Machine learning driven forward prediction and inverse design for 4D printed hierarchical architecture with arbitrary shapes

Liuchao Jin a,b,c, Shouyi Yu b,c, Jianxiang Cheng b,c, Haitao Ye c,d, Xiaoya Zhai e, Jingchao Jiang f, Kang Zhang a, Bingcong Jian g,h,c, Mahdi Bodaghi h, Qi Ge b,c,e,* Wei-Hsin Liao a,i,**

a Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China
b Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen, 518055, China
c Shenzhen Key Laboratory of Soft Mechanics & Smart Manufacturing, Southern University of Science and Technology, Shenzhen, 518055, China
d Department of Mechanical Engineering, City University of Hong Kong, Kowloon, Hong Kong, China
e School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026, China
f Department of Engineering, University of Exeter, Exeter, United Kingdom
g School of Mechanical Engineering, Tongji University, Shanghai, 200092, China
h Department of Mechanical Engineering, School of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS, UK
i Institute of Intelligent Design and Manufacturing, The Chinese University of Hong Kong, Hong Kong, China

A B S T R A C T

The forward prediction and inverse design of 4D printing have primarily focused on 2D rectangular surfaces or plates, leaving the challenge of 4D printing parts with arbitrary shapes underexplored. This gap arises from the difficulty of handling varying input sizes in machine learning paradigms. To address this, we propose a novel machine learning-driven approach for forward prediction and inverse design tailored to 4D printed hierarchical architectures with arbitrary shapes. Our method encodes non-rectangular shapes with special identifiers, transforming the design domain into a format suitable for machine learning analysis. Using Residual Networks (ResNet) for forward prediction and evolutionary algorithms (EA) for inverse design, our approach achieves accurate and efficient predictions and designs. The results validate the effectiveness of our proposed method, with the forward prediction model achieving a loss below $10^{-2}$ mm, and the inverse optimization model maintaining an error near 1 mm, which is low relative to the entire shape of the optimized model. These outcomes demonstrate the capability of our approach to accurately predict and design complex hierarchical structures in 4D printing applications.

1. Introduction

4D printing, an evolution of traditional 3D printing [1–5], has gained significant attention in the past decades due to its revolutionary potential in construction [6], textile [7,8], automotive [9], aerospace industry [10], and biomedical applications [11–15]. Unlike conventional additive manufacturing processes that yield static objects, 4D printing empowers the creation of dynamic structures capable of self-transformation over time in response to external stimuli [16,17], such as heat [18–24], light [25,26], humidity [27,28], pH [29,30], and electric or magnetic fields [31–34].

Advances in 4D printing depend on the development of robust forward prediction and inverse design methods [35–37]. Forward prediction involves predicting how a 4D-printed structure will behave when subjected to specific stimuli, such as temperature fluctuations, moisture absorption, or mechanical deformation [38]. In contrast, inverse design focuses on optimizing the material composition and structural configuration of a component to obtain predefined performance criteria or a desired response [39]. These two aspects of 4D printing design (forward prediction and inverse optimization) form the cornerstone of efficient structure development, providing insights into material behavior and guiding the design process to achieve desired results.

In traditional design workflows, finite element analysis (FEA) has served as the primary tool for predicting the behavior of 4D printed structures. However, the reliance on FEA poses several limitations, including time-consuming simulations and computational burdens, particularly in inverse design tasks requiring millions of iterations. Hamel

* Correspondence to: Room 321, Engineering Building (North), Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, China
** Correspondence to: Room 308, William M.W. Mong Engineering Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, 999077, China
E-mail addresses: geq@sustech.edu.cn (Q. Ge), whliao@cuhk.edu.hk (W.-H. Liao).
et al. [40] proposed a machine-learning approach that combines the finite element method with an evolutionary algorithm to design active composite structures capable of achieving target shape-shifting responses. Their method optimizes the distribution of passive and active materials within equally sized voxel units to attain the desired shape changes. However, the reliance on finite element analysis for forward prediction poses computational challenges due to its time-consuming nature. Similarly, Athinarayanan et al. [41] addressed the inverse design problem of 4D printing by presenting a computational framework based on finite element analysis and evolutionary algorithms. Their approach optimizes both the materials distribution and material layout within a design space through topology optimization, offering a highly capable tool for designing 4D-printed active composites. Nonetheless, like Hamel et al. [40]'s approach, the utilization of FEA for forward prediction presents computational bottlenecks.

