Surrogate Modeling of Acoustic Field-Assisted Particle Patterning Process with Physics-Informed Encoder-Decoder Approach

Yu Hui Lui[1], M Shahriar[1], Yayue Pan[2], Chao Hu[1], and Shan Hu[1*]

[1] Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA

[2] Department of Mechanical and Industrial Engineering, University of Illinois Chicago, Chicago, IL 60607, USA

*corresponding author: shanhu@iastate.edu

ABSTRACT

Manipulating the distribution of functional particles in a polymer matrix can enable the fabrication of multifunctional smart composite devices. Using an acoustic field for particle patterning is a promising technique to alleviate the need for electrically conductive particles or magnetically responsive particles. To better understand the acoustic particle patterning process, a 3D high-fidelity multiphysics model is generally utilized. However, thousands of forward simulations are often required to determine a suitable set of input parameters for a desired particle pattern. It is advantageous to replace the computationally expensive forward simulation model with a cheaper-to-evaluate surrogate model to optimize the acoustic particle patterning process. This work develops a physics-informed machine learning approach to build a surrogate model capable of predicting the acoustic pressure pattern, which is highly related to the particle pattern. The surrogate model has an encoder-decoder structure, and the model training uses simulation data generated from a 3D multiphysics model. The multiphysics model is validated against experimental data before the generation of the simulation data. Physical knowledge is incorporated into the encoder-decoder model through a physics-informed input derived from the output of a 2D multiphysics model. This 2D model is constructed based on a cut plane of the 3D model to preserved most of the acoustic pressure information from the complex 3D model while being more efficient to evaluate and suitable for online prediction. The proposed physics-informed encoder-decoder model can increase the quality of the acoustic pattern prediction by over 40% compared to the base encoder-decoder model. Incorporating the physics-informed input into the base encoder-decoder can significantly reduce the sample size and model complexity required for achieving a given acoustic pattern prediction accuracy. This work provides a guideline for developing physics-informed machine learning models for manufacturing processes.

Keywords: Surrogate modeling, Convolutional encoder-decoder networks, Deep learning, Physics-informed machine learning, Acoustic field particle patterning

1 INTRODUCTION

Heterogeneous mixing of multifunctional particles with polymer can form a particle-polymer composite with anisotropy properties that enable wider applications, such as biometric materials [1]–[6]. Acoustic field-assisted particle patterning (AFP) is a promising heterogeneous mixing method to control the spatial distribution of the functional particles (such as mechanical, electrical, chemical, and thermal functions) dispersed in a polymer matrix to fabricate multifunctional smart composite objects. To better understand the acoustic particle patterning process, a three-dimensional (3D) high-fidelity multiphysics model is generally utilized. Given a desired particle distribution, many iterations of experiment designs are needed to find the correct input parameters for the the particle patterning process. Even when utilizing optimization methods in model-based process design, such as gradient descent or genetic algorithm, thousands of forward simulations are often required to determine a suitable set of input parameters for a desired particle pattern. To alleviate the high computational cost of solving the partial differential equations (PDEs) in a physical model, it is often advantageous to replace the computationally expensive model with a cheaper-to-run mathematical model, i.e., a surrogate model, for the optimization of the acoustic particle patterning process. With a surrogate model, an optimization method can quickly search through a large process space with a large number of parameters and achieve a high throughput process design.

A surrogate model is often trained on a limited number of simulation-based data and then performs prediction tasks using the surrogate instead of solving the actual PDEs. Unfortunately, most of the existing surrogate models, such as Gaussian processes [7]–[9], polynomial chaos expansions [10], and support vector machines [11], are not designed to handle high-dimensional problems. In this study, for each set of input parameters, the output of the AFP system is the spatial distribution of the acoustic pressure, which is high-dimensional. The spatial distribution of the acoustic pressure can be represented as an image with acoustic pressure patterns, which allows us to convert our problem into an image prediction problem. Thus, the surrogate model requires the generation of the image that contains the acoustic pressure patterns given a set of input parameters.

Deep learninig approach has been widely used in computer vision tasks, such as image classification [12], [13], object detection [14], facial recognition [15] and human pose estimation [16], [17]. Most notably, the convolutional neural networks

(CNNs) with multiple convolutional layers have been found to be highly effective and efficient in these computer vision tasks. The main advantages of using CNNs over traditional artificial neural networks are 1) the automatic features extraction process, which alleviates the need for handcraft feature engineering, 2) the stack of multiple layers can learn highly representative, multiscale features from high-dimensional input, and 3) better generalization capability when the networks are well-designed [18], [19]. To enable CNNs for image prediction, an encoder-decoder network is utilized. An encoder network maps the high-dimensional input to a low-dimensional latent vector. This mapping is similar to using CNNs to classify an image, i.e., after inputting an image to a CNN, the CNN will output the class of the object in the image. A decoder network maps the encoded latent vector of the data to high-dimensional space. An autoencoder, which uses an encoder-decoder network to reproduce the input, is the neural network first used for dimension reduction [20] and then for other applications from gene ontology annotation predictions to structural damage identification [21], [22]. More recently, based on the encoder-decoder network structure, generative networks such as variational encoders [23] and Generative Adversarial Networks (GANs) [24], [25] have gained more and more interest due to their ability to generate desirable new samples based on the training data. However, note that in this work, the surrogate model should generate images of acoustic pressure based on the AFP system's input parameters, not some unknown variations of the training input. Thus, a traditional generative network, i.e., GAN, is unsuitable here. Although conditional GAN (cGAN) [26] might be a suitable candidate, we only consider a more traditional encoder-decoder network in this study to investigate the effect of adding physical input to the network.

In this work, we consider the surrogate modeling of an AFP system governed by multiphysics PDEs with different input parameters. The spatial discretization of the AFP system setup is the input to the model, and the corresponding acoustic pressure distribution is the output. A physics-informed encoder-decoder (PIED) surrogate model is developed and applied to predict the spatially discretized acoustic pressure distribution, i.e., the image of acoustic pressure distribution. The surrogate model is based on deep convolutional and transposed convolutional neural network architectures to enable efficient image prediction. There are many ways to incorporate physics into machine learning model, such as physics-guided loss function [27], [28], physics-guided initialization (e.g., transfer learning) [29], [30], and residual model [31], [32]. Physics-guided loss function techniques are to incorporate physical constraints into the loss function of machine learning models; physics-guided initialization techniques are to pretrain the machine learning model with one dataset and retrain on the dataset of interest; residual model techniques are to use a machine learning model to learn the error of physics model and combine the error prediction with a physics-based model's prediction to improve accuracy. In this study, we chose to incorporate physics into the machine learning model by utilizing a modified output (termed pseudo-output) from a physics-based model with lower dimension. The pseudo-output is used as an additional input (i.e., physics-informed input) to the machine learning model. Hence, our approach is similar to the physics-guided initialization method. The input consisted of the acoustic actuator setups, which were also the input to a 3D multiphysics model, and the pseudo output from a 2D physics-based model. The actuator setups of the model include the location, size, and vibration frequency of the actuator(s). The output of the surrogate model is the acoustic pressure pattern in the polymer matrix where the particle resides. Based on previous experience, the acoustic pressure pattern in the solution has a direct connection to the actual particle pattern [6]. The 2D model, which is derived from the cross-section of the 3D model, can run many orders of magnitude faster than the 3D model. Both the 2D and the 3D models are constructed and implemented in the COMSOL Multiphysics simulation software. To the best of our knowledge, this is the first attempt to incorporate physics into a machine learning model for surrogate modeling of an AFP system to improve the performance of the surrogate model in acoustics pattern distribution prediction.

In summary, our contributions are as follows:

- First, we propose the PIED model for surrogate modeling of the AFP system, which takes advantage of incorporating physics knowledge into deep learning for the image prediction task. The training and test samples involve 3,133 input-output pairs simulated from 2D and 3D multiphysics simulations.
- Second, we investigate the effects of adding the physics-informed input through feature maps analysis. The results show that the physics-informed input can provide meaningful structure to the feature maps of the ED network, which helps facilitate the training process.
- Third, we demonstrate the performance of ED and PIED models, designed to have different depths (numbers of convolutional layers) and use different activation functions and trained using datasets of different sizes. The results indicate that incorporating physics information allows the resulting physics-informed surrogate models to achieve similar prediction accuracy with fewer training samples and smaller model sizes (i.e., models with smaller depths) regardless of the activation functions used.

The rest of the paper is arranged in the following manner. Section 2 presents the methodology, which covers the surrogate modeling problem setup, the numerical model, and the overall framework of the surrogate model. Section 3 describes the generation of simulation data and the implementation details of training the surrogate model. Section 4 shows and discusses the results in predicting the acoustic pressure distribution. Section 5 summarizes the key ideas of this study, offers concluding remarks, and provides future recommendations for similar work

2 METHODOLOGY

2.1 Surrogate Modeling as Image Regression

The acoustic field-assisted particle patterning (AFP) system considered here is modeled by PDEs with solutions $y(s,x)$, i.e. the model response $\boldsymbol{y} \in \mathbb{R}^{d_y}$ at spatial location $\boldsymbol{s} \in \mathcal{I} \subset \mathbb{R}^{d_s}$ ($d_s = 1,2,3$), with one set of experimental condition $\boldsymbol{x} \in \mathbb{R}^{d_x}$, $\{\boldsymbol{x}(\omega), \omega \in \Omega\}$, where d_y is the number of outputs variables that we are interested in the AFP system, $\mathcal{I}$ is the index set, d_x is the number of input variables to realize one set of experimental condition, Ω is the sample space. This formulation allows for multiple output channels (i.e., $d_y > 1$) even though our interest here is on one output variable represented by the acoustic pressure distribution in the solution. In particular, we are interested in the acoustic pressure distribution on a 2D plane at a specific height, which we have $d_s = 1,2$. The number of input channels does not necessarily equal to d_x. We will discuss in a later section our physics-informed input allows for the use of only two channels to describe multiple input variables. We assume the computer simulation for the physical systems is performed over a given set of spatial grid locations $\mathcal{I} = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_{n_s}\}$(e.g., mesh nodes in finite element methods). In this case, the set experimental condition is defined over the fixed grids $\mathcal{I}$, where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^{d_x n_s}$. The corresponding response $\boldsymbol{y}$ is solved over $\mathcal{I}$, thus can be represented as a vector $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^{d_y n_s}$.

With the discretization described above and assuming fixed initial conditions, we consider the computation simulation as a black-box mapping of the form:

$$\eta: \mathcal{X} \to \mathcal{Y}. \tag{1}$$

In order to speed up the process optimization step, that is to find the desirable set of input conditions $\mathcal{X}$ for a given acoustic pattern, a surrogate model $\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$ is trained using limited simulation data $\mathfrak{D} = \{(\mathbf{x}^i, \mathbf{y}^i)\}_{i=1}^N$, to approximate the 'ground truth' simulation-induced function $\mathbf{y} = \eta(\mathbf{x})$, where θ are the surrogate model parameters, and N is the number of simulation runs (number of training simulation-based data).

