



# Efficient multiscale modeling of heterogeneous materials using deep neural networks

Fadi Aldakheel<sup>1</sup> · Elsayed S. Elsayed<sup>1</sup> · Tarek I. Zohdi<sup>2</sup> · Peter Wriggers<sup>3</sup>

Received: 31 October 2022 / Accepted: 21 March 2023 / Published online: 27 April 2023  
© The Author(s) 2023

## Abstract

Material modeling using modern numerical methods accelerates the design process and reduces the costs of developing new products. However, for multiscale modeling of heterogeneous materials, the well-established *homogenization techniques* remain *computationally expensive* for high accuracy levels. In this contribution, a machine learning approach, convolutional neural networks (CNNs), is proposed as a computationally efficient solution method that is capable of providing a high level of accuracy. In this work, the data-set used for the training process, as well as the numerical tests, consists of artificial/real microstructural images (“input”). Whereas, the output is the homogenized stress of a given representative volume element  $\mathcal{RVE}$ . The model performance is demonstrated by means of examples and compared with traditional homogenization methods. As the examples illustrate, high accuracy in predicting the homogenized stresses, along with a significant reduction in the computation time, were achieved using the developed CNN model.

**Keywords** Deep learning · Convolutional neural networks · Computational micro-to-macro approach · Heterogeneous materials

## 1 Introduction

Engineering design is an optimization process in which engineers identify and solve problems. The key elements of product design are functionality, cost, form, material properties, and behavior. In this context, heterogeneous materials that consist of different constituents such as polymer matrix composites, metallic-materials, wood, concrete, and ceramic composites, play a major role in the optimization process. Such advanced materials have been in use for a few decades in the automotive, aerospace industry, and constructions due to their superior properties, compared to their homogeneous counterparts. As the macroscopic properties of such heterogeneous materials are greatly influenced by their micro-constituents properties, the understanding of the

microstructure of such materials is an essential task to fully utilize their potential in the design process. In this regard, the direct use of the classical multiscale techniques for solving such problems is unfeasible due to the huge computational costs. Thus, a need for a computationally less expensive approach exists. One of the current possibilities introduced in literature is the employment of data-driven models to reduce the computation costs [1–5]. Machine learning, deep learning, or neural networks are examples of such data-driven models that help to reduce the model complexity and may surpass conventional constitutive modeling, see [6–26] and the citations therein.

The objective of this work is to introduce a computationally efficient and accurate solution to speed up multiscale modeling processes. The ability of the Convolutional Neural Networks (CNN) to capture the macroscopic constitutive properties from images of the microstructure will be utilized to predict the homogenized elastic stress. In comparison with published works [27–30], we highlight the remarkable efficiency of the proposed CNN model by employing different microstructures with various phases and comparing the results with the classical homogenization approaches. Furthermore, a key point is the illustration of advanced transfer learning of the CNN model to new realistic microstructures.

✉ Fadi Aldakheel  
fadi.aldakheel@ibnm.uni-hannover.de

<sup>1</sup> Institute of Mechanics and Computational Mechanics, Leibniz Universität Hannover, Appelstrasse 9a, 30167 Hannover, Germany

<sup>2</sup> Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA 94720-1740, USA

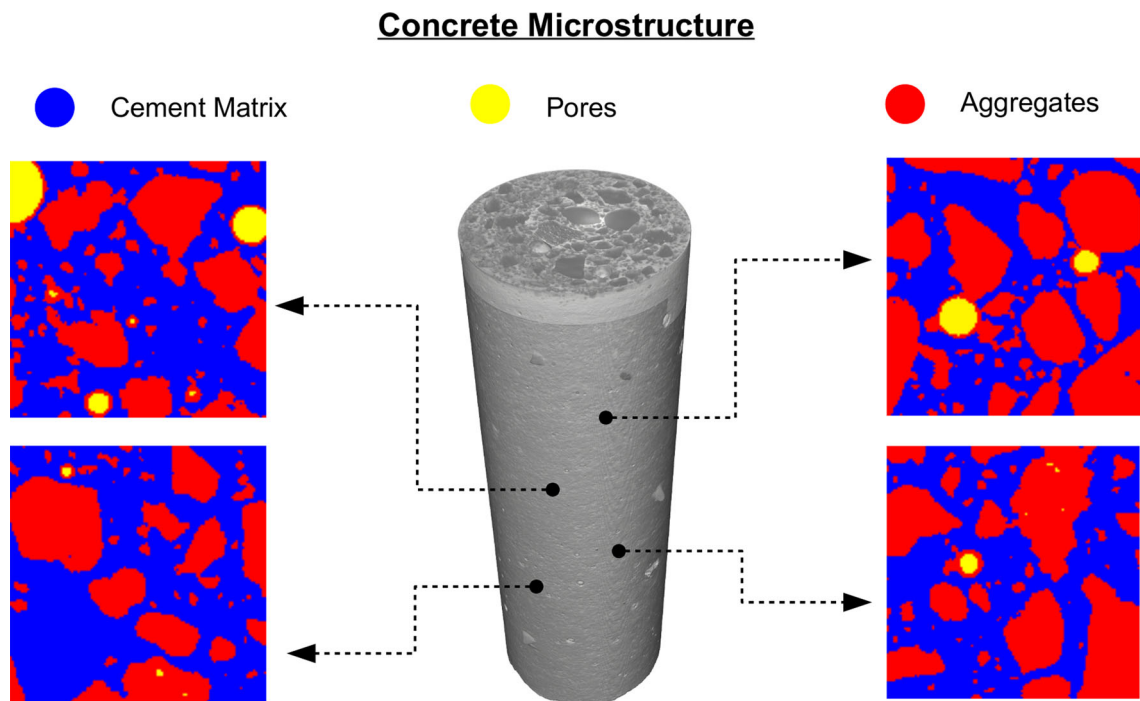
<sup>3</sup> Institute of Continuum Mechanics, Leibniz Universität Hannover, 30823 Garbsen, Germany

These detailed investigations have not been considered so far.

From the modeling point of view, the multiscale approach can be classified in two distinct classes denoted as *concurrent* and *hierarchical* multiscale techniques. These are defined by differentiation of the macro characteristic length scale  $\mathcal{L}_{macro}$  with its micro domain counterpart  $\mathcal{L}_{micro}$ . The concurrent multiscale method implies  $\mathcal{L}_{micro} \equiv \mathcal{L}_{macro}$ , as classified in [33, 35–38]. Whereas in the hierarchical multiscale method, the average size of the heterogeneous microdomain is much smaller than its macro specimen size, i.e.  $\mathcal{L}_{micro} \ll \mathcal{L}_{macro}$ , see [36, 39, 40]. This is often denoted as scale separation law, see computational homogenization approaches based on the Hill-Mandel principle; outlined for instance in [39, 41] among others. The aim of all those mentioned multiscale simulation techniques is the reduction of uncertainties and empirical assumptions while simultaneously increasing the accuracy of the solution. For a better understanding of scale bridging, we consider a real concrete specimen with different microstructures, illustrated in Fig. 1. On the micro-scale, the representative volume elements  $\mathcal{RVE}$ s consist of aggregates, pores and a cement matrix. At the macro-scale, construction applications in building and energy technologies are of interest, including the microscopic material behavior needed to develop the required system-output.

For determining the effective properties of heterogeneous materials at the micro-scale, there exist many approaches. For periodic composites with linear constitutive behavior, those

properties can be analyzed by solving a sufficient number of unit cell problems along with the corresponding boundary conditions. Furthermore, the asymptotic homogenization approach can be considered in this category related to scale separation, as well documented in [42, 43]. In the case of irregular microstructures, the required effective properties cannot be computed exactly. In the literature, the available methods in this category are limited to the computation of upper and lower bounds for the effective stiffness, as outlined in [44, 45]. This was further extended in the work of [46, 47] by considering variational principles, leading to better estimates. [48] developed a self-consistent method by embedding a single inclusion into an infinite domain of the initially unknown effective matrix material. Another important scheme is the well-known two-scale computational homogenization (FE<sup>2</sup>) which determines the effective properties by two nested BVPs (boundary-value-problems) along with the corresponding scale transition law, see for instance [49–61]. In this case, the material behavior at the microscopic level is analyzed by employing the representative volume element concept, whereas a homogenization technique is considered to compute the macroscopic response. We refer to [62–64] for fundamental homogenization principles of local mechanical responses. The above-introduced analytical and physically motivated mathematical models lead to pronounced computational costs. To accelerate micro-macro simulations of materials with complex micro-structures, the



**Fig. 1** Concrete real specimen with different microstructures (CT-images source: [www.baustoff.uni-hannover.de](http://www.baustoff.uni-hannover.de) related to [31–34])

search for advanced technologies that reduce expensive simulation time is critical.

In the present contribution, the focus will be put on the computational micro-to-macro approach of heterogeneous materials where the characteristic length of the microscale is much smaller than its macro counterpart [65–74]. In this contribution, a time-efficient and robust convolutional neural network (CNN) model will be utilized. A composite with 3 different phases is considered within this work (e.g. a hybrid composite laminates consisting of shape memory alloy and glass fiber as reinforced phases and epoxy resin as the host material of the composite; or concrete microstructure consisting of aggregates, voids and cement-paste matrix). The training data is generated by applying standard finite element simulations (FEM) on a cubical sample of a microstructure with different numbers of inclusions/phases. Such geometrical heterogeneities are obtained through random spatial distributions and arbitrary sizes of inclusions/phases, leading to random volume fractions in each microstructure. The network is then trained using the microstructures as an input and the homogenized stresses as labels. As an advantage, the proposed method has the potential for providing accurate and feasible approximations for various engineering applications. To this end, promising results that approach the exact solution are achieved.

The paper is organized as follows: In Sect. 2 a brief overview of the micro-to-macro transition concept is introduced. Next, the theory of the convolutional neural networks (CNNs) is presented in detail in Sect. 3. The CNN model is then employed to predict the homogenized macroscopic stress of a microstructure representing a heterogeneous composite in Sect. 4. The model capability is illustrated through various representative examples in Sect. 5 and compared with the traditional multiscale methods. The trained network is then applied to learn the constitutive behavior of elastic materials within the finite element application. Thereafter, the trained model is used to predict the macroscopic stress and through transfer learning, it is applied to a new structure. Section 6 presents a summary and outlook for extensions of this work.