To overcome these challenges and expedite the design process, machine learning emerges as a compelling alternative [5]. By leveraging machine learning algorithms, it becomes possible to establish relationships between material allocation and desired responses, facilitating fast and accurate predictions. The integration of machine learning into 4D printing design workflows aims to streamline the forward prediction and inverse optimization processes, offering a data-driven approach to structure development. Sun et al. [42] introduced a novel machine learning and evolutionary algorithm-based approach for designing 4D-printed active composite structures. They employed a recurrent neural network (RNN) based machine learning model for forward shape-change prediction and then use evolutionary algorithms to solve the inverse problem of finding the optimal design. Despite the efficiency demonstrated by their approach for multiple target shapes, it focuses primarily on forward prediction and optimization of 2D rectangular surfaces, limiting its applicability to arbitrary shapes. Similarly, Sun et al. [43] proposed an integrated machine learning and sequential subdomain optimization approach for ultrafast inverse design of 4D-printed active composite structures. However, like previous studies, its focus is limited to forward prediction and optimization of 2D rectangular surfaces, restricting its applicability to arbitrary shapes. Sun et al.[44] also focused on inverse design for active composite plates which were still rectangle. Their approach combines machine learning (ML) with gradient-descent (GD) and evolutionary algorithms (EA) to efficiently determine the material distribution necessary for achieving desired 3D shape changes. Besides, Jin et al. [45] designed a residual neural network-enabled forward prediction and genetic algorithm-based inverse optimization for the design and optimization of 4D-printed rectangular hierarchical architecture. In addition, Zhao et al. [46] proposed an optimization method for grayscale digital light processing (DLP) 3D printing based on machine learning and evolutionary algorithms. Their approach integrates automated finite element model-based evaluation and a machine learning model based on recurrent neural networks to efficiently predict and optimize grayscale distributions for desired deformations. However, similar to previous approaches, it is tailored for 2D rectangular surfaces and lacks versatility in handling arbitrary shapes.

Therefore, in this paper, we present a novel approach to forward prediction and inverse design for 4D printed hierarchical architecture with arbitrary shapes, driven by machine learning techniques. Our methodology aims to address the limitations of traditional design methods by harnessing the power of machine learning to predict structure behavior and optimize design parameters. By encoding non-rectangular shapes with unique identifiers and leveraging Residual Networks (ResNet) for forward prediction and evolutionary algorithms for inverse design, our proposed framework demonstrates promising results in predicting and designing complex hierarchical structures in 4D printing applications. The schematic diagram of the proposed loop pipeline is shown in Fig. 1. On the one hand, machine learning-enabled forward prediction is used to efficiently figure out the relationship between the design and the behavior of the hierarchical architectures, which answers the question of how to predict the behavior of this design when subjected to stimuli. On the other hand, the evolutionary algorithms enabled inverse optimization is applied to inverse design the hierarchical architecture, which answers the question that given desired behavior, how to design the part so that the designed part can reach the desired behavior when subjected to stimuli. Through empirical validation and case studies, we showcase the efficacy of our approach in achieving precise and efficient design outcomes, paving the way for the advancement of 4D printing technology.

The remainder of this paper is organized as follows. In Section 2, we will discuss the modeling and digital representation of 4D printed
hierarchical architecture with arbitrary shapes to facilitate training and optimization by machine learning algorithms. Section 3 will present the methodology of the forward prediction method based on machine learning. In Section 4, we will illustrate the inverse optimization method for 4D printed hierarchical architecture with arbitrary shapes using evolutionary algorithms. Section 5 will present the results obtained from our approach. Finally, in Section 6, we will draw conclusions based on our findings and discuss potential future works in this field.

2. Modeling and digital representation of hierarchical architecture

To facilitate machine learning training and understanding, we employ digital representation to encode the material allocation within the hierarchical architecture. For instance, in Fig. 2, depicting a hierarchical architecture with two distinct materials – blue and red – we assign numerical labels for representation. Specifically, we designate “1” to represent the blue material and “2” to denote the red material. This digital representation enables the entire structure to be inputted into machine learning algorithms for analysis and learning purposes.

It is important to note that within this representation, we introduce the definitions of active and passive materials to enhance clarity and differentiation between them. Active materials, in the context of 4D printing, refer to those with a significant thermal expansion ratio, meaning they undergo considerable dimensional changes in response to changes in temperature. These materials exhibit dynamic behavior, expanding or contracting when subjected to external stimuli such as heat. In contrast, passive materials either have no thermal expansion ratio or exhibit minimal dimensional changes in response to temperature variations.

The concept of hierarchical architecture with active and passive materials harnesses the unique properties of these materials to achieve the 4D printing effect. By strategically combining active and passive materials within a hierarchical structure, dynamic systems capable of self-transformation over time can be created. When exposed to external stimuli such as heat or moisture, the active materials undergo significant dimensional changes, while the passive materials remain relatively unchanged. This contrast in behavior between the active and passive components drives the shape-shifting capabilities of the hierarchical architecture, resulting in the observed 4D printing effect. The hierarchical arrangement ensures that specific regions of the structure respond differently to external stimuli, allowing for precise control over the overall shape transformation process.

The existing research commonly employs matrices of 0s and 1s to digitally represent the passive and active materials for data inputs of entire rectangular parts [40,42,43], facilitating machine learning analyses. However, challenges emerge when dealing with non-rectangular shapes, leading to either non-uniform input data for recurrent neural networks or non-rectangular input data for convolutional neural networks. These discrepancies in input data format pose obstacles to the application of machine learning techniques, highlighting the need for solutions to ensure uniformity and compatibility with neural network architectures.