For a 3D AFP system, we are interested in the acoustic pressure distribution over 2D regular grids of $H \times W$, where H and W denote the number of grid points in the two axes of the spatial domain (height and width), and $n_s = H \times W$. It is reasonable to arrange the simulation data as an image dataset $\mathfrak{D} = \{(\mathbf{x}^i, \mathbf{y}^i)\}_{i=1}^N$, where $\mathbf{x} \in \mathbb{R}^{d_x \times H \times W}$ is the input conditions, and $\mathbf{y} \in \mathbb{R}^{d_y \times H \times W}$ is the simulated acoustic pressure distribution for $\mathbf{x}$.

Therefore, we transform the surrogate modeling problem into an image-to-image regression problem, with the regression function as

$$\eta: \mathbb{R}^{d_x \times H \times W} \to \mathbb{R}^{d_y \times H \times W}. \tag{2}$$

The image regression problem is concerned with pixel-wise predictions, e.g., predicting the gray scale of each pixel in an image, or in our physical problem, predicting the acoustic pressure at each grid point. A common strategy for image regression problems is to use an encoder-decoder architecture [33], [34]. The intuition behind the encoder-decoder architecture is to go through a coarse-refine process, i.e., to extract high-level coarse features from the input images using an encoder network, and then refine the coarse features to output images through a decoder network. One characteristic of a typical encoder-decoder vision task is that the input and output images share the underlying structure. However, for our surrogate modeling tasks, the input and output images can appear to be different due to the physical modeling process defined by PDEs. Regardless, we will show that an encoder-decoder architecture with physics-informed input can be used as a surrogate of the original physics-based model.

2.2 Numerical Model

2.2.1 3D model

The numerical model of the AFP setup was implemented and solved by using finite element software COMSOL Multiphysics 5.2a. Fig. 1(a) shows the geometry and material choices for the finite element model (FEM) to reflect the experimental setup of the AFP system for the manufacturing of Polydimethylsiloxane (PDMS) and aluminum (Al) particle composite with controlled spatial distribution of Al particles inside the PDMS matrix. In the FEM model, a layer of PDMS fprecursors (l: 77 mm, w: 77 mm, h: 2mm) is at the top, followed by a Kapton film (l: 77 mm, w: 77 mm, h: 0.075 mm) that is directly underneath, and finally, the four piezoelectric transducer (PZT) plates (l: 20 mm, w: 15 mm, h: 1.5 mm) is placed at the bottom. The material choice for the PZT is PZT-4D from the material library in COMSOL. Different types of physics were applied to different areas of interest for simulation (simulation domain). In total, there are three simulation domains: Kapton film domain, PZT plates domain, and PDMS precursors domain. "Solid Mechanics-Linear Elastic Materials" physics was applied to the Kapton film domain, both "Solid Mechanics-Piezoelectric Materials" and "Electrostatics-Charge Conservation, Piezoelectric" physics were applied to the PZT domain, and the "Pressure Acoustics" physics was applied to the PDMS precursors domain. The boundary conditions of the model are as follows: the bottom surface of the PZT was set to "roller" condition, the interface between the PZT and Kapton film was coupled, the interface between Kapton film and PDMS precursors was set to "Acoustic-Structure Boundary," the edge of the Kapton film was set to be fixed in the z-axis within the

"Prescribed Displacement", the rest of the physical boundaries of Kapton film and PZT were set to "free" condition, and the rest of the boundaries of the PDMS precursors domain was set to "Sound Hard Boundary" condition.

The electrostatic boundary conditions for the four PZT plates were set in accordance with whether the PZT was turned on. To simulate the case when all four PZTs are "on," a "Electric Potential" boundary condition with an amplitude of 80 V was applied to the bottom surface of the PZTs, whereas a "Ground" boundary condition was applied to the top surface of the PZT. When using a frequency domain study, the model was solved for the displacement in the Kapton film and the acoustics pressure in the PDMS precursors matrix by simulating an applied sinusoidal potential to the PZT plates. According to [6], the distribution of suspended particles is controlled by the acoustic pressure in the PDMS precursors matrix, which is the medium where the suspended particles distributed. In other words, the pattern of the particles is similar to the pattern of the acoustic pressure in the PDMS precursors domain. Since the computation time of a particle trajectory simulation is much longer than that of acoustic pressure pattern simulation, the acoustic pressure pattern within the PDMS precursors domain in the x-y plane is used as the output from the FEM.

The physical properties of the components in the AFP system need to be determined. For the Kapton film, the Young's modulus, the density, and the Poisson's ratio were set to 2.5 GPa, 1420 kg/m^3, and 0.34, respectively. The speed of sound was set to 889 m/s, and the effective density of the PDMS precursors with the aluminum (Al) particles was set to 1000 kg/m^3.

2.2.2 2D model

The 2D model, a simplified version of the 3D model, is constructed based on the cross-section of the 3D model in the xz-plane (Fig. 1(a)). The dash-double dot line in Fig. 1(a) shows the cross-section location for the 2D model in the 3D model. Fig. 1(b) shows the 2D model we used in this study, which is the cross-section at the dash-double dot line shown in Fig. 1(a). At the bottom are the two PZT plates domain (l: 15 mm, h: 1.5 mm), attached to the top of the PZT-4D plate is a Kapton film domain (l: 108.89 mm, h: 0.075 mm), and at the top of the Kapton film is the PDMS precursor domain (l: 108.89 mm, h: 2 mm). All the boundary conditions are specified similarly to those from the 3D model. In the 2D model, only one of the PZT-4D is set to "on" with a sinusoidal voltage of 80 V at different frequencies.

The 2D model consisted of all the material components of the 3D AFP system. Note that the 2D model consisted of two PZTs, which allow the simulation of acoustic pressure resulting from the sinusoidal voltage applied to the PZTs. Due to the reduction of dimension and the omission of the two PZTs which do not intersect with the dash double-dot line in Fig. 1(a), the acoustic pressure pattern simulated from the 2D model is different from that of the 3D model. However, their acoustic pattern responses should correlate qualitatively to the changes in frequency and material properties in similar ways. Therefore, the output from the 2D model is modified to obtain the pseudo-output, which is then used as the physics-informed input in the proposed physics-informed surrogate model. Details on obtaining the pseudo-output is presented in section 2.3.

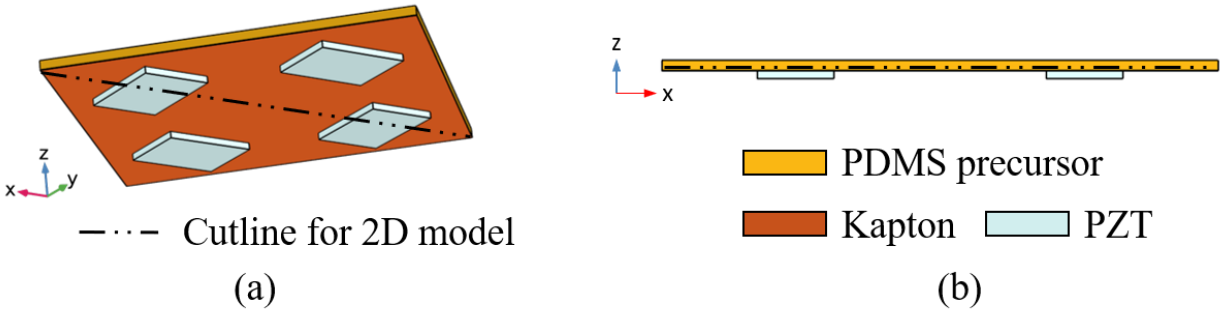

Fig. 1 (a) The 3D FEM model and (b) the 2D FEM model of the AFP setup. The material choice in each simulation domain is color-labeled for illustration purpose. The 2D model is derived from a cut-plane along the dash-double dot line presented in (a).

2.3 Input and Output Structures

The objective of this study is to accurately predict the acoustic pressure pattern of an AFP system for given experimental conditions, such as frequency of the PZT, location, and shape of the PZT, and material properties of each of the components in the system. Two surrogate modeling approaches are considered in this study (See Fig. 2). The first approach (Fig. 2 (a)) is termed the base surrogate model, which combines the PZT setup (number and locations) and frequency of the PZT to predict the resulting acoustic pressure pattern. The inputs to this base surrogate model are similar to those of a 3D FEM. The surrogate

model is served to learn the interaction of the PDEs and the inputs to simulate the output, which is the spatial acoustic pressure pattern. Specifically, the input to the base surrogate model is a single channel ($d_x = 1$) image with $p \times p$ pixels ($H = p$ and $W = p$) that show the location and shape of the PZT that is turned on (See Fig. 2). In Fig. 2, we show the case where only one PZT at the left is "on," and therefore, other PZTs are not visible. The frequency at which the PZT vibrates is set as an input to the decoder. The number of pixels for the input image, p, is determined based on the depth of the encoder-decoder model used (See Fig. 3). Fig. 3 illustrates a six convolutional layer PIED model architecture used in this study. p is determined such that the size of the feature map from the last convolutional layer of the encoder before the fully connected layer is always 4×4.

The second surrogate modeling approach is a physics-informed surrogate model. We proposed to include a physics-based input, which is modified from the output of the 2D physics-based model. We chose to incorporate the physics-informed input as an additional input channel, $d_x = 2$, with the same image size ($p \times p$). Any additional input channel represents a new variable (information) to be learned by the model. The idea of incorporating a new variable through an additional image channel originates from the fact that we need three channels to represent a colored image. Here, the additional input channel provides physical information about the AFP system. The process of constructing the physics-based input is as follows: 1) First, we simulate the acoustic pressure in the x-direction from the 2D model; 2) Then, we repeat this acoustic pressure in the y-direction to make an image that has the same size as the base surrogate model. We will show in the result section that the physics-based input is able to provide a favorable structure to the feature maps of the surrogate model, which facilitates the training of the surrogate model.

