IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Towards Optimal Design of Dielectric Elastomer Actuators Using a Graph Neural Network Encoder

Yangfan Li, Jun Liu, *Member, IEEE*, Wenyu Liang, *Member, IEEE*, Zhuangjian Liu

***Abstract*—Dielectric elastomer actuators (DEAs), a type of "artificial muscles", can generate significant deformations and offer speedy responses when exposed to voltage. Owing to their high electromechanical conversion efficiency and great flexibility, they have been extensively used in soft robot applications, such as soft grippers, walking robots, crawling robots, climbing robots, swimming robots, etc. Although previous research has explored the use of DEAs in soft robot locomotion, achieving optimal behavior is challenging due to the complexity of the constituent materials and the highly nonlinear nature of the problem. In this study, a simulation-based design optimization approach is proposed to address this challenge. The proposed approach involves developing a computational modeling framework that evaluates the electromechanical behavior of the DEA. A graph neural network (GNN) is employed as an encoder to extract the latent representation of the geometry in a low dimensional space, which is further used to construct a surrogate model for fast prediction of target responses. To achieve an optimal actuation capability under design constraints, a multi-objective optimization function is formulated to balance the actuation distance and the actuator size, where the Pareto front demonstrates the trade-off between the actuation distance and design constraints. Finally, three optimized designs are fabricated and tested, demonstrating a performance improvement of over 140% compared to an intuitive design. This framework can greatly benefit the design of DEA-based soft robotics.**

***Index Terms*—Soft robot materials and design; Modeling, Control, and Learning for Soft Robots; Optimization and Optimal Control.**

I. Introduction

Dielectric elastomer actuators (DEAs) are known for their ability to undergo significant deformations and provide fast responses when exposed to voltage. They are typically made of a DE membrane sandwiched by compliant electrodes and a flexible supporting frame. Upon application of voltage, an electric field is produced within the membrane, causing the generation of Maxwell stress and resulting in both an in-plane expansion and out-of-plane contraction of the membrane. The attached flexible frame or structure would also deform accordingly. To achieve substantial voltage-induced deformation [1] and suppress electromechanical instabilities-induced failure [2]–[4], the employed DE membrane is usually pre-stretched. Because of their easy activation, high electromechanical conversion efficiency, great flexibility, and lightweight, DEAs have gained significant attention in recent years and have been used in various soft robotics applications [5], [6], such as soft grippers [7], [8], walking robots [9], crawling robots [10], climbing robots [11], swimming robots [12], and so on. For instance, Gu *et al.* [11] created a soft wall-climbing robot that uses a DEA as its flexible body and two electro-adhesive pads as its feet. By controlling the voltage application, the DEA can quickly switch between a relaxed state and an active state, allowing fast changes in height and movement in confined spaces. Similarly, Li and their team [12] designed a self-powered soft swimming robot with DE membranes serving as the "muscles" for flapping fins. When voltage is applied, the DE membranes located between the support frame and the fins transition from a relaxed state to an active state, producing a flapping movement of the fins through a certain angle. These works focus on utilizing the DEAs to achieve targeted motions in soft robots. However, the performance of such a soft robotic system strongly relies on the designer's experience and a large number of trial-and-error generally required to achieve satisfactory performance. How to optimize the components to obtain the best-desired behaviors remains a challenge due to the highly nonlinear nature of the problem.

The performance of DEA depends on not only the inherent material properties of DE but also the shape and topology of supporting frames. To achieve the best-desired performance, one option is to perform topology optimization on the frames using the gradient-based topology optimization methods [13]–[15]. A recent work has utilized the Solid Isotropic Material with Penalization method (SIMP) method with finite element method (FEM) to optimize the DEA-based leg and abdomen for a crawling robot [16]. A single-objective function was formulated to maximize the output displacement of leg-DEA and abdomen-DEA. However, gradient-based topology optimization methods tend to encounter numerical instabilities and convergence issues due to the high nonlinearity of the problem [17], [18].

An alternative approach is to build a surrogate model to estimate the performance of different designs and then employ a non-gradient-based optimization algorithm to identify an optimized design. In this way, the convergence issue caused by numerical instabilities in gradient-based optimization will be circumvented, and it is more likely to find the optimal solution in problems with a large number of local minima. Another advantage is that multi-objective optimization problems could be formulated and quickly solved with the surrogate model as a fast estimator. However, building a reliable surrogate model can be a challenging task, as a reliable and sufficiently large database must be obtained either through experiments or simulations. More importantly, the responses for soft actuators are highly nonlinear and multi-physics, yielding that an arbitrary

Manuscript received: April, 20, 2023; Accepted August, 11, 2023.

This paper was recommended for publication by Editor Yong-Lae Park upon evaluation of the Associate Editor and Reviewers' comments. This research is supported by A*STAR C211518005. *(Corresponding author: Jun Liu)*

Yangfan Li, Jun Liu and Zhuangjian Liu are with Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mail: liuj@ihpc.a-star.edu.sg).

Wenyu Liang is with Institute for Infocomm Research (I^{2}R), Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mail: liang_wenyu@i2r.a-star.edu.sg).

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

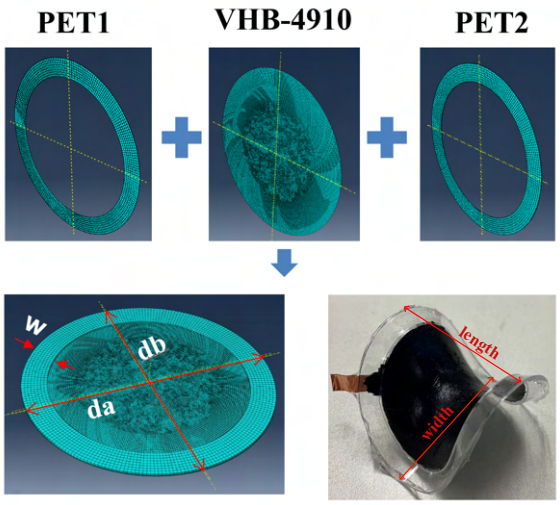

Fig. 1. A typical structure of an elliptical DE manipulator: it consists of two layers of DE membrane sandwiched by hollow PET frames. The carbon grease is applied on the top and bottom sides of the membrane as electrodes. After pre-stretch relaxation, the DEA buckles into a saddle shape with characteristic length and width denoted as red lines in the bottom right picture.

machine learning model may not be expressive for this kind of problem, and a careful design of deep learning neural network (DNN) with physics awareness is desired.

