness, as the hypervolume indicator improved by 67% over the optimization process, as shown in Fig. 25(b).

The initial and optimized models are compared in Fig. 26. The upper row of Fig. 26 illustrates the HF models, whereas the lower row depicts the stress distributions. Fig. 26(a) and (c) show the initial models with two different volumes of 5.4×10^4 and 5.9×10^4, whereas Fig. 26(b) and (d) show the optimized models with the same volumes as (a) and (c), respectively. In all depicted models, the stress concentrates on the inner corner of the L-bracket, which is physically reasonable. When comparing Fig. 26(a) and (b), the stress concentration in the optimized model is reduced by 16% compared to the initial model with the same volume. This large reducion in stress concentration with same volume is because the optimized model did not obtain a short rib in the most right hole of reinforcement but extended reinforcement to the loading area, leading to a more efficient reinforcement structure. The optimized model with a volume of 5.9×10^4 in Fig. 26(d) exhibits a 27% reduction in stress concentration compared to the initial model with the same volume in Fig. 26(c). This significant reduction in stress concentration is not only due to the reinforcement extention to the loading area but also due to the sharpening of the reinforcement boundary, leading to a smaller stress/volume ratio. Considering the optimized results described above, the proposed framework can effectively explore the multi-channel design space when optimizing spatially varying values for strongly non-linear problems.

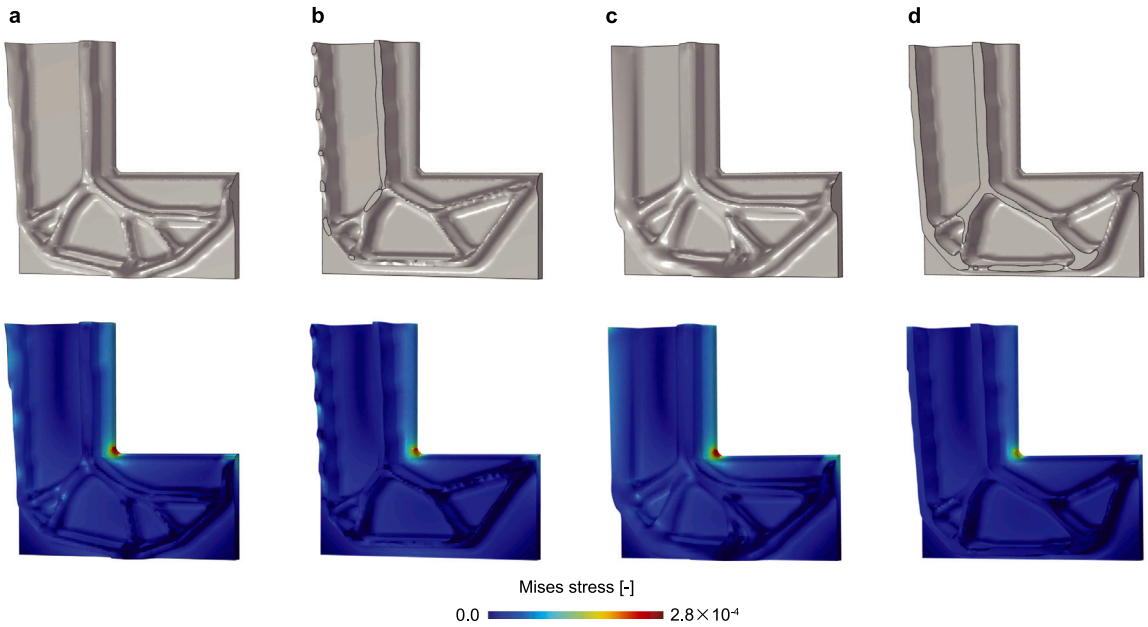

Fig. 26. Comparison of stress distribution between initial designs and optimized designs: (a) initial design with volume 5.4×10^4 and max stress 3.1×10^{-4}; (b) optimized design with volume 5.4×10^4 and max stress 2.6×10^{-4}; (c) initial design with volume 5.9×10^4 and max stress 3.0×10^{-4}; (d) optimized design with volume 5.9×10^4 and max stress 2.2×10^{-4}.