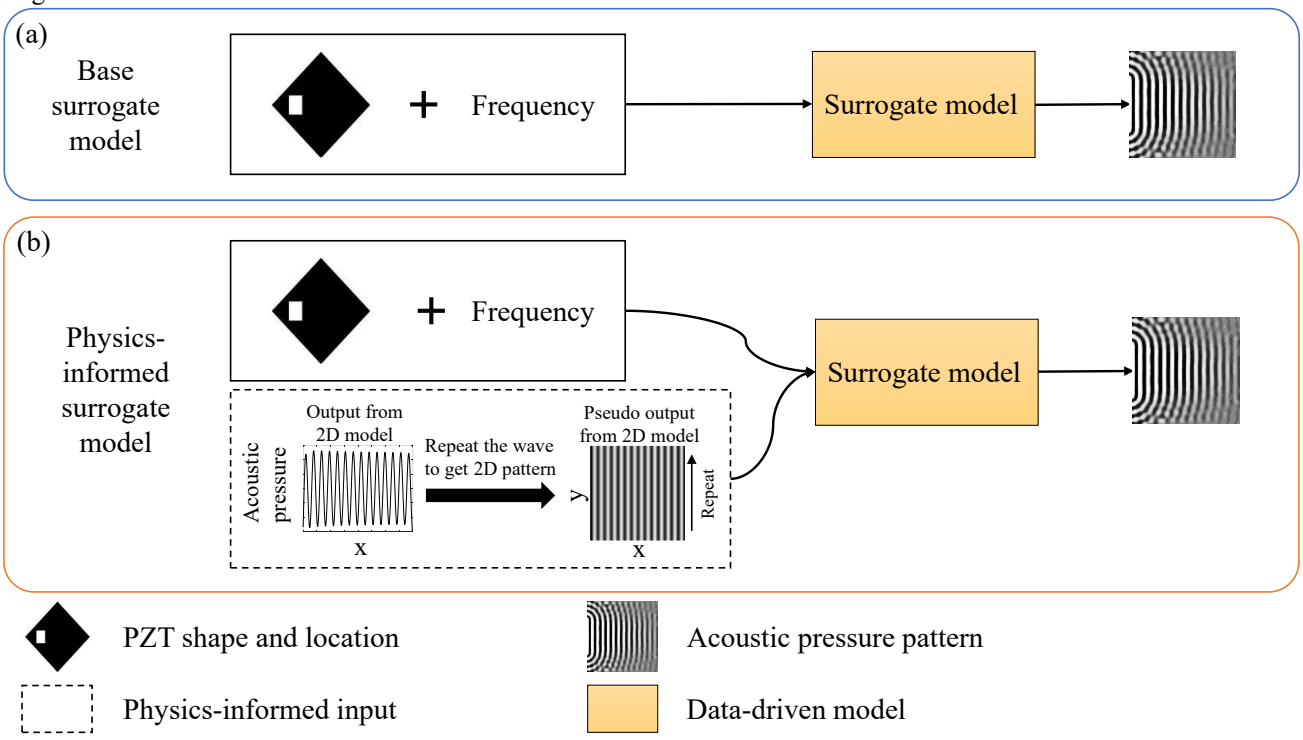

Fig. 2 The two surrogate models proposed in this study. (a) Base surrogate model; (b) Physics-informed surrogate model incorporating lower-order physics-based model's output as the input.

2.4 Overall Framework

This research will investigate using an encoder-decoder (ED) architecture as a surrogate model to replace a computationally expensive FEM for the AFP system. The overall structure of the surrogate model is shown in Fig. 3. The encoder comprises six convolutional (Conv) stages, each of which includes a Conv layer, a batch normalization (BN) layer, and an activation layer. In the context of deep learning, convolution is a linear operation that involves the multiplication of the inputs and kernels, which is a set of weights. The Conv stage is used to down-sample the input image to extract high-level characteristic features. The encoder is followed by a fully connected (FC) layer, which compresses the 2D features into a one-dimensional (1D) feature vector. The frequency at which the PZT vibrates is concatenated to this 1D feature vector before feeding into the decoder. The decoder comprises seven transposed convolutional (ConvT) stages, each including a transposed convolutional layer, a BN

layer, and an activation layer. The transposed convolutional stage is used to up-sample the feature vector to a desired output feature map to recover the input spatial resolution. In this study, the encoding layer halves the size of the feature maps, while the decoding layer doubles the size of the feature maps. A leaky rectified linear unit (LeakyReLU) is used as the activation function for all the convolutional stages and transposed convolutional stages except for the output layer. We choose to use sigmoid as the activation function for the output layer as it gives better output images compared to other activation functions we have tried. The LeakyReLU and sigmoid activation functions are used to introduce nonlinear properties into the neural networks. The BN layer is a widely used technique in deep learning for reducing the changes of the inputs' distribution in each layer and stabilizing the network training process. The implementation of the surrogate model is accomplished using the Pytorch package.

2.5 Training Algorithm

A deep neural network model contains many parameters (e.g., weights and bias of convolutional and FC layers) that need to be optimized during the training process. To properly identify the optimum values for these parameters in our physics-informed encoder-decoder model, a loss function L was defined to measure differences between the model predictions and the associated ground truth. We employed a widely used optimization method, Adam optimizer [35], to update the model parameters $\boldsymbol{\theta}$ (weights $\boldsymbol{\omega}$, and biases $\mathbf{b}$) to minimize the loss function. The optimization process is repeated many times, with each iteration executed on a small batch of training samples, until the loss converge to a small value. Given an input image $\mathbf{x}$, a target image $\mathbf{y}$, and the prediction $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$, the training loss function $L(\boldsymbol{\theta})$ is a mean square error with regularization term was defined as:

$$L(\theta) = \frac{1}{m \times n} \sum_{i=1}^{m \times n} (\mathrm{f}_i - \mathrm{y}_i)^2 + \lambda \Omega(\boldsymbol{\theta}), \tag{3}$$

where $\lambda \Omega(\boldsymbol{\theta}) = \frac{\boldsymbol{\lambda}}{\mathbf{2}} \boldsymbol{\theta}^{\mathbf{T}} \boldsymbol{\theta}$ is the regularization term, λ denotes the $\mathcal{L}_2$ regularization factor that weighed the relative contribution of the regularization term, m is the number of samples used in each iteration, and $n = d_y \times H_y \times W_y$ is the number of pixels in all channels of one output image. The parameters $\boldsymbol{\theta}$ includes the kernel weights in all the convolution and transposed convolution layers, the scale and shift parameters in all the batch normalization layers, and the weights and bias in an FC layer.

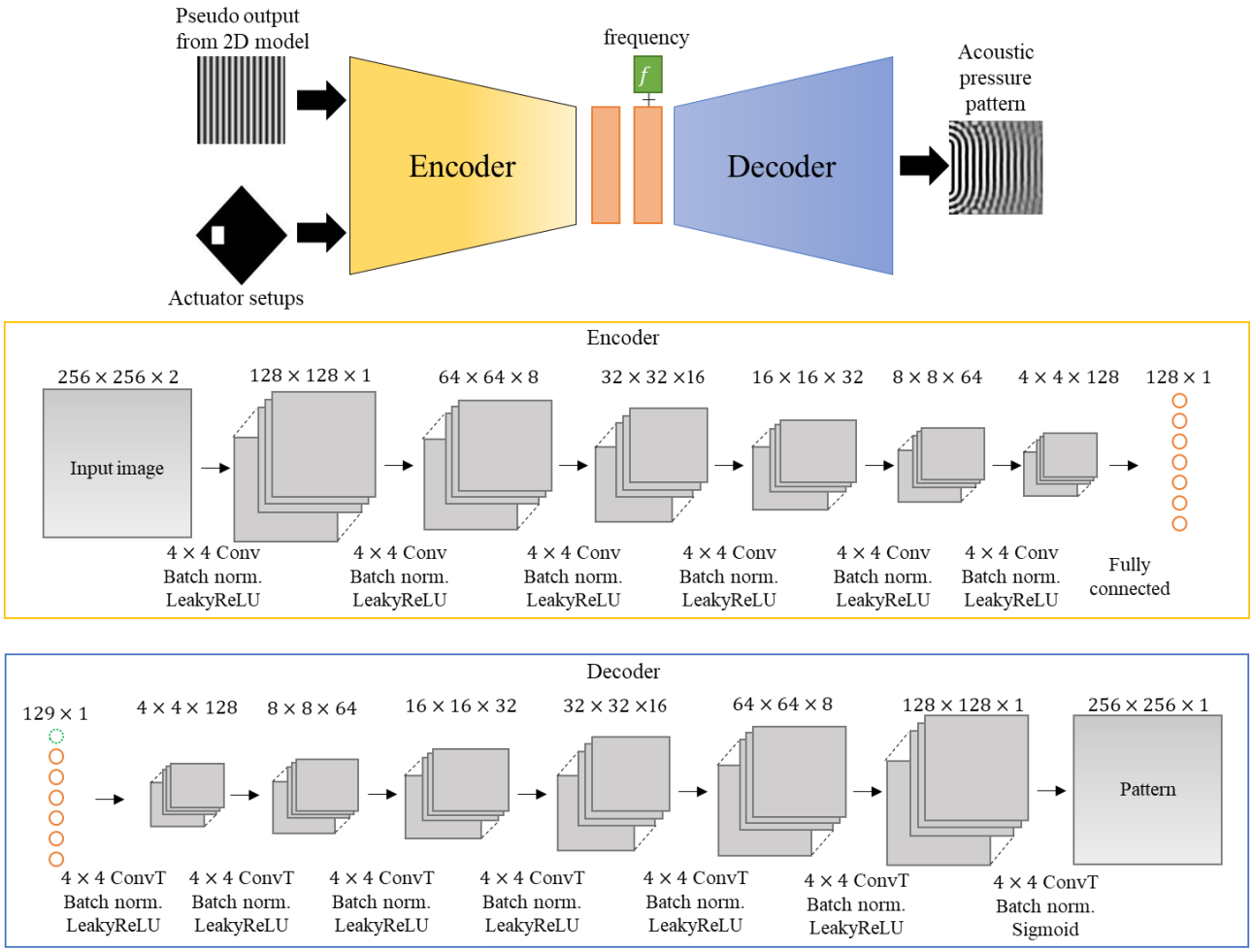

Fig. 3 Overview of the implemented physics-informed encoder-decoder (PIED) model in this study.

3 SIMULATION DATA AND IMPLEMENTATION DETAILS

3.1 Simulation data

In this study, the PIED model is trained with and tested on simulation data from the 3D FEM. For each simulation, we considered the variation of two different input parameters: the number and location of PZT and the vibrational frequency of the PZT. We considered the case when the PZT actuator could only exist at four different locations, as shown in Fig. 1(a). The numbers of PZT actuators that were "on" were set to be one, two, three, and four. For each PZT arrangement except the arrangement with four PZTs, we rotated the PZT and the resulting acoustic pressure image to 90, 180, and 270 degrees to increase the sample size. We used a full factorial experimental design to sample the frequency at which the PZT actuator vibrates from 10000 Hz to 160000 Hz. The sampling rate from 10000 Hz to 100000 Hz was 500 Hz, whereas the sampling rate from 100000 Hz to 160000 Hz was 1000 Hz. In total, there were 3133 unique combinations of PZT arrangements and frequencies sampled between 10000 Hz and 160000 Hz.

3.2 Image filtering

We observed a low-frequency acoustic pressure pattern caused by the excitation of a low-frequency vibration mode when the PZT vibrates at frequencies larger than ~100000 Hz (See Fig. 4). Fig. 4(a) shows the acoustic pressure pattern at 149000 Hz. It can be observed that the high-frequency acoustic pattern is convoluted by a low-frequency pattern, which makes it difficult to extract the high-frequency acoustic pattern of interest. In this study, the low-frequency vibration mode is of low interest because it either is an artifact from the finite element model simulation or has a negligible effect on the

final particle pattern distribution in the experiment, which was demonstrated in the experimental results (section 4.1). The process of removing this vibration mode is shown in Fig. 4. In essence, we applied a low pass filter in the frequency domain of the image by utilizing a fast Fourier transform (FFT). The resulting low-frequency vibration mode was removed from the original acoustic pressure pattern to reveal the high-frequency pattern.