In this work, a simulation-based data-driven design optimization approach is proposed to address the aforementioned challenges. The nonlinear field theory and the Yeoh strain energy potential using a user-defined material model (UMAT) in ABQUAS are first implemented for investigating the dynamic electromechanical behavior of DEs. Then the geometry is parameterized and a series of designs of experiments for simulation is generated. Without loss of generality, the flexible graph data structures are chosen to represent the topology of the DEAs, which consist of nodes and edges connecting them. Then, the graph data and desired responses are fed to a Graph Neural Network (GNN) for building a physics-aware surrogate model, which is later used in conjunction with the Non-dominated Sorting Genetic Algorithm II (NSGA-II) optimization method to optimize the DEA design. As a demonstration, an oval-shaped DEA design is optimized as shown in Fig. 1, which can serve as the basic "muscle" component in a soft robot. The proposed approach is validated experimentally and the results show that the target performance of the DEA was successfully improved by over 140%.

The main contributions of this paper include i) a systematic multi-objective optimization framework for the optimization design task of a DEA; ii) a GNN as an encoder to extract the latent representation of the geometry in a low dimensional space for fast prediction; iii) a concise and useful guidance on the design of elliptic and circular DEAs. Significantly, compared to Artificial Neural Networks (ANNs) or Convolutional Neural Networks (CNNs), GNNs can encode the physics-related information of the system as node or edge features, such as coordinates, neighbor connections, material properties, etc. The way to construct graph representation of DEA structures is highly extendable to any irregular topology that may be involved in other soft robotic designs. The proposed framework is also applicable to optimizing other constitutive units of soft robotics with multi-functionalities.

The rest of this paper is organized as follows. In Section II, the constitutive equations used to model the coupled electromechanical response of DEs are presented. In Section III, a data-driven workflow for the design optimization of DEAs is introduced, where a multi-objective optimization problem is formulated to simultaneously minimize the DEA size and maximize the actuation distance. Section IV presents the simulation, optimization, and experimental results. This is followed by the conclusion and discussion in Section V.

II. Constitutive Model

Researchers have developed theoretical and numerical tools for evaluating the electro-mechanical response of DE [19]–[24], which are valuable for understanding the physics of the electro-mechanical behavior of DE material. However, not all models are suitable for design optimization purposes, where the model is required to effectively capture the essential phenomena occurring under the working conditions. Simultaneously, the model should maintain simplicity for rapid performance evaluation, even if this comes at the expense of some accuracy. For this consideration, we build our model based on our previous work [23], but re-implement the model as a user-defined material model (UMAT) in Abaqus with enhanced stability.

Some definitions of the kinetics of constitutive theory are given as follows. Considering the position of material particles in the reference configuration $\boldsymbol{X}$, and its current deformed position at time t is given by $\boldsymbol{x}(\boldsymbol{X}, t)$. The deformation gradient tensor $\boldsymbol{F}$ is defined as $\boldsymbol{F} = \partial \boldsymbol{x}/\partial \boldsymbol{X}$, and $J = \det(\boldsymbol{F})$ represents the volume change at the material point. The right Cauchy-Green tensor $\boldsymbol{C}$ and the left Cauchy-Green tensor $\boldsymbol{b}$ are defined as $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$ and $\boldsymbol{b} = \boldsymbol{F}\boldsymbol{F}^T$, so the first invariant is defined by $I_1 = \text{tr}(\boldsymbol{C}) = \text{tr}(\boldsymbol{b})$.

A. Energy Potential

1) Mechanical Free Energy: The strain energy density function is usually used to describe the mechanical behavior of hyperelastic materials. There are several useful forms of strain energy functions that have been implemented for the elastomer modeling, such as the Gent model [23], Yeoh's model [25], Ogden model, and Mooney-Rivilin model [26]. Here we choose to implement Yeoh's model, as Yeoh's model provides good agreements with experiments when the material undergoes very large bi-axial stretches [26]. The strain energy density of compressible Yeoh's model is defined to be the sum of the isochoric part W_m^{iso} and volumetric part W_m^{vol} as:

$$W_m = W_m^{iso} + W_m^{vol}, \tag{1}$$

where

$$W_m^{iso} = c_1(\bar{I}_1 - 3) + c_2(\bar{I}_1 - 3)^2 + c_3(\bar{I}_1 - 3)^3, \tag{2}$$

$$W_m^{vol} = \frac{1}{2}K(J-1)^2, \tag{3}$$

c_i are material parameters that represent shear resistance and are usually determined from the uni-axial tensile tests. K is the bulk modulus that represents the resistance of volume change of the material. $\bar{I}_1 = tr(\bar{\boldsymbol{C}})$ is the first invariant of $\bar{\boldsymbol{C}}$, where $\bar{\boldsymbol{C}} = J^{-2/3}\boldsymbol{C}$ is the deviatoric part of the $\boldsymbol{C}$.

2) Electric Free Energy: The so-called *ideal dielectric elastomer* theory proposed by Zhao *et al.* [2]–[4] is adopted here. It is a widely-used thermodynamics theory-based model that can explain the electromechanical instabilities encountered in the experiment. In this model, the crosslinks of elastomers do not affect too much on the polarization of monomers, and the Maxwell stress generated from the applied voltage is nonlinear. Denote the electric field and electric displacement in the reference configuration as $\tilde{\boldsymbol{E}}$ and $\tilde{\boldsymbol{D}}$, then the true electric field and true electric displacement in the current configuration can be represented by

$$\boldsymbol{E} = \boldsymbol{F}^{-T} \cdot \tilde{\boldsymbol{E}} \text{ and } \boldsymbol{D} = \frac{1}{J}\boldsymbol{F} \cdot \tilde{\boldsymbol{D}}. \tag{4}$$

For an "ideal dielectric elastomer", the true electric field $\boldsymbol{E}$ and the true electric displacement $\boldsymbol{D}$ in the current configuration are assumed to have a linear relationship $\boldsymbol{D} = \varepsilon \boldsymbol{E}$ [4], where the dielectric permittivity of the elastomer ε is assumed to be a constant independent of deformation. Consequently, we have

$$\tilde{\boldsymbol{D}} = J\varepsilon \boldsymbol{C}^{-1} \cdot \tilde{\boldsymbol{E}}. \tag{5}$$

Therefore, the free energy due to electrical polarization is computed by

$$W_e(\boldsymbol{F}, \tilde{\boldsymbol{E}}) = -\frac{1}{2} J\varepsilon \tilde{\boldsymbol{E}} \cdot \boldsymbol{C}^{-1} \tilde{\boldsymbol{E}}, \tag{6}$$

where the minus symbol represents that the electric polarization usually causes relaxation to the system.