## 2 Micro-to-macro transition concept

The aim of this section is to give an overview of the multi-scale mechanics of materials. A material point  $\bar{x}$  within the solid  $\bar{B}$  at the macro level is considered. In addition to kinematics and balance relations, given by continuum mechanics, the constitutive behavior is needed to solve a boundary value problem in  $\bar{B}$ . In cases when the material body consists of a heterogeneous microstructure that determines its behavior, a constitutive assumption posed a priori can only provide a coarse estimate of the effective constitutive behavior. The

homogenization approach relies on the assumption that the two considered scales are well separated, i.e. a typical size on the macroscale is much larger than a typical size of the underlying microstructure,  $l_{macro} \gg l_{micro}$  as depicted in Fig. 2. At the microscale, the so-called representative volume element  $\mathcal{RVE}$  models the microstructure that corresponds to the macroscopic material point. Under boundary conditions which are determined by the macroscopic state of deformation, a boundary-value-problem is defined on this representative volume element. At this level, both the governing balance relations and the constitutive behavior are known. The goal is to return this constitutive information from a finer scale to the macro level. The  $\mathcal{RVE}$  acts as a statistically representative portion of the heterogeneous microstructure (grains separated by grain boundary, voids, inclusion, crack, and other similar defects), see [64]. Its size must be chosen such that it is large enough to be representative or rather such that it sufficiently accounts for the character and distribution of heterogeneities. Nevertheless, it should be much smaller than the specimen considered at the macro level to ensure a scale separation on the one hand, and to achieve an increased efficiency on the other hand. If the material or geometric properties of the underlying microstructures vary spatially within a macro specimen, as e.g. in functionally graded materials, the representative volume element can be chosen differently in different macro regions if only a local periodicity is required.

### 2.1 Macroscopic boundary value problem

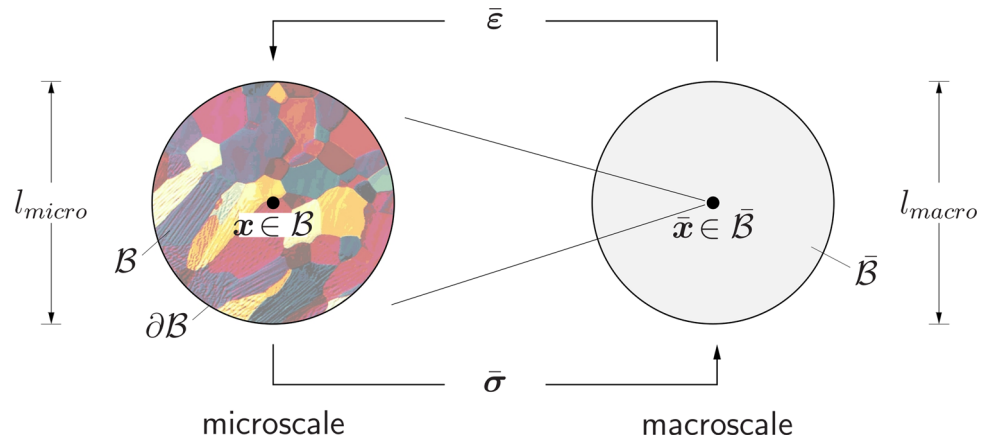
Let  $\bar{B} \subset \mathcal{R}^\delta$  denote a macroscopic body with dimension  $\delta \in [2, 3]$ , as sketched in Fig. 2. We study the mechanical deformation of the body under quasi-static loading in the time interval  $\bar{T} \subset \mathcal{R}_+$ . In what follows,  $\bar{\nabla}(\cdot) := \partial_{\bar{x}}(\cdot)$  and  $(\dot{\cdot}) := \partial_t(\cdot)$  denote the gradient and time derivative of the macroscopic field  $(\cdot)$ , respectively. The primary variable field is the displacement field  $\bar{u}$  of the material point  $\bar{x} \in \bar{B}$  at time  $\bar{t} \in \bar{T}$

$$\bar{u} : \begin{cases} \bar{B} \times \bar{T} \rightarrow \mathcal{R}^\delta \\ (\bar{x}, \bar{t}) \mapsto \bar{u}(\bar{x}, \bar{t}) \end{cases} \quad (1)$$

The kinematic relation arising from the considered framework is the linear strain tensor  $\bar{\epsilon} := \bar{\nabla}_s \bar{u} = \frac{1}{2}(\bar{\nabla} \bar{u} + \bar{\nabla}^T \bar{u})$ . The strains are assumed to be small, i.e.  $|\bar{\nabla} \bar{u}| < \bar{\epsilon}$  is bounded by a small number  $\bar{\epsilon}$ .

In what follows we assume the existence of a macroscopic potential density functional  $\bar{\Pi}$ , which is obtained by a variational principle of homogenization for the underlying microstructure. With this assumption, we can formulate a variational structure resulting with the minimization problem for the determination of macroscopic primary variable

**Fig. 2** Computational micro-to-macro transition approach of heterogeneous materials



$$\{\bar{\mathbf{u}}\} = \text{Arg} \left\{ \inf_{\bar{\mathbf{u}}} \bar{\Pi}(\bar{\mathbf{u}}) \right\}, \tag{2}$$

where  $\bar{\Pi}$  is defined as

$$\bar{\Pi}(\bar{\mathbf{u}}) = \int_{\bar{\mathcal{B}}} \bar{\Psi}(\bar{\boldsymbol{\varepsilon}}) dV - \bar{\mathcal{P}}_{ext}(\bar{\mathbf{u}}), \tag{3}$$

hereby, the potential energy  $\bar{\Psi}$  is fully determined at the microscopic  $\mathcal{RVE}$  by a variational formulation principle and the mechanical loading contribution

$$\bar{\mathcal{P}}_{ext}(\bar{\mathbf{u}}) = \int_{\bar{\mathcal{B}}} \bar{\boldsymbol{\gamma}} \cdot \bar{\mathbf{u}} dV + \int_{\partial \bar{\mathcal{B}}_T} \bar{\mathbf{t}}_N \cdot \bar{\mathbf{u}} dA. \tag{4}$$

The given macroscopic body force per unit volume is introduced as  $\bar{\boldsymbol{\gamma}}$ , whereas the tractions on the Neumann boundary  $\partial \bar{\mathcal{B}}_T$  are depicted as  $\bar{\mathbf{t}}_N$ . Furthermore, consider a decomposition of the surface  $\partial \bar{\mathcal{B}} = \partial \bar{\mathcal{B}}_{\bar{\mathbf{u}}} \cup \partial \bar{\mathcal{B}}_T$  into a part  $\partial \bar{\mathcal{B}}_{\bar{\mathbf{u}}}$  where the displacements are prescribed and a part  $\partial \bar{\mathcal{B}}_T$  with given tractions, along with  $\partial \bar{\mathcal{B}}_{\bar{\mathbf{u}}} \cap \partial \bar{\mathcal{B}}_T = \emptyset$ .

The necessary condition of (2) results in the equilibrium equations describing the macroscopic problem for the quasi-static case under consideration. The Euler–Lagrange equation is the balance of linear momentum

$$\text{div}[\bar{\boldsymbol{\sigma}}] + \bar{\boldsymbol{\gamma}} = \mathbf{0} \quad \text{in } \bar{\mathcal{B}}, \tag{5}$$

along with the Neumann-type boundary conditions  $\bar{\boldsymbol{\sigma}} \cdot \bar{\mathbf{n}} = \bar{\mathbf{t}}_N$  on  $\partial \bar{\mathcal{B}}_T$ . The macroscopic stresses  $\bar{\boldsymbol{\sigma}}$  are

$$\bar{\boldsymbol{\sigma}} = \partial_{\bar{\boldsymbol{\varepsilon}}} \bar{\Psi}(\bar{\boldsymbol{\varepsilon}}), \tag{6}$$

governed by the macroscopic potential energy function  $\bar{\Psi}$  obtained by the homogenization of the microstructure.

### 2.2 Microscopic boundary value problem

Let  $\mathcal{B} \subset \mathcal{R}^d$  denote a periodic microstructure ( $\mathcal{RVE}$ ) as depicted in Fig. 2. In what follows,  $\nabla(\cdot) := \partial_x(\cdot)$  and

$\dot{(\cdot)} := \partial_t(\cdot)$  denote the gradient and the time derivative of the microscopic field  $(\cdot)$ , respectively. The primary variable field is the displacement field  $\mathbf{u}$  of the material point  $\mathbf{x} \in \mathcal{B}$  at time  $t \in \mathcal{T}$

$$\mathbf{u} : \begin{cases} \mathcal{B} \times \mathcal{T} \rightarrow \mathcal{R}^d \\ (\mathbf{x}, t) \mapsto \mathbf{u}(\mathbf{x}, t) \end{cases} \tag{7}$$

The microscopic linear strain tensor  $\boldsymbol{\varepsilon}$  is the symmetric part of the displacement gradient  $\boldsymbol{\varepsilon} := \nabla_s \mathbf{u} = \frac{1}{2}(\nabla \mathbf{u} + \nabla^T \mathbf{u})$ . We now postulate a variational principle of homogenization that determines the macroscopic potential energy  $\bar{\Psi}$  introduced above, as follows

$$\bar{\Psi}(\bar{\boldsymbol{\varepsilon}}) = \inf_{\boldsymbol{\varepsilon}} \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} \Psi(\boldsymbol{\varepsilon}) dV. \tag{8}$$

This definition is conceptually in line with the formulations outlined in [52]. At the micro-level, we identify the stresses as

$$\boldsymbol{\sigma}(\boldsymbol{\varepsilon}) := \partial_{\boldsymbol{\varepsilon}} \Psi(\boldsymbol{\varepsilon}), \tag{9}$$

governed by the constitutive functions  $\Psi$ . With this definition at hand, the variation of the principle (8) gives the condition

$$\begin{aligned} & \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} (-\text{div}[\partial_{\boldsymbol{\varepsilon}} \Psi]) \cdot \delta \dot{\mathbf{u}} dV \\ & + \frac{1}{|\mathcal{B}|} \int_{\partial \mathcal{B}} (\partial_{\boldsymbol{\varepsilon}} \Psi \cdot \mathbf{n}) \cdot \delta \dot{\mathbf{u}} dA = 0. \end{aligned} \tag{10}$$

Hence, the Euler–Lagrange equation for the variational principle (8) is

$$\text{div}[\boldsymbol{\sigma}] = \mathbf{0} \quad \text{in } \mathcal{B} \tag{11}$$

on the microstructure  $\mathcal{B}$ .