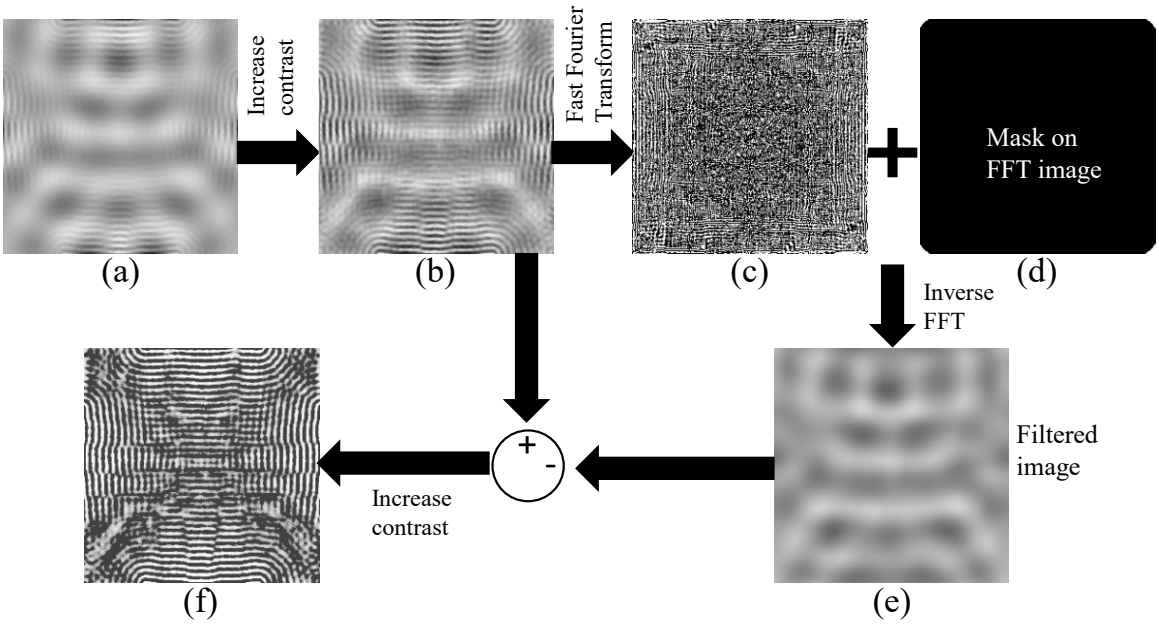

Fig. 4 Processes of removing the low-frequency vibration mode in the acoustic pressure pattern. (a) Acoustic pressure pattern from 3D FEM, (b) Acoustic pressure pattern with higher contrast by using an *adapthisteq* function in Matlab, (c) FFT of image (b), (d) the low pass filter for image (c), (e) the inverse FFT of the filtered image (c), and (f) is the resulting acoustic pressure pattern after the removal of low-frequency vibration mode.

3.3 Implementation detail

The objective of the Adam optimizer is to minimize the loss function in equation (3). The number of training epochs is set to be 1000 with a mini-batch size of 64 samples. We chose an initial learning rate of 0.0005 and decreased this rate by a factor of 10 for every 500 epochs. Table 1 presents the important parameters used in training the PIED.

The proposed PIED model consists of five convolutional layers, five transposed convolutional layers, and two FC layers. The model parameters $\boldsymbol{\theta}$ (weights and biases) are randomly initialized with a mean of 0 and a standard deviation of 1.

It is important to prevent overfitting (better generalization) when training a deep neural network. There are many ways to ensure the generalization of the model. We employed early stopping. That is, we stopped the training process before the preset number of epochs when the validation mean square error (MSE) was larger than or equal to the smallest validation MSE for 10 epochs consecutively.

The training algorithm and PIED model are implemented in PyTorch, which is an optimized tensor library used for deep learning applications, with GPU acceleration

Table 1 List of parameters used in the surrogate model training

Parameter	Value
Number of epochs	1000
Initial learning rate	0.0005
$\mathcal{L}_2$ regularization	0.01

3.4 Image quality metric

The accuracy of the PIED model is evaluated on the test data separated from the simulation data from the 3D FEM. 20% of all samples are randomly removed from the simulation data set as the test data (627 samples), and the remaining 80% of all samples are used for the training set (2506 samples). The same training and test set are used for all the analyses unless otherwise specified. A well-known quality metric named structural similarity index measure (SSIM) was used to measure the similarity between the true image (simulated from the 3D FEM) and the predicted image (output from the surrogate model). It was developed by Wang et al. [36], and is considered correlated with the quality perception of the human visual system. The SSIM measures image distortion as a combination of three factors, namely luminance distortion, contrast distortion, and structural distortion, expressed collectively as

$$SSIM(\mathbf{y},\hat{\mathbf{y}}) = l(\mathbf{y},\hat{\mathbf{y}})c(\mathbf{y},\hat{\mathbf{y}})s(\mathbf{y},\hat{\mathbf{y}}) \tag{4}$$

where

$$\begin{cases} l(\mathbf{y},\hat{\mathbf{y}}) = \frac{2\mu_{\mathbf{y}}\mu_{\hat{\mathbf{y}}}+C_1}{\mu_{\mathbf{y}}^2+\mu_{\hat{\mathbf{y}}}^2+C_1} \\ c(\mathbf{y},\hat{\mathbf{y}}) = \frac{2\sigma_{\mathbf{y}}\sigma_{\hat{\mathbf{y}}}+C_2}{\sigma_{\mathbf{y}}^2+\sigma_{\hat{\mathbf{y}}}^2+C_2} \\ s(\mathbf{y},\hat{\mathbf{y}}) = \frac{\sigma_{\mathbf{y}\hat{\mathbf{y}}}+C_3}{\sigma_{\mathbf{y}}\sigma_{\hat{\mathbf{y}}}+C_3} \end{cases} \tag{5}$$

The term $l(\mathbf{y},\hat{\mathbf{y}})$ measures the closeness of the two images' mean luminance, $\mu_{\mathbf{y}}$ and $\mu_{\hat{\mathbf{y}}}$, which maximized and equal to one only if $\mu_{\mathbf{y}} = \mu_{\hat{\mathbf{y}}}$. The second term $c(\mathbf{y},\hat{\mathbf{y}})$ measures the similarity of the two images' contrast by the standard deviation. The second term is maximal when $\sigma_{\mathbf{y}} = \sigma_{\hat{\mathbf{y}}}$. The third term $s(\mathbf{y},\hat{\mathbf{y}})$ measures the correlation coefficient between two images, where $\sigma_{\mathbf{y}\hat{\mathbf{y}}}$ is the covariance between $\mathbf{y}$ and $\hat{\mathbf{y}}$. SSIM index goes from 0 to 1, which corresponds to no correlation between two images to $\mathbf{y} = \hat{\mathbf{y}}$. The three positive constants $C_1, C_2,$ and C_3 are used to avoid null denominator.

4 RESULTS AND DISCUSSION

4.1 Experimental validation

We validated the 3D FEM with experimental results. Fig. 5 compares the pattern of Al particles in PDMS with the acoustic pressure pattern simulated from the 3D FEM. Visually comparing the particle pattern from the experiment and acoustic pressure patterns from the simulation can be challenging. Here, we chose to use the distance between two adjacent particle patterns and the distance between two adjacent acoustic nodes (locations with minimum acoustic pressure) to quantify the validity of the simulation model. This is because the particles in this experiment will aggregate at the locations with minimum acoustic pressure (i.e. acoustic nodes) due to their positive acoustic contrast factor [6]. The distance values averaged over ten measurements on the same sample are listed in Fig. 5. It can be observed that the distance between the acoustic nodes under two different frequencies, 80 kHz and 120 kHz, matched the distance of the particle patterns (lines or islands of particles). In general, the shape of the particle pattern matches very well with the simulated acoustic pressure. The results in Fig. 5 show that the FEM used in this study is validated with an actual experimental setup. It is thus reasonable to use the FEM to generate data for training and testing the surrogate model.

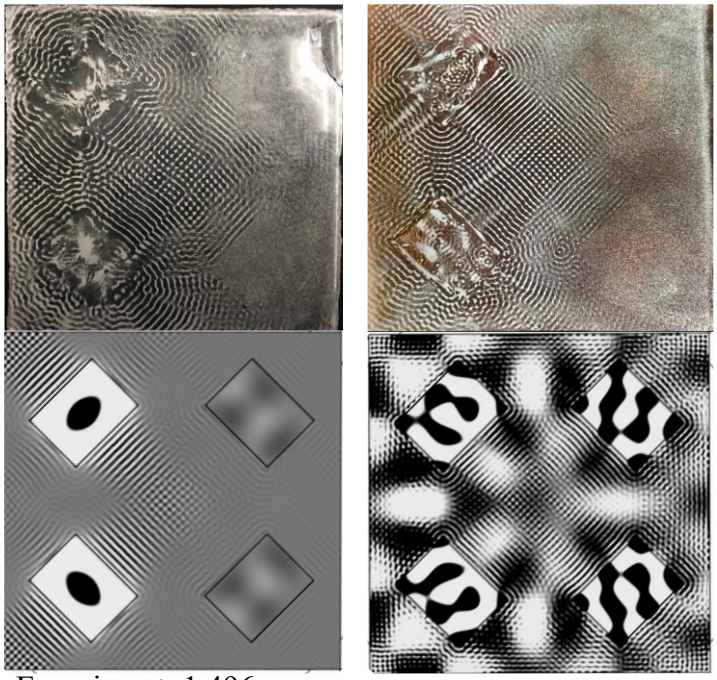

Fig. 5 Validation of simulation model with experimental results by comparing the distance between the locations of particle accumulation with the distance between the location where acoustic pressure is minimum for two different frequencies (a) 80 kHz and (b) 120 kHz.

4.2 Acoustic pattern prediction from the surrogate model

Figure 6 shows the prediction results from the two surrogate modeling approaches of five selected AFP setups. The first setup has one vibrating PZT located at 0 degrees (we set this angle to increase as the PZT moves in the clockwise direction), and it was set to vibrate at 11 kHz. The second setup has two vibrating PZTs located at 0 and 270 degrees, and they vibrate at 15.5 kHz. The third setup has four vibrating PZTs located at 0, 90, 180, and 270 degrees, and they vibrate at 73 kHz. The fourth setup consisted of three vibrating PZTs located at 0, 180, and 270 degrees, and they vibrated at 125 kHz. The fifth setup consisted of one vibrating PZT located at 90 degrees, and it vibrated at 133 kHz. From the figure, we observed that the acoustic pressure pattern at higher frequencies is less regular due to the interference from a low-freqeuency vibration mode. Although the filtering process can effectively remove the low-frequency acoustic pattern, it cannot regenerate the corrupted high-frequency patterns. As a result, there are pixelated regions in some filtered acoustic patterns. These pixelated regions consisted mostly the artifacts from the image filtering process and can hardly be learned by the surrogate model. In Fig. 6(a), we show the prediction results by the base surrogate (ED) model. It can be observed that for most of the AFP setup, the ED model can only accurately predict parts of the acoustic pressure pattern. The SSIM metric can be presented in two ways. In one way, we presented the average SSIM for the quality of the whole image. In another way, we presented the full SSIM image, where a darker pixel indicates lower quality. The full SSIM image is calculated using a moving average of a square kernel (default at seven pixels on each side when using a function named *compare_ssim* from the scikit-image (a.k.a. skimage) python package) through the two images being compared. The average SSIM is calculated by taking the mean of the full SSIM image. The full SSIM image allowed us to visualize the image quality of the prediction. From Fig. 6(a), the SSIM value of the ED model is generally less than 0.5, and based on the SSIM image, the quality of the predicted acoustic pattern does not match the true pattern for most cases, even when the acoustic patterns are clean and regular (first and second setups).