3) Total Free Energy: The total free energy stored in the elastomer is a sum of the mechanical energy and electric energy, which is given by

$$W(\boldsymbol{F}, \tilde{\boldsymbol{E}}) = W_m(\boldsymbol{F}) + W_e(\boldsymbol{F}, \tilde{\boldsymbol{E}}), \tag{7}$$

which shows that electromechanical coupling is a purely geometric phenomenon since the stretching and polarization contribute independently to the overall free energy function.

B. Constitutive Equations

The variation of the free energy function can be computed as:

$$\delta W(\boldsymbol{F}, \tilde{\boldsymbol{E}}) = \boldsymbol{P}\delta\boldsymbol{F} - \tilde{\boldsymbol{D}}\delta\tilde{\boldsymbol{E}}, \tag{8}$$

where $\boldsymbol{P}$ is work conjugate to the deformation gradient $\delta\boldsymbol{F}$ and also known as the first Piola-Kirchhoff (PK1) stress. $\tilde{\boldsymbol{D}}$ is work conjugate to the electric deformation gradient $\delta\tilde{\boldsymbol{E}}$ in the reference configuration. Mathematically, they can be written as

$$\boldsymbol{P} = \frac{\partial W\left(\boldsymbol{F}, \tilde{\boldsymbol{E}}\right)}{\partial \boldsymbol{F}} \text{ and } \tilde{\boldsymbol{D}} = \frac{\partial W\left(\boldsymbol{F}, \tilde{\boldsymbol{E}}\right)}{\partial \tilde{\boldsymbol{E}}}. \tag{9}$$

Correspondingly, the Cauchy stress is calculated by

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_m^{iso} + \boldsymbol{\sigma}_m^{vol} + \boldsymbol{\sigma}_e, \tag{10}$$

Algorithm 1 Gibbs sampling for DEA design

Step 1:
Initialize d_a, d_b and w at starting point $t = 0$.
Step 2:
Set the state transition threshold n_1 and the required number of samples n_2.
Step 3:
For $t = 0, ..., n_1 + n_2$ Do
Sample d_a, d_b and w as:

$$d_a^{(t+1)} \sim p(d_a^{(t+1)}|d_a^{(t)}) = U(50, 100) \tag{11}$$

$$d_b^{(t+1)} \sim p(d_b^{(t+1)}|d_a^{(t)}) = Tri(d_a^{(t)} - 30, d_a^{(t)}, d_a^{(t)} - 30) \tag{12}$$

$$w^{(t+1)} \sim p(w^{(t+1)}|d_b^{(t)}) = \mathcal{N}(0.12d_b^{(t)}, 0.03d_b^{(t)}) \tag{13}$$

End For
Step 4:
Choose the sampling results from $n_1 + 1$ to $n_1 + n_2$.

where

$$\boldsymbol{\sigma}_m^{iso} = \frac{1}{J}\left[2c_1 + 4c_2(\bar{I}_1 - 3) + 6c_2(\bar{I}_1 - 3)^2\right]\left(\bar{\boldsymbol{b}} - \frac{1}{3}\bar{I}_1\boldsymbol{I}\right),$$

$$\boldsymbol{\sigma}_m^{vol} = K\left(J - 1\right)\boldsymbol{I},$$

$$\boldsymbol{\sigma}_e = \varepsilon\left[\boldsymbol{E} \otimes \boldsymbol{E} - \frac{1}{2}\boldsymbol{I}(\boldsymbol{E} \cdot \boldsymbol{E})\right],$$

and $\bar{\boldsymbol{b}} = J^{-2/3}\boldsymbol{b}$ is the deviatoric part of the left Cauchy-Green tensor $\boldsymbol{b}$.

III. Design and Optimization

The DEA to be optimized consists of two layers of pre-stretched elastomer membrane (VHB 4910 manufactured by 3M) sandwiched by two identical elliptical polyethylene terephthalate (PET) frames, as shown in Fig. 1. The electrode coating on the membrane is carbon grease (846-80G manufactured by MG Chemicals).

A. Design of Experiments

The frame configuration can be characterized by four variables, namely the outer major and minor diameter d_a and d_b, frame width w, and thickness of PET frames. Due to the limited commercially available thicknesses of PET frames t, the thickness is kept constant at $t = 0.4$ mm. These design variables are not entirely independent because the relationship $d_a \geq d_b$ must be fulfilled to prevent redundant simulations. Based on prior experience, when d_a and d_b are specified, w must fall within a reasonable range to ensure the DEA can be activated. To achieve this, the Gibbs Sampling method is employed to sequentially sample d_a, d_b, and w. Specifically, d_a is uniformly distributed over a given range $[50, 100]$ mm, while $d_a - d_b$ follows a triangular distribution and w/d_b follows a normal Gaussian distribution. The details are shown in Algorithm 1.

It can be observed that during the sampling process, each sample is always influenced by the previous one. Therefore, the sample sequence obtained through Gibbs sampling forms a Markov chain. Based on this, the properties of Markov chains, similar to the Metropolis-Hasting algorithm, can be used to prove that the stationary distribution of this Markov

chain is the target joint probability distribution. This sampling process ensures that the resulting design is not excessively disproportionate. A total of $n_2 = 600$ designs are generated.