To address the challenge posed by non-rectangular shapes in machine learning models, we propose a solution wherein we fill the gaps between the non-rectangular shape and a rectangular shape with zeros. Subsequently, we assign non-zero numbers to represent the material-containing parts of the non-rectangular shape. This approach ensures that the input data fed into the machine learning model becomes uniform and rectangular, thereby facilitating the training process and enabling efficient analysis and prediction.

As depicted in Fig. 2, the non-rectangular shapes such as triangles, circles, and pentagons represent the regions requiring material allocation optimization. Conventional approaches may use binary values (0 for passive materials, 1 for active materials) to represent such shapes, leading to irregular input data shapes, posing challenges for machine learning models. Our proposed solution involves enclosing these irregular shapes within a rectangular frame, filling the empty spaces with zeros, and assigning numeric values (e.g., 1 for passive materials, 2 for active materials) to the corresponding regions. Additional numeric values can be utilized to represent further material variations. This method ensures that the input data shape fed into the machine learning model becomes uniform or rectangular, thereby enhancing model compatibility and efficiency.

3. Forward prediction methods

In the context of forward prediction methods for hierarchical architecture, the primary objective revolves around harnessing machine
learning algorithms to establish correlations between the allocation of different materials within components and the resultant behavior of the printed parts. Traditionally, finite element analysis has served as the predominant method for forward prediction, involving intricate simulations to replicate real-world scenarios. However, the inherent time-consuming nature of FEA, particularly in the context of inverse design tasks, presents a significant bottleneck. With inverse design necessitating millions of cases for thorough forward prediction evaluation, the limitations of FEA become apparent, especially when dealing with large or complex parts. Consequently, the advancement of inverse design methodologies is hindered by the computational overhead associated with traditional FEA approaches. To address this challenge, our proposal advocates for the integration of machine learning as a viable alternative to FEA. By leveraging machine learning, which can assimilate data from FEA outcomes or experiment data in reality, predictive models can be developed to establish robust relationships between material allocation and behavior. These models offer a substantially faster alternative to traditional FEA simulations, thereby expediting the forward prediction process and facilitating the exploration of diverse material configurations.

In this paper, we utilize ResNet, short for Residual Networks, as our forward prediction model due to its remarkable capability to capture intricate patterns and features within complex datasets. ResNet is a deep learning architecture that was introduced to address the vanishing gradient problem in deep convolutional neural networks [47]. The vanishing gradient problem occurs when training deep networks with many layers, as the gradients can become extremely small, leading to slow convergence or even a complete halt in learning. ResNet introduces skip connections or shortcut connections that enable the flow of gradients directly from earlier layers to later layers, bypassing a few layers in between. These skip connections allow for the training of deeper networks by alleviating the vanishing gradient problem. By propagating the gradients more effectively, ResNet enables the training of models with hundreds or even thousands of layers. The skip connections in ResNet are implemented through the use of residual blocks. A residual block consists of a convolutional layer followed by batch normalization and a non-linear activation function. The output from the block is then added to the input of the block, forming the skip connection. This addition operation allows the network to learn the residual or the difference between the input and the desired output. The ResNet structure we have constructed is visually depicted in Fig. 3, providing a clear representation of its architectural layout and components. In the subsequent subsections, we embark on a detailed exploration of
the ResNet model, elucidating its operational process and underlying methodologies.

3.1. Data acquisition

The data acquisition process for the machine learning model relies on the utilization of Abaqus simulation software to generate the requisite datasets. This process unfolds in several key stages. Firstly, we employ Python scripting to automate the creation of Abaqus input files tailored to our specific requirements. These input files encapsulate the parameters and conditions necessary for simulating the behavior of our hierarchical architectural structures. Through this automation, we generate a substantial volume of input files, ensuring an ample dataset for subsequent analysis and model training. Subsequently, we initiate Abaqus simulations in parallel using Python, a crucial step that accelerates the data generation process. By harnessing parallel computing capabilities, we can execute multiple simulations concurrently, thereby maximizing computational efficiency and expediting the acquisition of output databases. During the simulation phase, Abaqus calculates and records various output metrics, including deformation data, which are integral to our machine learning task. Upon completion of the simulations, we access the output databases containing the pertinent information generated during the computational experiments. To extract the requisite data for our machine learning models, we employ Python scripts to read and parse the output databases generated by the Abaqus simulations. This step allows us to systematically retrieve the material allocation data and corresponding deformation data for each simulation case, facilitating the assembly of comprehensive datasets for analysis and model training. Upon completion of the data acquisition process, we possess the necessary material allocation data and deformation data required to train and validate our machine learning models. These datasets serve as the foundation for the machine learning model, enabling it to explore the complex relationships between material allocation and structural behavior within hierarchical architectural designs.

