

DEEP LEARNING-BASED MODELING AND SIMULATION OF HEAT CONDUCTION

ONDŘEJ ŠPERL, JAN SÝKORA*

*Czech Technical University in Prague, Faculty of Civil Engineering, Department of Mechanics,
Thákurova 2077/7, 160 00 Prague 6 – Dejvice, Czech Republic*

* corresponding author: jan.sykora.1@fsv.cvut.cz

ABSTRACT. The present study focuses on applying deep neural networks (DNNs) to surrogate modeling of heat conduction problems. Deep learning algorithms, valued for their ability to learn hierarchical data representations through multi-layered networks, excel at identifying complex patterns. In this work, the U-Net architecture – widely recognized for its effectiveness in image segmentation – is adapted to model stationary heat transfer, providing a novel approach to a critical challenge in engineering and physics. Specifically, we propose a deep learning-based surrogate model to predict stationary temperature fields in 2D rectangular domains representing two-phase heterogeneous materials.

KEYWORDS: Stationary heat transport, convolutional neural network, U-Net architecture.

1. INTRODUCTION

Surrogate modeling is a popular computational strategy employed to approximate complex, high-fidelity models with simpler, more efficient representations, while aiming to preserve the essential characteristics of the original forward model [1]. This approach is particularly advantageous in scenarios where the original model is computationally intensive, making it impractical for tasks such as optimization, uncertainty quantification, or real-time simulations. The primary objective of surrogate modeling is to achieve a balance between accuracy and computational efficiency, thereby facilitating more rapid and cost-effective analyses. Surrogate modeling has found widespread application across various scientific and engineering disciplines, including aerospace [2], civil engineering [3], materials science [4], and chemistry [5].

There are various types of surrogate models, each characterized by distinct advantages and tailored to specific application domains. Polynomial response surfaces [6], such as quadratic or cubic models, are simple and efficient but may struggle with highly nonlinear systems. Radial basis functions [7] and Kriging [8] (or Gaussian processes) offer more flexibility in capturing complex behaviors but can be more computationally demanding. Artificial neural networks [9] and support vector machines [10] are powerful tools for high-dimensional and nonlinear problems, though they require careful tuning and can be data-intensive. Polynomial chaos expansions [11] are particularly useful for uncertainty quantification problems, as they provide a structured way to represent uncertainty in the model outputs.

There has been extensive research on constructing surrogate models based on the deep learning approach. These can be generally classified into physics-based, see [12, 13], and data-driven surrogate models [14]. Physics-based surrogate models rely on mathematical

equations derived from physical laws to describe system behavior, requiring less data and offering greater interpretability but potentially struggling with complex, nonlinear phenomena. In contrast, data-driven surrogate models use statistical and machine learning techniques to learn from large datasets, providing rapid predictions and good generalization.

Here, we concentrate on the construction of data-driven surrogate models utilizing deep neural networks. Each DNN contains multiple layers of interconnected nodes enabling one to learn complex data representations by identifying hierarchical patterns and features, as detailed in [14]. A typical DNN consists of three types of layers:

- (1.) An input layer that processes input data into the model,
- (2.) hidden layers that constitute the core of the DNN, applying weights to the inputs and passing them through the activation functions to generate outputs, and
- (3.) an output layer that computes the final probability scores of the desired outputs.

A specialized form of DNN is the convolutional neural network (CNN), designed for image classification and visual data interpretation [15]. CNNs operate on the principles of mathematical convolution, extracting salient features from images using learnable filters. They are primarily composed of convolutional layers and max pooling layers, which reduce the number of weights in the neural network.

Surrogate models founded on the principles of DNN and CNN architectures and constructed for heat transfer modeling represent the main idea of this contribution. Our framework extends the work detailed in [16, 17], examining this aspect from various perspectives. The paper is structured as follows: The

Methodology Section 2 concisely outline the fundamentals of non-stationary heat transport and the subsequent discretization using the finite element method. The effectiveness of the proposed DNN-based surrogate model is evaluated in Section 3 through its application to the stationary heat conduction problem with Dirichlet boundary conditions. Our conclusions will be summarized in the final Section 4.