B. Simulation Automation

Due to the symmetry nature of the DEA designs, a quarter model is built to reduce the computational cost. The simulation procedure has two distinct stages: pre-strain relaxation and electric actuation. During the pre-strain relaxation phase, a pre-defined deformation gradient tensor is applied to the UMAT, resulting in the gradual imposition of a pre-strain field with principal stretch ratios of $\lambda_1 = \lambda_2 = 4$, and $\lambda_3 = 0.0625$. Ideally, the DE actuator will buckle into a saddle shape, with the principal bending axes coinciding with the x and y axes. Based on our convention, it holds that $da \geq db$, the maximum distances between two opposing points along the principal axes are not equal, which we define as the characteristic width w_1 and length l_1 of the saddle shape. The characteristic length is observed to align with the major axis of the ellipse, while the width aligns with the minor axis. During the electric actuation stage, a voltage is gradually applied to the electrodes of the DE actuator, resulting in its radial expansion. As a result, both the characteristic width and length increase.

At the end of each stage, we calculate the characteristic width and length of the deformed DE actuator, and the difference between the two is determined as the actuation distance. We observe that the actuation distance along the minor axis is typically greater than that along the major axis, which is of interest to optimize. The characteristic length at the end of the electric actuation stage determines the maximum physical size of the DEA, limiting its application in a confined space, which is used as a geometric constraint in the following optimization. Python scripts are used to automatically generate simulation models, submit jobs, and extract and post-process data.

To validate the simulation model presented in Section II, a comparison between the simulation results and experimental data obtained for an initial circular DEA design is performed. The initial design, derived using the trial-and-error approach through extensive and repeated attempts and experiments, featured a diameter of $d_a = d_b = 90$ mm and a width of $w = 8.0$ mm. The measured characteristic width of this design was approximately $w_1 = 35.0$ mm, with an actuation distance of $\Delta U_{EXP} = 5.0$ mm and an average deviation of 0.08 mm. The simulated characteristic width is $w_1 = 32.1$ mm and a simulated actuation distance of $\Delta U_{FEM} = 4.6$ mm. These results demonstrate a good agreement with the experimental data, indicating the effectiveness of our simulation model.

C. Surrogate Model with GNN as An Encoder

One challenge in building up a surrogate model is to reasonably reduce the problem into a lower-dimensional representation. An intuitive way is to directly use shape parameters such as diameters, material thickness, etc. However, they may only be able to describe simple geometries and are difficult to be generalized. A more expressive tool to represent the inherent geometry feature of the model is desirable. GNN is a type of

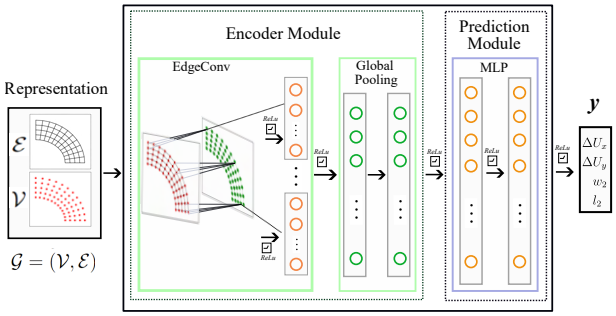

Fig. 2. The graph neural network architecture used for accelerated modeling, which consists of an encoder module and a prediction module. GNN takes the graph representations of geometry as input and gives four response parameters as output, including actuation distance in x and y directions ΔU_x and ΔU_y, and the characteristic width w_2 and length l_2.

neural network that operates on graph data structures, which consist of nodes and edges connecting them [27]. The graph data could be easily constructed from mesh-based simulations, such as the FEM, which are applicable to various types of domains. In GNNs, each node is represented by a vector that encodes its features, and information is propagated between nodes by passing messages along the edges. This message-passing operation is implemented using neural network layers, allowing the GNN to capture complex dependencies and local neighborhood structure. GNNs have been successfully applied to various tasks, such as node classification, graph classification, and recommendation systems, and have found wide applications in fields such as social networks, drug discovery, protein interface prediction, and traffic prediction [28].

The graph representation can be created in two ways, either by directly extracting the finite element mesh nodes and edges as graph nodes and edges or by using mathematical equations to generate nodes and edges that can mesh the domain. For reducing computational cost without losing generality, the second method is employed in this work. For each generated design, we construct the graph structure with 36 nodes and 114 edges for the quarter model correspondingly.

A general GNN model is defined on a graph data structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, ..., v_n\}$ is the node set and $\mathcal{E} = \{e_1, ..., e_m\}$ is the edge set contains the topology connection of nodes. $\mathcal{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ is node feature set where $\boldsymbol{x}_i$ is the feature vector associated with node v_i. $\mathcal{N}(i)$ represents the neighborhood node set that has an edge connection to node i. The key mechanism of GNN is *message passing*, which is formulated as

$$\boldsymbol{x}_i^{(l)} = \gamma^{(l)} \left(\boldsymbol{x}_i^{(l-1)}, \bigoplus_{j \in \mathcal{N}(i)} \phi^{(l)} \left(\boldsymbol{x}_i^{(l-1)}, \boldsymbol{x}_j^{(l-1)}, \boldsymbol{e}_{j,i} \right) \right). \tag{14}$$

It means that to update the embedding of node features $\boldsymbol{x}_i^{(l)}$ on current layer l, we need information of node embedding $\boldsymbol{x}_i^{(l-1)}$ on previous layer $l-1$, and its neighbourhood node embedding $\boldsymbol{x}_j^{(l-1)}$, as well as edge feature embedding $\boldsymbol{e}_{j,i}$ represents the edge connection between node i and j. γ and ϕ are differentiable functions, and $\oplus$ is a permutation invariant aggregation operator. Various types of GNN models have been proposed in the past few years, such as GCN [29], GraphSAGE

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

[30], GAT [31], etc., to design proper $\oplus$, γ, and ϕ operators, for catering to different graph architecture and representation requirements. In this work, the Edge Convolution layer [32] is adopted to encode both the node coordinates and the distances between neighboring nodes as node features, which allows passing more geometry-related information to the neural network, which can be mathematically defined as:

$$\boldsymbol{x}_i^{(l)} = \max_{j \in \mathcal{N}(i)} \phi^{(l)} \left(\boldsymbol{x}_i^{(l-1)}, \boldsymbol{x}_j^{(l-1)} - \boldsymbol{x}_i^{(l-1)} \right). \quad (15)$$