### 2.3 Homogenization quantities and macro-homogeneity conditions

The microscopic boundary-value-problem is linked to the macroscopic framework by a scale bridging approach accounting for homogenized quantities defined on the boundary/volume of the representative volume element. The macroscopic strain  $\bar{\boldsymbol{\epsilon}}$  and microscopic strain  $\boldsymbol{\epsilon}$  along with their associated work-conjugate stress measures  $\bar{\boldsymbol{\sigma}}$  and  $\boldsymbol{\sigma}$  are defined as

$$\begin{aligned} \bar{\boldsymbol{\epsilon}} &:= \frac{1}{|\mathcal{B}|} \int_{\partial\mathcal{B}} \text{sym}(\mathbf{u} \otimes \mathbf{n}) \, dA = \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} \boldsymbol{\epsilon} \, dV \quad \text{and} \\ \bar{\boldsymbol{\sigma}} &:= \frac{1}{|\mathcal{B}|} \int_{\partial\mathcal{B}} \text{sym}(\mathbf{t}_N \otimes \mathbf{x}) \, dA = \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} \boldsymbol{\sigma} \, dV, \end{aligned} \quad (12)$$

where  $|\mathcal{B}|$  is the volume of the micro-structure  $\mathcal{B}$ . In order to derive appropriate boundary conditions for the  $\mathcal{RVE}$ , we consider a *Hill-Mandel macro-homogeneity condition* proposed by [62] stating the equivalence of the macroscopic stress power with the volume average of its microscopic counterpart

$$\bar{\boldsymbol{\sigma}} : \dot{\bar{\boldsymbol{\epsilon}}} = \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} \boldsymbol{\sigma} : \dot{\boldsymbol{\epsilon}} \, dV = \frac{1}{|\mathcal{B}|} \int_{\partial\mathcal{B}} \mathbf{t}_N \cdot \dot{\mathbf{u}} \, dA. \quad (13)$$

To solve the microscopic boundary-value-problem we need to set up appropriate boundary conditions for the  $\mathcal{RVE}$ . For the micro-elastic model under consideration we focus on three types of boundary constraints:

- (i) *Dirichlet-type boundary conditions*: linear displacement constraint on the boundary  $\partial\mathcal{B}$ . In the first approach, a homogeneous strain is imposed in the full microstructure, i.e.  $\boldsymbol{\epsilon} = \bar{\boldsymbol{\epsilon}}$  in  $\mathcal{B}$ . Hence, the microscopic strain is identical to the macro-strain at any point  $\mathbf{x} \in \mathcal{B}$  of the microstructure. This is called *Voigt–Taylor assumption* referring to [44] and yields an upper bound of the stiffness of the  $\mathcal{RVE}$ .
- (ii) *Neumann-type boundary conditions*: constant stress on the boundary  $\partial\mathcal{B}$ , zero micro-tractions. The so-called *Reuss–Sachs bound*, [45], gives a lower bound of the stiffness of the microstructure. Here, a homogeneous stress is applied in the full domain  $\mathcal{B}$  in the form  $\boldsymbol{\sigma} = \bar{\boldsymbol{\sigma}}$ .
- (iii) *Periodic boundary conditions*: periodicity of all primary fields on opposite surfaces  $\partial\mathcal{B}^+$  and  $\partial\mathcal{B}^-$  of the microstructure.

The upper Voigt–Taylor and lower Reuss–Sachs bounds are only mentioned for completeness and do not play a crucial role in the subsequent treatment of the boundary conditions. The periodic constraints (iii) are only applicable for perfectly periodic microstructures, but for many cases it turns

out that they provide better results than either the Dirichlet-type constraints (i) or the Neumann-type conditions (ii). Here, we refer to the work of [74] and the references cited therein.

For the periodic boundary conditions plotted in Fig. 3, the surface of the  $\mathcal{RVE}$  decomposes into two parts  $\partial\mathcal{B} = \partial\mathcal{B}^+ \cup \partial\mathcal{B}^-$  with normals  $\mathbf{n}^+$  and  $\mathbf{n}^- = -\mathbf{n}^+$  at associated points  $\mathbf{x}^+ \in \partial\mathcal{B}^+$  and  $\mathbf{x}^- \in \partial\mathcal{B}^-$ . The deformation will be extended by a fine scale fluctuation field marked in what follows with a tilde ( $\tilde{\cdot}$ )

$$\mathbf{u} = \bar{\boldsymbol{\epsilon}} \cdot \mathbf{x} + \tilde{\mathbf{u}}, \quad (14)$$

around the macro-modes. As an additional constraint for the fluctuation fields, the Hill condition reads

$$\frac{1}{|\mathcal{B}|} \int_{\partial\mathcal{B}} \mathbf{t}_N \cdot \dot{\tilde{\mathbf{u}}} \, dA = 0. \quad (15)$$

This constraint can be satisfied for *periodic fluctuations* and *anti-periodic tractions* at the boundary  $\partial\mathcal{B}$  resulting in

$$\tilde{\mathbf{u}}^+ = \tilde{\mathbf{u}}^- \quad \text{and} \quad \mathbf{t}_N^+ = -\mathbf{t}_N^- \quad \text{on} \quad \partial\mathcal{B}, \quad (16)$$

at associated points  $\mathbf{x}^+ \in \partial\mathcal{B}^+$  and  $\mathbf{x}^- \in \partial\mathcal{B}^-$ . Such boundary conditions are the most reasonable choice for the homogenization analysis under consideration, even for microstructures which are non-periodic.

As a *summary of the computational framework*:

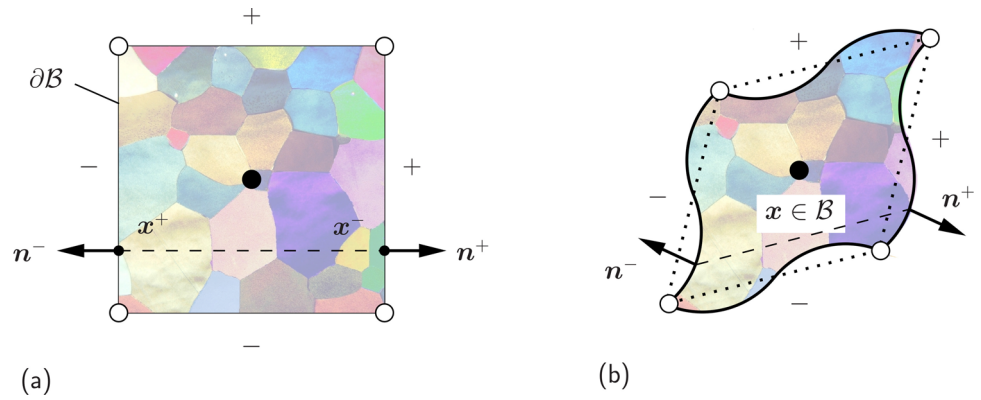
1. Evaluate the macroscopic deformation at each material point.
2. The necessary boundary conditions are applied at the microscopic  $\mathcal{RVE}$ .
3. The microscopic boundary value problem is computed under the macroscopic loading.
4. The volumetric averaged microscopic quantities are transferred to the dedicated material macroscopic points.
5. The macroscopic boundary value problem is solved.

For further details on computational homogenization method for different engineering applications, the interested reader is referred to the references in the introduction part of this work.

### 3 Theory of convolutional neural network (CNN)

To overcome the issue of large computational costs to solve the multiscale problem on the micro-to-macro level, summarized in Sect. 2, an approach to replace the calculation with a

**Fig. 3** Periodic microstructure, where the surface of the  $\mathcal{RVE}$  in **a** decomposes into two parts  $\partial\mathcal{B} = \partial\mathcal{B}^+ \cup \partial\mathcal{B}^-$  with normals  $\mathbf{n}^+$  and  $\mathbf{n}^- = -\mathbf{n}^+$  at associated points  $\mathbf{x}^+ \in \partial\mathcal{B}^+$  and  $\mathbf{x}^- \in \partial\mathcal{B}^-$ . **b** Periodic boundary conditions for the displacement



machine learning model is introduced. Applying deep learning (DL) is an empirical, highly iterative process that requires training several models in order to reach satisfactory results. During this process, a combination of different parameters and hyper-parameters is tested.

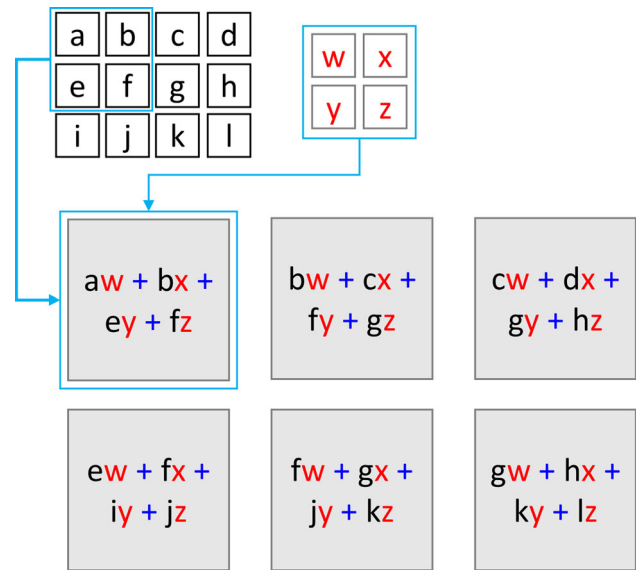
The purpose of this section is to give a brief insight into the structure and operation of the Convolutional Neural Networks (CNN), without going into specific variants and manifestations of this technique. CNN is a kind of Artificial Neural Network and is associated with deep learning techniques. More specific, the CNN model is a specialized Feed Forward Neural Network (FFNN) that excels at working with data that have a grid like structure, such as images or time-series data. CNN has success at computer vision applications, such as image classification, object detection, text detection and recognition, see [75]. The weight-sharing techniques achieved by convolutional and pooling layers are the main differentiator between CNN and FFNN. Those two essential operations, along with other deep learning functions will be explained next.

### 3.1 The convolution operation

The convolution operation could be defined as a linear multiplication between an input and a set of weights. In this work, the input is a gray-scale image and the set of weights is a matrix, known as a **filter** or **kernel**. Multiplying the filter with the input image at different points allows it to capture the same feature regardless of its location in the image. The output of the convolutional operation is usually referred to as a **feature map**. For an image the feature value at a position  $(i, j)$  in a feature map is given by the equation

$$z_{i,j} = \sum_m \sum_n I_{i+m,j+n} k_{m,n}, \tag{17}$$

where  $I$  is the image, and  $k$  is the applied filter with  $m, n$  dimensions. This convolution operation is visualized in Fig. 4. Nonlinearity is introduced to each feature value in the feature map element-wise by the activation function



**Fig. 4** An example of a convolution operation

$$a_{i,j} = f(z_{i,j}), \tag{18}$$

Note that, the activation functions, also known as transfer functions, have a big impact on the performance of deep learning models. They introduce nonlinearity to computations done in the neuron and determine whether or not to activate it. Hereby, we employed two different non-linear activation functions, namely the **Rectified linear unit (ReLU)** [76] and **Swish function** [77]. As they both produced similar results we selected ReLU in this contribution. ReLU is considered to be the default activation function in modern neural networks architecture, usually used in the hidden layers with another activation function at the output layer. As it represents an almost linear function, it is computationally efficient while allowing for backpropagation. ReLU can be represented mathematically as

$$f(z) = \max(0, z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0. \end{cases} \tag{19}$$

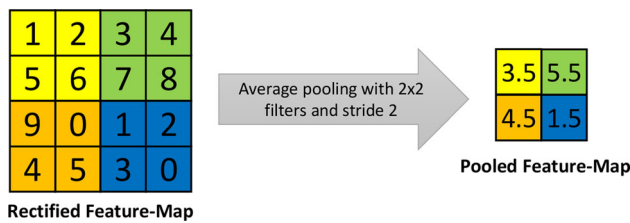


Fig. 5 Example of Average Pooling operation

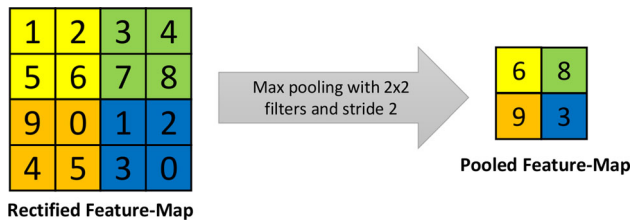


Fig. 6 Example of Max Pooling operation

### 3.2 Pooling layers

Another main component of convolutional neural networks is pooling layers, usually placed between two convolutional layers. They reduce the dimensions of the feature map resulting from the convolution operation, thus reducing the learning parameters as well as the computational cost and speeding up the training process. The pooling layer does not have learning parameters, but the size of the pooling window and the type of the pooling function performed on the pixels in this window are hyper-parameters that need to be tuned during the training process. There are many kinds of pooling operations used in CNN models, max pooling and average pooling being the most commonly used ones. In an average pooling operation, the average value of the pixels in the pooling window is calculated as sketched in Fig. 5, while in the max pooling operation, the highest pixel value in the pooling window is considered, as shown in Fig. 6.