2. METHODOLOGY

Initially, we consider a bounded body $D \subset \mathbb{R}^3$ (reference configuration) with a piecewise smooth boundary Γ . Specifically, the Dirichlet, Neumann, and Robin boundary conditions are applied to $\Gamma_D \subset \Gamma$, $\Gamma_N \subset \Gamma$, and $\Gamma_R \subset \Gamma$, respectively, such that $\Gamma = \Gamma_D \cup \Gamma_N \cup \Gamma_R$. To analyze the temporal behavior of D , we consider a time interval $[0, T_s] \subset \mathbb{R}_+$. The temperature evolution in D is given by the function

$$\theta : D \times [0, T_s] \rightarrow \mathbb{R}^3, \quad (1)$$

where θ (in degrees Celsius) represents the temperature. The heat transport is then governed by the transient heat balance equation with initial and boundary conditions.

$$\left\{ \begin{array}{l} c_v(x) \frac{\partial \theta}{\partial t}(x, t) - \nabla \cdot (\lambda(x) \nabla \theta(x, t)) = 0, \\ \quad \quad \quad x \in D, t \in (0, T_s), \\ \theta(x, t) = \theta_D(x, t), \\ \quad \quad \quad x \in \Gamma_D, t \in (0, T_s), \\ \lambda(x) \frac{\partial \theta}{\partial n}(x, t) = q_N(x, t), \\ \quad \quad \quad x \in \Gamma_N, t \in (0, T_s), \\ \alpha(\theta(x, t) - \theta_\infty(x, t)) = \lambda(x) \frac{\partial \theta}{\partial n}(x, t), \\ \quad \quad \quad x \in \Gamma_R, t \in (0, T_s), \\ \theta(x, 0) = \theta_{in}(x), \\ \quad \quad \quad x \in D. \end{array} \right. \quad (2)$$

The equations involve several key parameters to define the thermal behavior of the system. The thermal conductivity is represented by $\lambda(x)$ [$\text{W m}^{-1} \text{K}^{-1}$], while the volumetric heat capacity, $c_v(x)$ [$\text{J m}^{-3} \text{K}^{-1}$], is derived as the product of the volumetric mass density $\rho_s(x)$ [kg m^{-3}] and the specific heat capacity $c_p(x)$ [$\text{J kg}^{-1} \text{K}^{-1}$], expressed as $c_v(x) = \rho_s(x) c_p(x)$. The final simulation time is denoted by T_s [s], and the ambient temperature is given by $\theta_\infty(x, t)$ [$^\circ\text{C}$]. Furthermore, the heat transfer coefficient is α [$\text{W m}^{-2} \text{K}^{-1}$], with $\theta_D(x, t)$ [$^\circ\text{C}$] and q_N [W m^{-2}] representing the prescribed temperature and the heat flux, respectively.

We begin by restricting our analysis to the stationary case, which significantly simplifies data management in the neural network training process. This approach allows us to focus on a comprehensive and detailed investigation of the problem. The governing equation for the stationary heat problem is given by:

$$\left\{ \begin{array}{ll} \nabla \cdot (\lambda(x) \nabla \theta(x)) = 0, & x \in D, \\ \theta(x) = \theta_D(x), & x \in \Gamma_D, \\ \lambda(x) \frac{\partial \theta}{\partial n}(x) = q_N(x), & x \in \Gamma_N, \\ \alpha(\theta(x) - \theta_\infty(x)) = \lambda(x) \frac{\partial \theta}{\partial n}(x), & x \in \Gamma_R. \end{array} \right. \quad (3)$$

The finite element method (FEM) is a powerful numerical technique used to approximate solutions to complex partial differential equations. It works by transforming continuous functions into discrete vectors, making them easier to handle with numerical tools. One of the most common FEM approaches is the Galerkin method, which we use in this context to discretize the stationary heat conduction equation, see [18]. The method's core idea is to minimize the error between the exact and approximate solutions by ensuring the residual – the difference between them – is orthogonal to the chosen finite-dimensional space. The Galerkin method's orthogonality condition for the residual guarantees that the approximation error is minimized, yielding a mathematically consistent solution, see [19].

