

ARTICLE OPEN



Teaching solid mechanics to artificial intelligence—a fast solver for heterogeneous materials

Jaber Rezaei Mianroodi ¹✉, Nima H. Siboni ² and Dierk Raabe ¹

We propose a deep neural network (DNN) as a fast surrogate model for local stress calculations in inhomogeneous non-linear materials. We show that the DNN predicts the local stresses with 3.8% mean absolute percentage error (MAPE) for the case of heterogeneous elastic media and a mechanical contrast of up to factor of 1.5 among neighboring domains, while performing 103 times faster than spectral solvers. The DNN model proves suited for reproducing the stress distribution in geometries different from those used for training. In the case of elasto-plastic materials with up to 4 times mechanical contrast in yield stress among adjacent regions, the trained model simulates the micromechanics with a MAPE of 6.4% in one single forward evaluation of the network, without any iteration. The results reveal an efficient approach to solve non-linear mechanical problems, with an acceleration up to a factor of 8300 for elastic-plastic materials compared to typical solvers.

npj Computational Materials (2021)7:99; <https://doi.org/10.1038/s41524-021-00571-z>

INTRODUCTION

The mechanical response of materials depends highly on the microstructure and its heterogeneity, including all defects, phases and chemical features. An inseparable task in modeling a corresponding coupled multi-physics material problem that links mechanical properties to microstructure, consists in solving the underlying differential equations for mechanical equilibrium. This task is particularly challenging for the case of non-linear material response, highly inhomogeneous material properties and complex microstructure topologies. Several methods to numerically solve such non-linear solid mechanics problems have already been developed and are in use, such as spectral solvers and the finite element method. However, ever-increasing sophistication of material models by including more physics and coupled problems in multi-scale methods is bringing these numerical solvers to their limits.

Therefore, a number of attempts were made to speed up material modeling using artificial intelligence (AI) in general, and specifically using deep neural networks (DNN). For example, Aydin et al.¹ devised a multi-fidelity computational framework to train fully connected neural networks and applied it to predict the deformation of a thin elastic membrane. Although the method appears promising for elastic scenarios, it might be less suited for cases with non-linear material behavior, such as envisaged here. The reason is that for high fidelity simulations of heterogeneous and high mechanical contrast materials with inelastic constitutive response, high accuracy is needed to cope with non-linear effects such as stress and strain localization, shear banding and onset of yielding.

Cecen et al.² developed a data-driven approach to efficiently link three-dimensional microstructures to their homogenized properties using convolutional neural networks with improved accuracy in property predictions as well as reduction in the computation time compared to conventional microstructure quantification. Fernandez et al.³ employed artificial neural networks as a surrogate model to transfer atomistic grain boundary decohesion data to continuum scale modeling of intergranular fracture in Aluminum. Li et al.⁴ developed a

hierarchical neural hybrid method to efficiently compute failure probabilities in high dimensional problems employing the multi-fidelity approach introduced by Aydin et al.¹ They showed that for achieving an accurate estimate of the rare failure probability, a traditional Monte Carlo method needs to solve the equations significantly more frequently than the proposed hierarchical neural hybrid method. In addition, Monte Carlo models are generally numerically not very efficient, due to their discrete event probing and associated generation and comparison of values against random numbers.

Although numerous works on the use of machine learning (ML) in materials science have been published, these are often geared towards predicting an average (homogenized) behavior of the system, based on large input data sets. In contrast to these mean-field approaches, a few other publications have recently appeared with the goal of predicting also some spatially resolved features of solution fields in complex material systems. For example, Yang et al.⁵ employed conditional generative adversarial neural network (cGAN) to predict stress and strain fields directly from the material microstructure geometry. Wang et al.⁶ introduced a genomic flow network (GFNet) and a mosaic flow predictor to estimate the solution of Laplace and Navier–Stokes equations in domains of unseen shapes and boundary conditions. Pandey et al.⁷ proposed a machine learning based surrogate model for predicting the spatially resolved microstructure evolution in materials exposed to uniaxial tensile loading. Similarly, in this work, we focus on resolving the local response of the system. More specific, we devise a general AI-based solver for predicting the local stresses in heterogeneous solids with high mechanical contrast features and non-linear material response. This solver can be used to replace or augment conventional numerical approaches like finite element methods. In this work, we demonstrate the ability of ML to calculate mechanical stress field in complex microstructures with both, elastic and elasto-plastic material behavior. For the sake of simplicity, isotropic elastic and elastic-perfect-plastic response (zero hardening) are adopted here. We focus here only on stress, since it appears explicitly in the partial differential equation for mechanical equilibrium. In principle, the same framework could be

¹Microstructure Physics and Alloy Design, Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany. ²DeepMetis, Berlin, Germany. ✉email: j.mianroodi@mpie.de

used to predict strain as well. Furthermore, when knowing the constitute behavior of the material, in specific cases, it may be possible to calculate strain components from the predicted stress field through post-processing: the elastic strain is through Hooke's law always linearly related to the local stress but the plastic strain is path-dependent and may be more challenging to resolve. In such cases an equivalent machine learning approach as used here for stress can also be applied without any loss in generality.

