1 2

# OPERATOR INFERENCE FOR ELLIPTIC EIGENVALUE PROBLEMS\*

#### 3

## HAOQIAN LI<sup>†</sup>, JIGUANG SUN<sup>‡</sup>, AND ZHIWEN ZHANG<sup>§</sup>

Abstract. Eigenvalue problems for elliptic operators play an important role in science and engi-4 neering applications, where efficient and accurate numerical computation is essential. In this work, we 5 propose a novel operator inference approach for elliptic eigenvalue problems based on neural network 6 7 approximations that directly maps computational domains to their associated eigenvalues and eigen-8 functions. Motivated by existing neural network architectures and the mathematical characteristics 9 of eigenvalue problems, we represent computational domains as pixelated images and decompose 10 the task into two subtasks: eigenvalue prediction and eigenfunction prediction. For the eigenvalue prediction, we design a convolutional neural network (CNN), while for the eigenfunction prediction, 11 we employ a Fourier Neural Operator (FNO). Additionally, we introduce a critical preprocessing 12 13 module that integrates domain scaling, detailed boundary pixelization, and main-axis alignment. 14 This preprocessing step not only simplifies the learning task but also enhances the performance of the neural networks. Finally, we present numerical results to demonstrate the effectiveness of the 15 proposed method. 16

Key words. Operator inference, elliptic eigenvalue problems, convolutional neural network,
 Fourier Neural Operator.

19 MSC codes. 35J15, 65N25, 68T07, 65T50.

1. Introduction. Eigenvalue problems for partial differential equations (PDEs) arise in a wide range of disciplines, including mathematics, physics, and engineering. Classical numerical methods—such as the finite element method, finite difference method, and spectral method—have been extensively studied and successfully applied [6, 11, 3]. While these methods are effective and reliable, they are often computationally intensive, rendering them less suitable for time-sensitive applications such as real-time simulation and inverse problems [18, 16].

Recently, data-driven and deep learning approaches have gained increasing at-27 tention for the computation of eigenvalue problems. In [7], Han et al. construct 28two deep neural networks to represent eigenfunctions in high-dimensional settings. 29 Li and Ying [13] propose a semigroup-based method using neural networks to solve 30 high-dimensional eigenvalue problems. In [22], Wang and Xie develop a tensor neural 31 network approach to compute multiple eigenpairs. Ji et al. [9] introduce a deep Ritz 32 33 method for approximating multiple elliptic eigenvalues in high dimensions. Ben-Shaul et al. [1] design an unsupervised neural network to compute multiple eigenpairs, and 34 Dai et al. [4] propose a subspace method utilizing basis functions generated by neural 35 networks. 36

<sup>\*</sup>Submitted to the editors DATE.

**Funding:** JS was partially supported by NSF Grant DMS-2109949 and SIMONS Foundation Collaboration Grant 711922. ZZ was partially supported by the National Natural Science Foundation of China (Project 92470103), the Hong Kong RGC grant (Project 17304324), an R&D Funding Scheme from the HKU-SCF FinTech Academy, the Outstanding Young Researcher Award of HKU (2020-21), and Seed Funding for Strategic Interdisciplinary Research Scheme 2021/22 (HKU).

<sup>&</sup>lt;sup>†</sup>Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China. lihaoqianlhq@connect.hku.hk.

<sup>&</sup>lt;sup>‡</sup>Corresponding author. Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, U.S.A. jiguangs@mtu.edu.

<sup>&</sup>lt;sup>§</sup>Corresponding author. Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China. AND Materials Innovation Institute for Life Sciences and Energy (MILES), HKU-SIRI, Shenzhen, P. R. China. zhangzw@hku.hk

In contrast with the above studies, which treat one eigenvalue problem (the domain is fixed), we are interested in the development of neural operators. Focusing on elliptic eigenvalue problems, we aim to build deep neural networks that take a general domain as input and output eigenvalues and/or eigenfunctions. Such operator learning is crucial for many query scenarios, e.g., shape optimization, inverse spectral problems, and real-time simulations, for which fast and reliable predictions of eigenvalues and eigenfunctions of different domains are necessary.

The first challenge lies in reformulating the eigenvalue problem as an operator 44 learning task. This reformulation can be approached in various ways—for instance, by 45representing the domain boundary as a continuous function. In this paper, we adopt 46a neural-network-oriented strategy. Specifically, we choose an operator representation 47 48 that aligns with existing neural networks known to perform well on similar problems. Leveraging the strengths of Convolutional Neural Networks (CNNs, [12]) and Fourier 49Neural Operators (FNOs, [14]), we represent both the computational domains and 50 their associated eigenfunctions as pixelated images. This image-based representation explicitly preserves the geometric features of the domains, which is beneficial for 53 learning and generalization.

The second challenge lies in generating a suitable dataset. Our dataset includes 54both random polygonal domains and smooth domains constructed using random 55Bézier curves. To support learning from pixel-based representations, we introduce 56 a preprocessing module designed to address several key issues. First, we scale each domain to fit within the unit square, ensuring consistent size across samples. To ac-58 count for the rotational and translational invariance of eigenvalues, we apply a main axis alignment (ma) technique to normalize domain orientation. Since pixelization 60 can introduce representation errors, particularly along rapidly varying boundaries, 61 we apply a detailed pixelization (dp) technique that refines boundary representation 62 without increasing image resolution or computational cost. This preprocessing step 63 standardizes the input data and enables the model to focus on intrinsic geometric 64 65 features, rather than being distracted by variations in scale, position, or orientation. The proposed method primarily falls within the framework of neural operator 66 learning [15, 2, 23]. Since the introduction of Fourier Neural Operators (FNOs), 67 there has been growing interest in employing them for learning mappings between 68 infinite-dimensional function spaces, with demonstrated success in applications such 69 as weather forecasting and inverse problems [10, 20, 17]. FNOs excel in learning com-70 71 plex mappings between function spaces due to their ability to capture both local and global features through Fourier transforms. Our proposed method extends this capa-72 bility by using CNNs and FNOs to respectively predict eigenvalues and eigenfunctions. 73 Specifically, we employ CNNs to learn the mapping from the domain representations 7475 (pixelated images) to eigenvalues, and FNOs to learn the mapping from domains to

reigenfunctions. This separation of tasks not only enhances the model's interpretability, but also allows for flexible model expansion [5]. A key strength of our method is its generalizability to unseen domain geometries, which is particularly useful for shape optimization and real-time design queries, where rapid evaluation of diverse configurations is critical.