2.1. DNN-BASED SURROGATE MODELS

Generally, the process of constructing a data-driven surrogate model involves several key steps. Initially, a set of design points is sampled – usually from the multidimensional parameter space – often using techniques like Latin Hypercube Sampling or Sobol sequences to ensure a representative distribution. This set of design points is then evaluated using the high-fidelity FEM-based model to generate corresponding output/sampled data. The input-output (labelled) pairs are subsequently used to train the specific type of surrogate model. However, the generation of this synthetic dataset through high-fidelity FEM simulations often represents the most computationally expensive bottleneck in the workflow, and the proper size of the dataset needs to be carefully considered.

The surrogate models proposed here are based on the CNN architecture, with the U-Net architecture chosen as the most suitable option for further development. Introduced by Ronneberger et al. [17], the U-Net builds on the fully convolutional network framework proposed by Long et al. [20]. Its innovation lies in augmenting the standard contracting network with upsampling layers, replacing pooling operations to progressively increase output resolution. This design allows convolutional layers to produce precise outputs by leveraging contextual information. A key feature is the use of extensive feature channels during upsampling, which ensures effective context propagation to higher-resolution layers. The resulting architecture is nearly symmetrical, forming its characteristic “U” shape. The specific implementation for our numerical example is depicted in Figure 1.

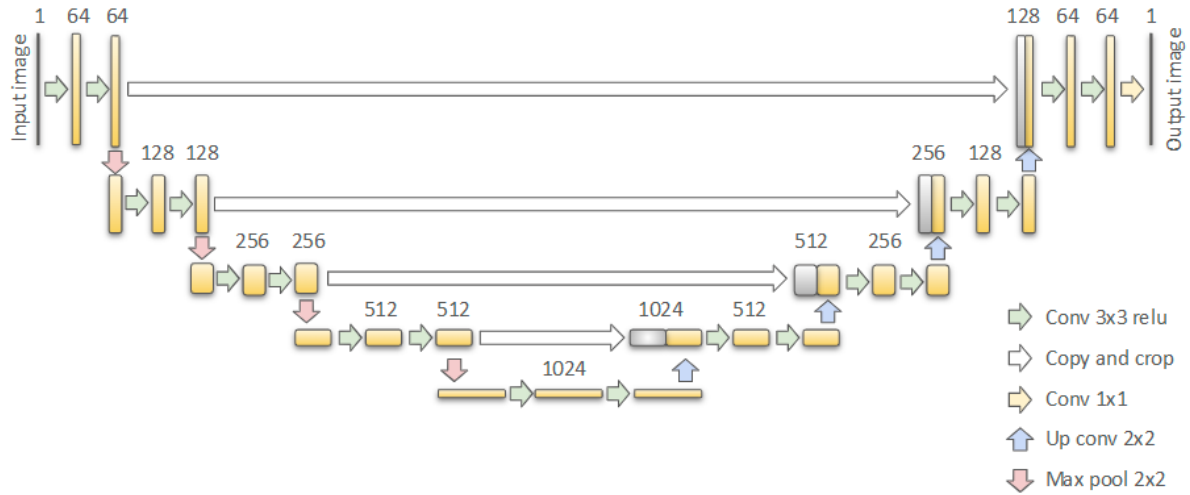


FIGURE 1. This U-Net diagram illustrates a convolutional network that downsamples the input in the encoder via pooling and convolutions, then upsamples it in the decoder using up-convolutions. Skip connections bridge corresponding encoder and decoder layers, preserving spatial information lost during downsampling.