Here, we present the details for the finite element analysis. For meshing, we employed a detailed meshing strategy to enhance the accuracy of our results. Specifically, each voxel in our model was subdivided into a $10 \times 10$ mesh. This means that every voxel, which represents a discrete volume element in the simulation space, is further divided into a grid of 100 smaller finite elements.

The boundary conditions are essential for ensuring realistic simulations in finite element method. In our simulation setup using Abaqus, we implemented a fixed boundary condition on the left edge of the model during the temperature increasing step. This boundary condition restricts displacement in the $x$-direction to zero, simulating a scenario where the left edge of the structure is fixed or clamped. The entire part is subjected to a temperature change of 30°C.

In this paper, we used TPU as the passive material and SMP (with a glass transition temperature of 55 degrees Celsius) from SMP Technology, Japan, as the active material. The materials were then printed using a dual-nozzle Polarbear 3D printer. The material properties for the active material are Young’s modulus of 3.26 MPa, Poisson’s ratio of 0.35, and coefficient of thermal expansion of $-0.01$ (1/°C). Conversely, the passive material has Young’s modulus of 15.4 MPa, Poisson’s ratio of 0.35, and coefficient of thermal expansion of 0.0 (1/°C). These properties delineate the mechanical and thermal characteristics of the materials used in the beam. For simplification and efficiency, both materials were assigned elastic properties. We used two-state properties to simulate the performance of the printed part, including the properties at low and high temperatures. The thermal expansion coefficient is calculated by the difference in length at low and high temperatures. The Young’s modulus and the Poisson’s ratio are taken from the properties at the high temperature because most of the deformation happens above the glass transition temperature. For TPU, although the real thermal expansion coefficient is positive, the shrinkage of the part due to residual stress and strain after 3D printing effectively compensates for the thermal expansion. This compensation results in a near-zero effective thermal expansion coefficient. Therefore, for simplification, using a zero thermal expansion coefficient for TPU in the simulation is a reasonable approximation to reflect the actual behavior observed in the printed samples. For SMP, the real thermal expansion coefficient is also positive. However, the large amount of residual stress and strain that exists post-3D printing leads to significant shrinkage of the part when the temperature increases and these stresses and strains are released. To accurately represent this effect in our simulations, we regard the thermal expansion coefficient as negative. This adjustment allows us to model the thermal behavior of SMP more accurately, reflecting the observed shrinkage due to the release of residual stresses.

3.2. Data preprocessing

The first step in preparing the data for the ResNet model is data preprocessing. This involves normalizing both the input and output data to ensure they are in a suitable range for training. For the input data, which represents active (1), passive (0), and no (0) allocation matrices, we apply a normalization equation: $x_n = x - \bar{x}$, where $x$ is the input data and $\bar{x}$ is normalized input data. This converts the original values of 0, 1, and 2 to $-1$, 0, and 1 respectively, allowing for better convergence during training. On the other hand, the output data is normalized based on the mean and standard deviation. By using the equation $y_n = (y - y_{\text{mean}})/y_{\text{std}}$, where $y$ is the output data, $y_{\text{mean}}$ is the normalized output data, and $y_{\text{mean}}$ and $y_{\text{std}}$ represent the mean and the standard deviation of all output data, respectively, we transform the output data to have a mean of 0 and a standard deviation of 1, ensuring consistency and stability in the training process.

After normalizing the data, the next step is to split it into training and validation sets. The purpose of this split is to evaluate the performance of the model on unseen data and prevent overfitting. We allocate 80% of the normalized data to the training set and the remaining 20% to the validation set. This division ensures that the model is trained on a significant amount of data while still having a separate set for evaluation.

3.3. Model architecture

The ResNet model architecture is designed to capture complex patterns and features in the data. As demonstrated in Fig. 3, it consists of convolution layers, max-pooling, and residual blocks. The input layer starts with a convolution layer that has 4 filters and a (3, 3) kernel size, followed by max-pooling to downsample the data. The core of the model consists of five residual blocks, each containing convolution layers with 8 filters. These blocks contribute to the depth and complexity of the model, enabling it to extract intricate features from the data. Finally, the output layer is a fully connected layer with a linear activation function. This layer shapes the model’s output to the desired form, ready for prediction or classification.