We conduct extensive experiments to evaluate the robustness, generalization capability, and predictive accuracy of the proposed method. The results demonstrate that our approach achieves high accuracy in both regression and function approximation tasks for the first few eigenpairs. Notably, the relative error in eigenvalue prediction remains stable across different eigenvalue indices, yielding uniformly low errors for the first 20 eigenvalues. In contrast, the relative error in eigenfunction prediction tends to increase with higher eigenvalue indices, reflecting the growing complexity and oscillatory nature of the corresponding eigenfunctions.

The rest of the paper is organized as follows. In Section 2, we discuss the model setup and data generation. We provide details for the main axis alignment and detailed pixelization. Section 3 presents the structures of the CNN and FNO for eigenvalue and eigenfunction prediction, respectively. The loss functions and training of the networks are also discussed. In Section 4, we present various experiments to demonstrate the effectiveness of networks. We end up with conclusions and future work in Section 5.

2. Model Setup and Data Generation. In this section, we present the modal problem, the choice of the operator, and the generation of dataset. We consider the representative Dirichlet eigenvalue problem in two dimensions. Let  $\Omega \subset \mathbb{R}^2$  be a simply connected bounded Lipschitz domain. The eigenvalue problem is to find  $\lambda \in \mathbb{R}$ and  $u \in H^1(\Omega)$  such that

101 (2.1) 
$$-\Delta u = \lambda u \text{ in } \Omega \text{ and } u = 0 \text{ on } \partial \Omega$$

102 As stated above, we shall represent  $\Omega$  by an image such that the shape information 103 is encoded in pixel values, 1 for pixels inside the domain and 0 outside. This represen-104 tation allows for the incorporation of geometric information. To further simplify the 105 task, we shall treat the eigenvalues and the eigenfunctions separately by introducing 106 two subtasks:  $\Omega \to \lambda$  and  $\Omega \to u$ .

107 The eigenvalues are invariant to both position and rotation of  $\Omega$ . In addition, 108 they are scale-dependent. If we multiply the x and y coordinates by a factor k, the 109 eigenvalues are scaled by a factor of  $\frac{1}{k^2}$ . Hence, we first scale a domain such that it fits 110 into the unit square  $[0, 1]^2$ . Taking these properties into consideration, we introduce 111 a preprocessing module to simplify the task.

**2.1. Domain Generation.** To generate a domain, we select n random points in  $[0,1]^2$ , enforcing a minimum distance threshold c between them. These points are translated so that the centroid is (0,0) and then reordered according to their principal arguments, denoted by  $p_1, p_2, ..., p_n$ . Polygons are obtained by connecting adjacent points with line segments

117 (2.2) 
$$L_{i,i+1}(t) = t \cdot p_i + (1-t) \cdot p_{i+1}, \quad t \in [0,1].$$

118 To obtain a smooth domain, a cubic Bézier curve segment is generated to connect 119  $p_i$  and  $p_{i+1}$ . Then  $\partial\Omega$  is obtained by concatenating all the Bézier curve segments. We 120 briefly describe how to construct a cubic Bézier curve as follows. The angles between 121 each pair of consecutive points are calculated. A weighted average is then computed

122 (2.3) 
$$\theta_{i,i+1}^* = w \cdot \theta_{i-1,i} + (1-w) \cdot \theta_{i,i+1}, \quad i = 1, 2, \dots, n-1,$$

123 where  $\theta_i$  is the principle argument of  $p_i$  and  $w = \arctan(\epsilon)/\pi + 0.5$ . Here  $\epsilon$  is a 124 parameter that controls the "smoothness" of the curve with  $\epsilon = 0$  being the smoothest. 125 Two control points between  $p_i$  and  $p_{i+1}$  are generated

126 (2.4) 
$$p_{i}^{*} = p_{i} + r \cdot (\cos(\theta_{i,i+1}^{*}), \sin(\theta_{i,i+1}^{*})), \\ p_{i+1}^{*} = p_{i+1} - r \cdot (\cos(\theta_{i,i+1}^{*}), \sin(\theta_{i,i+1}^{*})), \\ \end{cases}$$

127 where the positive number  $r \in [0, 1]$  controls the curvature. For r = 0,  $p_i^*$  and 128  $p_{i+1}^*$  coincide, respectively, with  $p_i$  and  $p_{i+1}$ , and the curve has larger curvature at the control points. Intermediate values of r produce smoother curves, with maximal smoothness when r = 0.5. When increasing further toward r = 1, sharp features start to appear near the crossing of the initial and final curve tangents. The points  $p_i^*$  and  $p_{i+1}^*$  are used to define a cubic Bézier curve segment connecting  $p_i$  and  $p_{i+1}$ :

133 (2.5) 
$$B_{i,i+1}(t) = (1-t)^3 p_i + 3(1-t)^2 t p_i^* + 3(1-t)t^2 p_{i+1}^* + t^3 p_{i+1}, \quad t \in [0,1].$$

The above method for generating smooth domain boundaries is inspired by a discussion on Stack Overflow <sup>1</sup> and has also been employed in [21] for the generation of random shapes.

**2.2. Dataset Generation.** The eigenvalue problem (2.1) has several properties that bring challenges to the design of effective neural networks. Firstly, it is rotationally invariant. Traditional computer vision techniques typically address this by employing data augmentation, where images are rotated by various angles to increase the diversity of the training data. Another approach is to introduce global pooling layers or more complex architectures to capture rotational invariance. However, these methods do not fully address the difficulty.