The entire surrogate machine learning model is illustrated in Fig. 2, which includes an encoder module and a prediction module. The encoder module comprises an EdgeConv layer, which was an implementation of the Edge Convolution layer in PyG library [33], two fully connected linear layers, and one global pooling layer. The input $\boldsymbol{x}_i \in \mathbb{R}^2$ to the encoder module is node features containing the nodal coordinates (x, y), which is defined on $\mathcal{V}$. Since the edge connection remains unchanged during message-passing and network evolution, only the node features require updating throughout the network. These node features are encoded through the GNN Edge Convolution layer, and the output is a new expanded set of feature vectors (embeddings) for each node. These output feature vectors are then input to a subsequent global pooling layer. Another linear layer is included after the global pooling layer and converts node features to graph-level information and transmits them to the prediction module.

The prediction module is constructed with a Multi-Layer Perceptron (MLP) architecture [34], which includes multiple layers of interconnected nodes, with each node in a layer connected to every node in the previous layer. Each node in an MLP performs a weighted sum of its inputs, applies an activation function to impose the nonlinear behavior, and sends the output to the next layer. The weights and biases of the MLP are learned through a training process using backpropagation, which adjusts the weights to minimize the error between the network's predictions and the actual output. The output of the prediction module yields four distinct responses, namely the actuation distance in both the x and y directions ΔU_x and ΔU_y, as well as the characteristic width w_2 and length l_2 at the end of the electric actuation phase. These four outputs are directly related to the design objective and constrain for this DEA design optimization problem.

D. Optimization

Once the surrogate model is trained and validated, it can be leveraged for solving inverse design problems. With a given PET thickness, the goal is to maximize the actuation distance while minimizing the characteristic length and width. Thus the optimization problem can be defined to find the optimal $\boldsymbol{\xi} = \{d_a, d_b, w\}$ resulting responses of the greatest actuation distance ΔU_x or ΔU_y and the minimal characteristic length l_2 and width w_2. Equivalently, we choose to minimize a loss function $f(x)$ which is mathematically computed as

$$\begin{aligned} \min \quad & f_1(x) = -\max(\Delta U_x, \Delta U_y) \\ \min \quad & f_2(x) = \max(l_2, w_2) \end{aligned} \quad (16)$$

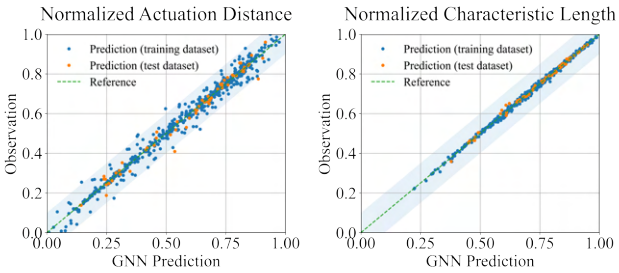

Fig. 3. Scatter plot of GNN prediction versus ground truth for normalized actuation distance ΔU_y (left) and characteristic length l_2 (right). The shaded region represents an error within 10% of the maximum value as the reference error range. Almost all the points are within the shaded region, indicating the high accuracy achieved from the trained network.

where ΔU_x, ΔU_y is the actuation distance along y direction, w_2, l_2 are the characteristic width and length at the end of the electric actuation phase. The geometric constraints are set as $50 \leq d_a \leq 100$, $50 \leq d_b \leq 100$, $3 \leq d_a \leq 25$. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) [35] is employed to solve the multi-objective optimization problem.

IV. RESULTS

A. Simulation and Surrogate Model

The PET frames were modeled as Neo-Hookean hyperelastic material with a density of 1380 kg/m^3, and a shear modulus of 230 MPa. The response of the DE membrane is simulated by the UMAT developed in Section II. The material parameters for the DE membrane are chosen as $\rho = 1000$ kg/m^3, $C_{10} = 0.0608$ MPa, $C_{20} = -8.88 \times 10^{-4}$ MPa, $C_{30} = 1.67 \times 10^{-5}$ MPa and $\varepsilon = 3.98 \times 10^{-11}$ F/m [26].

With the Gibbs Sampling method, we generate 600 designs The actuation distances and characteristic length and width are normalized before being fed to the surrogate model. These data points are split into a training set (80%) and a testing set (20%) for training the surrogate model. Fig. 3 shows that the trained surrogate model can make a reasonable prediction for all the training and testing data with a correlation of $R^2 = 0.977$ for ΔU_y and $R^2 = 0.996$ for l_2.

In addition, it is noticed that the trained GNN model is capable of making accurate predictions even when inner nodes are randomly displaced while the boundary nodes remain fixed in their original positions. In one case, 10 inner nodes with random displacements within a range of 1 mm were perturbed and it was observed that the predicted response on the new graph data was invariant to the one on the original data. This observation is expected based on the physics of the problem. From a geometric perspective, the relocation of inner nodes does not affect the topology of the structure as long as node connections are maintained. Thus, this test demonstrates that the GNN model can effectively reflect the underlying physics of the system, even without additional training. It should be noted that if node connections are altered, the GNN model must be trained with relevant data to ensure accurate predictions. For instance, if inner edges are removed, resulting in a hole in the structure, such cases must be included in the training data to ensure the GNN's representative capabilities.

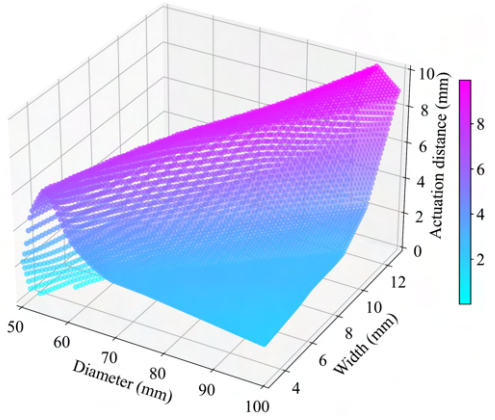

Fig. 4. Scatter plot of actuation distance ΔU as a function of the diameter and width for cases with $d_a = d_b = d$. The results are obtained with the help of the GNN model. The plot shows that an optimal diameter exists for a given PET width, and this optimal diameter increases nearly linearly with the width of the PET. The color bar indicates the magnitude of the actuation distance.