3.4. Hyperparameters

To train the ResNet model effectively, several hyperparameters need to be specified. The learning rate is set to 0.0005, determining the step size for adjusting the model weights during optimization. A suitable learning rate ensures that the model converges efficiently without over-shooting or getting stuck in local minima. The batch size is set to 50, which determines the number of samples processed in each mini-batch during training. A larger batch size can lead to faster convergence, but it requires more memory. The max epochs parameter is set to 500, defining the maximum number of training iterations. This parameter prevents the model from training indefinitely and helps control training time. Additionally, a loop decay factor of 0.6 per 25 epochs is applied to decrease the learning rate gradually over time, allowing the model to fine-tune its performance.
3.5. Training

Once the data and model architecture are prepared, the training process begins. The model is compiled using the Adam optimizer, a popular optimization algorithm that adapts the learning rate for each parameter. Mean squared error (MSE) is chosen as the loss function, which measures the dissimilarity between the predicted output and the actual output. Additionally, early stopping is employed to monitor the validation loss. If the validation loss fails to improve for a set number of consecutive epochs, the training process is halted to prevent overfitting. Throughout the training process, the model learns patterns and relationships within the input data and its corresponding output, aiming to make accurate predictions or classifications on new, unseen data.

4. Inverse design methods

Inverse optimization is employed to determine the optimal material allocation within the hierarchical architecture to ensure that the resulting printed part exhibits predefined performance characteristics. In essence, it involves working backward from the desired performance requirements to identify the material distribution that will best achieve those objectives.

In this paper, we employ an evolutionary algorithm to perform inverse optimization for 4D printed hierarchical architectures with arbitrary shapes. The evolutionary algorithm is a computational optimization technique inspired by the process of natural selection and evolution in biology. It mimics the principles of natural evolution to search for optimal solutions to complex problems iteratively. Unlike topology optimization, which often relies on gradient-based methods and can get trapped in local optima [48–50], evolutionary algorithms perform a global search. This ability to explore the entire solution space increases the likelihood of finding the true global optimum, especially in complex and high-dimensional problems. At the core of the evolutionary algorithm is a population of candidate solutions, represented as individuals or chromosomes. Each individual in the population encodes a potential solution to the optimization problem. In our case, the individuals represent different configurations of the 4D printed hierarchical architecture with arbitrary shapes.

The steps for the evolutionary algorithms are depicted in Fig. 4. The algorithm begins by generating an initial population comprising 5000 individuals. Each individual represents a potential solution for the material allocation and stimuli distribution in 4D printing. These solutions are randomly generated to provide a diverse range of configurations for the 4D-printed structures. Random initialization helps explore a wide range of design possibilities and prevents the algorithm from getting stuck in local optima.

Once the initial population is generated, each individual undergoes fitness evaluation using a fitness function. The fitness function quantifies the performance of the 4D-printed structure based on the Root Mean Square Error (RMSE) between the predicted and target coordinates. The RMSE serves as a measure of how well the material allocation and stimuli distribution align with the desired deformation. Lower RMSE values indicate better alignment and higher fitness. The fitness evaluation provides a numerical measure of each individual’s effectiveness in achieving the desired deformation.

After fitness evaluation, the population is sorted based on the fitness scores. The top 10% of individuals, which exhibit the highest fitness scores, are selected as elite individuals. These elite individuals possess favorable characteristics in terms of material allocation and stimuli distribution that contribute to achieving the desired deformation. By preserving the best-performing individuals, the algorithm ensures that promising solutions are carried over to the next generation, increasing the chances of finding optimal or near-optimal solutions.

The remaining 70% of the population is used for crossover operations. Crossover is a genetic operation that combines genetic material from two parent individuals to create two offspring. In this implementation, a two-point crossover is employed. Two random crossover columns, referred to as a and b, are selected. The genetic material between these columns, which is circled in green in Fig. 4, is exchanged between the parents, creating two children. This process introduces diversity and potential improvements into the population by exploring different combinations of genetic material. By combining genetic information from different individuals, crossover helps in the exploration of the solution space and can lead to the discovery of novel and better solutions.

The remaining 20% of the population is allocated for mutation operations. Mutation introduces random changes to the genetic material of individual members within the population. It helps introduce diversity and prevents the algorithm from prematurely converging to suboptimal solutions. In this implementation, a random column in the digital chromosome is selected, and the corresponding bit is flipped (from 2 to 1 or vice versa), as illustrated in Fig. 4. This operation adds randomness to the population and aids in exploring the solution space by introducing small, random variations to the genetic material.

The combination of elite individuals, offspring generated through crossover, and individuals subjected to mutation forms the new population for the next generation. This new population represents the updated generation and undergoes the same process of fitness evaluation, selection of elite individuals, crossover, and mutation. By iteratively refining the population, the algorithm explores the design space and aims to converge towards optimal or near-optimal solutions for the 4D printing design optimization problem.

The iterative process continues until a satisfactory level of convergence or desired performance is achieved. Convergence refers to the point where the algorithm has reached a stable state, and further iterations do not significantly improve the solutions. The stopping criteria for convergence can be predefined based on the desired level of performance or a specific number of generations. Once the algorithm converges or achieves the desired performance, the best solution(s) found in the final population can be selected as the optimal or near-optimal solutions for the 4D printing design optimization problem.

5. Results and discussions

In this section, we present the results of both the forward prediction and inverse optimization models developed for the 4D-printed hierarchical architecture with arbitrary shape, as outlined in the previous sections.