We choose a different approach that simplifies the problem: main axis alignment, which aligns the domains along their principal axes. This preprocessing step eliminates the rotational invariance by adjusting the domain's orientation before feeding it into the model. Specifically, given a series of points on  $\partial\Omega$ , denoted by

148 
$$Q = \begin{pmatrix} x_0 & x_1 & \dots & x_n \\ y_0 & y_2 & \dots & y_n \end{pmatrix}$$
, we compute the covariance matrix of  $Q$ 

149 (2.6) 
$$M = \frac{1}{n-1} \sum_{i=1}^{n} (Q - \bar{Q})^T (Q - \bar{Q}).$$

150 Then, we apply the eigenvalue decomposition of M

151 (2.7) 
$$M = D_M \cdot v_M,$$

where the eigenvectors  $v_M$  are sorted in ascending order according to the associated eigenvalues (diagonal of  $D_M$ ). The principal axes are given by these eigenvectors. We then rotate the region coordinates about its centroid using  $v_M$ 

155 (2.8) 
$$Q_{\text{rotated}} = (Q - \bar{Q}) \cdot v_M.$$

156 Secondly, the problem is flipping invariant about x or y axis. To address this 157 issue, we impose an additional constraint by requiring that the x-direction standard 158 deviation in the left half of the region is less than or equal to that of the right half, 159 and that the y-direction standard deviation in the lower half is less than or equal to 160 that of the upper half.

161 Thirdly, we need to represent a domain as a low-resolution pixelated image. The 162 value of a pixel is 0 if it is completely outside  $\partial\Omega$  and 1 if it is completely inside. The 163 situation for pixels on the boundary is more complicate. The eigenvalues are sensitive 164 to the pixel value on the boundary if the domain is rather narrow. A simple 0/1 165 choice inevitably reduces accuracy.

This issue is particularly subtle and cannot be resolved through conventional hyperparameter tuning, as illustrated by the following example. Consider four rectangular domains  $\Omega_i = (0, w_i) \times (0, h)$  with height h = 1 and width  $w_i = \frac{2 + \frac{i}{4}}{32}, i = 1, 2, 3, 4$ .

<sup>&</sup>lt;sup>1</sup>https://stackoverflow.com/questions/50731785/create-random-shape-contour-using-matplotlib

169 Using  $32 \times 32$  pixels, they have the same image but different eigenvalues. The first 170 Dirichlet eigenvalues are

171 (2.9) 
$$\lambda_i = \pi^2 \left( \frac{1}{w_i^2} + \frac{1}{h^2} \right), \quad i = 1, 2, 3, 4,$$

172 which approximately equal to 2006.21, 1626.91, 1346.26, and 1132.81.

173 While increasing the image resolution can help, it significantly increases the com-174 putational cost. To this end, we employ edge-sensitive pixelization. Specifically, we 175 first generate a high-resolution image and then use average pooling to down-sample 176 the image to the target resolution. The pixel value near the boundary is no longer 177 binary (0/1) but instead a value between 0 and 1, reflecting the proportion of the 178 region contained within the pixel. Although this technique is simple, it significantly 179 improves accuracy without increasing the training cost or the size of the model.

To generate the dateset, i.e., eigenpairs of (2.1), we use a linear Lagrange finite element code on a triangular mesh with mesh size  $h \approx 0.01$  (Chp. 3 of [19]). The errors of the computed eigenpairs can be ignored for our purposes. Figure 1 illustrates the workflow of data generation. Figure 2 displays some samples from the dataset. For each domain, it shows the first ten eigenvalues and the associated eigenfunctions.

### 185 **3. Methodology.**

**3.1. CNN for Eigenvalue Prediction.** For eigenvalue prediction, we propose a convolutional neural network. CNNs are particularly well-suited for the task due to their powerful ability to capture both global and local geometric features. The local receptive fields of CNNs allow them to effectively extract fine-grained spatial information, while the deeper layers capture global patterns that are critical for understanding the overall shape of the domain.

In comparison to traditional Deep Neural Networks (DNNs), CNNs exhibit a 192specialized architecture tailored for processing grid-like data, such as images. One 193 of the primary distinctions lies in how CNNs handle spatial relationships. Unlike 194 DNNs, which treat input data as a flattened vector, CNNs preserve the spatial ar-195rangement of pixels. Furthermore, CNNs reduce the computational burden compared 196 to DNNs by employing weight sharing and pooling operations. The shared weights in 197 convolutional layers drastically reduce the number of parameters compared to fully 198 connected layers in DNNs. This not only enhances the model's ability to generalize 199but also makes CNNs more computationally efficient. Note that CNNs possess the 200 universal approximation property [8]. 201

The proposed CNN eigenvalue predictor consists of several key components: convolutional layers, max-pooling layers, batch normalization layers, fully connected layers, activation functions (ReLU), and skip connections. Below is a brief overview of each component and its role.

The convolutional layer is the core building block of a CNN. It applies a set of learnable filters (or kernels) to the input data in order to extract local features. For a 2D image input, the convolution operation is given by

209 (3.1) 
$$\operatorname{Conv}_{k}(\mathbf{X})(i,j) = (\mathbf{X} * \mathbf{K})(i,j) = \sum_{m=0}^{k} \sum_{n=0}^{k} \mathbf{X}(i+m,j+n) \mathbf{K}(m,n)$$

where  $\mathbf{X} \in \mathbb{R}^{d \times d}$  is the input image, **K** is the convolutional kernel (filter), k is the size of the filter.



FIG. 1. Diagram of dataset generation.

Max-pooling is applied after a convolutional layer to downsample the spatial dimensions of the feature map. It reduces the computational complexity and helps prevent overfitting. The max-pooling is defined as

215 (3.2) 
$$\operatorname{Pool}_{k'}(\mathbf{X})(i,j) = \max_{m,n \in [0,k']} \mathbf{X}(i+m,j+n),$$

where k' is the size of the pooling window. The operation selects the maximum value within each local region, producing a downsampled output.

Batch normalization (Bn) is used to normalize the inputs by adjusting and scaling the activations. This helps to stabilize and speed up the training process. The



FIG. 2. Samples in the dataset. The first 10 eigenvalues and eigenfunctions are shown, as well as their corresponding regions.

220 normalized output is given by

221 (3.3) 
$$\operatorname{Bn}(\mathbf{X}) = \gamma \left(\frac{\mathbf{X} - \mu}{\sigma}\right) + \beta,$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation of the batch, respectively, and  $\gamma$ and  $\beta$  are learnable scaling and shifting parameters, respectively.