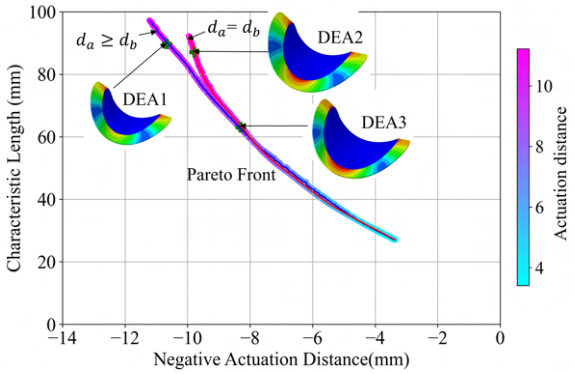

Fig. 5. Pareto optimal front for the multi-objective optimization problem with geometric constraints $d_a = d_b$ and $d_a \geq d_b$. The color bar indicates the magnitude of the actuation distance.

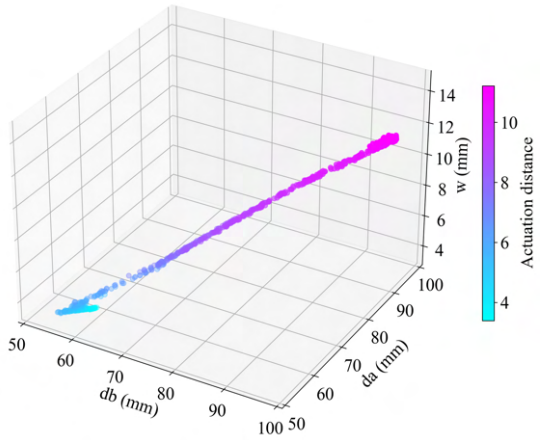

Fig. 6. Optimal design parameters corresponding to the Pareto front with $d_a \geq d_b$: the color bar indicates the magnitude of the actuation distance.

B. Parametric Study

Leveraging the well-crafted surrogate model, it is convenient to perform a parametric study to examine the impact of individual design parameters on the expected output. We ad-hoc set $d_a = d_b = d$ based on the consideration that: i) to illustrate the design points on a 3D coordinate system, and ii) circular designs are naturally preferred for facilitating fabrication. The computational time for a FEM simulation is around 300 to 900 seconds depending on its geometry and mesh size. But with the GNN model, it can be greatly reduced to less than 0.1 seconds. The computational speed-up by the GNN model is estimated at an order of 10^3 to 10^5.

Fig. 4 depicts the actuation distance as a function of the diameter d and width w of PET frames. The figure illustrates that an optimal diameter exists for a given PET width, and this optimal diameter increases nearly linearly with the width of the PET. In the case of a fixed width w, the actuation distance initially increases for small diameters, reaching a peak value before decreasing rapidly. This trend can be explained as follows: For small w values, the entire structure's stiffness is relatively low, and the DEA system can be easily bent. However, the strain energy stored in the PET is not sufficient to enable the system to recover its original shape, and the system cannot be actuated to a large distance after applying voltage. As the width w increases, the strain energy stored in PET becomes larger, enabling the system to achieve greater distances. However, once w surpasses a critical value, the system's stiffness becomes too high, making it difficult to bend. Consequently, the strain energy stored in the PET decreases again, leading to a smaller actuation distance. The conducted experiments confirm these phenomena, and Fig. 4 can be used as guidance to select the optimal design point for achieving the desired actuation distance by selecting the desired diameter d and optimal width w.

C. Multi-Objective Optimization

For a more general case for $d_a \neq d_b$, a simple parametric study can not guarantee an optimal solution for multiple objectives. Instead, the approach of multi-objective optimization can be utilized to address these complex design scenarios. The Pareto front, which represents the set of optimal solutions for the multi-objective optimization problem, is depicted as a blue line with dots in Fig. 5. It shows that the maximal actuation distance increases almost linearly with the characteristic length. With the information provided by the Pareto front, we can efficiently identify the optimal design that maximizes actuation capability within the given constraints. To investigate the effectiveness of the established model in facilitating the design of a circular DEA with optimal actuation performance, we also impose the additional constraint, $d_a = d_b = d$, and perform a search for optimal values of w. The corresponding Pareto front is also plotted as a red line in Fig. 5. It shows that there is an overlap region of two Pareto fronts, where the minimized characteristic length is below 60 mm and the actuation distance is less than 8 mm, meaning that the optimal DEAs within this region are all circular designs. Above this region, the optimal DEAs take elliptical shapes and the imposition of the constraint $d_a = d_b$ would result in a

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

TABLE I
RESULTS OF THE OPTIMIZED AND INITIAL DEA DESIGNS

Name (unit:mm)	d_a	d_b	w	ΔU_y @ 5 kV			Improvement
				GNN	FEM	EXP	%
DEA1	95.0	86.8	11.6	10.7	10.5	12.3	146
DEA2	95.0	95.0	13.0	9.8	9.4	11.5	130
DEA3	72.0	72.0	7.2	7.9	8.0	8.8	76
DEA0	90.0	90.0	8.0	4.2	4.6	5.0	N.A.

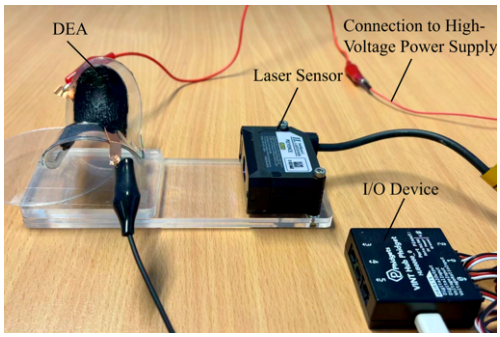

Fig. 7. Experimental system for measuring the actuation distance of DEA

smaller actuation distance and/or longer characteristic length.