### 3.3 Loss function

The optimum parameters for the CNN model are found by optimization of an objective function. The **loss function** is a term referring to the objective function which is the case for an optimization process using minimization. A loss function measures the error between the model prediction and the ground truth. Thus, it indicates how well the model is performing and should be able to represent this error. That's why different predicting problems require different types of loss functions. For example, in the case of binary classification, binary cross-entropy is a suitable choice. In the case of regression prediction problems, an appropriate loss function

could be square error loss or absolute error loss. The loss function is known as the cost function when applied to the whole data set. In this contribution, we use the mean squared error loss (MSE). It is the sum of squared differences between predicted and ground truth values, which are expressed by the equation

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \tag{20}$$

where  $N$  is the number of training examples,  $y_i$  is the ground truth, and  $\hat{y}_i$  are the predicted values.

A back-propagation and optimization algorithm are considered in this work. Here, gradient descent is one of the most commonly used optimization algorithms in neural networks that minimize a loss function. Furthermore, the adaptive momentum estimation (Adam) has shown successful results while dealing with learning rate problems occurring with the adaptive gradient algorithm (Adagrad).

## 4 CNN aided multiscale mechanics of materials

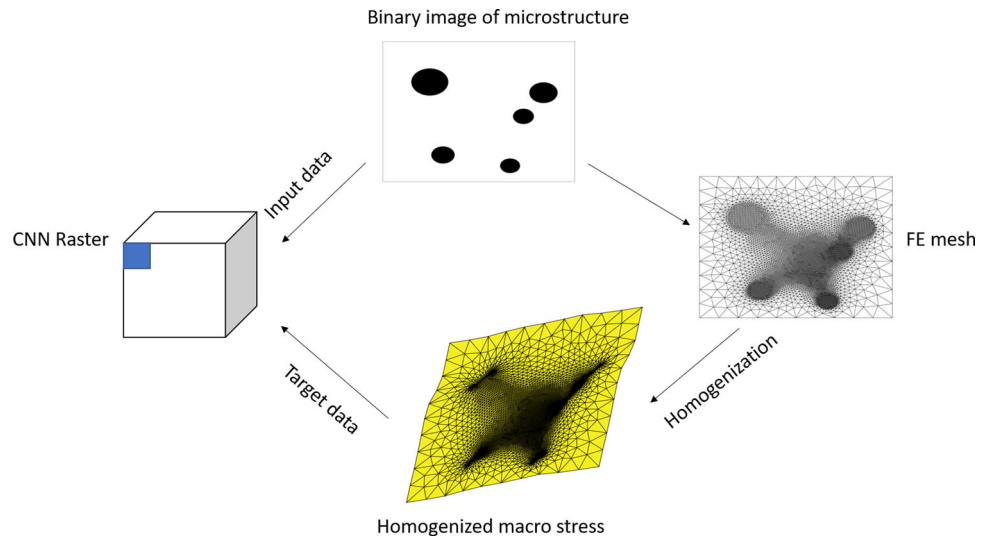
In this section, a convolutional neural network model is used to predict the homogenized macroscopic stress  $\bar{\sigma}$  of a microstructure, which represents a heterogeneous composite, consisting of three-phases, under a predefined macroscopic strain  $\bar{\epsilon}$ . The proposed model falls under the *supervised-learning* category, i.e. the data used for the training and testing processes are labeled. The data generation process, the design of the CNN model architecture, and the results will be discussed in the following subsections.

### 4.1 Dataset generation

The macroscopic performance of materials often relies on their microscopic structures, which can not be seen by the naked eye. In recent years, actual complex microstructures can be captured with high resolution using modern non-destructive imaging techniques such as a micro-computed tomography scan  $\mu$ -CT or scanning electron microscopes. In this work, the microstructures are synthetically generated. A reinforced composite that consists of three different materials is considered (e.g. a hybrid composite laminates consisting of shape memory alloy and glass fibers as reinforced phases and epoxy resin as host material of the composite; or a concrete microstructure consisting of aggregates, voids and cement-paste matrix), see [34, 78, 79]. The workflow of generating the dataset is depicted in Fig. 7.

The generated dataset consists of 10800 square-shaped  $\mathcal{RVE}$ s with a side length of 1 mm. It is divided into 11

**Fig. 7** The work-flow for data-generation process



groups of 900  $\mathcal{RVE}$ s depending on various volume fractions and inclusions. Each  $\mathcal{RVE}$  contains a different number of inclusions with circular and ellipsoidal shapes that are randomly distributed: The circles vary in radius between 0.10–0.15 mm, whereas the ellipsoids have a small radius of 0.008 mm and a maximum radius of 0.08 mm; resulting in different volume fractions ranging from a minimum of 10% to a maximum of 35%.

## 4.2 Images of the microstructure

These  $\mathcal{RVE}$ s are randomly generated using various center-coordinates  $(x, y)$  and radius  $(r)$  of the circles and ellipsoids that represent the inclusions in a certain predefined bound, which represents the edges of the  $\mathcal{RVE}$ . The resulting geometrical information is then used to produce gray-scale images, where the inclusions are assigned a pixel value of 128 for ellipsoids and 255 for the circular inclusions, and the matrix is assigned a pixel value of 0. The dimensions of the images are  $(256, 256, 1)$ , where the first two numbers represent height and width and the last number represents the color channels. These gray-scale images are the input data of the CNN model. Figure 8 depicts representative volume elements ( $\mathcal{RVE}$ ) of the microstructure with various volume fraction distributions. To demonstrate the efficiency of the proposed CNN model, we transfer the learning model to a new  $\mathcal{RVE}$  structure and compare the results and computation costs with learning from scratch. For that reason, we started with simple two-phases microstructure as plotted in Fig. 9, then we test a complex structure at the end of this work. For the construction of those artificial microstructures, we refer to our work [34] related to the random allocation of the heterogeneities.

**Table 1** Material proprieties of  $\mathcal{RVE}$  constituents

Materials	Young's modulus (GPa)	Poisson's ratio
Matrix	3	0.38
Phase I (Ellipsoid)	70	0.20
Phase II (Circles)	50	0.35

## 4.3 Calculation of macroscopic stresses

Acquiring the labels for each one of the generated  $\mathcal{RVE}$  is the second step of generating the dataset, on which the proposed CNN model will be trained and tested. These labels are the components of the macroscopic stress  $\bar{\sigma}$

$$\bar{\sigma} = \hat{y} = [\bar{\sigma}_{11} \quad \bar{\sigma}_{22} \quad \bar{\sigma}_{12}]^T. \quad (21)$$

The computational multiscale framework is applied using the standard finite element method (FEM) as outlined in [80]. In this contribution, the generation of the idealized microstructures was done in Matlab and Gmsh, see [34]. Then the meshed microstructure was exported to Abaqus for solving the boundary value problem of the microstructure. The computations are carried out under the plane strain assumption while imposing periodic deformation and antiperiodic traction, which was introduced in Sect. 2. The composite consists of isotropic linear elastic constituents with properties shown in Table 1.

The micro-to-macro approach is based on a deformation-driven microstructure with a *constant* macroscopic strain. The dataset is then divided into three subsets training, validation, and testset 8640:1080:1080. The model uses the training set to learn the parameters (weights, biases). The validation set is used to fine-tune the model's hyper-parameters (learning rate, number of hidden layers,...) and as an indicator of