6. Conclusions

We proposed a data-driven MFTD framework that expands the search space for HF parameters, previously constrained to a single constant value across the whole optimization, aiming to reduce user dependency on finding optimal solutions. This framework leverages a multi-channel image representation of material distributions and HF parameters, enabling the exploration of both material distributions and HF parameters in a single optimization process. The effectiveness of this framework is demonstrated through four numerical examples: stiffness maximization and maximum stress minimization problems with constant thickness, and stiffness maximization and maximum stress minimization problems with spatially varying thickness.

The first example for stiffness maximization revealed that the proposed framework can successfully identify promising solutions with structural performance that is superior to that of the reference solutions by efficiently exploring both material distributions and HF parameters. Additionally, within the data-driven MFTD's crossover process for multi-channel images, the implementation of an oversampling technique significantly improved the overall search performance by enriching the dataset that includes samples with underrepresented HF parameters.

Within the second example, the proposed framework demonstrated its ability to identify promising solutions for maximum stress minimization problems, even in the presence of significant non-linearity. It showed a distinct advantage in globally searching for optima across material distribution and HF parameter spaces compared to the traditional data-driven MFTD method employing a SC-VAE. The proposed framework significantly reduces computational demands and user reliance on finding optimal solutions by enabling a more comprehensive optimization in a single run, unlike traditional methods that require extensive parametric studies with varying HF parameters.

Additionally, solutions for the maximum stress minimization were derived through two distinct combinations of LF optimization and HF evaluation, where the combination of LF maximum stress minimization and HF maximum stress evaluation is more straightforward and the combination of the LF stiffness maximization and HF maximum stress evaluation is more indirect yet widely utilized in mechanical engineering practices. As a result, the combination of LF maximum stress minimization and HF maximum stress evaluation led to more promising solutions by refining the initial dataset derived from a more direct LF topology optimization approach, which suggests that the synergy between LF topology optimization and HF evaluation critically influences the search efficiency for optimal solutions.

The proposed framework was further validated through two additional examples: stiffness maximization and maximum stress minimization problems with spatially varying thickness. The results of the stiffness maximization with spatially varying thickneses example demonstrated that the proposed framework can effectively explore the multi-channel design space, leading to superior performance compared to 3D topology optimization with closely related casting constraints. The maximum stress minimization with spatially varying thickness example revealed that the proposed framework can efficiently navigate the solution space by concurrently optimizing two design variable fields, material distributions, and spatially varying thickness, for strongly non-linear problems.

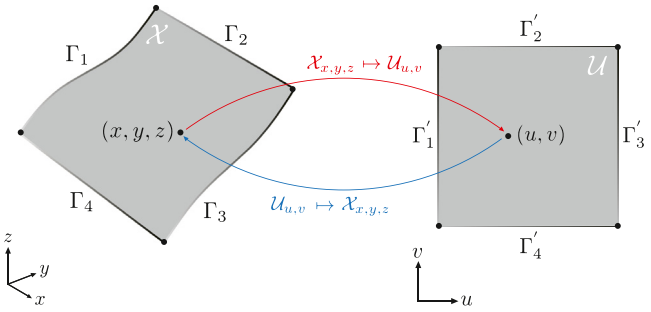

Fig. A.27. The concept of design domain mapping with notations.

CRediT authorship contribution statement

Hiroki Kawabe: Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kentaro Yaji:** Writing – review & editing, Software, Resources, Funding acquisition. **Yuichiro Aoki:** Supervision, Software, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by JSPS, Japan KAKENHI, Grant Number 21J22284 and 23H03799.

Appendix. Design domain mapping

Let $\boldsymbol{x}_i = (x_i, y_i, z_i)$ represent the 3D position of the ith node within a mesh patch $\mathcal{X}$, and $\boldsymbol{u}_i = (u_i, v_i)$ denote the 2D position (parameter value) of the corresponding node within the 2D unit plane mesh $\mathcal{U}$, as illustrated in Fig. A.27. The boundaries of the original mesh are divided into $\Gamma_1 - \Gamma_4$, corresponding to the boundaries of the unit plane mesh denoted as $\Gamma_1' - \Gamma_4'$, respectively.