There are two scenarios for the inverse design of 4D printed hierarchical architecture with arbitrary shapes as illustrated in Fig. 5. The first scenario is the fixed design domain with material allocation. In this scenario, the design domain remains fixed, which is defined within the green line in Fig. 5, and the objective is to determine the optimal allocation of active and passive materials within this design domain. The goal is to find the material distribution that achieves the desired deformation or shape change when subjected to the appropriate stimuli. The evolutionary algorithm is utilized to search for the optimal allocation of materials within the given design domain. By iteratively evaluating and refining the population of material allocations, the algorithm aims to converge towards the optimal or near-optimal solution that maximizes the desired deformation while satisfying any imposed constraints. The second scenario is the random design domain with topology optimization. In the second scenario, the design domain is not fixed, and each location within the structure can be allocated as no material, active material, or passive material. This scenario resembles topology optimization, where the goal is to determine the optimal distribution of material throughout the entire structure. The evolutionary algorithm is employed to explore the design space and find the optimal material distribution that achieves the desired deformation or shape change. This scenario offers a higher degree of design freedom, as each location can be assigned a specific material allocation based on
certain constraints and objectives. The algorithm iteratively evaluates and evolves the population of material distributions, aiming to converge towards the optimal or near-optimal solution that maximizes the desired deformation while satisfying the given constraints.

5.1. Fixed design domain

To demonstrate the effectiveness of the forward prediction and inverse optimization approach for 4D-printed hierarchical architectures with arbitrary shapes in fixed design domain, we will present two examples.

In the first example, we consider a non-rectangular beam composed of two opposite triangles. The objective is to design the material allocation within the beam structure so that it deforms to a predefined parabolic curve when subjected to the appropriate stimuli. To achieve this, we utilize a ResNet (Residual Neural Network) for forward prediction, which can accurately predict the deformation behavior of the beam for different material allocations. We generate a dataset of...
The results for the inverse design of 4D-printed hierarchical architecture with arbitrary shape (parabola case). (a). Optimized results obtained from the evolutionary algorithms and corresponding simulation results of the parts with the optimized material allocation. (b). The training loss and the validation loss for the forward prediction model. (c). The fitness for the best individual in each generation. (d). The comparison between the target deformation and simulated deformation.

20,000 instances, each consisting of a different material allocation and the corresponding resulting deformation. This dataset is used to train the ResNet model.

Once the ResNet model is trained, we employ an evolutionary algorithm for inverse optimization. The EA explores the design space of material allocations within the non-rectangular beam to find the optimal or near-optimal solution that deforms the beam to the predefined parabolic curve. The EA iteratively evaluates and evolves the population of material allocations, utilizing the fitness function based on the RMSE between the predicted and target parabolic curves. Through the iterative process of selection, crossover, and mutation, the EA converges towards an optimized material allocation that achieves the desired deformation behavior.

The desired curve for this case is that the irregular-shaped beam will deform into a parabola with an end displacement of \(-100\) mm. The results for the parabola curve are shown in Fig. 6. Fig. 6(a) displays the optimized results obtained from the evolutionary algorithms and the corresponding simulation results of the parts with the optimized material allocation, which met our expectations and requirements. In Fig. 6(b), the training loss and the validation loss for the forward prediction model are depicted, indicating that the forward prediction model can establish a robust relationship between the material allocation and behavior, with loss reaching below \(10^{-2}\). Fig. 6(c) shows the fitness for the best individual in each generation, demonstrating the good performance of our forward prediction and inverse optimization model. Fig. 6(d) displays the optimized results obtained from the evolutionary algorithms and Fig. 6(d) illustrates corresponding simulation results of the parts with the optimized material allocation. Fig. 7(c) and (d) give a real hand demonstration of close and open hand movements. Finally, the results are validated by experiment as shown in Fig. 7(f) and (h). The simulation results and experimental results align with the target shape, showing that the optimized result can replicate the movements demonstrated by the real hand.

From the above two examples, it is evident that our design method exhibits good universality for 4D-printed hierarchical architecture with irregular shapes, where the design domain is fixed. By employing the method of utilizing 0, 1, and 2 to represent different material allocations, we have effectively overcome the previous limitation where only rectangular structures could be optimized using 1 and 2. This expansion of the application of machine learning in forward prediction and inverse optimization has been successful in achieving satisfactory results.

### 5.2. Random design domain

In this subsection, we delve into the inverse design of 4D-printed hierarchical architecture with random design domains. Despite the randomness of these design domains, certain restrictions must be adhered to ensure the structural integrity and functionality of the printed parts, which is explored in Fig. 8. Firstly, we enforce the constraint that there cannot be any “isolated islands” within the structure, where an area is surrounded entirely by empty space, which is circled with blue boxes in Fig. 8. Such isolated regions serve no purpose and can compromise the overall structural integrity. Secondly, we prohibit the existence of “diagonal lands”, characterized by a \(2 \times 2\) matrix where one diagonal
Fig. 7. The results for the inverse design of 4D-printed hierarchical architecture with arbitrary shape (hand shape). (a) The training loss and the validation loss for the forward prediction model. (b) The fitness for the best individual in each generation. (c) Optimized results obtained from the evolutionary algorithms. (d) Corresponding simulation results of the part with the optimized material allocation. (e) Demonstration of closed hand. (f) Printed part using two-nozzle FDM machine. (g) Demonstration of open hand. (h) Shape morphing after heating.