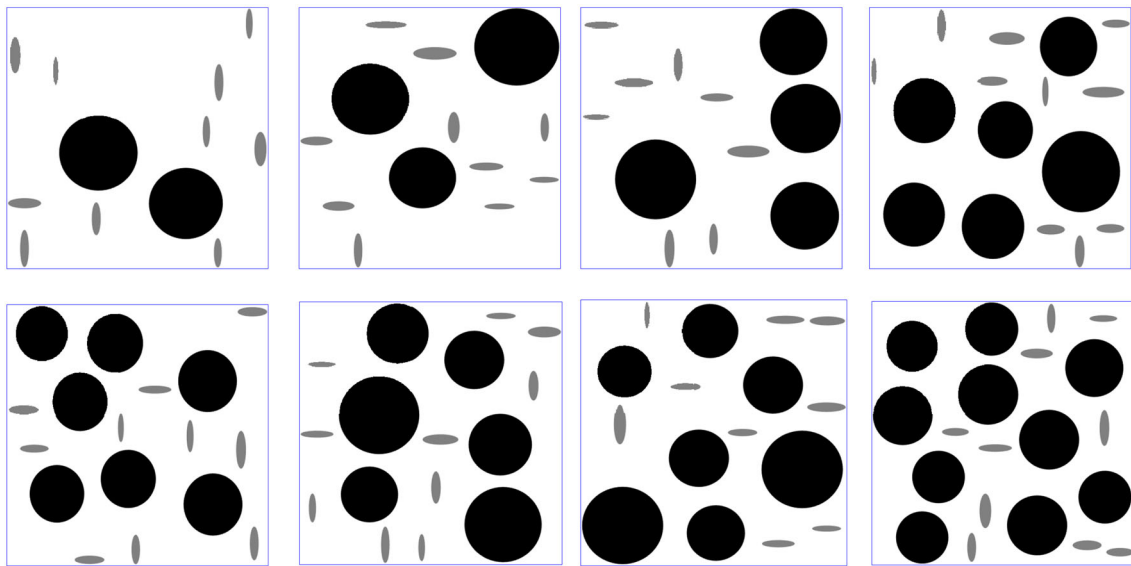


Fig. 8 Data set generation: Images of the idealized microstructure

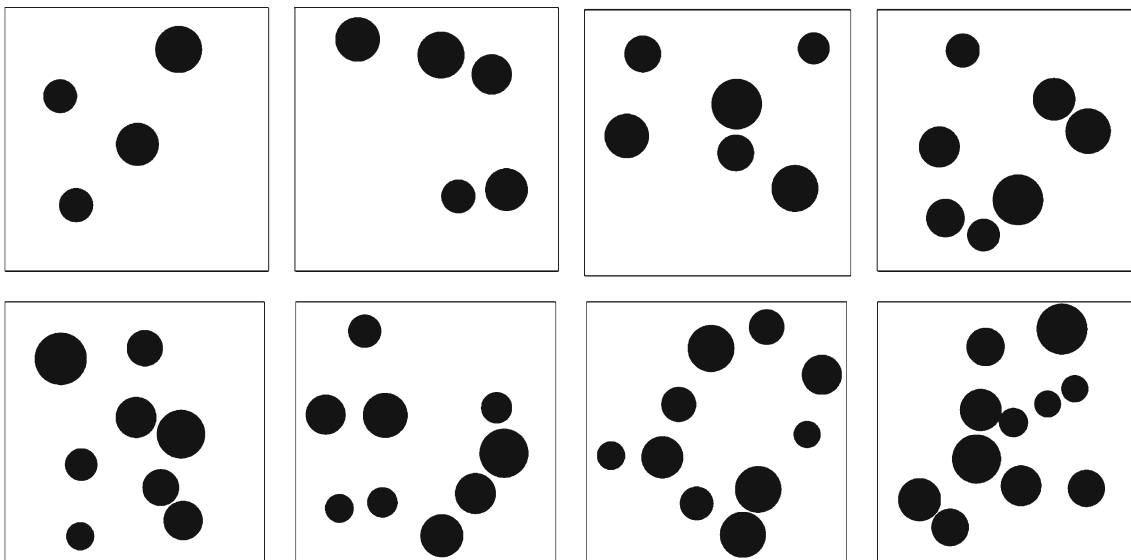


Fig. 9 Data set generation: transfer learning

over-fitting. Finally, the test set gives an unbiased estimation of how well the model generalizes on cases that it has not encountered before. Normalizing the input features and the output of neural networks is usually done to stabilize and speed up the training process. One of the common ways is to rescale the data to be in a range between [0, 1]. In the case of the presented dataset, the input data are already normalized in the range [0, 1], the output data has noticeable differences in values between the first two components, which represent the normal stresses, and the third one representing the shear stress, so they need to be scaled separately through the following equation

$$\bar{y} = \frac{y - \min(y)}{\max(y) - \min(y)}, \tag{22}$$

with the normalized  $\bar{y}$  components of  $y$ .

Neural networks are known for their strong ability to memorize the training data examples, which leads to over-fitting, especially in the cases of small datasets. To overcome this problem, several regularization techniques are applied, such as ( $L_2$  regularization, dropout layers, and early stopping).  $L_2$  regularization, also known as weight decay regularization, imposes a squared  $L_2$  norm on the weights. The regularized cost function in this case is

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^N w_i^2, \quad (23)$$

where  $\lambda$  is a regularization parameter. Early stopping is another regularization technique, which works by stopping the training process after a predefined number of epochs (once the error of the validation set starts to increase, as it is a sign of over-fitting). Then, the parameters of the model with the best performance on the validation set are restored. Dropout regularization works by randomly deactivating a certain number of the nodes in the fully connected layers at each iteration during the training process.

#### 4.4 CNN-model architecture

Applying Deep learning is a highly iterative process, in which the first step is to define an idea for solving the task at hand (such as deciding the model type and structure). The second step is to implement this idea and the last step is to experiment and see if the results are satisfactory. If this is not the case one has to iterate until reaching the desired results. Although having good domain knowledge helps to go through this cycle efficiently, it is very unlikely to get the right parameters without iterations.

Transfer learning, where a model that works well on the same task already exists, is used directly or retrained to tune the hyper-parameters. This is one method to speed up the model design process and will be investigated in this work. Another method is to borrow ideas from models that proved to be successful in similar tasks. The proposed model architecture is implemented using Keras 2.7 and Python 3.7.12. The training process was done using Google virtual machines through Google-COLAB, the machine was equipped with a NVIDIA Tesla P100-PCIE-GBU with 16 GB memory.

Usually, CNN models are applied in fields like image classification or object detection, but in this application, it is used for a regression task. That's why the mean square error (MSE) between the CNN predictions and the ground truth (results from the finite element simulations) is chosen as a cost function. CNN models consist of convolutional, pooling, and fully connected layers. The choice of these layer's parameters decides the capacity and the power of representation of the model. To have a sufficient representation, different filter sizes up to (5,5) are chosen through the convolutional layers for the evaluated model. Furthermore, a stride of 1 and implicit zero padding (same padding) is applied to reduce the effect of narrower output dimension resulting from the convolution operation. As regularization techniques,  $L_2$  regularization with a factor of 0.01, and early stop with 40 epochs as a predefined number to stop the training if the validation set error does not decrease, are considered. The activation function used through all the layers is ReLU due to its com-

**Table 2** MAPE and  $R^2$  values on the test dataset

Stress components	$\bar{\sigma}_{11}$	$\bar{\sigma}_{22}$	$\bar{\sigma}_{12}$
MAPE	0.87%	0.87%	0.36%
$R^2$	0.98	0.98	0.91

putational efficiency. The Adam optimizer with a learning rate of 0.01 is chosen for training the model while applying a learning rate decay factor of 0.1 if the validation error does not decrease for 20 consecutive epochs.

As a starting point of this work, we employed already existing CNNs with deeper network architectures [27–29] furthermore, we utilize additional operations, including skip connections in e.g. ResNet50, DenseNet, and MobileNetV2, filter concatenation in InceptionV3, and other operations like squeeze and excitation blocks in MobileNetV3 (consume more time in contrast to AlexNet with only convolution and pooling operations).

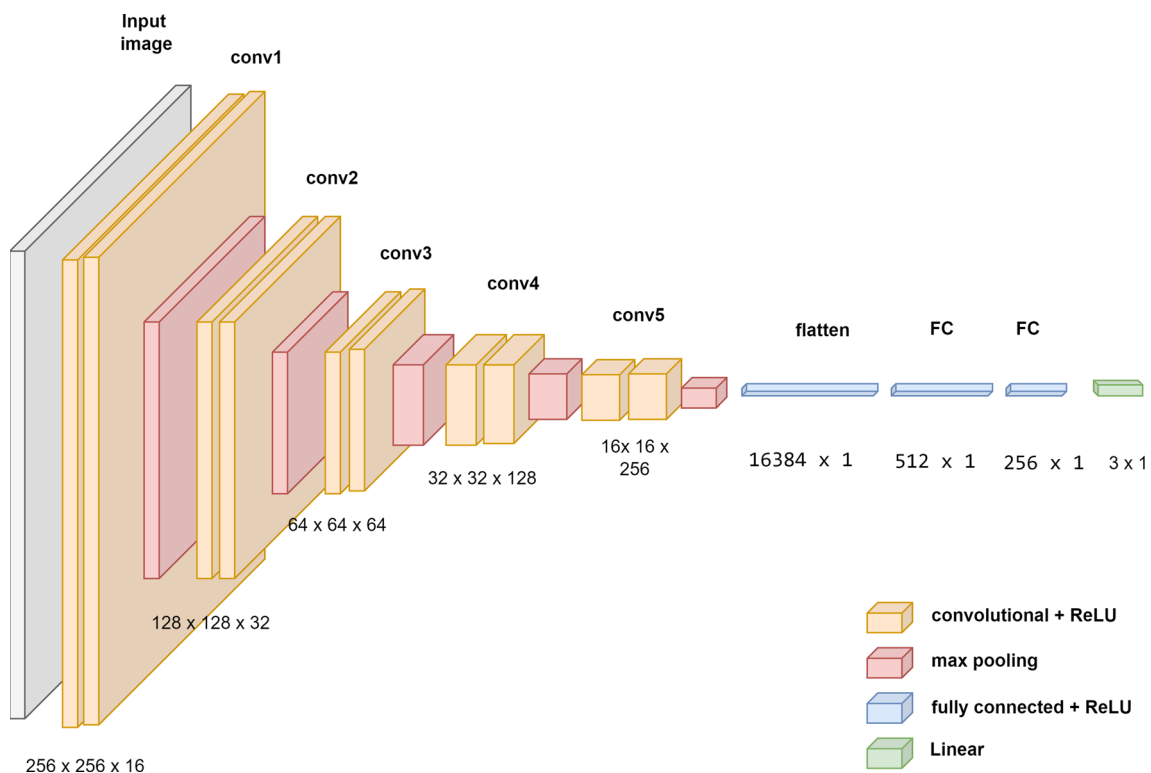
Deeper architectures were designed to handle complex features in a dataset containing several hundred class labels. However, in our study, the dataset contains fewer variations in pixel intensities between samples and no color channels were involved. Therefore a simpler architectural construction for the material property prediction would be appropriate. Hence, it is valuable to develop a model from scratch based on the insights we gathered from initial experimentation, which will certainly add concrete advantage in comparing the performance of different models for the dataset and application used in the study. The proposed model architecture is sketched in Fig. 10.

Neural networks are regarded as black boxes, as it is hard to understand the reasoning behind the specific decisions they make and the values of the weights they choose. The filters in the convolutional neural networks are weights that the network learns. Since CNN works with images these filters allow for visualizing the feature maps, which could lead to a better understanding of the features that the model learns. Figure 11 shows the different feature maps through the model. It can be seen that they represent most of the details of the microstructure, but as they go deeper, the learned features seem to be more abstract and hard to be interpreted.

## 5 Results and discussion

### 5.1 Three-phases microstructure

The first model problem is concerned with predicting the macroscopic stresses of the three-phases idealized microstructures introduced in Fig. 8 with various volume fractions. The details of the CNN-model along with the required data for the



**Fig. 10** CNN-model architecture for predicting homogenized stress

**Table 3** Average CPU time needed for computing the homogenized stress of  $\mathcal{RV}\mathcal{E}$ s containing different numbers of inclusions

Volume Fraction	FEM-Model (s)	CNN-Model (s)
12.6	82.0	0.07
20.7	98.0	0.07
25.4	152.0	0.07
31.8	200.0	0.07

FEM-model are described in Sect. 4. The CNN model will be evaluated in an unbiased way using a test dataset consisting of 1080  $\mathcal{RV}\mathcal{E}$ s, which was randomly picked from the whole dataset. The model predictions are shown in Fig. 12 as a scatter plot, where the red-line represents the actual values obtained by finite element simulations (FEM-model). The figure shows the ability of the model to successfully predict the three components of the homogenized stresses.