The boundary conditions of the unit plane mesh are divided as follows:

$$\begin{aligned}
\Gamma_1' &:= \{(u,v) | u = 0, 0 \le v \le 1\},\\
\Gamma_2' &:= \{(u,v) | 0 \le u \le 1, v = 1\},\\
\Gamma_3' &:= \{(u,v) | u = 1, 0 \le v \le 1\},\\
\Gamma_4' &:= \{(u,v) | 0 \le u \le 1, v = 0\}.
\end{aligned} \tag{A.1}$$

Then the mapping $\mathcal{X}_{x,y,z} \mapsto \mathcal{U}_{u,v}$ is computed by solving the following Laplace equation in the original mesh:

$$\begin{aligned}
\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0,\\
\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} = 0,
\end{aligned} \tag{A.2}$$

under the following Dirichlet boundary conditions:

$$\begin{aligned}
(u,v) &= (0, t_1(x,y,z)) \quad \text{on} \quad \Gamma_1,\\
(u,v) &= (t_2(x,y,z), 1) \quad \text{on} \quad \Gamma_2,\\
(u,v) &= (1, t_3(x,y,z)) \quad \text{on} \quad \Gamma_3,\\
(u,v) &= (t_4(x,y,z), 0) \quad \text{on} \quad \Gamma_4.
\end{aligned} \tag{A.3}$$

Herein, t_1 denotes the arc length ratio of $\Gamma_1^{x,y,z}$ and Γ_1, where $\Gamma_1^{x,y,z}$ denotes the boundary from the connection point of Γ_4 and Γ_1 to a point (x, y, z) on Γ_1. Solving the Laplace equation indicates the mapping minimizes the Dirichlet energy for a smooth map $\boldsymbol{u}$ from a differential surface patch $\mathcal{X}$ to its image $\mathcal{U}$, defined as follows:

$$E_D = \frac{1}{2} \int_{\mathcal{X}} |\nabla \boldsymbol{u}|^2 dA, \tag{A.4}$$

where $\nabla \boldsymbol{u}$ means the unit mesh $\mathcal{U}$'s gradient. This projection distorts the mesh to accurately represent the surface's three-dimensional contours on a two-dimensional plane, ensuring all mesh elements fit within a unit square boundary. It can be said that DDM is

formulated as a *harmonic mapping* [46], which minimizes distortion in the mapping by mapping each node on the surface so that the Dirichlet energy is minimized given fixed boundary conditions. Harmonic mapping can smoothly map the interior of the surface while maintaining specified boundary conditions.

Harmonic mapping is equivalent to *conformal mapping* [47] when applied under Dirichlet boundary conditions that secure all boundaries within the design domain. The primary goal of conformal mapping is to maintain local angles throughout the mapping process by minimizing the *conformal energy*, which includes the Dirichlet energy as defined in Eq. (A.4), alongside the area of the design domain. The conformal mapping has been widely used in topology optimization to realize the topology optimization problem on complex surfaces [48–50]. Given that the design domain's area remains constant under fixed boundary conditions, minimizing conformal energy effectively reduces the Dirichlet energy, thereby decreasing distortion in the mapping. This rationale underpins the selection of the conformal mapping algorithm, favored for its straightforward implementation in Python, the primary development environment for this study. The implementation of the conformal mapping function leverages the IGL Python library based on several works [51–53].

Data availability

Data will be made available on request.