We presented the prediction results by PIED model for the same five AFP setups in Fig. 6(b). From the SSIM image, we can see that the quality of the predicted acoustic patterns closely matched the true acoustic pattern when the acoustic patterns are clean and regular. When there exists pixelated regions, similar to those shown in fourth and fifth AFP setups, the PIED model can still accurately capture regions with regular and clean acoustic patterns. Fig. 6 clearly contrasted the PIED model from the ED model. With the same model architecture of six convolution stages and seven transposed convolution stages, same

parameters initialization, same training epochs, and learning rate, the PIED model clearly demonstrates the benefit of introducing physics-informed input to the ED model. We use PIED6 and ED6 to represent models with six convolution stages. The average SSIMs over all test data are 0.77 and 0.53 for PIED6 and ED6 models, respectively. To verify whether the prediction accuracy improvement is caused by additional input variable (PIED6 model uses an input image of two channels while the ED6 model uses an input image of one channel), we introduce an non-physics-informed input to the ED6 model. The non-physics-informed input contains the material properties and physical dimensions of all simulation domains (See Fig. A1 in Appendix A). Details on how to contruct the non-physics-informed input are described in Appendix A. Based on the prediction result, the surrogate model (ED6) with non-physics-informed input do not show any prediction accuracy improvement over the base surrogate model. The result indicates the prediction accuracy improvement arises from the introduction of physics-informed input, rather than having an additional input variables.

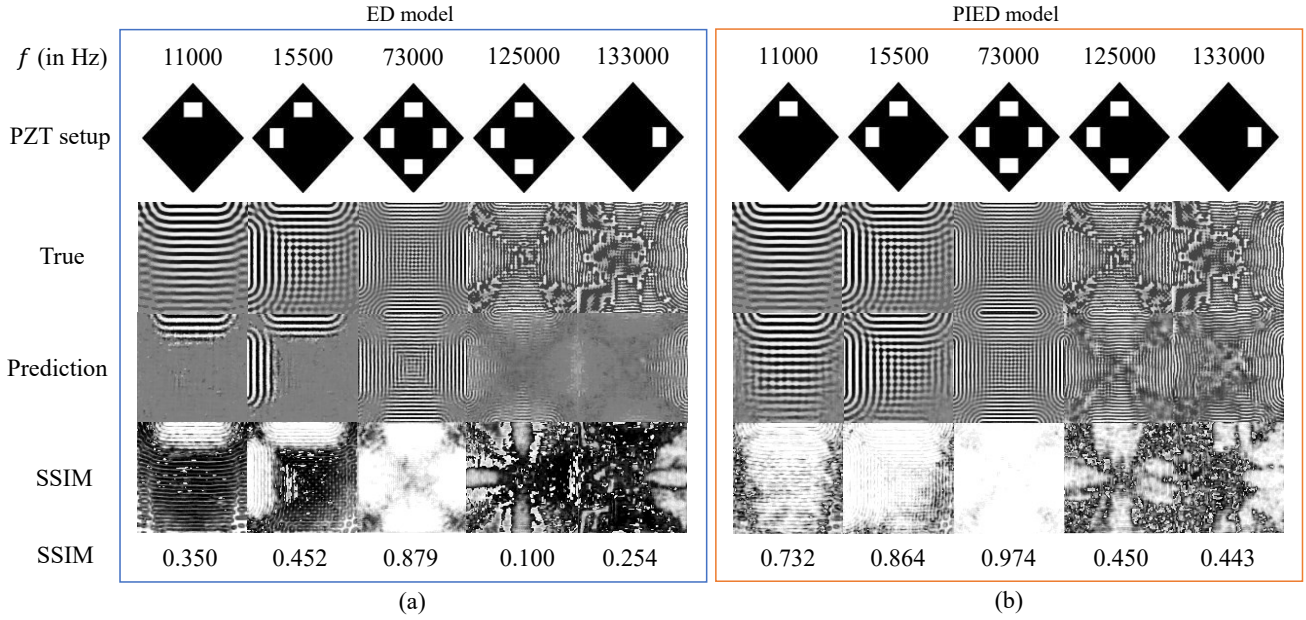

Fig. 6 Acoustic pressure patterns predicted by (a) ED6 model and (b) PIED6 model.

4.3 Feature maps analysis

To understand the reason behind the better performance of the PIED model, we looked into the feature maps of the convolutional layers. A feature map is the output of one kernel/filter applied to the previous layer. It can be used to visualize the types of features that the model learned and provide insight into the internal representation that the model has of a specific input at a given point in the model. In Fig. 7, we show part of the feature maps from the first three layers of the encoder for the first AFP setup shown in Fig. 6. It can be observed that without a physics-informed input, the feature maps from the ED model only learned the PZT setup in the convolutional layer, which is the location of the vibrating PZT and the shape of the Kapton film. In contrast, the feature maps from the PIED model show line patterns closely match the pseudo-output from the 2D FEM. These line patterns give the model an underlying structure, which will help the decoder better reconstruct the acoustic pattern of the 3D model (see Fig. 6).

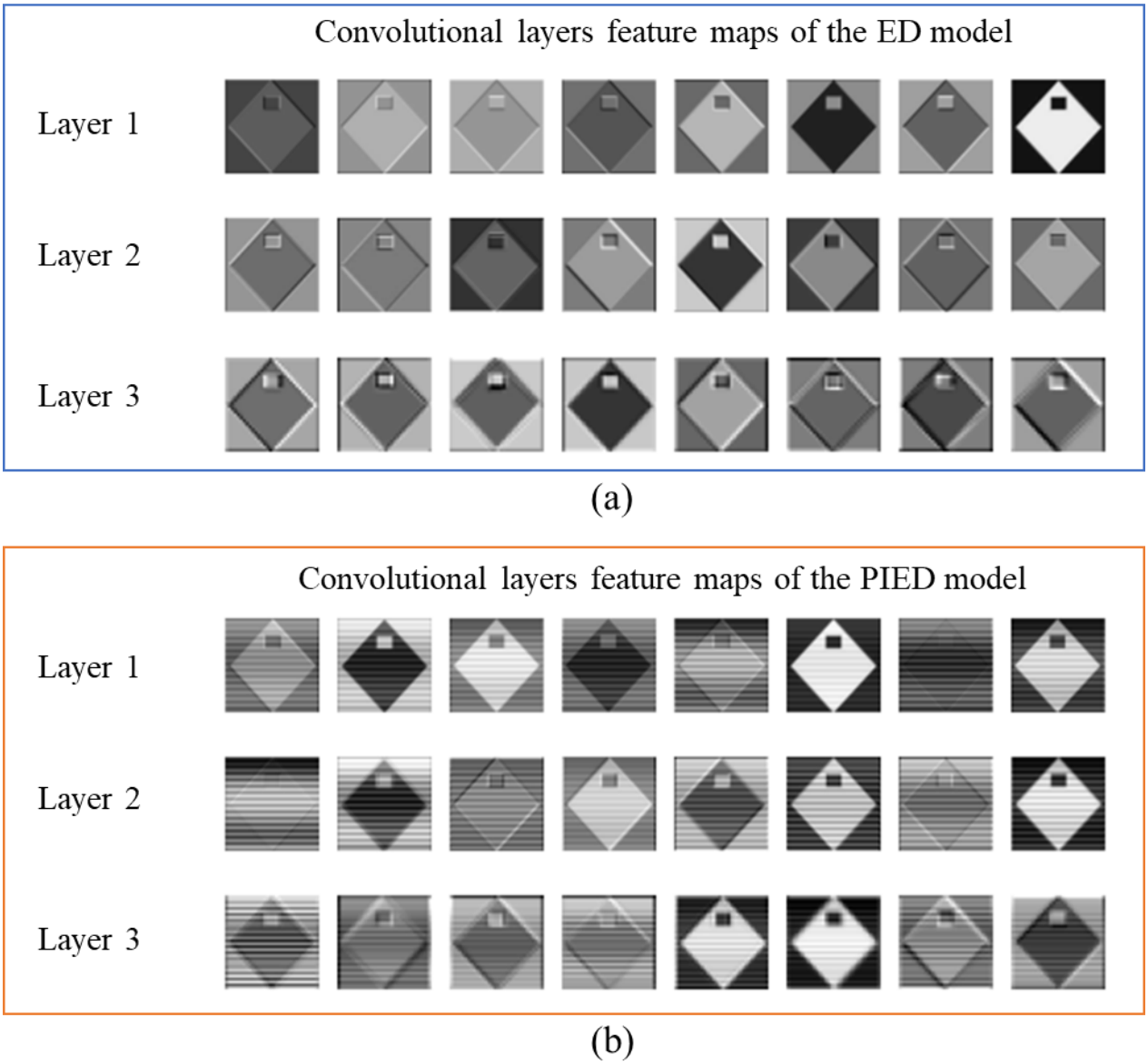

Fig. 7 Feature maps in the first three convolutional layers of (a) ED model and (b) PIED model.

4.4 Effects of convolutional layers and number of training samples

It is commonly known that the optimal number of convolutional layers is closely related to the amount and complexity of the available data. A deeper network (i.e., more layers) is generally better suited for more complex tasks, such as computer vision tasks. However, if insufficient data is used for training, a deeper network has a larger chance of suffering from overfitting. In this study, we implemented an early stopping strategy to prevent overfitting (See section 3.3). We conducted a parametric study to empirically investigate the effect of the number of convolutional layers and the number of training data on the acoustic pressure prediction of the surrogate model. In Fig. 8, we compared the training time, training error, test time, and test accuracy of the two surrogate modeling approaches. We use PIEDm and EDm to label the model with m number of convolutional layers and $m+1$ number of transposed convolutional layers. To account for the down-sampling nature of convolutional layers, we set the feature maps size at the last convolutional layer to $C_{Convlast} \times 4 \times 4$, by changing the size of the input image accordingly with the number of convolutional layers. $C_{Convlast}$ is the number of kernels at the last convolutional layer. The maximum input image size is limited to 512×512 pixels due to the memory constraints of the graphic card. To evaluate the effect of the training sample size, we purposely reduce the sample size to 25%, 50%, and 75% of the original training set without altering the test data. Based on Fig. 8, four observations can be made:

- The training time (Fig. 8(a)) and test time (Fig. 8(c)) of PIED7 and ED7 are much longer compared to other models with fewer Conv layers. These two models are relatively inefficient to train and test. Compared to the model with six Conv layers (PIED6/ED6), the number of parameters in the PIED7/ED7 model is 3 times greater, with a total of 7.7 million model parameters. To decide whether the extra time used for training and testing is worthwhile, we have to look more closely at the improvement in surrogate modeling accuracy.
- Second, for each number of convolutional layers, the PIED model always achieved a lower training error when compared to the ED model. The training error for the ED7 model does not decrease monotonically with the training sample size due to the model requiring a larger epoch number to converge during training while we stop the training process at 1000 epochs. Therefore, some ED7 models are only partially trained with partially optimized model parameters. All PIED models are sufficiently trained before 1000 epochs, which indicates that the underlying feature maps' structure (Fig. 7) provided by the physics-informed input can facilitate the training of the surrogate model.
- Third, the test accuracies of the surrogate models in terms of SSIM increased with a larger number of Conv layers and a larger training sample size. Notably, the PIED models performed much better than ED models, with the PIED5-7 models having significantly higher test accuracy compared to other models. The accuracy improvement from PIED5 to PIED7 slows down. To confirm the optimum number of Conv layers, models PIED8, ED8, PIED9, and ED9 are tested. Note that these two models are executed on a different computer system, with a more capable graphic card. Therefore, the computational time is not suitable for comparison, and it is omitted here. Based on Fig. 9, the prediction accuracies increase rapidly from 3 to 5 Conv layers. The improvement slows down from 5 to 7 Conv layers and eventually reduced from 8 to 9 Conv layers. From Fig. 9, we can conclude that the optimum number of Conv layers is 7 for the PIED model. While the general trends hold true for the ED model, larger model parameters prove to be challenging when there is no physics-informed input, as can be observed by the inconsistent prediction performance for the ED model with a higher number of Conv layers (7 or above).
- Fourth, models with a smaller number of convolutional layers are less sensitive to the number of training data (See Fig. 8(d)). As the model size and depth increase, the improvement of prediction accuracy with the increase of the number of training data also increases. The results show that the more complex the model, the more data it needs for training to achieve the optimum results the model is capable of.

Based on the above observations, it is tempting to pick the model with the highest test accuracy, which is PIED7. However, we need to know whether the extra computational burden is worthwhile. We can compare the degree of test accuracy improvement to the amount of training/test time increase. When doing the calculations, we found that the improvement of the test accuracy from PIED6 to PIED7 is only 2.93%, while it required an extra 61.67% training time and 29.6% test time. The extra training and test time do not bring much improvement in the acoustic pressure prediction. From PIED5 to PIED6, the accuracy improvement is 11.06%, while requiring only an extra 30.93% training time and 28.67% of test time. Hence, when considering both the accuracy improvement and the test time growth, it is obvious that PIED6 gives a much better efficiency compared to PIED7 and considerable accuracy improvement over PIED5. Therefore, the final surrogate model was chosen to contain six Conv layers, which is used to generate the results in Fig. 6 and Fig. 7. The test SSIMs from different models are also summarized in Table 2. The model with the highest SSIM is highlighted in bold. Surrogate models with six Conv layers are also used for other analyses hereafter.

Besides the effects of the number of convolutional layers and training sample size, we also investigated how the choice of an activation function affects the prediction performance. The results are presented in Appendix B. We found from the results that LeakyReLU is among the best-performing activation functions. Regardless of the activation function, the proposed physics-informed surrogate model always outperforms the base surrogate model.

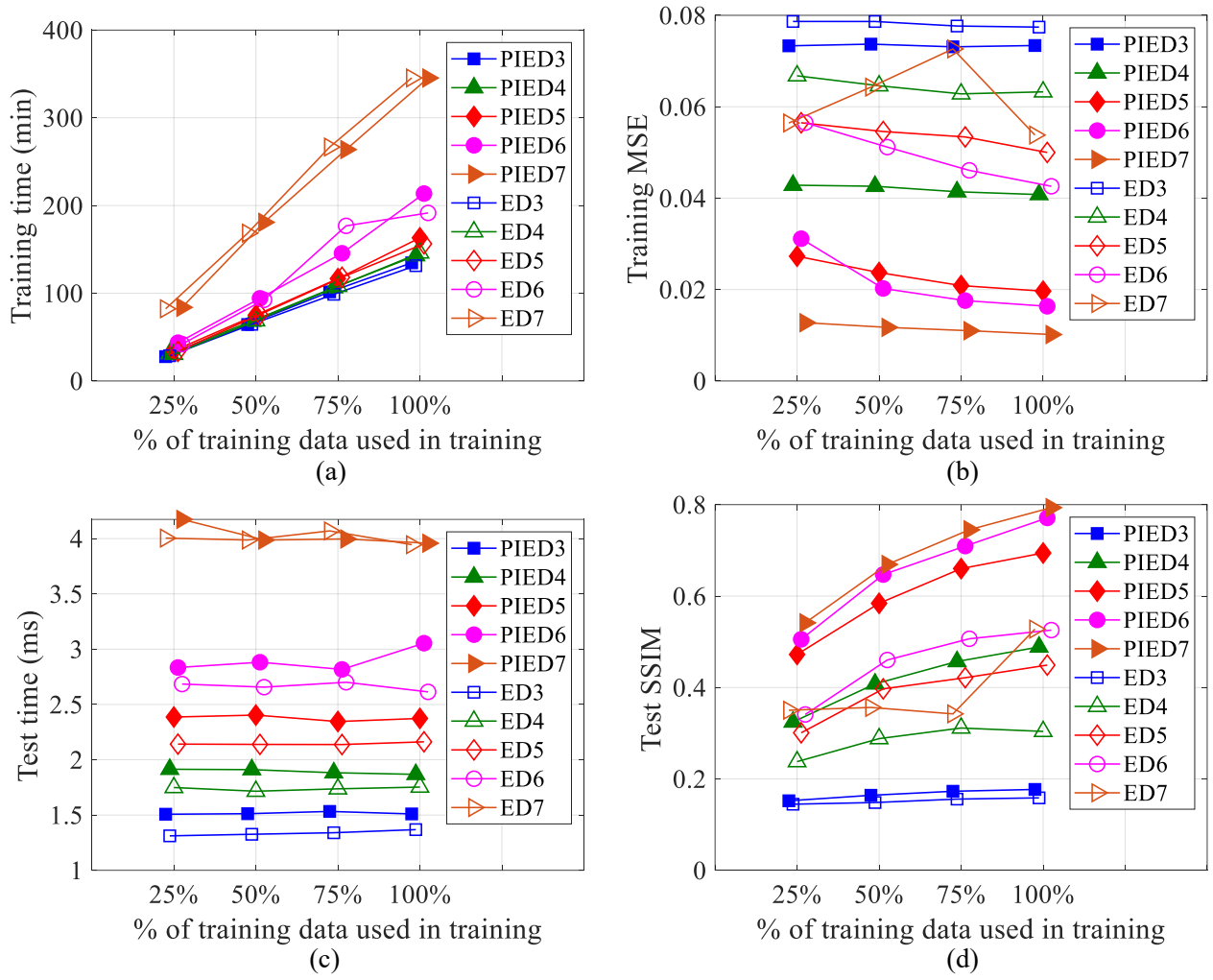

Fig. 8 Parametric study results on the effect of the number of training data and the number of Conv layers. (a) Training time comparisons, (b) training MSE comparisons, (c) test time comparisons, and (d) test SSIM comparisons.

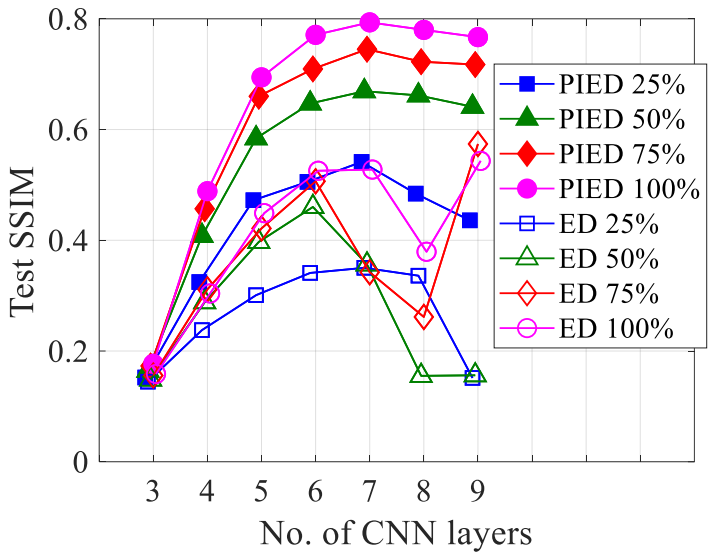

Fig. 9 Effect of the number of Conv layers on prediction accuracies.

Table 2 Test SSIM for models with different convolutional layers and amount of training data. (The number of convolutional layers, m, is limited by the graphic card's memory. $m = 7$ is the largest number of convolutional layers for our graphic card of NVDIA GTX 1070 8GB. For $m > 7$, we run the models using another system with a graphic card of NVDIA GTX Titan XP 12GB)

Training data	PIEDm						
	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$
25%	0.15	0.32	0.47	0.51	**0.54**	0.48	0.44
50%	0.16	0.41	0.58	0.65	**0.67**	0.66	0.64
75%	0.17	0.46	0.66	0.71	**0.75**	0.72	0.72
100%	0.18	0.49	0.69	0.77	**0.79**	0.78	0.77
	EDm						
	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$
25%	0.14	0.24	0.30	0.34	0.35	0.34	0.15
50%	0.15	0.29	0.40	0.46	0.36	0.16	0.16
75%	0.16	0.31	0.42	0.51	0.34	0.26	0.57
100%	0.16	0.30	0.45	0.53	0.53	0.38	0.54

4.5 Out-of-distribution sample prediction

The performance of the proposed PIED model is further evaluate on out-of-distribution (OOD) sample prediction. To evaluate OOD prediction performance, the simulation data are separated into training and test sets based on the frequency at which the simulation data is simulated in COMSOL. To check if the test data is OOD of the training data, we employed t-Distributed Stochastic Neighbor Embedding (t-SNE), an unsupervised, nonlinear technique used for visualizing high dimensional data. The *tsne* function in MATLAB is used for analyzing the t-SNE of the acoustic patterns in the training set and test set. As the *tsne* function only takes vectors as inputs, the acoustic patterns' image data (2D) are flattened into a vector before using the function.

The proposed PIED model is trained on data without certain frequencies and is tested on data with the missing frequencies. In this case, we specify the test set to contain all the simulation data with frequencies ranging from 70 kHz-90 kHz. The

surrogate model is trained with the remaining data (without frequencies of 70 kHz-90 kHz). The idea is to evaluate if the surrogate model can accurately predict the acoustic pattern in the frequency range of 70 kHz-90 kHz by interpolating from the training data. The OOD prediction results are presented in Fig. 10. The t-SNE plot visualizes the high-dimensional image data as a two-dimensional embeddings considering the similarities of the data. Based on the figure, the test data (red circles) clearly has its own distribution, which is separated from the training data (colored dots). This separation in the data distribution should create huge challenges for the surrogate model to accurately predict the test data. As a result, the predicted acoustic patterns (hollow black square) from either PIED or ED models do not capture the distribution of the test data. In terms of the average SSIM value (Fig. 10), PIED6 is slightly better than ED6. In the t-SNE plot, the PIED model can produce predictions close to the embeddings of acoustic pressure patterns at around 70 kHz and 90 kHz, whereas the ED model's predictions are only close to embeddings of acoustic pressure patterns at around 70 kHz. Visually, the embeddings of the predictions from both models are trying to bridge the gap between the embeddings of frequencies 50-70 kHz and 90-110 kHz. However, the true embeddings from the test data (frequencies from 70-90 kHz) seem to move away from the gap between the embeddings from frequencies 50-70 kHz to 90-110 kHz. The difference between the two models is that the PIED model predictions' embeddings appear scattered, while the ED model predictions' embeddings show more regular trends. The scattered behavior of the PIED predictions can be explained by the embeddings of the physics-informed input (See Appendix C), which appear to be scattered.