The model performance is measured in a quantitative way by the mean absolute percentage error (MAPE) for each component of the homogenized stresses. The MAPE is defined as

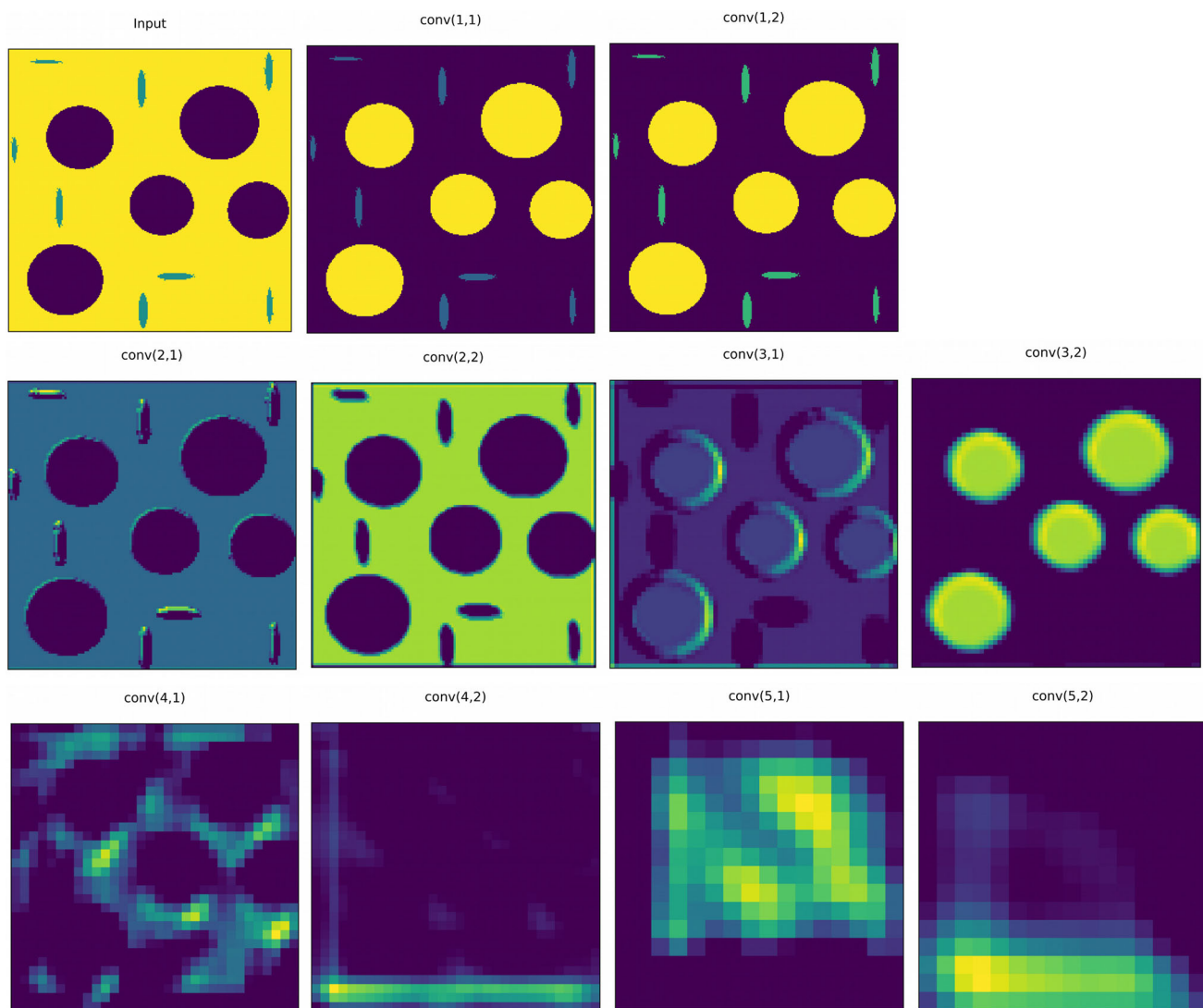
$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{\hat{y}_j - y_j}{y_j} \right|, \tag{24}$$

where  $\hat{y}_j$  is the predicted value and  $y_j$  is the actual value obtained by FEM simulations. Another method to evaluate the performance of the model is the  $R^2$ -score, also known as the coefficient of determination, which is a statistical measure that shows how well the CNN-model approximates the actual data.  $R^2$ -score usually has a value in range between 0 and 1, defined as

$$R^2 = 1 - \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y}_j)^2}, \tag{25}$$

in terms of the mean value  $\bar{y}_j$ , where the values closer to 1 represent a model with better performance. The MAPE and  $R^2$ -score values for each component of the homogenized stresses are given in Table 2. The results of the proposed CNN-model are very promising with accurate macroscopic stress prediction, that have a good MAPE and coefficient of determination close to 1.

One of the main aspects to consider implementing deep learning in the mechanical field is the high computational efficiency, thus the ability to speed up the design process. Recent advancements in the computational power, specially the GPU parallelization ability, allow shorter training times for deep learning models. In this work, the training process of the model took around 30.0 min for a total of 300 epochs. Once the model is trained, the CNN-model advantages start



**Fig. 11** Visualisation of the input image (three-phases idealized microstructure) and the feature maps learned through the convolutional layers of the CNN-model

to kick in. It can be seen in Table 3 that the computation time of the standard finite element method increases with the increase of the volume fraction of inclusions (increase number of inclusions). This is due to the increasing number of elements needed to accurately compute the homogenized stresses. On the other hand, the increase in number of inclusions or volume fraction does not affect the prediction time of the CNN model, see Table 3. The CNN-model is 1171 times faster than the FEM-model in the case of a  $\mathcal{RVE}$  with 12.6 volume fraction of inclusions, and up to 2857 times in the case of computationally demanding  $\mathcal{RVE}$ s with volume fraction of 31.8. The computations shown in Table 3 were executed on an Intel(R) Core(TM) i7-8750H CPU @ 2.2 GHZ and 16 GB of RAM.

## 5.2 Transfer learning model versus training from scratch model

Finally, we demonstrate the capacity of the CNN-model further by transferring the trained model of Sect. 5.1 to a new geometry (data). The goal here is to predict the macroscopic stresses using the best trained model and compare that with a training from scratch model (same procedure as described in the previous example). To this end, we used less number of data (new microstructures) as sketched in Fig. 9 and only 100 epochs were considered in the learning process, to demonstrate the CNN transfer-learning efficiency.

As a first comparison, we plot the macroscopic stresses predictions of both models in Fig. 13 together with the actual values obtained by FEM-model. Hereby, the transfer learning model shows a better performance close to the FEM

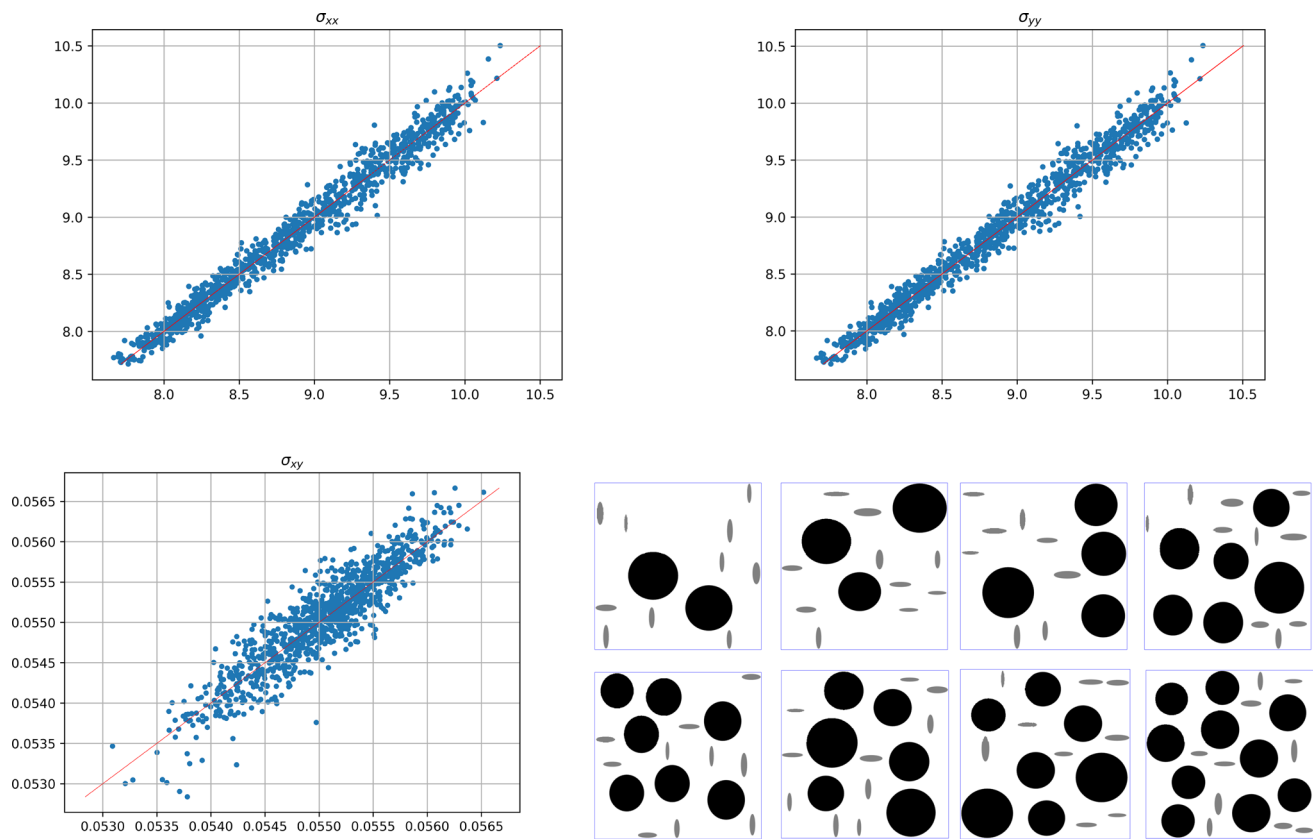


Fig. 12 CNN-predictions (y-axis) versus ground-truth values (FEM x-axis) of the stress components:  $\bar{\sigma}_{xx}$ ,  $\bar{\sigma}_{yy}$ ,  $\bar{\sigma}_{xy}$