References

[1] M.P. Bendsøe, N. Kikuchi, Generating optimal topologies in structural design using a homogenization method, Comput. Methods Appl. Mech. Engrg. 71 (1988) 197–224.
[2] O. Sigmund, K. Maute, Topology optimization approaches, Struct. Multidiscip. Optim. 48 (2013) 1031–1055.
[3] J.D. Deaton, R.D. Grandhi, A survey of structural and multidisciplinary continuum topology optimization: post 2000, Struct. Multidiscip. Optim. 49 (2014) 1–38.
[4] J. Zhu, W. Zhang, L. Xia, Topology optimization in aircraft and aerospace structures design, Arch. Comput. Methods Eng. 23 (2016) 595–622.
[5] C. Le, J. Norato, T. Bruns, C. Ha, D. Tortorelli, Stress-based topology optimization for continua, Struct. Multidiscip. Optim. 41 (2010) 605–620.
[6] E. Holmberg, B. Torstenfelt, A. Klarbring, Stress constrained topology optimization, Struct. Multidiscip. Optim. 48 (2013) 33–47.
[7] P. Duysinx, M.P. Bendsøe, Topology optimization of continuum structures with local stress constraints, Internat. J. Numer. Methods Engrg. 43 (8) (1998) 1453–1478.
[8] C.A.C. Coello, G.B. Lamont, D.A. Van Veldhuizen, Evolutionary Algorithms for Solving Multi-Objective Problems, Springer, 2007.
[9] C.Y. Hu, K.Y. Tseng, Topology optimization of structures using modified binary differential evolution, Struct. Multidiscip. Optim. 42 (2010).
[10] J.F.A. Madeira, H.L. Pina, H.C. Rodrigues, GA topology optimization using random keys for tree encoding of structures, Struct. Multidiscip. Optim. 40 (2010).
[11] D.J. Munk, G.A. Vio, G.P. Steven, Topology and shape optimization methods using evolutionary algorithms: a review, Struct. Multidiscip. Optim. 52 (2015) 613–631.
[12] O. Sigmund, On the usefulness of non-gradient approaches in topology optimization, Struct. Multidiscip. Optim. 43 (2011) 589–596.
[13] M. Pelikan, D.E. Goldberg, E. Cantú-Paz, A survey of optimization by building and using probabilistic models, IEEE Trans. Evol. Comput. 6 (2) (2002) 123–141.
[14] D. Foster, Generative Deep Learning: Teaching Machines to Paint, Write, and Play, O'Reilly Media, 2019.
[15] L. Regenwetter, A.H. Nobari, F. Ahmed, Deep generative models in engineering design: A review, J. Mech. Des. 144 (7) (2022) 071704.
[16] D. Lee, W.W. Chen, L. Wang, Y.-C. Chan, W. Chen, Data-driven design for metamaterials and multiscale systems: A review, Adv. Mater. 36 (8) (2024) 2305254.
[17] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, 2013, ArXiv Preprint.
[18] T. Guo, D.J. Lohan, R. Cang, M.Y. Ren, J.T. Allison, An indirect design representation for topology optimization using variational autoencoder and style transfer, in: 2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020).
[20] S. Oh, Y. Jung, S. Kim, I. Lee, N. Kang, Deep generative design: Integration of topology optimization and generative models, J. Mech. Des. 141 (11) (2019) 111405.
[21] L. Wang, S. Tao, P. Zhu, W. Chen, Data-driven topology optimization with multiclass microstructures using latent variable Gaussian process, J. Mech. Des. 143 (3) (2020) 031708.
[22] S. Yamasaki, K. Yaji, K. Fujita, Data-driven topology design using a deep generative model, Struct. Multidiscip. Optim. 64 (3) (2021) 1401–1420.
[23] K. Yaji, S. Yamasaki, K. Fujita, Data-driven multifidelity topology design using a deep generative model: Application to forced convection heat transfer problems, Comput. Methods Appl. Mech. Engrg. 388 (114284) (2022).
[24] A.I.J. Forrester, A. Sobester, A.J. Keane, Multi-fidelity optimization via surrogate modelling, in: Proceedings of the Royal Society a: Mathematical, Physical and Engineering Sciences, Vol. 463, The Royal Society, 2007, pp. 3251–3269.
[25] K. Yaji, S. Yamasaki, S. Tsushima, K. Fujita, A framework of multi-fidelity topology design and its application to optimum design of flow fields in battery systems, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 59186, American Society of Mechanical Engineers, 2019, V02AT03A059.
[26] K. Yaji, S. Yamasaki, K. Fujita, Multifidelity design guided by topology optimization, Struct. Multidiscip. Optim. 61 (3) (2020) 1071–1085.
[27] H. Kobayashi, K. Yaji, S. Yamasaki, K. Fujita, Freeform winglet design of fin-and-tube heat exchangers guided by topology optimization, Appl. Therm. Eng. 161 (114020) (2019).
[28] M. Kato, T. Kii, K. Yaji, K. Fujita, Tackling an exact maximum stress minimization problem with gradient-free topology optimization incorporating a deep generative model, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineering, Boston, Massachusetts, USA, 2023.
[29] T. Kii, K. Yaji, K. Fujita, Z. Sha, C. Conner Seepersad, Latent crossover for data-driven multifidelity topology design, J. Mech. Des. 146 (5) (2024) 051713.
[30] S. Tsutsui, M. Yamamura, T. Higuchi, Multi-parent recombination with simplex crossover in real coded genetic algorithms, in: Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation - Volume 1, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999, pp. 657–664.
[31] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377.