Based on the results, the proposed PIED model shows little benefit over the ED model on OOD prediction. The advantages of the PIED model over the ED model are its ability to produce accurate predictions with less training data (25% training data on PIED6 has comparable prediction accuracy against 100% on ED6) and smaller model (prediction accuracy of PIED4 is comparable to that of ED6 and ED 7). Thus, the introduction of physics-based input improves the efficiency of the surrogate model in terms of the sample size required for training and the model size required for accurate prediction.

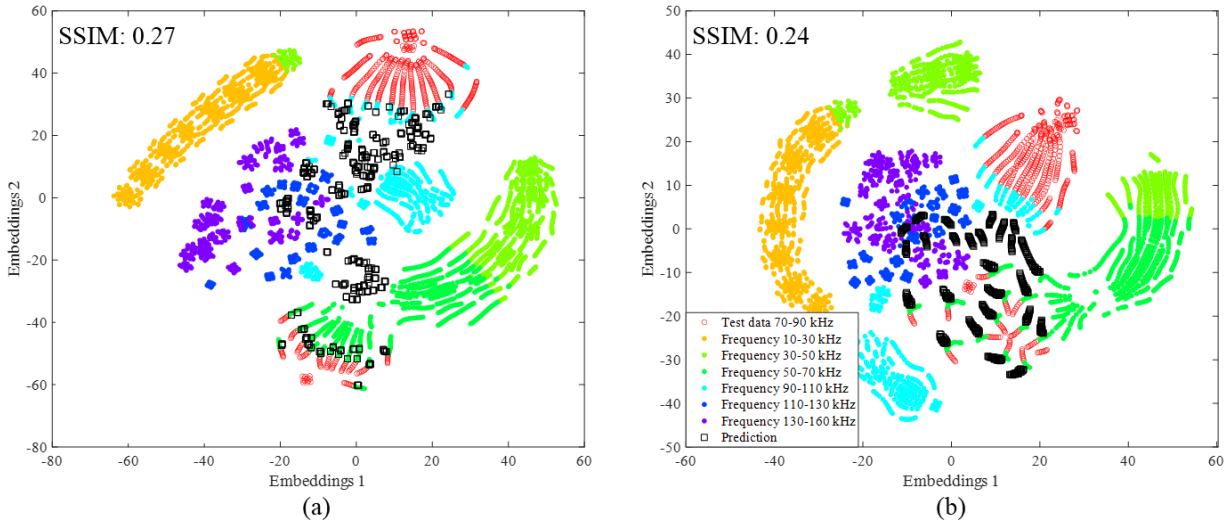

Fig. 10 t-SNE plot of the acoustic patterns used in training set, test set, and predictions by (a) PIED6 and (b) ED6 models.

5 CONCLUSION AND FUTURE RECOMMENDATIONS

This study has demonstrated that the construction of a surrogate model for a 3D finite element model based on encoder-decoder architecture can accurately predict acoustic pressure patterns by leveraging a physics-informed input derived from the output of a 2D finite element model. The resulting physics-informed encoder-decoder (PIED) model not only exhibits improved accuracy (~44% improvement) compared to the encoder-decoder (ED) model without the physics-informed input, but also exhibits improved run time (i.e., the test time) compared to the 3D finite element model. In turn, this model can hasten the process design where the optimum input parameters to the acoustic field-assisted particle patterning (AFP) system need to be determined given an image of the desired particle pattern.

This work investigated the reason behind the better performance of the PIED model over the ED model. The physics-informed input was found to enable favorable structure in the convolutional layers' feature maps, which facilitate the training of the PIED model. The optimum number of convolutional layers for the PIED model is found to be seven through a parametric study. However, a model with six convolutional layers is adopted for other analyses considering the accuracy and efficiency tradeoff. Although the PIED model shows little benefit over the ED model when performing out-of-distribution (OOD)

prediction, we have demonstrated that the introduction of physics-based input to the surrogate model can significantly reduce the number of samples and complexity of the model required to accurately train the surrogate model. This approach of incorporating physics as the input to the machine learning model is easy to implement and can potentially provide guidance to future research in the development of physics-informed machine learning models for other engineered systems.

Despite the advantages, we have identified a few directions for future improvements in the physics simulation model and the proposed methodology for incorporating physics into machine learning based surrogate modeling for acoustic field-assisted particle patterning application:

- First, despite having a good agreement on the high-frequency acoustic pressure patterns between the physics simulation results and experimental results, we observed a low-frequency vibrational mode in the simulation results, which was not observed in the experimental results. Even after our best efforts in filtering out the low-frequency vibrational mode, it has generated some unwanted artifacts, which has increased the difficulty in surrogate modeling. The low-frequency vibrational mode might originate from the assumption of perfect couplings between PZT actuators, Kapton film, and PDMS precursors matrix. Future simulation can consider the imperfect couplings through energy loss mechanisms such as introducing a damping factor.
- Second, the physical information was incorporated into the surrogate model through a physics-informed input. Although the prediction accuracies of PIED are much better than the ED model, it fails to provide similar prediction accuracies improvement when an OOD dataset is used. Future work can explore other techniques for incorporating physics into machine learning models, such as physics-guided loss function to incorporate physics model constraints to the loss function to improve the prediction accuracy on OOD samples.
- Third, the encoder-decoder network used for the surrogate model can be replaced by more complex networks such as conditional GAN (cGAN) [26] and U-Net [33] to evaluate if the strategy of using a physics-informed input can provide similar improvement in acoustic pressure distribution prediction.
- Lastly, the training data used in this study only consider the simulation of different AFP system setups without considering material variations. Future work can consider datasets with different material properties to further evaluate the effectiveness of the proposed method.

6 ACKNOWLEDGEMENTS

This research was in part supported by the US National Science Foundation (NSF) Grant Nos. CMMI-1663509. Any opinions, findings, or conclusions in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agency.

7 CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

8 REPLICATION OF RESULTS

Physics-informed encoder-decoder models presented in this paper have been implemented in Python. The Python code and simulated and measured image datasets are available upon request to the corresponding author.

9 REFERENCES

[1] Y. Yang *et al.*, "Biomimetic Anisotropic Reinforcement Architectures by Electrically Assisted Nanocomposite 3D Printing," *Adv. Mater.*, vol. 29, no. 11, p. 1605750, Mar. 2017.

[2] D. Kokkinis, M. Schaffner, and A. R. Studart, "Multimaterial magnetically assisted 3D printing of composite materials," *Nat. Commun. 2015 61*, vol. 6, no. 1, pp. 1–10, Oct. 2015.

[3] B. G. Compton and J. A. Lewis, "3D-Printing of Lightweight Cellular Composites," *Adv. Mater.*, vol. 26, no. 34, pp. 5930–5935, Sep. 2014.

[4] J. J. Martin, B. E. Fiore, and R. M. Erb, "Designing bioinspired composite reinforcement architectures via 3D magnetic printing," *Nat. Commun. 2015 61*, vol. 6, no. 1, pp. 1–7, Oct. 2015.

[5] C. Zhao *et al.*, "Layered nanocomposites by shear-flow-induced alignment of nanosheets," *Nat. 2020 5807802*, vol. 580, no. 7802, pp. 210–215, Apr. 2020.

[6] L. Lu, X. Tang, S. Hu, and Y. Pan, "Acoustic field-assisted particle patterning for smart polymer composite fabrication in stereolithography," *3D Print. Addit. Manuf.*, vol. 5, no. 2, pp. 151–159, Jun. 2018.

[7] T. W. Simpson, T. M. Mauery, J. J. Korte, and F. Mistree, "Kriging models for global approximation in simulation-based multidisciplinary design optimization," *AIAA J.*, vol. 39, no. 12, pp. 2233–2241, May 2001.

[8] Y. H. Lui, M. Li, M. Sadoughi, C. Hu, and S. Hu, "Physics-Based State of Health Estimation of Lithium-Ion Battery Using Sequential Experimental Design," in *Volume 2B: 44th Design Automation Conference*, 2018.

[9] M. Sadoughi, C. Hu, C. A. MacKenzie, A. T. Eshghi, and S. Lee, "Sequential exploration-exploitation with dynamic trade-off for efficient reliability analysis of complex engineered systems," *Struct. Multidiscip. Optim.*, vol. 57, no. 1, pp. 235–250, 2018.

[10] I. Bilionis and N. Zabaras, "Multi-output local Gaussian process regression: Applications to uncertainty quantification," *J. Comput. Phys.*, vol. 231, no. 17, pp. 5718–5746, Jul. 2012.

[11] X. Zhang, R. Srinivasan, and M. Van Liew, "Approximating SWAT Model Using Artificial Neural Network and Support Vector Machine1," *JAWRA J. Am. Water Resour. Assoc.*, vol. 45, no. 2, pp. 460–474, Apr. 2009.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, 2015, vol. 1, pp. 448–456.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[15] S. Balaban, "Deep learning and face recognition: the state of the art," in *Biometric and Surveillance Technology for Human and Activity Identification XII*, 2015, vol. 9457, p. 94570B.

[16] T. Pfister, K. Simonyan, J. Charles, and A. Zisserman, "Deep convolutional neural networks for efficient pose estimation in gesture videos," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9003, pp. 538–552.

[17] T. Pfister, J. Charles, and A. Zisserman, "Flowing convnets for human pose estimation in videos," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 1913–1921.

[18] S. Shen, M. Sadoughi, X. Chen, M. Hong, and C. Hu, "A deep learning method for online capacity estimation of lithium-ion batteries," *J. Energy Storage*, vol. 25, no. March, p. 100817, 2019.

[19] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification," *J. Comput. Phys.*, vol. 366, pp. 415–447, Jan. 2018.

[20] W. Wang, Y. Huang, Y. Wang, and L. Wang, "Generalized autoencoder: A neural network framework for dimensionality reduction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 496–503.

[21] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and P. Ni, "Structural damage identification based on autoencoder neural networks and deep learning," *Eng. Struct.*, vol. 172, pp. 13–28, Oct. 2018.

[22] D. Chicco, P. Sadowski, and P. Baldi, "Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions," *Proc. 5th ACM Conf. Bioinformatics, Comput. Biol. Heal. Informatics*.

[23] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014.

[24] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[25] J. Choe, S. Park, K. Kim, J. H. Park, D. Kim, and H. Shim, "Face Generation for Low-Shot Learning Using Generative Adversarial Networks," in *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2017, vol. 2018-Janua, pp. 1940–1948.