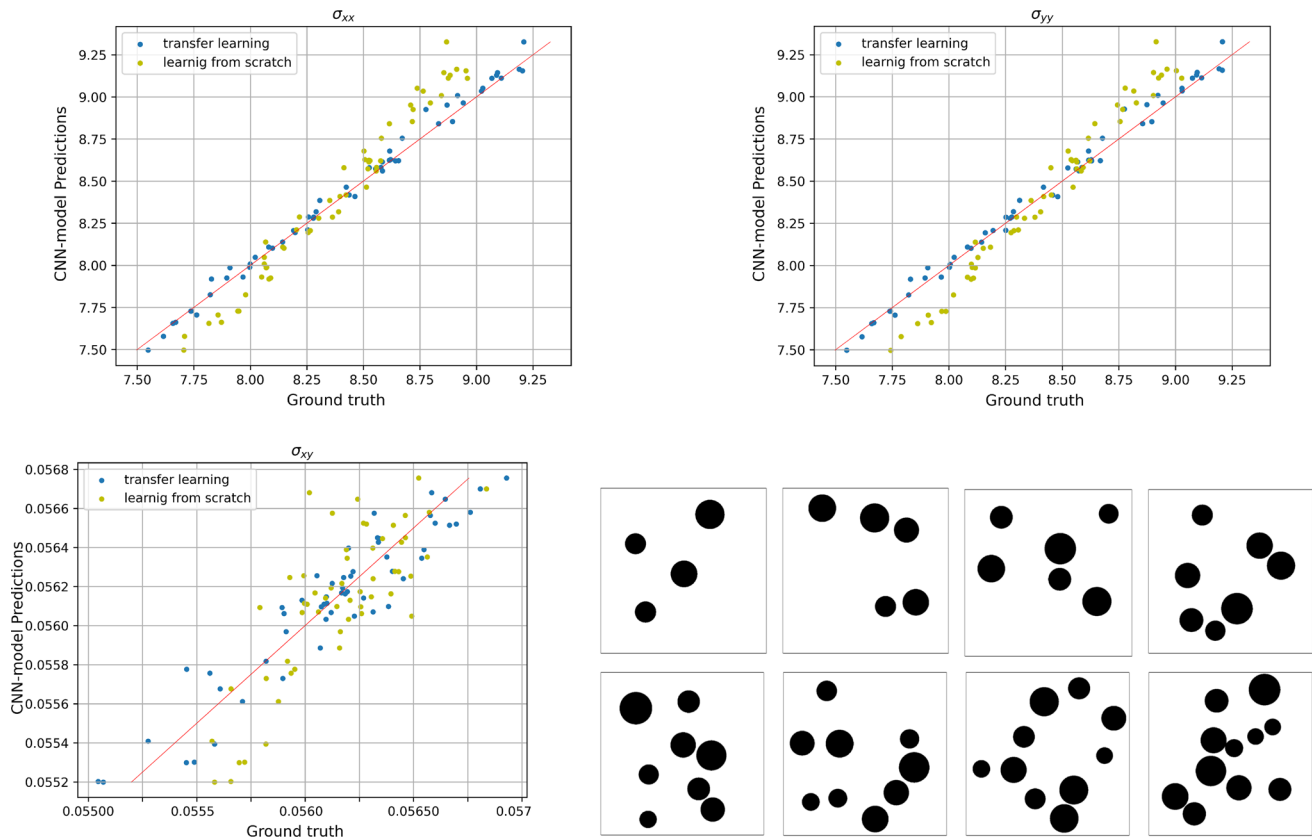
Table 4 Performance comparison between transfer learning model and training from scratch model: MAPE and  $R^2$  values on the test dataset

Training from scratch model	$\bar{\sigma}_{11}$	$\bar{\sigma}_{22}$	$\bar{\sigma}_{12}$
MAPE	1.0%	0.9%	0.5%
$R^2$	0.87	0.87	0.55
Transfer learning model	$\bar{\sigma}_{11}$	$\bar{\sigma}_{22}$	$\bar{\sigma}_{12}$
MAPE	0.21%	0.22%	0.35%
$R^2$	0.99	0.99	0.80

red-line compared with the training from scratch model. The good prediction of the transfer learning model required much less number of epochs (around 20 epochs) and smaller MSE when compared with the training from scratch model as illustrated in Fig. 14. Thus, we were able to further accelerate the micro-to-macro simulations of materials with complex micro-structures using the transfer learning approach. Next, we also compare the mean absolute percentage error (MAPE) and the  $R^2$ -score for both models in Table 4. As expected, the transfer learning model illustrates a better MAPE and coefficient of determination compared with the scratch training model.

### 6 Conclusion

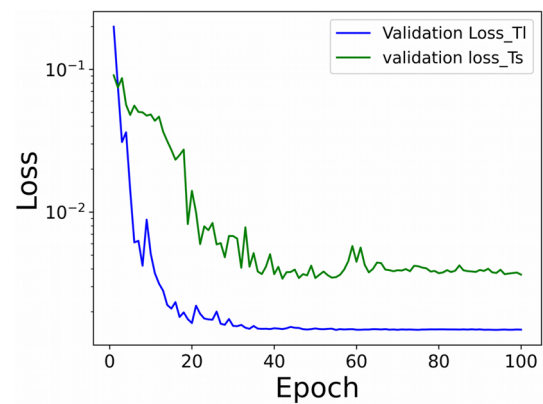
In this contribution, a Convolutional Neural Networks (CNN) model was developed to predict the homogenized stress in the case of elastic small deformation for a synthetically generated microstructure representing a heterogeneous material. The considered material was a composite with 3 different phases (e.g. a hybrid composite laminate consisting of shape memory alloy and glass fiber as reinforced phases and epoxy resin as the host material of the composite; or concrete microstructure consisting of aggregates, voids and cement-paste matrix). The positions of inclusions and their sizes varied randomly in each  $\mathcal{RV}\mathcal{E}$  of the generated microstructure. The data set used for the training process and the numerical tests consisted of images of the generated microstructures, which were considered the CNN model’s input, and the homogenized stresses resulting from applying finite element simulations were considered to be the labels. The proposed CNN model was chosen after several numerical tests performed on different model architectures, these were inspired by the work in different papers taking the CNN approach to solve computational mechanics tasks. The architecture of the CNN model consists of five sets of 2 convolutional layers followed by a max-pooling layer



**Fig. 13** Performance comparison between transfer learning model and training from scratch model: CNN-predictions versus ground-truth values of the stress components

then a flatten layer and two fully connected layers, while utilizing  $L_2$  and early stopping as regularization methods. The obtained results when employing the CNN model and its transfer learning show high accuracy in predicting the homogenized stress and a significant decrease in computation time.

In this work, we focused our investigations on linear elastic material response, however, the proposed CNN model can efficiently be extended toward inelastic (time-dependent) material response by coupling the CNN structure with Recurrent Neural Networks (RNN) or adding the deformed configurations as an input. These topics await investigation. Furthermore, our model performance was evaluated using numerical results. As a future step, real experimental data of concrete microstructures (DFG Priority Program SPP 2020 Experimental-Virtual-Lab) will be employed as future training data in the CNN approach.



**Fig. 14** Transfer learning losses (in blue) versus training from scratch losses (in green). (Color figure online)

**Acknowledgements** Fadi Aldakheel (FA) gratefully acknowledges support for this research by the “German Research Foundation” (DFG) in the PRIORITY PROGRAM SPP 2020 within its second funding phase. FA would like to thank N. Noii and A. Khodadadian for the support with mesh generation related to our paper [34].

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Nguyen LTK, Keip M-A (2018) A data-driven approach to nonlinear elasticity. *Comput Struct* 194:97–115
- Eggersmann R, Kirchdoerfer T, Reese S, Stainier L, Ortiz M (2019) Model-free data-driven inelasticity. *Comput Methods Appl Mech Eng* 350:81–99
- Carrara P, De Lorenzis L, Stainier L, Ortiz M (2020) Data-driven fracture mechanics. *Comput Methods Appl Mech Eng* 372:113390
- Lu X, Yvonnet J, Papadopoulos L, Kalogeris I, Papadopoulos V (2021) A stochastic fe2 data-driven method for nonlinear multiscale modeling. *Materials* 14:2875
- Bahmani B, Sun W (2021) A kd-tree-accelerated hybrid data-driven/model-based approach for poroelasticity problems with multi-fidelity multi-physics data. *Comput Methods Appl Mech Eng* 382:113868
- Bessa MA, Glowacki P, Houlder M (2019) Bayesian machine learning in metamaterial design: Fragile becomes supercompressible. *Adv Mater* 31:1904845
- Fuchs A, Heider Y, Wang K, Sun W, Kaliske M (2021) Dnn2: a hyper-parameter reinforcement learning game for self-design of neural network based elasto-plastic constitutive descriptions. *Comput Struct* 249:106505
- Zohdi T (2022) Machine-learning and digital-twins for rapid evaluation and design of injected vaccine immune-system responses. *Comput Methods Appl Mech Eng* 401:115315
- Fernández M, Rezaei S, Rezaei Mianroodi J, Fritzen F, Reese S (2020) Application of artificial neural networks for the prediction of interface mechanics: a study on grain boundary constitutive behavior. *Adv Model Simul Eng Sci* 7:1–27
- Zhang X, Garikipati K (2020) Machine learning materials physics: multi-resolution neural networks learn the free energy and nonlinear elastic response of evolving microstructures. *Comput Methods Appl Mech Eng* 372:113362
- Lu L, Meng X, Mao Z, Karniadakis GE (2021) Deepxde: a deep learning library for solving differential equations. *SIAM Rev* 63:208–228
- Vlassis NN, Sun W (2022) Component-based machine learning paradigm for discovering rate-dependent and pressure-sensitive level-set plasticity models. *J Appl Mech* 89
- Lopez E, Gonzalez D, Aguado J, Abisset-Chavanne E, Cueto E, Binetruy C, Chinesta F (2018) A manifold learning approach for integrated computational materials engineering. *Arch Comput Methods Eng* 25:59–68
- Heider Y (2021) Multi-field and multi-scale computational fracture mechanics and machine-learning material modeling, Ph.D. thesis, Habilitation thesis, Rheinisch-Westfälische Technische Hochschule Aachen. <https://publications.rwth-aachen.de/record/835190>
- Henkes A, Wessels H, Mahnken R (2022) Physics informed neural networks for continuum micromechanics. *Comput Methods Appl Mech Eng* 393:114790
- Zohdi T (2022) A note on rapid genetic calibration of artificial neural networks. *Comput Mech* 70:819–827
- Fuhg JN, Bouklas N (2022) On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling. *Comput Methods Appl Mech Eng* 394:114915
- Fernández M, Fritzen F, Weeger O (2022) Material modeling for parametric, anisotropic finite strain hyperelasticity based on machine learning with application in optimization of metamaterials. *Int J Numer Methods Eng* 123:577–609
- As’ad F, Avery P, Farhat C (2022) A mechanics-informed artificial neural network approach in data-driven constitutive modeling. *Int J Numer Methods Eng* 123:2738–2759
- Benaïmeche MA, Yvonnet J, Bary B, He Q-C (2022) A k-means clustering machine learning-based multiscale method for anelastic heterogeneous structures with internal variables. *Int J Numer Methods Eng* 123:2012–2041
- Cueto E, Chinesta F (2022) Thermodynamics of learning physical phenomena. *arXiv preprint arXiv:2207.12749*
- Stöcker J, Fuchs A, Leichsenring F, Kaliske M (2022) A novel self-adversarial training scheme for enhanced robustness of inelastic heterogeneous descriptions by neural networks. *Comput Struct* 265:106774
- Zhang L, Lu Y, Tang S, Liu WK (2022) Hidenn-td: Reduced-order hierarchical deep learning neural networks. *Comput Methods Appl Mech Eng* 389:114414
- Niekamp R, Niemann J, Schröder J (2023) A surrogate model for the prediction of permeabilities and flow through porous media: a machine learning approach based on stochastic brownian motion. *Comput Mech* 71:563–581
- Wessels H, Böhm C, Aldakheel F, Hüpgen M, Haist M, Lohaus L, Wriggers P (2022) Computational homogenization using convolutional neural networks. In: Current trends and open problems in computational mechanics. Springer, Berlin, pp 569–579
- Kalina KA, Linden L, Brummund J, Kästner M (2023) Fe<sup>ANN</sup>: an efficient data-driven multiscale approach based on physics-constrained neural networks and automated data mining. *Comput Mech*. <https://doi.org/10.1007/s00466-022-02260-0>
- Rao C, Liu Y (2020) Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. *Comput Mater Sci* 184:109850
- Yang Z, Yabansu YC, Al-Bahrani R, Liao W-K, Choudhary AN, Kalidindi SR, Agrawal A (2018) Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. *Comput Mater Sci* 151:278–287
- Beniwal A, Dadhich R, Alankar A (2019) Deep learning based predictive modeling for structure-property linkages. *Materialia* 8:100435
- Eidel B (2023) Deep CNNs as universal predictors of elasticity tensors in homogenization. *Comput Methods Appl Mech Eng* 403:115741
- Aldakheel F (2020) A microscale model for concrete failure in poro-elasto-plastic media. *Theoret Appl Fract Mech* 107:102517