The fully connected (FC) layer is applied after the convolutional and pooling layers to make predictions based on the extracted features. The FC layer takes the flattened output from the previous layer and computes a weighted sum, followed by a bias,

228 (3.4) 
$$\operatorname{FC}_{d_{in},d_{out}}(\mathbf{x}) = \mathbf{W}_{d_{in} \times d_{out}} \mathbf{x} + \mathbf{b}_{d_{out}},$$

where  $\mathbf{W}_{d_{in} \times d_{out}}$  is the weight matrix,  $\mathbf{x} \in \mathbb{R}^{d_{in}}$  is the flattened input vector, and b<sub>d<sub>out</sub> is the bias term.</sub>

The activation function introduces non-linearity into the network, enabling it to learn complex patterns. We use the Rectified Linear Unit (ReLU)

233 (3.5) 
$$\operatorname{ReLU}(x) = \max(0, x).$$

This function replaces all negative values with zero, introducing sparsity and reducing the risk of vanishing gradients during the training.

Skip connections (or residual connections) allow the model to bypass certain layers and directly pass the output from one layer to a deeper layer. In our case, since we do not need a very deep network, we use skip connections to pass downsampled original input to the deep convolution layers to allow the model to capture global features. These connections also help prevent the vanishing gradient problem and allow for more efficient training. In the context of downsampling, we use a 1 × 1 convolution to reduce the spatial dimensions

243 (3.6) 
$$\operatorname{Skip}_{s}(\mathbf{X}) = \mathbf{X} * \mathbf{K}'_{1 \times 1}^{(s)},$$

where s is the stride and  $\mathbf{K}'_{1\times 1}^{(s)}$  is a 1 × 1 conventional kernel with stride s, which corresponds to the downsampling factor of the input. This operation effectively reduces the spatial dimensions of the input without changing the depth (number of channels) of the feature maps, making it suitable for skip connections that maintain spatial information while reducing computational cost. The result of this operation is added element-wise to the output of a convolutional layer, combining both local and global information.

For different image sizes, the network needs to be slightly adjusted. For the 251 $32 \times 32$  case, we use 3 convolutional blocks, each containing a convolutional layer 252with 64/128/256 channels, kernel size 7/5/3, padding 3/2/1, and stride 1. Each 253convolutional layer is followed by a batchnorm layer, an activation function, and a 254max-pooling operation with size  $2 \times 2$ , which downsamples the width and height of the 255feature map by 2. Next, a skip connection of the downsampled input is added to the 256output of the max-pooling operation of each convolutional block. The convolutional 257block can be written as 258

259 (3.7) 
$$\operatorname{ConvBlock}(\mathbf{X}) = \operatorname{Pool}_{k'} \left( \operatorname{ReLU}(\operatorname{Bn}(\operatorname{Conv}_k(\mathbf{X}))) \right) + \operatorname{Skip}(\mathbf{X}).$$

The output of the last convolutional block is then flattened and passed through three fully connected layers and two activation functions in between to make predic-

tions. For  $64 \times 64$ , we simply add another conventional block and adjust the channels, kernel size, and padding accordingly. The rest of the network remains unchanged.

264 The overall structure is illustrated in Figure 3.



FIG. 3. Structures of the proposed CNNs. Left:  $32 \times 32$  image. Right:  $64 \times 64$  image.

**3.2. FNO for Eigenfunction Prediction.** We propose a Fourier Neural Operator (FNO) to learn the mapping between images and the eigenfunctions. Domains with complex boundaries or irregular shapes have subtle influences on the eigenfunctions, and FNO's ability to operate in the frequency domain can capture these features efficiently.

An FNO is organized as a sequence of layers that alternately operate in the spa-270tial domain and the frequency domain. In each Fourier block, the input is first passed 271through a linear mapping layer  $\mathcal{P}$  to lift the dimensionality of the channels. Then, 272the lifted input passes through multiple Fourier blocks, in which a discrete Fourier transform  $\mathcal{F}$  is applied, with a weight tensor multiplication  $\mathcal{R}$ , which introduces learn-274able parameters, followed by an inverse Fourier transform  $\mathcal{F}^{-1}$ . In the Fourier block, 275the lifted input is also processed in parallel by a local linear transformation  $\mathcal{W}$ . The 276output of the above two paths is added element-wise and finally mapped back at the 277last Fourier block to the output channel dimension via another linear mapping  $\mathcal{Q}$ . We 278 describe each layer in more detail as follows. 279

The first layer of the FNO applies a linear mapping  $\mathcal{P}$ , which operates along the channel dimension of the input. This layer is to elevate the input channels to a higher-dimensional space, enabling the model to learn richer representations of the input data. The transformation is implemented as a fully connected (FC) layer

284 (3.8) 
$$\mathcal{P}(\mathbf{X}) = \mathrm{FC}_{c_{in},c_{hd}}(\mathbf{X}),$$

where  $\mathbf{X} \in \mathbb{R}^{c_{in} \times d \times d}$  is the input with  $c_i n$  channels,  $c_{in}$  and  $c_{hd}$  are the number of channels in the input and hidden space, respectively. It allows for increased flexibility and capacity in the subsequent layers of the model, as it projects the input features into a higher-dimensional space where more complex interactions can be captured.

Once the input is mapped to a higher-dimensional space, the discrete Fast Fourier Transform (FFT) is applied to compute the frequency components of the input data, which are crucial for understanding periodic or spatially extended features of the domain

293 (3.9) 
$$(\mathcal{F}_M(\mathbf{X}))_c = \sum_{x,y=0}^{d-1} \mathbf{X}_c(x,y) e^{-\frac{2i\pi(M_1x+M_2y)}{d}}, \quad c \in [1,c_{hd}],$$

where  $M = (M_1, M_2)$  are the largest discrete modes for dimensions 1 and 2 (namely x and y direction) and  $\mathbf{X}_c$  is the *c*th channel of  $\mathbf{X}$ .