[32] S. Yamasaki, K. Yaji, K. Fujita, Knowledge discovery in databases for determining formulation in topology optimization, Struct. Multidiscip. Optim. 59 (2018) 595–611.
[33] L. Antelmi, N. Ayache, P. Robert, M. Lorenzi, Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data, in: Proceedings of the 36th International Conference on Machine Learning, Vol. 97, 2019, pp. 302–311.
[34] K. Shang, H. Ishibuchi, L. He, L.M. Pang, A survey on the hypervolume indicator in evolutionary multiobjective optimization, IEEE Trans. Evol. Comput. 25 (1) (2020) 1–20.
[35] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.
[36] B.S. Lazarov, O. Sigmund, Filters in topology optimization based on Helmholtz-type differential equations, Internat. J. Numer. Methods Engrg. 86 (2011) 765–781.
[37] B. Bourdin, Filters in topology optimization, Internat. J. Numer. Methods Engrg. 50 (9) (2001) 2143–2158.
[38] F. Wang, B.S. Lazarov, O. Sigmund, On projection methods, convergence and robust formulations in topology optimization, Struct. Multidiscip. Optim. 43 (2011) 767–784.
[39] Z. Yuan, L. Geng, N. Wang, T. Wu, W. Qi, Y. Dai, J. Huang, Topology optimization method of stamping structures based on the directional density field, Materials 17 (3) (2024).
[40] K. Svanberg, The method of moving asymptotes - a new method for structural optimization, Internat. J. Numer. Methods Engrg. 24 (2) (1987) 359–373.
[41] R.J. Yang, C.J. Chen, Stress-based topology optimization, Struct. Optim. 12 (1996) 98–105.
[42] P. Duysinx, O. Sigmund, New developments in handling stress constraints in optimal material distribution, in: Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, American Institute of Aeronautics and Astronautics, 1998.
[43] L. Li, K. Khandelwal, Volume preserving projection filters and continuation methods in topology optimization, Eng. Struct. 85 (2015) 144–161.
[44] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, ArXiv Preprint.
[45] C.X. Ling, C. Li, Data mining for direct marketing: Problems and solutions, in: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, 1998, pp. 73–79.
[46] J. Eells, J.H. Sampson, Harmonic mappings of Riemannian manifolds, Amer. J. Math. 86 (1) (1964) 109–160.
[47] X.D. Gu, W. Zeng, F. Luo, S.T. Yau, Numerical computation of surface conformal mappings, Comput. Methods Funct. Theory 11 (2012) 747–787.
[48] P. Vogiatzis, M. Ma, S. Chen, X.D. Gu, Computational design and additive manufacturing of periodic conformal metasurfaces by synthesizing topology optimization with conformal mapping, Comput. Methods Appl. Mech. Engrg. 328 (2018) 477–497.
[49] W. Huo, C. Liu, Z. Du, X. Jiang, Z. Liu, X. Guo, Topology optimization on complex surfaces based on the moving morphable component method and computational conformal mapping, J. Appl. Mech. 89 (5) (2022) 051008.
[50] Q. Ye, Y. Guo, S. Chen, N. Lei, X.D. Gu, Topology optimization of conformal structures on manifolds using extended level set methods (x-LSM) and conformal geometry theory, Comput. Methods Appl. Mech. Engrg. 344 (2019) 164–185.
[51] M. Desbrun, M. Meyer, P. Alliez, Intrinsic parameterizations of surface meshes, Comput. Graph. Forum 21 (3) (2002) 209–218.
[52] L. Bruno, S. Petitjean, N. Ray, J. Maillot, Least squares conformal maps for automatic texture atlas generation, ACM Trans. Graph. 21 (3) (2002) 362–371.
[53] P. Mullen, Y. Tong, P. Alliez, M. Desbrun, Spectral conformal parameterization, Comput. Graph. Forum 27 (5) (2008) 1487–1494.