Fig. 6 shows the corresponding d_a, d_b, w for the solutions located on the Pareto front with $d_a \geq d_b$, and the color represents the magnitude of the actuation distance. It demonstrates that there exists an almost linear relation between the optimal width and the optimal d_a, d_b, meaning that the ratio of w/d_a or w/d_b is near a constant in the optimal solution space. Such an observation is in agreement with the parametric study results in Fig. 4. The larger characteristic lengths and actuation distances are obtained with larger DEA sizes. Specifically, the design with the maximum actuation distance has a $d_a = 99.9$ mm that is very close to the size constraint of 100.0 mm, while its d_b is around 88.6 mm, and w is approximately 13.0 mm. As w approaches its lower limit of 3 mm, several design points with d_a and d_b near 60 mm provide the optimal actuation distance within a range of $[3, 7]$ mm. These findings shed light on the complex interplay between the design parameters of the DEA and the resulting actuation performance and characteristic length, which is of great importance for the development of advanced soft robotics and related applications.

D. Validation with Experiments

The experimental system setup for data collection and model validation is shown in Fig. 7. The setup consists of a fabricated DEA, an adjustable high-voltage power supply (0 to 10 kV), a KEYENCE IL-S065 laser sensor, and a Phidget I/O device. The laser sensor provides precise displacement measurements of the DEA deformation, with a repeatability of 2 μm. Meanwhile, the I/O device includes various types of I/Os, such as digital-to-analog converters (DACs) and analog-to-digital converters (ADCs). The I/O device is utilized to send the control signal to the adjustable power supply to control the voltage output and to receive the measurement output from the laser sensor. Furthermore, the I/O device

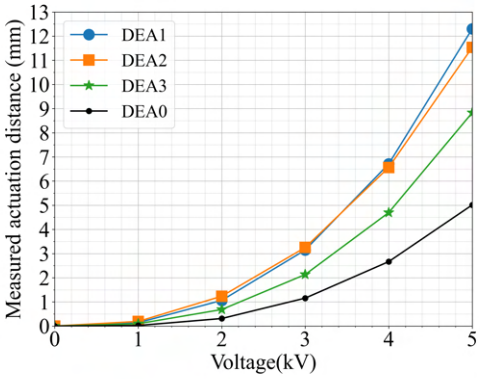

Fig. 8. Measured actuation distance at different input voltages showing that the value has been substantially increased by the optimized designs

is connected and communicated to a computer via USB and Robot Operating System (ROS) with a sampling rate of 100 Hz. The voltage control and data collection process are implemented on the computer via MATLAB. During the DEA displacement/deformation measurements, the measurement is recorded five seconds after a new voltage is applied to the DEA.

Three optimized designs of different size constraints are fabricated and marked as green crosses in Fig. 5. DEA1 and DEA2 have comparable sizes with the initial design DEA0 while DEA3 has a smaller diameter. These designs are tested in the laboratory, and their details are presented in Table I. The predicted actuation distances from GNN and FEM are quite close for all the cases, and the trend coincides with Pareto front in Fig. 5. When compared with the experimental data, the GNN and FEM predictions tend to underestimate the actuation distance with errors ranging from 8% to 18%. These sim-to-real gaps can be attributed to several factors. One explanation is that the fabrication process can introduce inaccuracies. For example, during the pre-stretching process, the actual stress/strain fields may deviate from the simulated models. Despite these discrepancies, the established framework is able to discover a more optimized and efficient design compared to the original intuitive design.

Fig. 8 illustrates the relationship between the actuation distance and applied voltage ranging from 1 kV to 5 kV. It shows that the actuation distance increases non-linearly with increasing applied voltage for all designs. Furthermore, all the optimized designs exhibit larger actuation distances than the initial designs across all voltage levels. The DEA with the largest characteristic length exhibits the most significant increase in actuation distance. Specifically, compared to the initial design, DEA1 doubles the actuation distance across all voltage levels. Specifically, at a voltage of 5 kV, the measured actuation distance of DEA1 was significantly enhanced by 146%.

V. CONCLUSION AND DISCUSSION

In this study, a simulation-aided design and optimization framework that enhances the performance of DEAs are presented. Firstly, a high-fidelity simulation model is established

to evaluate the electromechanical behaviors of DEAs during both the relaxation and actuation stages. The model allows for the generation of a database that was used to construct a surrogate model leveraging the state-of-the-art GNN. The GNN accepts the direct geometric representation of the designs, i.e., nodes and edges, as inputs, and produces the desired responses as its output. Then a multi-objective optimization problem was formulated to achieve a trade-off between the maximal actuation distance and characteristic length/width and NSGA-II was employed to solve the problem. The Pareto optimal front showed that larger characteristic lengths and actuation distances were obtained with larger DEA sizes. Finally, three optimized circular DEAs were fabricated and tested. Experimental results confirmed that the optimized designs achieved significantly improved actuation distances when compared with an intuitive design, demonstrating the effectiveness of the proposed framework.

In this study, no perforations were introduced to the current DEA design due to two primary reasons. Firstly, the inclusion of perforations results in a reduction of structural stiffness that is akin to a decrease in width. Secondly, the absence of perforations is favorable in terms of streamlining the fabrication process. Nonetheless, the GNN model can effectively incorporate such variations as node or edge features. Our future study aims to demonstrate the GNN's potential in designing more irregular soft robot component modules with sophisticated node and edge features. Likewise, the framework can also be used to design integrated soft robotic systems. Thus it is expected to greatly benefit the development of powerful and efficient soft robotics and related applications.

REFERENCES

[1] R. Pelrine, R. Kornbluh, Q. Pei, and J. Joseph, "High-speed electrically actuated elastomers with strain greater than 100%," *Science*, vol. 287, pp. 836–839, Feb. 2000.

[2] X. Zhao, W. Hong, and Z. Suo, "Electromechanical hysteresis and coexistent states in dielectric elastomers," *Phys. Rev. B Condens. Matter*, vol. 76, p. 134113, Oct. 2007.