Fig. 8. The constraints of material allocation for the random design domain.

contains material while the other diagonal does not. This constraint is circled with yellow boxes in the picture. This configuration introduces unpredictability and instability into the system, rendering it unsuitable for practical applications. Finally, along the direction of the optimized curve, it is essential to have a continuous “continent” connecting one end of the curve to the other, which is illustrated by the green line in Fig. 8. The green line represents the break between these two “continents”. This constraint ensures the coherence of the part and prevents it from being fragmented into disjointed segments, maintaining its functionality and usability. By adhering to these constraints, we aim to optimize the design of 4D-printed hierarchical architectures with random design domains, ensuring their reliability and effectiveness in real-world applications.

In order to verify the optimization process for 4D-printed hierarchical architectures with random irregular design domains, we conduct the following example. Firstly, for each column in the design domain, we randomly select a number between 1 and 4 to determine the quantity of materials to be filled. Starting from the centerline of the column, we then distribute the selected number of materials evenly in both upward and downward directions. By repeating this process for each column, we ensure that the three aforementioned constraints are not violated. Indeed, while the method described provides a straightforward approach to address the constraints, there are numerous alternative strategies that can be explored.

Based on this method, we conducted optimization of the 4D-printed hierarchical architecture with random design domains. Our task is to distribute the materials, represented by 0, 1, and 2, within the design domain of 8 by 64 using the aforementioned method. The objective is to design a component that, upon receiving external stimulation, will deform into a parabolic shape with an end displacement of 80 mm. The outcomes are depicted in Fig. 9. Fig. 9(a) showcases the optimized results obtained from the genetic algorithms alongside the...
Fig. 9. The results for the inverse design of 4D-printed hierarchical architecture with arbitrary shape for random design domain. (a) Optimized results obtained from the evolutionary algorithms and corresponding simulation results of the parts with the optimized material allocation. (b) The training loss and the validation loss for the forward prediction model. (c) The fitness for the best individual in each generation. (d) The comparison between the target deformation and simulated deformation.

Corresponding simulation outcomes of the parts with the optimized material allocation. The optimized results obtained from the genetic algorithms reveal a precisely tailored material allocation within the 4D-printed hierarchical architecture. This allocation demonstrates a seamless integration of passive and active materials, resulting in a structure that exhibits the desired shape-shifting behavior. The corresponding simulation results validate the effectiveness of the optimized material distribution, showcasing the accurate realization of the intended design. Fig. 9(b) illustrates the training loss and validation loss for the forward prediction model. The training loss and validation loss for the forward prediction model depict a robust learning process, with both losses converging to low values. This convergence indicates that the forward prediction model has successfully captured the complex relationship between material allocation and structural behavior. The consistently low losses affirm the model’s ability to accurately predict deformations based on digital material inputs. Fig. 9(c) represents the fitness for the best individual in each generation. The fitness plot for the best individual in each generation showcases the evolutionary algorithm’s iterative optimization process. With each generation, the RMSE steadily decreases, indicating progressive refinement of the material allocation strategy. Fig. 9(d) provides a comparison between the target deformation and simulated deformation. The comparison between the target deformation and simulated deformation highlights the efficacy of the inverse design methodology. The simulated deformation closely matches the target deformation, demonstrating the accuracy and reliability of the designed hierarchical architecture. This alignment between the desired and simulated outcomes validates the effectiveness of the proposed approach in achieving the intended structural behavior.

Compared with the fixed design domain, the random design domain not only has a higher degree of freedom, but can also include the fixed design domain. For instance, consider the two opposing triangles illustrated in Fig. 6. If optimization is required for these triangles, the initial population in the evolutionary algorithm can be set to comprise these two opposing triangles. Subsequently, these triangles can be inputted into the random design domain’s forward prediction model for evaluation. The evaluations obtained are then used for selection, crossover, and mutation operations. Over successive generations, the shapes of each generation of offspring will be two opposing triangles. In other words, for random design domain algorithms, any shape within the design boundaries can be forward predicted and inverse optimized.

5.3. Discussion

The hierarchical structure studied in our research presents a design space smaller than 1000 units, which is modest compared to many topology optimization (TO) problems in the literature that involve significantly larger design spaces, often exceeding 100 by 100 dimensions. The primary challenge in such large-scale optimization problems is the computational intensity and the time required for each inverse optimization process. However, our method is designed to be scalable. The hierarchical approach we use allows us to break down the larger problem into smaller, more manageable sub-problems. This decomposition not only simplifies the complexity of each optimization task but also enables parallel processing, which significantly enhances computational efficiency. By distributing the computational load across multiple CPU cores, we can perform numerous simulations concurrently, thereby reducing the overall time required for optimization.