FFT transforms the spatial domain data into a frequency domain representation. Then the frequency domain representation is processed by a weight tensor  $\mathcal{R} \in \mathbb{C}^{M_{max} \times c_{hd} \times c_{hd}}$  to introduce learnable parameters that enable the model to capture specific patterns or features in the input data

300 (3.10) 
$$(\mathcal{R} \cdot \mathcal{F}_M(\mathbf{X}))_{m,c} = \sum_{j=1}^{c_{hd}} \mathcal{R}_{m,c,j}(\mathcal{F}_M(\mathbf{X}))_j, \quad m \in [1, M_{max}], \quad c \in [1, c_{hd}].$$

Afterward, an Inverse Fast Fourier Transform (IFFT) is applied to map  $\tilde{\mathbf{X}} = \mathcal{R} \cdot \mathcal{F}_M(\mathbf{X})$  to the spatial domain

303 (3.11) 
$$(\mathcal{F}_{M}^{-1}(\tilde{\mathbf{X}}))_{c} = \sum_{\tilde{x}=0}^{d_{\tilde{x}}-1} \sum_{\tilde{y}=0}^{d_{\tilde{y}}-1} \tilde{\mathbf{X}}_{c}(\tilde{x},\tilde{y}) e^{2i\pi \left(\frac{(M_{1}\tilde{x})}{d_{\tilde{x}}} + \frac{(M_{2}\tilde{y})}{d_{\tilde{y}}}\right)}, \quad c \in [1, c_{hd}].$$

304 Finally, the following Fourier integral operator  $\mathcal{K}$  is applied

305 (3.12) 
$$\mathcal{K}_M(\mathbf{X})(x,y) = \mathcal{F}^{-1}(\mathcal{R} \cdot \mathcal{F}_M(\mathbf{X}))(x,y).$$

306 While the data passes through the Fourier integral operator, it also passes through

 $_{307}$  a local linear transformation  $\mathcal{W}$  in parallel, which is implemented as a two-stacked

 $1 \times 1$  convolution with an activation function in between. It operates in the spatial

domain and allows the model to perform fine-grained adjustments to the features. The transformation  $\mathcal{W}$  is defined as

310 The transformation VV is defined as

311 (3.13) 
$$\mathcal{W}(\mathbf{X}) = \operatorname{Conv}_{1 \times 1}(\operatorname{GeLU}(\operatorname{Conv}_{1 \times 1}(\mathbf{X}))),$$

where GeLU is the Gaussian Error Linear Unit (3.14)

313 GeLU(x) = 
$$x \cdot \Phi(x)$$
, where  $\Phi(x) = 0.5x \left( 1 + \tanh\left(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)\right) \right)$ ,

a smooth approximation of ReLU that has been shown to improve performance in various deep-learning tasks.

A Fourier block combining  $\mathcal{W}$  and  $\mathcal{K}_M(\mathbf{X})$  is defined as

317 (3.15) FourierBlock<sub>M</sub>(
$$\mathbf{X}$$
) = GeLU ( $\mathcal{K}_M(\mathbf{X}) + \mathcal{W}(\mathbf{X})$ ).

After several Fourier blocks, the output is passed through a second linear mapping Q, which converts the dimensionality of the channels back to the desired output dimension. This layer is implemented as a fully connected layer and is responsible for mapping the high-dimensional representations learned in the previous layers back to the output space. The mapping Q is defined as

323 (3.16) 
$$\mathcal{Q}(\mathbf{X}) = \mathrm{FC}_{c_{hd}, c_{out}}(\mathbf{X}).$$

324 The complete network representation is as follows:

325 (3.17) 
$$\operatorname{FNO}(\mathbf{X}) = \mathcal{Q} \circ \operatorname{FourierBlock}_{M}^{(n)} \circ \ldots \circ \operatorname{FourierBlock}_{M}^{(1)} \circ \mathcal{P}(\mathbf{X}).$$

For different input image sizes, the same network architecture is used. Only the 326 number of hidden channels changes. Specifically, for  $d \times d$  images, the number of 327 hidden channels is d. In the experiments, we use 4 Fourier blocks and 16 frequency 328 modes in both directions. To better capture the spatial domain information after the 329 330 Fourier transform, we incorporate positional encoding into the input. In particular, we augment the original single-channel input by adding x and y coordinate values of 331 332 each sampling point as additional channels. Consequently, the input image becomes a three-channel representation. The overall architecture of the proposed FNO is illustrated in Figure 4. 334

**335 3.3. Network Training.** We describe the loss functions for the networks and the training parameters. For the prediction of the eigenvalue, the loss function is

337 (3.18) 
$$\mathcal{L}_1(\hat{y}, y) = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|,$$

where N is the number of data points,  $\hat{y}_i$ 's are the predicted values, and  $y_i$ 's are the actual values. The MSE penalizes large deviations between predicted and actual values.

We use the Adam optimizer ( $\beta_1 = 0.9, \beta_2 = 0.999$ ) for training. The initial learning rate is set to 0.001, and the learning rate is halved every 10 steps during training.



FIG. 4. Structure of the proposed FNO.

This allows the model to start with a higher learning rate for faster convergence and
gradually reduce it for finer adjustments as the model approaches the optimal solution.
The dataset consists of 20,000 domain-eigenvalue-eigenfunction triplets, with
10,000 smooth domains and 10,000 non-smooth domains. We use 80% of the dataset
(16,000 input-output pairs) for training and the remaining 20% (4,000 input-output
pairs) for testing. The batch size is set to 32. The model is trained for 100 epochs.

For the prediction of the eigenfunction, the design of the loss function and the training process are more challenging. One needs to take the following key elements into account. (1) Domain of definition. The output image should be 0 outside  $\Omega$ . (2) Normalization. It is important that the predicted eigenfunctions are normalized. This prevents the model from learning arbitrary scaling factors. (3) Sign ambiguity. Even if the predicted eigenfunctions are normalized, they may still differ by a sign.

To address the above difficulties, we propose an adaptive relative masked  $L_2$  loss

356 (3.19) 
$$\mathcal{L}_{ARM2}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} \min \left\{ \| \hat{\mathbf{y}}_i \otimes X_i - \mathbf{y}_i \|_2, \| - \hat{\mathbf{y}}_i \otimes X_i - \mathbf{y}_i \|_2 \right\} / \| \mathbf{y}_i \|_2,$$

where  $\otimes$  denotes the Hadamard product (element-wise multiplication),  $X_i$  is a  $d \times d$ binary mask indicating the domain interior,  $\hat{\mathbf{y}}$  and  $\mathbf{y}$  are the  $d \times d$  predicted and true eigenfunctions, and  $\|\cdot\|_2$  denotes the  $L_2$  norm.