3. EXAMPLE

We investigate a 2D rectangular domain representing two-phase heterogeneous materials under Dirichlet boundary conditions, with the left and right boundaries held at 20 °C and 5 °C, respectively. The domain is meshed into quadrilateral finite elements, where each element corresponds to a pixel of a specific material phase. The domain includes two material phases with thermal conductivities of $\lambda_1 = 1.0 \text{ W m}^{-1} \text{ K}^{-1}$ and $\lambda_2 = 5.0 \text{ W m}^{-1} \text{ K}^{-1}$. Heterogeneity is introduced by randomly distributing overlapping λ_1 circles within the predominantly λ_2 domain, as depicted in Figure 2.

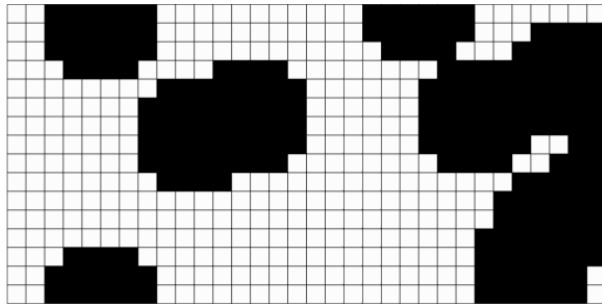


FIGURE 2. One particular sample from FEM-based dataset depicting the morphology of a heterogeneous material.

A well-constructed synthetic dataset is vital for training neural networks effectively, ensuring both efficient learning and high performance. We generated a dataset of 20 000 FEM simulations with randomly created domain morphologies to capture critical features. Given the simplicity of the problem, managing this large number of simulations was straightforward, allowing the network to model complex relationships and adapt to various material fields.

Since the data preprocessing is almost the most important part of machine learning, the FEM-based

dataset is further normalized to temperatures varying between 0 °C and 1 °C. This normalization offers several advantages for neural networks: It accelerates learning for gradient-based algorithms, helps mitigate the issue of local minima that can trap neural networks during optimization, enhances model performance by accommodating the scale sensitivity of activation functions, reduces overfitting by promoting better generalization and lower sensitivity to specific input feature values, and ensures numerical stability by preventing computational issues related to extremely large or small numbers.

After normalization, the dataset is divided into training, validation, and testing subsets to ensure the development of a robust and generalizable deep learning model. This partitioning is essential for several reasons: it enables an unbiased assessment of model performance on unseen data (testing set), helps prevent overfitting by monitoring training progress, and identifies if the model is memorizing noise or outliers. In addition, the validation set is critical for optimizing hyperparameters – such as learning rate, batch size, or network architecture – and for comparing different models or algorithms. For these reasons, our dataset is split into three distinct parts: 15 000 samples for training, 2 500 for validation, and 2 500 for testing.

During the training process, the unknown parameters of the DNN-based surrogate model are determined using an optimization solver based on the training dataset. Simultaneously, the loss is evaluated on the validation dataset to monitor overfitting and underfitting issues and to estimate the appropriate number of training epochs. The loss values for both the training and validation datasets in our specific example are illustrated in Figure 3. The close alignment between the two curves suggests that the model is neither underfitting nor overfitting, achieving a stable and reliable