The optimization outcomes presented in Figs. 6, 7, and 9 demonstrate that the shapes activated through our approach closely align with the intended target shapes at a macroscopic level. However, a closer inspection at the microscopic level reveals that the surfaces of the deformed materials do not exhibit the same smoothness as the target shapes. This discrepancy primarily arises due to the size of the pixels used in the design. Larger pixels can result in less smooth curves, giving the surface a rough appearance. This issue could potentially pose a problem in practical applications where surface smoothness is critical. To address the observed unsmoothness, we propose a couple of strategies. First, by decreasing the size of the pixels used in the design, we can achieve finer resolution and smoother curves. This approach involves increasing the computational resolution of the design process,
which, although computationally more intensive, would result in a more accurate representation of the target shapes. Smaller pixels allow for more detailed and precise control over the material distribution, thereby enhancing the smoothness of the deformed surfaces. Additionally, the surface quality can also be enhanced by optimizing the parameters of the FDM printing process. Adjustments such as layer height, extrusion width, and print speed can significantly impact the surface finish. For instance, reducing the layer height can lead to finer layers being deposited, which smoothens the surface. Similarly, optimizing the extrusion width and print speed can help in achieving better layer adhesion and reducing surface roughness. By fine-tuning these parameters, it is possible to produce smoother surfaces even with the existing design resolutions.

6. Conclusions and future works

In this paper, we have presented a novel approach for forward prediction and inverse design tailored to 4D printed hierarchical architecture with arbitrary shapes. By integrating machine learning techniques, specifically Residual Networks for forward prediction and evolutionary algorithms for inverse design, our methodology addresses the challenges posed by non-rectangular shapes in traditional design workflows. Our approach enables the prediction of complex hierarchical structures’ behavior and the optimization of material allocation to achieve desired shape transformations efficiently. Through empirical validation and case studies, we have demonstrated the effectiveness and versatility of our proposed framework in accurately predicting and designing 4D printed structures with arbitrary shapes, paving the way for advancements in 4D printing technology.

This research contributes to the broader field of 4D printing design optimization by offering a data-driven approach that leverages machine learning to streamline the design process. By encoding non-rectangular shapes with unique identifiers and employing ResNet for forward prediction, our methodology overcomes the limitations of traditional design methods and enhances the efficiency and accuracy of shape prediction. Furthermore, the use of evolutionary algorithms for inverse design enables the optimization of material allocation for achieving predefined performance criteria, facilitating the creation of adaptive and functional structures. Overall, our work underscores the potential of machine learning-driven approaches in advancing the capabilities of 4D printing technology, opening new avenues for designing complex and dynamic structures for various applications across industries.

The future work would be to extend the design and optimization methods from 2D surfaces to 3D plates or objects. While the current focus has been on optimizing material allocation for 2D surfaces, moving towards 3D structures would introduce additional challenges and complexities. Designing 3D plates or objects with specific deformation behaviors requires considering not only the material allocation in the surface plane but also the material distribution in the depth dimension. This would involve optimizing the material properties and arrangement in three dimensions to achieve desired deformation characteristics.

CRediT authorship contribution statement

Liucao Jin: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Shouyi Yu: Writing – original draft, Visualization, Formal analysis, Data curation. Jianxiang Cheng: Writing – original draft, Validation, Formal analysis. Haitao Ye: Writing – original draft, Visualization, Validation, Formal analysis. Xiaoya Zhai: Writing – review & editing, Visualization, Validation, Investigation. Jingchao Jiang: Writing – review & editing, Visualization, Validation, Formal analysis. Kang Zhang: Writing – review & editing, Validation, Investigation. Bingcong Jian: Writing – review & editing, Visualization, Formal analysis, Data curation. Mahdi Bodaghi: Writing – review & editing, Visualization, Supervision. Qi Ge: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. Wei-Hsin Liao: Writing – review & editing, Validation, Supervision, Resources, Formal analysis, Conceptualization.

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.

Data availability

The data supporting the findings of this study are available within the article.

Acknowledgments

W. L. acknowledges Research Grants Council (C4074-22G), Hong Kong Special Administrative Region, China, and The Chinese University of Hong Kong, Hong Kong (Project ID: 3110174). Q. G. acknowledges the National Natural Science Foundation of China (No. 12072142), the Key Talent Recruitment Program of Guangdong Province, China (No. 2019QD012458), and the support by the Science, Technology and Innovation Commission of Shenzhen Municipality, China under grant no. ZDSYS2021062309005017. M. B. acknowledges the support by the UK Engineering and Physical Sciences Research Council (EPSRC) (grant number EP/YO11457/1). For the purpose of open access, the author has applied a ‘Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References