360 Noted that the output **y** is already normalized such that

361 (3.20) 
$$\|\mathbf{y}_i\|_2 = C, \quad i \in \{1, 2, \dots, N\}.$$

The loss function (3.19) ensures that the value of the predicted eigenfunction within the domain is optimized, while the value outside is ignored. It also takes care of the sign ambiguity by considering both positive and negative versions of the predicted eigenfunction. Finally, the denominator  $||\mathbf{y}_i||_2$  normalizes the loss by the true eigenfunction norm within the region. It ensures that the predicted eigenfunctions have the same scale, reducing the difficulty of model training.

Again, the Adam optimizer is used. The initial learning rate is 0.001 and the learning rate decay factor is 0.8. We set the batch size to be 128 and train for 50 epochs. The remaining hyperparameters are the same as the eigenvalue case.

4. Numerical Experiments. We present experimental results, including an ablation study, to show the performance of the proposed networks. Special attention is devoted to the prediction of the first eigenpair, which plays a critical role in applications such as spectral geometry and quantum mechanics. Meanwhile, the successful prediction of other eigenpairs demonstrates the robustness and versatility of

4.1. Eigenvalue Prediction. To evaluate the performance of the eigenvalue prediction, we use several metrics, including the Root Mean Squared Error (RMSE), R-squared  $(R^2)$ , and Mean Absolute Percentage Error (MAPE).

Let  $y_i$  and  $\hat{y}_i$  be the true eigenvalues and the predicted eigenvalues, respectively. The RMSE measures the average magnitude of the errors between the predicted and actual values

383 (4.1) 
$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}.$$

<sup>384</sup> It is particularly sensitive to large errors, as it squares the differences before averaging.

385 A lower RMSE indicates better prediction accuracy.

Let  $\bar{y}$  be the mean of the true eigenvalues. The R-squared metric

387 (4.2) 
$$R^{2} = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_{i} - y_{i})^{2}}{\sum_{i=1}^{N} (y_{i} - \bar{y})^{2}}$$

measures the proportion of variance in the true eigenvalues that is explained by the model's predictions. It indicates how well the model captures the underlying relationship between the input and the eigenvalue. An  $R^2$  value closer to 1 indicates that the model explains most of the variance in the data, while a value closer to 0 or even smaller than 0 suggests poor performance.

393 Let  $\epsilon$  be a small positive number. The MAPE is defined as

394 (4.3) 
$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{\hat{y}_i - y_i}{y_i + \epsilon} \right|,$$

which measures the average absolute percentage error between the predicted and true eigenvalues. The MAPE is particularly useful for comparing predictions across different scales.

Figure 5 shows the loss curves for the first eigenvalue prediction for  $32 \times 32$  and 64 × 64 images in log scale. Both the training and testing losses of the two models stop decreasing significantly around 80 epochs. Figure 6 shows the predictions of the first eigenvalues of some domains, which are close to the true values.

Table 1 shows different error metrics for the prediction of the first eigenvalue. The detailed pixelization and main axis alignment significantly improve the model accuracy. The transition from a  $32 \times 32$  grid to a  $64 \times 64$  yields a slight improvement in the MAPE, from 1.15% to 0.73%, and a modest increase in the RMSE. The examples in the rest of the paper use the  $32 \times 32 + dp + ma$  setting.

Figure 7 presents several samples with large errors. They are elongated domains.
Such domains are more sensitive to small pixelization errors. More intricate shapes
tend to introduce greater approximation errors, leading to higher MAPEs.

410 Next, we consider more eigenvalues. Table 2 shows the performance metrics for 411 the 1st, 2nd, and 3rd eigenvalues. It can be seen that RMSE becomes larger for larger 412 eigenvalues, while  $R^2$  and MAPE are relatively stable. Figure 8 plots RMSE,  $R^2$ , and 413 MAPE for the first 20 eigenvalues. RMSE increases uniformly with the eigenvalue 414 index, and the growth rate aligns with the standard deviation of the corresponding 415 eigenvalues.  $R^2$  and MAPE remain nearly constant.

376

our approach.



FIG. 5. Losses in log scale for the first eigenvalue. Left:  $32 \times 32$  image. Right:  $64 \times 64$  image.



FIG. 6. Sample predictions of the first eigenvalue. Top:  $32 \times 32$ . Bottom:  $64 \times 64$ .

This manuscript is for review purposes only.

### TABLE 1

Performance metrics for the first eigenvalue prediction with different settings.  $d \times d+dp+ma$  stands for  $d \times d$  image with detailed pixelization on the boundaries and main axis alignment.  $d \times d+ma$  stands for  $d \times d$  image with main axis alignment.

| Setting                  | RMSE   | $R^2$  | MAPE  |
|--------------------------|--------|--------|-------|
| $32 \times 32$ +dp+ma    | 2.2121 | 0.9988 | 1.15% |
| $32 \times 32 + ma$      | 5.9933 | 0.9917 | 1.75% |
| $32 \times 32$           | 7.3343 | 0.9864 | 1.67% |
| $64 \times 64 + dp + ma$ | 2.2113 | 0.9988 | 0.73% |
| $64 \times 64 + ma$      | 3.0124 | 0.9979 | 1.57% |
| $64 \times 64$           | 3.6447 | 0.9966 | 1.78% |



FIG. 7. Samples with the largest errors for first eigenvalue predictions. Top: prediction samples with the largest RMSEs. Bottom: prediction samples with the largest MAPEs. "T", "P", and "E" refer to the true, predicted, and error values, respectively. All samples are from the  $32 \times 32 + dp + ma$  setting.

TABLE 2 Performance metrics for the first three eigenvalues  $(32 \times 32 + dp + ma)$ .



FIG. 8. Prediction metrics for the first 20 eigenvalues. Left: RMSE. Middle:  $R^2$ . Right: MAPE

416 **4.2. Eigenfunction Prediction.** We now present results for the eigenfunction 417 prediction. Three metrics are used: Maximum Absolute Error (MaxAE), Peak Signal-418 to-Noise Ratio (PSNR), and Relative L1 Error (RelL1).