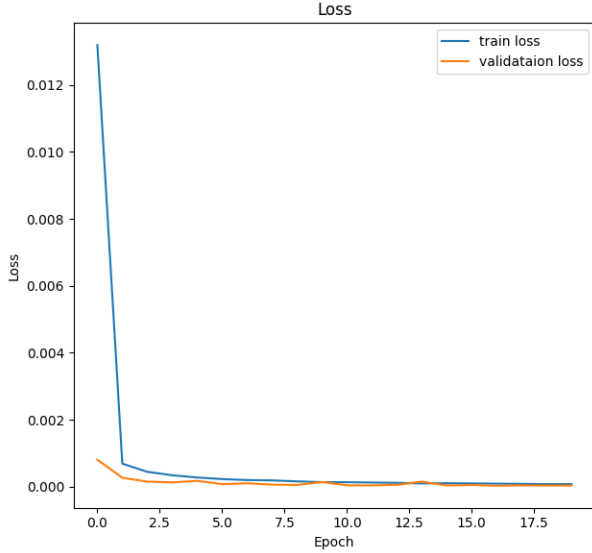


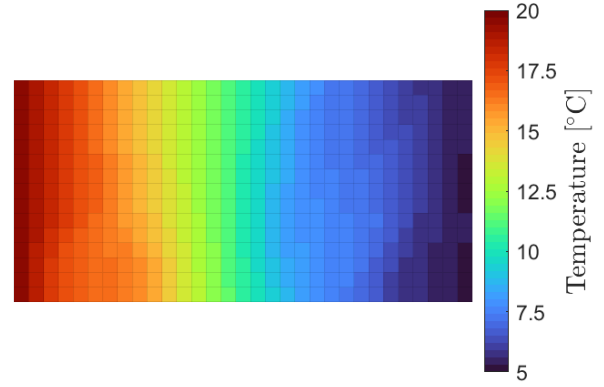
FIGURE 3. Convergence graph.

performance by the end of the training. The presented graph indicates that achieving a relatively well-trained model requires an epoch count exceeding 10.

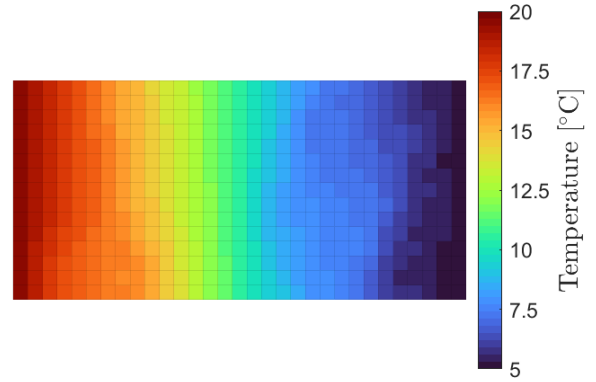
Once, DNN-based surrogate model is trained, a qualitative assessment can be conducted using unseen data. Prior to this assessment, Figure 4 illustrates the temperature fields corresponding to a particular thermal conductivity field from the unseen dataset. Figure 4a presents the ground truth temperature field obtained through the FEM simulation, whereas Figure 4b displays the temperature field predicted by the DNN-based surrogate model. Upon visual comparison, the two results appear nearly identical. However, when the error values are visualized in Figure 4c, differences ranging from -0.2°C to 0.3°C become evident. The evaluation across the entire unseen dataset is presented in Figure 5, which displays the mean (Figure 5a), standard deviation (Figure 5b), and maximum value of errors (Figure 5c). These graphs are derived from the dataset comprising the absolute temperature differences between the FEM-based simulations and the predictions generated by the DNN-based surrogate model. The graph of average absolute error (Figure 5a) demonstrates that the proposed DNN-based surrogate model achieves a relatively high degree of accuracy. This observation is supported by the second graph (Figure 5b), which illustrates the standard deviation of the absolute error and shows that it remains consistently within similar bounds.

4. CONCLUSION

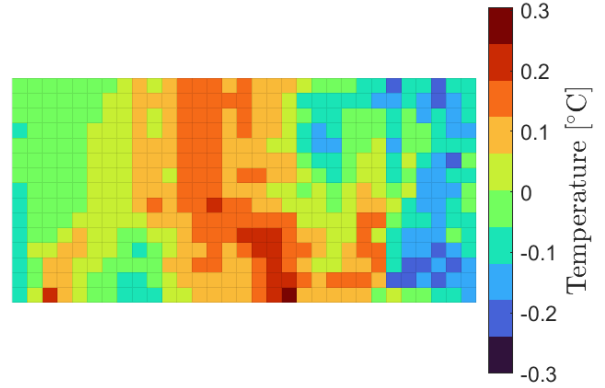
This study demonstrates the efficiency of a DNN-based surrogate model in predicting stationary temperature fields within a heterogeneous 2D rectangular domain subjected to Dirichlet boundary conditions. The model was trained using an extensive dataset of 15 000 FEM simulations, with careful attention paid



(A). FEM-based temperature field.