[26] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," Nov. 2014.

[27] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations,” Nov. 2017.
[28] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris, “Physics-Constrained Deep Learning for High-dimensional Surrogate Modeling and Uncertainty Quantification without Labeled Data,” *J. Comput. Phys.*, vol. 394, pp. 56–81, Jan. 2019.
[29] D. M. Hurtado, K. Uziela, and A. Elofsson, “Deep transfer learning in the assessment of the quality of protein models,” Apr. 2018.
[30] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles,” *Springer Proc. Adv. Robot.*, vol. 5, pp. 621–635, 2018.
[31] A. Daw, A. Karpatne, W. D. Watkins, J. S. Read, and V. Kumar, “Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling,” in *Knowledge-Guided Machine Learning*, 2022, pp. 353–372.
[32] E. J. Parish and K. Duraisamy, “A paradigm for data-driven predictive modeling using field inversion and machine learning,” *J. Comput. Phys.*, vol. 305, pp. 758–774, Jan. 2016.
[33] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241.
[34] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[35] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Dec. 2014.
[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

APPENDIX

A. Effects of input variables

For a machine learning model, more input variables typically lead to better prediction accuracy. In this study, the ED6 model only contains one input variable, whereas the PIED6 model contains two input variables. A fair comparison between the two approaches requires similar numbers of input variables. Here, a non-physics-informed input can be used in the base surrogate model (the ED6 model). We construct the non-physics-informed input variables based on the material properties and dimension information in each simulation domain used in the physics simulation (Fig. A1). In the non-physics-informed input matrix, each column represents a different simulation domain (i.e., PZT domain, Kapton film domain, and PDMS precursors domain), and each row represents the material properties and dimensions of the simulation domains. The first three rows are the length, thickness, and width of the simulation domain. The next four rows are the density, Young's modulus, Poisson's ratio, and the speed of sound. The material property value is set to zero if a simulation domain does not require a specific material property. To obtain the non-physics-informed input, the 7×3 matrix is resized to match the input image size of the surrogate model using MATLAB function, *imresize*.

The average SSIM over all test data for this model is calculated to be 0.52, slightly lower than the SSIM of the base surrogate model (ED6). This lower SSIM might be because the additional input variable fails to provide useful information to the surrogate model, while requiring the training of more model parameters.

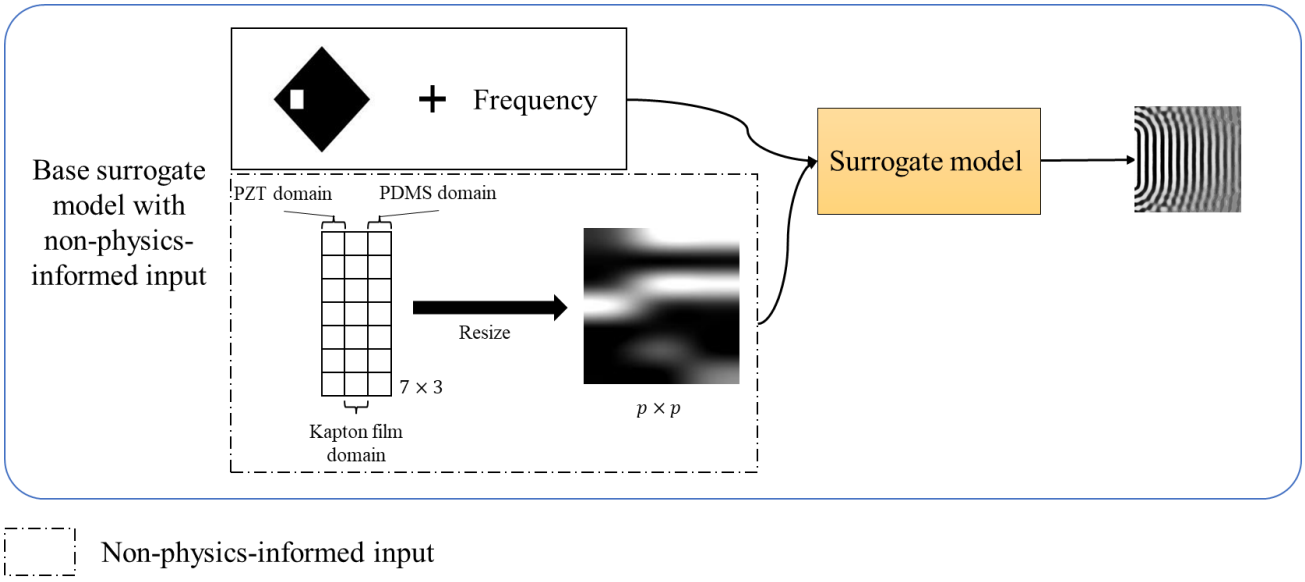

Fig. A1 Base surrogate model with non-physics-informed input.

B. Effects of activation functions

In deep neural networks, activation functions introduce nonlinear properties into the neural networks. Throughout this study, we have implemented the commonly used LeakyReLU activation function and are able to achieve good acoustic pressure prediction accuracies. In order to check if the use of different activation functions would produce different acoustic pressure pattern prediction results, we tested different activation functions on both the proposed physics-informed surrogate model and the base surrogate model. The activation functions tested here include sigmoid, hyperbolic tangent (Tanh), rectified linear unit (ReLU), Gaussian error linear unit (GeLU), and hard Swish (HardSwish).

The prediction results are summarized in Table B1 and Fig. B1. From the results, physics-informed surrogate models achieved higher prediction accuracy when compared to base surrogate models regardless of the activation functions.

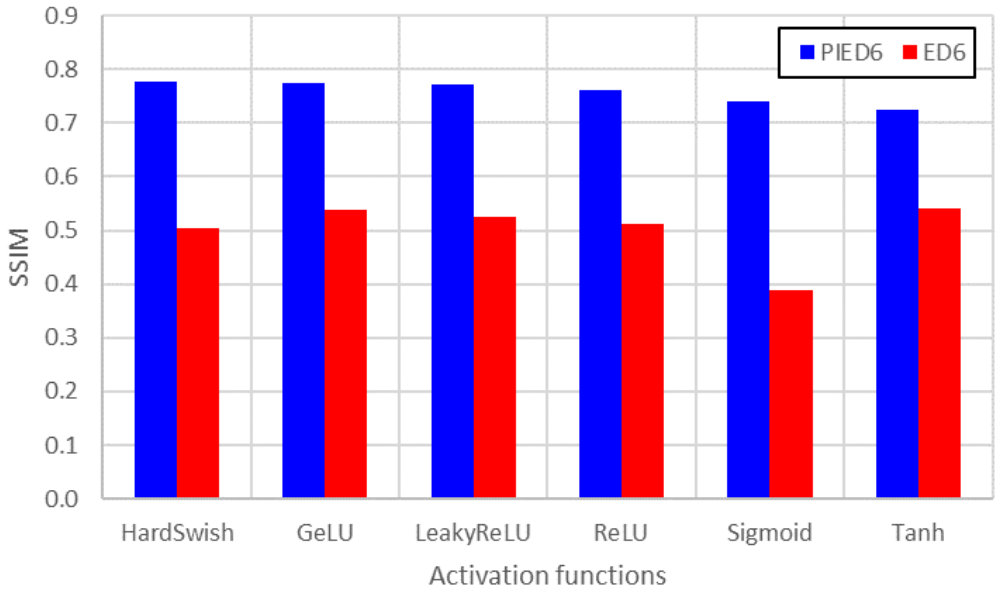

Fig. B1 Comparison of acoustic pressure pattern prediction accuracy by PIED6 and ED6 using different activation functions.

Table B1 Comparison of acoustic pressure pattern prediction accuracy (measured by SSIM) by PIED6 and ED6 using different activation functions. The activation function yielding the highest SSIM value for each type of surrogate model is highlighted in bold.

Activation function	PIED6	ED6
HardSwish	**0.778**	0.503
GeLU	0.773	**0.538**
LeakyReLU	0.771	0.526
ReLU	0.761	0.513
Sigmoid	0.739	0.388
Tanh	0.724	0.540

C. t-SNE of physics-informed inputs

In this section, we plot the t-SNE of the physics-informed inputs for out-of-distribution (OOD) prediction study.

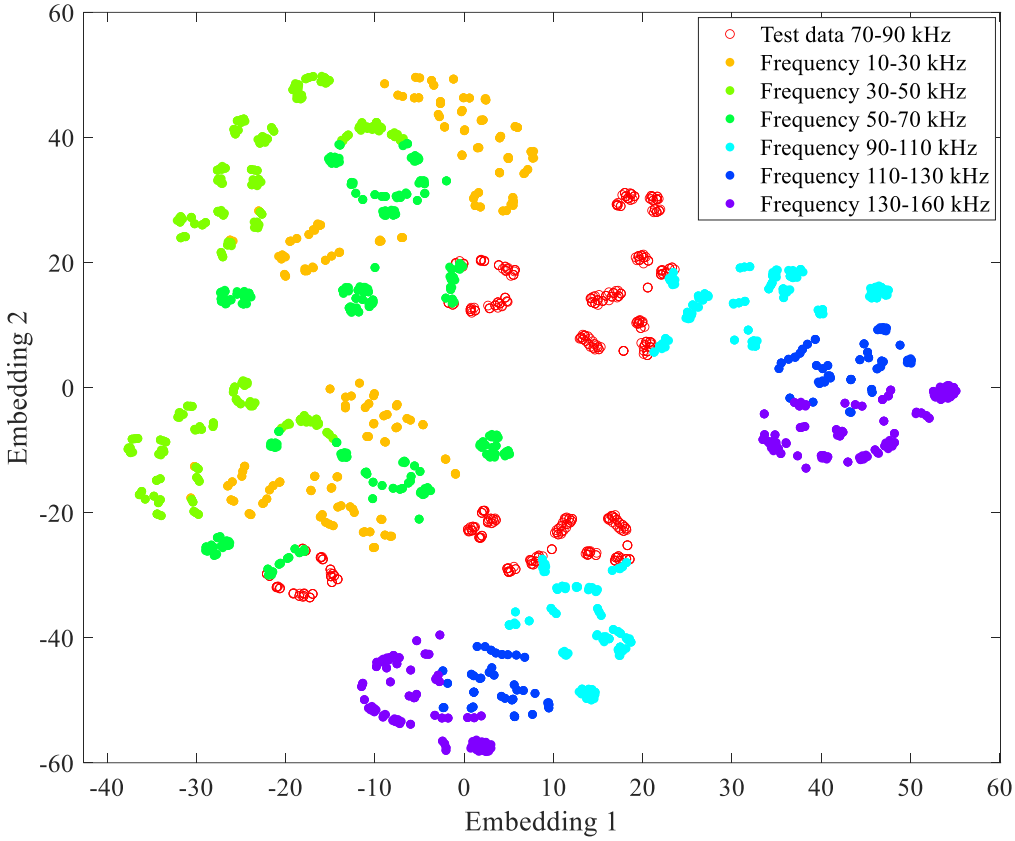

Fig. C1 t-SNE plot of the physics-informed inputs for OOD prediction study.