[3] X. Zhao and Z. Suo, "Method to analyze electromechanical stability of dielectric elastomers," *Appl. Phys. Lett.*, vol. 91, p. 061921, Aug. 2007.

[4] Z. Suo, "Theory of dielectric elastomers," *Acta Mechanica Solida Sinica*, vol. 23, pp. 549–578, Dec. 2010.

[5] U. Gupta, L. Qin, Y. Wang, H. Godaba, and J. Zhu, "Soft robots based on dielectric elastomer actuators: a review," *Smart Mater. Struct.*, vol. 28, p. 103002, Oct. 2019.

[6] Y. Guo, L. Liu, Y. Liu, and J. Leng, "Review of Dielectric Elastomer Actuators and Their Applications in Soft Robots," *Advanced Intelligent Systems*, vol. 3, no. 10, p. 2000282, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aisy.202000282.

[7] G. Kofod, W. Wirges, M. Paajanen, and S. Bauer, "Energy minimization for self-organized structure formation and actuation," *Appl. Phys. Lett.*, vol. 90, p. 081916, Feb. 2007.

[8] Y. Wang, U. Gupta, N. Parulekar, and J. Zhu, "A soft gripper of fast speed and low energy consumption," *Sci. China Technol. Sci.*, vol. 62, pp. 31–38, Jan. 2019.

[9] Q. Pei, M. Rosenthal, S. Stanford, H. Prahlad, and R. Pelrine, "Multiple-degrees-of-freedom electroelastomer roll actuators," *Smart Mater. Struct.*, vol. 13, pp. N86–N92, Oct. 2004.

[10] Y. Guo, J. Guo, L. Liu, Y. Liu, and J. Leng, "Bioinspired multimodal soft robot driven by a single dielectric elastomer actuator and two flexible electroadhesive feet," *Extreme Mechanics Letters*, vol. 53, p. 101720, May 2022.

[11] G. Gu, J. Zou, R. Zhao, X. Zhao, and X. Zhu, "Soft wall-climbing robots," *Sci. Robot.*, vol. 3, p. eaat2874, Dec. 2018.

[12] G. Li, X. Chen, F. Zhou, Y. Liang, Y. Xiao, X. Cao, Z. Zhang, M. Zhang, B. Wu, S. Yin, Y. Xu, H. Fan, Z. Chen, W. Song, W. Yang, B. Pan, J. Hou, W. Zou, S. He, X. Yang, G. Mao, Z. Jia, H. Zhou, T. Li, S. Qu, Z. Xu, Z. Huang, Y. Luo, T. Xie, J. Gu, S. Zhu, and W. Yang, "Self-powered soft robot in the mariana trench," *Nature*, vol. 591, pp. 66–71, Mar. 2021.

[13] M. P. Bendsøe and O. Sigmund, *Topology Optimization*. Springer Berlin Heidelberg, 2004.

[14] X. Huang and M. Xie, *Evolutionary Topology Optimization of Continuum Structures: Methods and Applications*. Wiley & Sons, Limited, John, 2010.

[15] M. Y. Wang, X. Wang, and D. Guo, "A level set method for structural topology optimization," *Comput. Methods Appl. Mech. Eng.*, vol. 192, pp. 227–246, Jan. 2003.

[16] B. Chen, N. Wang, R. Wang, B. Zhu, X. Zhang, W. Sun, and W. Chen, "Automatic design of dielectric Elastomer-Based crawling robots using shape and topology optimization," *J. Mech. Robot.*, vol. 15, p. 021006, June 2022.

[17] Y. Luo, M. Y. Wang, and Z. Kang, "Topology optimization of geometrically nonlinear structures based on an additive hyperelasticity technique," *Comput. Methods Appl. Mech. Eng.*, vol. 286, pp. 422–441, Apr. 2015.

[18] R. Ortigosa, J. Martínez-Frutos, A. J. Gil, and D. Herrero-Pérez, "A new stabilisation approach for level-set based topology optimisation of hyperelastic materials," *Struct. Multidiscip. Optim.*, vol. 60, pp. 2343–2371, Dec. 2019.

[19] J. Zhou, W. Hong, X. Zhao, Z. Zhang, and Z. Suo, "Propagation of instability in dielectric elastomers," *Int. J. Solids Struct.*, vol. 45, pp. 3739–3750, June 2008.

[20] H. S. Park, Q. Wang, X. Zhao, and P. A. Klein, "Electromechanical instability on dielectric polymer surface: Modeling and experiment," *Comput. Methods Appl. Mech. Eng.*, vol. 260, pp. 40–49, June 2013.

[21] D. L. Henann, S. A. Chester, and K. Bertoldi, "Modeling of dielectric elastomers: Design of actuators and energy harvesting devices," *J. Mech. Phys. Solids*, vol. 61, pp. 2047–2066, Oct. 2013.

[22] Z.-Q. Zhang, C. C. Foo, and G. R. Liu, "A Semi-Explicit Finite Element Method for Dynamic Analysis of Dielectric Elastomers," *Int. J. Comput. Methods*, vol. 12, p. 1350108, Feb. 2015.

[23] J. Liu, C. C. Foo, and Z.-Q. Zhang, "A 3D multi-field element for simulating the electromechanical coupling behavior of dielectric elastomers," *Acta Mechanica Solida Sinica*, vol. 30, pp. 374–389, Aug. 2017.

[24] T. Lu, C. Ma, and T. Wang, "Mechanics of dielectric elastomer structures: A review," *Extreme Mechanics Letters*, vol. 38, p. 100752, July 2020.

[25] M. Wissler and E. Mazza, "Modeling and simulation of dielectric elastomer actuators," *Smart Mater. Struct.*, vol. 14, pp. 1396–1402, Oct. 2005. Publisher: IOP Publishing.

[26] M. Wissler and E. Mazza, "Modeling of a pre-strained circular actuator made of dielectric elastomers," *Sensors and Actuators A: Physical*, vol. 120, pp. 184–192, Apr. 2005.

[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017.

[28] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.

[29] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[30] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[31] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[32] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[33] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[35] J. Blank and K. Deb, "pymoo: Multi-objective optimization in python," *IEEE Access*, vol. 8, pp. 89497–89509, 2020.