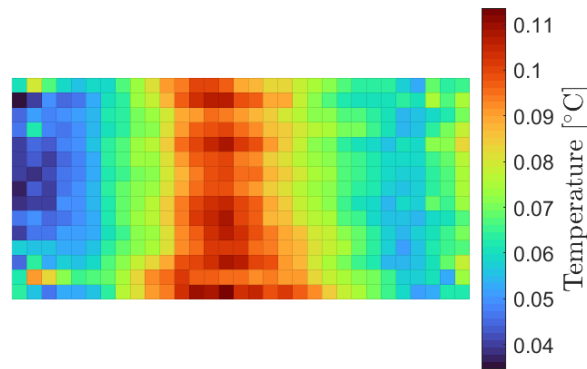
(B). Temperature field predicted from the DNN-based surrogate model.



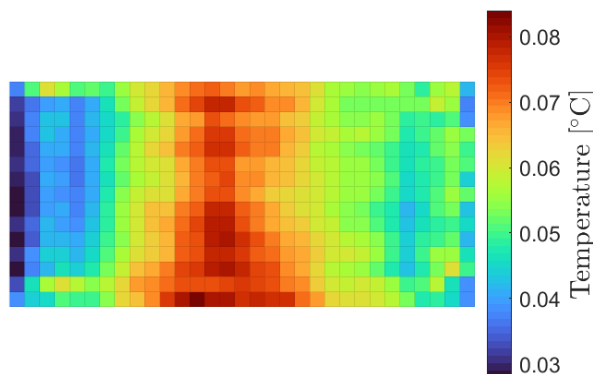
(C). Difference between FEM-based and surrogate model-based temperature fields.

FIGURE 4. Comparison of temperature fields computed for one specific thermal field.

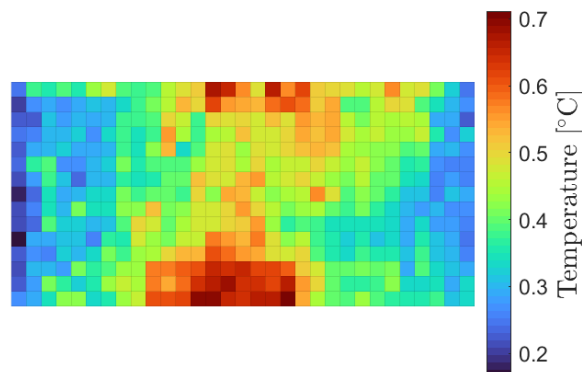
to data preprocessing and normalization to enhance learning and generalization. The evaluation of the model on unseen data revealed a high degree of accuracy, with average absolute errors remaining consistently low and standard deviations within acceptable bounds. The successful application of DNN-based surrogate model highlights its potential for efficient and accurate prediction of complex physical phenomena, offering a promising alternative to computationally intensive FEM simulations.



(A). Average absolute error of temperature in $^{\circ}\text{C}$.



(B). Standard deviation of the absolute error in $^{\circ}\text{C}$.



(C). Maximum value of errors in $^{\circ}\text{C}$.

FIGURE 5. Qualitative assessment of DNN-based surrogate model conducted using unseen dataset.

ACKNOWLEDGEMENTS

The authors are thankful for financial support from the Student Grant Competition of CTU, project No. SGS23/152/OHK1/3T/11 and the Czech Science Foundation, project No. 22-35755K.