The Maximum Absolute Error (MaxAE) is defined as the largest difference between the predicted and true eigenfunctions

421 (4.4) 
$$\operatorname{MaxAE} = \max |u_{\operatorname{true}}(x) - u_{\operatorname{pred}}(x)|,$$

422 where  $u_{\text{pred}}(x)$  and  $u_{\text{true}}(x)$  are the predicted and true eigenfunctions evaluated at x, 423 respectively.

The Peak Signal-to-Noise Ratio (PSNR), often used in image and signal processing, is defined as

426 (4.5) 
$$\operatorname{PSNR} = 20 \log_{10} \left( \frac{\max_{x} |u_{\text{true}}(x)|}{\text{RMSE}} \right),$$

427 where  $\max_{x} |u_{\text{true}}(x)|$  is the maximum value of the true eigenfunction and RMSE is 428 the Root Mean Squared Error between the predicted and true functions.

PSNR provides a measure of the quality of the predicted eigenfunction by comparing the noise (or error) relative to the maximum signal (the true function). A larger PSNR indicates better quality, as the error is smaller relative to the signal's strength. This metric is especially useful for evaluating the overall quality of the predicted eigenfunction, particularly when many small errors may not be captured by MaxAE.

435 The Relative L1 Error (RelL1) is defined as

436 (4.6) 
$$\operatorname{RelL1} = \frac{\sum_{x} |u_{\operatorname{true}}(x) - u_{\operatorname{pred}}(x)|}{\sum_{x} |u_{\operatorname{true}}(x)|}.$$

The use of MaxAE, PSNR, and RelL1 allows for a comprehensive evaluation of
the eigenfunction prediction: MaxAE gives insight into the worst-case errors, PSNR
provides a relative measure of the prediction quality, and RelL1 offers an overall
picture of the error function.

Figure 9 shows the loss curves for the first eigenfunction prediction on the  $32 \times 32$ grid and the  $64 \times 64$  grid in log scale. The decreases in training and testing losses slow down significantly after around 40 epochs, indicating that the models have been sufficiently trained. Figure 10 shows sample predictions of the first eigenfunction for both  $32 \times 32$  and  $64 \times 64$  grids. The predictions are generally close to the actual circumstions and the acels of amount of both mid-sizes is relatively small





FIG. 9. Losses in log scale for the first eigenfunction. Left:  $32 \times 32$ . Right:  $64 \times 64$ .

Table 3 lists MaxAE, PSNR, and RelL1 for the first eigenfunction prediction under different settings. Similar to Table 1, the use of detailed pixelization (dp) and main axis alignment (ma) improves the accuracy. For the  $32 \times 32$  grid, dp+ma reduces MaxAE from 0.26 to 0.12, improves PSNR from 42.64 to 52.00, and significantly reduces RelL1 from 3.63% to 1.30%. Similar results are observed for the  $64 \times 64$  grid. MaxAE decreases from 0.14 to 0.09, RelL1 decreases from 1.55% to 0.72%, and PSNR increases from 49.88 to 57.54.

TABLE 3

Performance metrics for the first eigenfunction prediction with different settings.  $d \times d+dp+ma$ stands for  $d \times d$  grid with detailed pixelization and main axis alignment.  $d \times d+ma$  stands for  $d \times d$ grid with main axis alignment.

| Setting                  | MaxAE | PSNR  | RelL1 |
|--------------------------|-------|-------|-------|
| $32 \times 32$ +dp+ma    | 0.12  | 52.00 | 1.30% |
| $32 \times 32 + ma$      | 0.27  | 43.14 | 3.57% |
| $32 \times 32$           | 0.26  | 42.64 | 3.63% |
| $64 \times 64 + dp + ma$ | 0.09  | 57.54 | 0.72% |
| $64 \times 64 + ma$      | 0.14  | 50.18 | 1.56% |
| $64 \times 64$           | 0.14  | 49.88 | 1.55% |

Table 4 shows the performance metrics for different eigenfunctions. In contrast to Table 2, where the relative error (MAPE) remains stable for different eigenvalues, the prediction of eigenfunctions exhibits a clear trend: both the absolute error (MaxAE)

<sup>457</sup> and the relative error (RelL1) increase, and PSNR decreases.

|             |         |     |     | TABLE     | 2 4            |                             |  |
|-------------|---------|-----|-----|-----------|----------------|-----------------------------|--|
| Performance | metrics | for | the | different | eigenfunctions | $(32 \times 32 + dp + ma).$ |  |

| Index of Eigenvalue | MaxAE | PSNR  | RelL1  |
|---------------------|-------|-------|--------|
| 1                   | 0.11  | 53.04 | 1.22%  |
| 2                   | 0.37  | 42.13 | 7.04%  |
| 3                   | 0.68  | 35.88 | 13.01% |

Figure 11 presents several samples of the predicted second and third eigenfunctions. Although the prediction errors are larger than the first eigenfunction, the



FIG. 10. Sample predictions of the first eigenfunction. Top:  $32 \times 32$ . Bottom:  $64 \times 64$ .

460 predicted and true eigenfunctions remain close. For higher eigenfunctions, the model 461 can capture the structure and key features.

Figure 12 shows MaxAE, PSNR, and RelL1 for the first 20 eigenfunctions. The performance declines as the eigenfunction index increases. It is interesting that MaxAE, which reflects the absolute error, aligns with the average gradient norm of the eigenfunctions. This correlation indicates that higher eigenfunctions, with more complex structures and oscillations, pose greater challenges for accurate prediction. The simultaneous decrease in PSNR and increase in RelL1 can partially be attributed



FIG. 11. Predictions of the second (top) and third (bottom) eigenfunctions.

468 to the resolution.

Figure 13 shows predicted results for higher eigenfunctions (10th and 20th). For some samples, the predicted eigenfunctions preserve the overall structures, despite non-negligible errors. Other predicted eigenfunctions differ significantly from the exact ones.