32. Wriggers P, Aldakheel F, Lohaus L, Heist M (2020) Water-induced damage mechanisms of cyclically loaded high-performance concretes. *Bauingenieur* 95:126–132
33. Aldakheel F (2021) Simulation of fracture processes using global-local approach and virtual elements, Hannover: Institut für Kontinuumsmechanik, Leibniz Universität Hannover. <https://doi.org/10.15488/11367>
34. Nohi N, Khodadadian A, Aldakheel F (2022) Probabilistic failure mechanisms via monte Carlo simulations of complex microstructures. *Comput Methods Appl Mech Eng* 399:115358
35. Lloberas-Valls O, Rixen DJ, Simone A, Sluys LJ (2012) Multiscale domain decomposition analysis of quasi-brittle heterogeneous materials. *Int J Numer Methods Eng* 83:1337–1366
36. Fish J (2014) *Practical multiscale modeling*. Wiley, Chichester
37. Fish J, Wagiman A (1993) Multiscale finite element method for a locally nonperiodic heterogeneous medium. *Comput Mech* 12:164–180
38. Aldakheel F, Nohi N, Wick T, Wriggers P (2020) A global-local approach for hydraulic phase-field fracture in poroelastic media. *Comput Math Appl*. <https://doi.org/10.1016/j.camwa.2020.07.013>
39. Michel J, Moulinec H, Suquet P (1999) Effective properties of composite materials with periodic microstructure: a computational approach. *Comput Methods Appl Mech Eng* 172:109–143
40. Aldakheel F, Soyarslan C, Palanisamy HS, Elsayed ES (2023) Machine learning aided multiscale magnetostatics. *arXiv preprint arXiv:2301.12782*
41. Hill R (1965) A self-consistent mechanics of composite materials. *J Mech Phys Solids* 13:213–222
42. Bensoussan A, Lions JL, Papanicolaou G (1978) Asymptotic analysis for periodic structures. In: Lions JL, Papanicolaou G, Rockafellar RT (eds) *Studies in mathematics and its applications*, vol 5. North-Holland Publishing Company, Amsterdam
43. Sanchez-Palencia E (1980) Non-homogeneous media and vibration theory, volume 127 of lecture notes in physics. Springer, Berlin
44. Voigt W (1887) Über die Beziehung zwischen den beiden Elastizitätskonstanten isotroper Körper. *Ann Phys* 38:573–587
45. Reuss A (1929) Berechnung der Fließgrenze von Mischkristallen aufgrund der Plastizitätsbedingung für Einkristalle. *Z Angew Math Mech* 9:49–58
46. Hashin Z, Shtrikman S (1962) On some variational principles in anisotropic and nonhomogeneous elasticity. *J Mech Phys Solids* 10:335–342
47. Hashin Z, Shtrikman S (1963) A variational approach to the theory of the elastic behaviour of multiphase materials. *J Mech Phys Solids* 11:127–140
48. Hill R (1965) A self-consistent mechanics of composite materials. *J Mech Phys Solids* 13:213–222
49. Smit RJM, Brekelmans WAM, Meijer HEH (1998) Prediction of the mechanical behavior of nonlinear heterogeneous systems by multi-level element modeling. *Comput Methods Appl Mech Eng* 155:181–192
50. Miehe C, Schotte J, Schröder J (1999) Computational micro-macro transitions and overall moduli in the analysis of polycrystals at large strains. *Comput Mater Sci* 16:372–382
51. Miehe C, Schröder J, Schotte J (1999) Computational homogenization analysis in finite plasticity. Simulation of texture development in polycrystalline materials. *Comput Methods Appl Mech Eng* 171:387–418
52. Miehe C (2002) Strain-driven homogenization of inelastic microstructures and composites based on an incremental variational formulation. *Int J Numer Methods Eng* 55:1285–1322
53. Feyel F, Chaboche J-L (2000) FE<sup>2</sup> multiscale approach for modelling the elastoviscoplastic behaviour of long fibre SiC/Ti composite materials. *Comput Methods Appl Mech Eng* 183:309–330
54. Terada K, Kikuchi T (2001) A class of general algorithms for multi-scale analyses of heterogeneous media. *Comput Methods Appl Mech Eng* 190:5427–5464
55. Kouznetsova V, Geers MGD, Brekelmans WAM (2002) Multi-scale constitutive modelling of heterogeneous materials with a gradient-enhanced computational homogenization scheme. *Int J Numer Methods Eng* 54:1235–1260
56. Geers MGD, Kouznetsova VG, Brekelmans WAM (2010) Multi-scale computational homogenization: trends and challenges. *J Comput Appl Math* 234:2175–2182
57. Schröder J, Keip M-A (2012) Two-scale homogenization of electromechanically coupled boundary value problems. *Comput Mech* 50:229–244
58. Schröder J (2000) *Homogenisierungsmethoden der nichtlinearen Kontinuumsmechanik unter Beachtung von Stabilitätsproblemen*, Habilitationsschrift, Institut für Mechanik (Bauwesen), Lehrstuhl, I, Universität Stuttgart. Bericht Nr. I-7
59. Chatzigeorgiou G, Javili A, Steinmann P (2014) Unified magnetomechanical homogenization framework with application to magnetorheological elastomers. *Math Mech Solids* 19:193–211
60. Schmidt F, Krüger M, Keip M-A, Hesch C (2022) Computational homogenization of higher-order continua. *Int J Numer Methods Eng* 123:2499–2529
61. Javili A, Chatzigeorgiou G, Steinmann P (2013) Computational homogenization in magneto-mechanics. *Int J Solids Struct* 50:4197–4216
62. Hill R (1972) On constitutive macro-variables for heterogeneous solids at finite strain. *Proc R Soc Lond (Ser A)* 326:131–147
63. Suquet PM (1987) Elements of homogenization for inelastic solid mechanics. In: Sanchez-Palencia E, Zaoui A (eds) *Lecture notes in physics: homogenization techniques for composite materials*. Springer, Berlin, pp 193–278
64. Nemat-Nasser S, Hori M (1999) *Micromechanics: overall properties of heterogeneous materials*, volume 36 of North-Holland series in Applied Mathematics and Mechanics, 2nd edn. Elsevier Science Publisher B.V., New York
65. Zohdi TI, Oden JT, Rodin GJ (1996) Hierarchical modeling of heterogeneous bodies. *Comput Methods Appl Mech Eng* 138:273–298
66. Zohdi T, Wriggers P (1999) A domain decomposition method for bodies with heterogeneous microstructure based on material regularization. *Int J Solids Struct* 36:2507–2525
67. Zohdi TI (2008) On the computation of the coupled thermo-electromagnetic response of continua with particulate microstructure. *Int J Numer Methods Eng* 76:1250–1279
68. Hain M, Wriggers P (2008) Numerical homogenization of hardened cement paste. *Comput Mech* 42:197–212
69. Temizer I, Wriggers P (2011) Homogenization in finite thermoelasticity. *J Mech Phys Solids* 59:344–372
70. Temizer I, Wriggers P (2008) A multiscale contact homogenization technique for the modeling of third bodies in the contact interface. *Comput Methods Appl Mech Eng* 198:377–396
71. Temizer I, Wriggers P (2007) An adaptive method for homogenization in orthotropic nonlinear elasticity. *Comput Methods Appl Mech Eng* 196:3409–3423
72. Wriggers P, Moftah SO (2006) Mesoscale models for concrete: homogenisation and damage behaviour. *Finite Elem Anal Des* 42:623–636
73. Temizer I, Zohdi TI (2007) A numerical method for homogenization in non-linear elasticity. *Comput Mech* 40:281–298
74. Zohdi TI, Wriggers P (2005) An introduction to computational micromechanics. In: Pfeiffer F, Wriggers P (eds) *Lecture notes in applied and computational mechanics*, vol 20. Springer, Berlin
75. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J et al (2018) Recent advances in convolutional neural networks. *Pattern Recogn* 77:354–377



76. Nair V, Hinton G (2010) Rectified linear units improve restricted boltzmann machines vinod nair vol 27, pp 807–814
77. Eger S, Youssef P, Gurevych I (2019) Is it time to swish? Comparing deep learning activation functions across NLP tasks. arXiv preprint [arXiv:1901.02671](https://arxiv.org/abs/1901.02671)
78. Segurado J, Llorca J (2002) A numerical approximation to the elastic properties of sphere-reinforced composites. *J Mech Phys Solids* 50:2107–2121
79. Wessels H, Weißenfels C, Wriggers P (2020) The neural particle method-an updated Lagrangian physics informed neural network for computational fluid dynamics. *Comput Methods Appl Mech Eng* 368:113127
80. Wriggers P (2008) *Nonlinear finite elements*. Springer, Berlin

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.