REFERENCES

- [1] B. Sudret, S. Marelli, J. Wiart. Surrogate models for uncertainty quantification: An overview. In *2017 11th European conference on antennas and propagation* (EUCAP), pp. 793–797. IEEE, 2017. <https://doi.org/10.23919/EuCAP.2017.7928679>
- [2] A. Prabhakar, J. Fisher, R. Bhattacharya. Polynomial chaos-based analysis of probabilistic uncertainty in hypersonic flight dynamics. *Journal of guidance, control, and dynamics* **33**(1):222–234, 2010. <https://doi.org/10.2514/1.41551>
- [3] A. Kučerová, J. Sýkora, P. Havlásek, et al. Efficient probabilistic multi-fidelity calibration of a damage-plastic model for confined concrete. *Computer Methods in Applied Mechanics and Engineering* **412**:116099, 2023. <https://doi.org/10.1016/j.cma.2023.116099>
- [4] Y. M. Marzouk, H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics* **228**(6):1862–1902, 2009. <https://doi.org/10.1016/j.jcp.2008.11.024>
- [5] K. McBride, K. Sundmacher. Overview of surrogate modeling in chemical process engineering. *Chemie Ingenieur Technik* **91**(3):228–239, 2019. <https://doi.org/10.1002/cite.201800091>
- [6] A. I. Khuri, S. Mukhopadhyay. Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(2):128–149, 2010. <https://doi.org/10.1002/wics.73>
- [7] M. D. Buhmann. Radial basis functions. *Acta Numerica* **9**:1–38, 2000. <https://doi.org/10.1017/S0962492900000015>
- [8] M. A. Oliver, R. Webster. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System* **4**(3):313–332, 1990. <https://doi.org/10.1080/02693799008941549>
- [9] J. Zou, Y. Han, S.-S. So. Overview of artificial neural networks. In *Artificial Neural Networks: Methods and Applications*, pp. 14–22. Humana Press, 2008. https://doi.org/10.1007/978-1-60327-101-1_2
- [10] M. Pecha, D. Horák. Analyzing l_1 -loss and l_2 -loss Support Vector Machines implemented in PERMON toolbox. In *AETA 2018-Recent Advances in Electrical Engineering and Related Sciences: Theory and Application*, pp. 13–23. Springer, 2020. https://doi.org/10.1007/978-3-030-14907-9_2
- [11] E. Janouchová, J. Sýkora, A. Kučerová. Polynomial chaos in evaluating failure probability: A comparative study. *Applications of Mathematics* **63**(6):713–737, 2018. <https://doi.org/10.21136/AM.2018.0335-17>
- [12] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, et al. Physics-informed machine learning. *Nature Reviews Physics* **3**(6):422–440, 2021. <https://doi.org/10.1038/s42254-021-00314-5>
- [13] M. Raissi, P. Perdikaris, G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* **378**:686–707, 2019. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [14] F. Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

- [15] J. Gu, Z. Wang, J. Kuen, et al. Recent advances in convolutional neural networks. *Pattern Recognition* **77**:354–377, 2018.
<https://doi.org/10.1016/j.patcog.2017.10.013>
- [16] Y. Wang, J. Zhou, Q. Ren, et al. 3-D steady heat conduction solver via deep learning. *IEEE Journal on Multiscale and Multiphysics Computational Techniques* **6**:100–108, 2021.
<https://doi.org/10.1109/JMMCT.2021.3106539>
- [17] O. Ronneberger, P. Fischer, T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241. Springer, 2015.
https://doi.org/10.1007/978-3-319-24574-4_28
- [18] J. Havelka, J. Sýkora. Application of Calderón’s inverse problem in civil engineering. *Applications of Mathematics* **63**(6):687–712, 2018.
<https://doi.org/10.21136/AM.2018.0323-17>
- [19] O. C. Zienkiewicz, R. L. Taylor. *The finite element method set*. Elsevier, 2005.
- [20] E. Shelhamer, J. Long, T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence* **39**(4):640–651, 2016.
<https://doi.org/10.1109/TPAMI.2016.2572683>