5. Conclusions. In this paper, we proposed a novel operator learning framework for the efficient prediction of eigenvalues and eigenfunctions of elliptic equations on various two-dimensional domains. Unlike most existing neural network methods



FIG. 12. Prediction metrics for the first 20 eigenfunctions. Left: MaxAE. Middle: PSNR. Right: RelL1.

that focus on single eigenvalue problems or specific high-dimensional settings, our 476 477 method takes different domains as inputs and outputs the corresponding eigenvalues and eigenfunctions. The key innovations of our approach are threefold. Firstly, we 478 479 adopt a divide-and-conquer strategy to decompose the task into two subtasks: eigenvalue prediction and eigenfunction prediction. Secondly, we reformulate the problem 480in a neural-network-oriented manner, enabling the effective application of CNNs for 481 eigenvalue prediction and FNOs for eigenfunction prediction. Thirdly, we simplify the 482 input problem before passing it into the neural networks, which helps to improve the 483 training efficiency and prediction accuracy. 484

To validate the effectiveness of our proposed method, we conducted extensive numerical experiments on a variety of two-dimensional elliptic domains. The results demonstrate that our method achieves significant improvements in terms of prediction accuracy and computational efficiency. In the future, we will extend our approach to more challenging problems, including non-self-adjoint eigenvalue problems, nonlinear eigenvalue problems and parameterized eigenvalue problems.

491 **Declarations.** The authors have no competing interests to declare that are rel-492 evant to the content of this article.

I. BEN-SHAUL, L. BAR, D. FISHELOV, AND N. SOCHEN, Deep learning solution of the eigenvalue
 problem for differential operators, Neural Computation, 35 (2023), pp. 1100–1134.

REFERENCES

493

- 496 [2] K. BHATTACHARYA, B. HOSSEINI, N. B. KOVACHKI, AND A. M. STUART, Model reduction and 497 neural networks for parametric PDEs, The SMAI journal of computational mathematics, 498 7 (2021), pp. 121–157.
- [3] I. BUBUŠKA AND J. OSBORN, Eigenvalue problems. Handbook of numerical analysis, Vol. II, 500 North-Holland, Amsterdam, 1991.
- [4] X. DAI, Y. FAN, AND Z. SHENG, Subspace method based on neural networks for eigenvalue
   problems, arXiv preprint arXiv:2410.13358, (2024).
- [5] H. DU, Z. LI, J. LIU, Y. LIU, AND J. SUN, Divide-and-conquer DNN approach for the inverse point source problem using a few single frequency measurements, Inverse Problems, 39 (2023), pp. Paper No. 115006, 19 pp.
- [6] M. FEIT, J. FLECK JR, AND A. STEIGER, Solution of the Schrödinger equation by a spectral method, Journal of Computational Physics, 47 (1982), pp. 412–433.
- [7] J. HAN, J. LU, AND M. ZHOU, Solving high-dimensional eigenvalue problems using deep neural networks: A diffusion Monte Carlo like approach, Journal of Computational Physics, 423 (2020), p. 109792.
- 511 [8] J. HE, L. LI, AND J. XU, Approximation properties of deep ReLU CNNs, Research in the 512 mathematical sciences, 9 (2022), p. 38.
- [9] X. JI, Y. JIAO, X. LU, P. SONG, AND F. WANG, Deep Ritz method for elliptical multiple
   eigenvalue problems, Journal of Scientific Computing, 98 (2024), p. 48.
- [10] M. A. KHABOU, L. HERMI, AND M. B. H. RHOUMA, Shape recognition using eigenvalues of the Dirichlet Laplacian, Pattern recognition, 40 (2007), pp. 141–153.





FIG. 13. Sample predictions of the 10th and 20th eigenfunctions. Top: the 10th eigenfunction. Bottom: the 20th eigenfunction.

- [11] J. R. KUTTLER AND V. G. SIGILLITO, Eigenvalues of the Laplacian in two dimensions, Siam
   Review, 26 (1984), pp. 163–193.
- [12] Y. LECUN, L. BOTTOU, Y. BENGIO, AND P. HAFFNER, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86 (1998), pp. 2278–2324.
- [13] H. LI AND L. YING, A semigroup method for high dimensional elliptic PDEs and eigen value problems based on neural networks, Journal of Computational Physics, 453 (2022),
   p. 110939.
- [14] Z. LI, N. KOVACHKI, K. AZIZZADENESHELI, B. LIU, K. BHATTACHARYA, A. STUART, AND
   A. ANANDKUMAR, Fourier neural operator for parametric partial differential equations, arXiv preprint arXiv:2010.08895, (2020).
- [15] L. LU, P. JIN, AND G. E. KARNIADAKIS, DeepOnet: Learning nonlinear operators for identifying
   differential equations based on the universal approximation theorem of operators, arXiv
   preprint arXiv:1910.03193, (2019).
- [16] T. NGUYEN, J. BRANDSTETTER, A. KAPOOR, J. K. GUPTA, AND A. GROVER, Climax: A foundation model for weather and climate, arXiv preprint arXiv:2301.10343, (2023).
- [17] N. PALLIKARAKIS AND A. NTARGARAS, Application of machine learning regression models to
   inverse eigenvalue problems, Computers & Mathematics with Applications, 154 (2024),
   pp. 162–174.
- [18] J. PATHAK, S. SUBRAMANIAN, P. HARRINGTON, S. RAJA, A. CHATTOPADHYAY, M. MARDANI,
   T. KURTH, D. HALL, Z. LI, K. AZIZZADENESHELI, ET AL., Fourcastnet: A global data-driven
   high-resolution weather model using adaptive Fourier neural operators, arXiv preprint
   arXiv:2202.11214, (2022).
- 539 [19] J. SUN AND A. ZHOU, Finite element methods for eigenvalue problems, Chapman and 540 Hall/CRC, 2016.
- [20] B. T. THODI, S. V. R. AMBADIPUDI, AND S. E. JABARI, Fourier neural operator for learning solutions to macroscopic traffic flow models: Application to the forward and inverse problems, Transportation research part C: emerging technologies, 160 (2024), p. 104500.
- [21] J. VIQUERATA AND E. HACHEMA, A supervised neural network for drag prediction of arbitrary
   2D shapes in low reynolds number flows, arXiv preprint arXiv:1907.05090, (2019).
- 546 [22] Y. WANG AND H. XIE, Computing multi-eigenpairs of high-dimensional eigenvalue problems 547 using tensor neural networks, Journal of Computational Physics, 506 (2024), p. 112928.
- [23] Z. WANG, J. XIN, AND Z. ZHANG, DeepParticle: learning invariant measure by a deep neural network minimizing wasserstein distance on data generated from an interacting particle method, Journal of Computational Physics, 464 (2022), p. 111309.