The results and their discussion for all trained scenarios with different constitutive response and microstructure topologies are presented in the next section, placing attention on both, the performance of the network on geometries and mechanical contrast ranges similar to the trained ones and also on those beyond the training range. Next, a general discussion and future opportunities are presented. This is followed by the methods section with a brief review of the governing equations for mechanical equilibrium in elasto-plastic solids and an overview of the machine learning algorithm we use in this study. Finally, the simulation setup and the generation of the mechanical stress field data that serve for training and evaluating the machine learning network are presented.

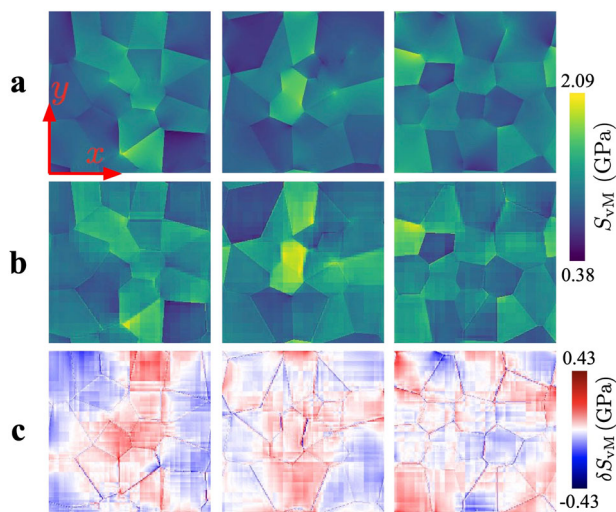


Fig. 1 Stress distribution in case of elastic material response. Simulation results obtained by DAMASK in conjunction with a spectral solver (a) and the corresponding AI predicted (b) von Mises stress distribution for three Voronoi tessellations of test microstructures (not included in the training dataset) using the modified U-Net approach. The error, calculated as deviation between the two von Mises stress distributions, δS_{VM} , is shown in (c).

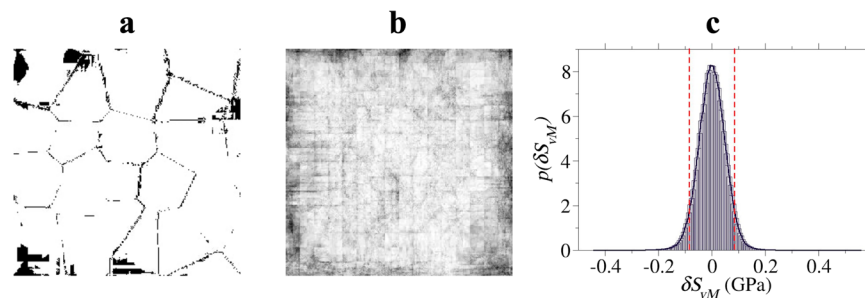


Fig. 2 Error distribution analysis for the case of elastic material response. Distribution of the absolute values of the deviations between the reference data obtained from the full-field DAMASK based spectral solver simulations and the U-Net predictions (a), showing only regions with errors above 83 MPa (4% relative error) in black color for one of the test cases. b The local probability of occurrence of errors larger than 83 MPa. The probability is shown by a gray-scale color-code (the white and black colors represent the probability of zero and one, respectively). This quantity is calculated by averaging the observations similar to (a) over all the test data-sets. In this procedure for each of the test samples, a value of 1 is assigned to regions with an error above 83 MPa, and 0 otherwise. c Error distribution for all points in all of the 50 cases in the test data-set with 4% relative error indicated as red dashed lines.

RESULTS

Isotropic elastic case

Once the network is trained using the training dataset outlined in “Setup and Training” section below, we evaluate it with sets of input material property distributions in topological configurations and parameter combinations which had not been used for training the network. More specific, the evaluation on the test dataset is done both (i) for geometries similar to the training data (i.e., Voronoi tessellations), and (ii) geometries which are completely different from those that had been used for training. For the geometries in (ii), we included cases where the surface of the inclusion (i.e., the internal domain) is curved, as a test for the model's capability to generalize beyond Voronoi-based representative volume elements, where straight, faceted interface shapes were used. These results are presented in the following sub-sections.

U-Net predictions for domain topologies similar to the training data. Von Mises stress (see “Methods” section below for a summary of the formulations) predictions based on the U-Net architecture for three Voronoi tessellations from the test data set are shown in Fig. 1. The figure reveals that the AI-based solution, obtained by the modified U-Net method, qualitatively and in part even quantitatively agrees well with the DAMASK⁸ solution, although there are regions with a higher relative error of up to 25%. As will be shown in the following, these regions of high deviation are limited to only a few rare points and for most of the grid points in the domain the relative error is below 4%. In particular, the surrogate model is able to adequately reproduce the sites of stress concentrations and the distribution at the domain junctions, which are due to local contrast in elastic constants.

To further investigate the source of the errors reported above, we re-plot Fig. 1c right, in Fig. 2a, showing only regions with error values higher than 83 MPa (4% relative error) as black. This analysis, firstly, suggests that there are localized regions of higher error between the domain boundaries where the material properties are discontinuous. This is expected since those regions are typically also the most challenging ones for conventional solvers to cope with as well. A possible mitigation strategy for such zones of high local deviation might be the use of filtering schemes in the network, quite similar as it is done in spectral solvers, where the underlying Fourier series can create non-physical high frequency noise in solutions.

As observed in Fig. 2a, in addition to the domain boundaries, there are patches of high error values seemingly close to the boundary of the calculation box. To confirm this, an averaging operation is performed, by simply adding images similar to the

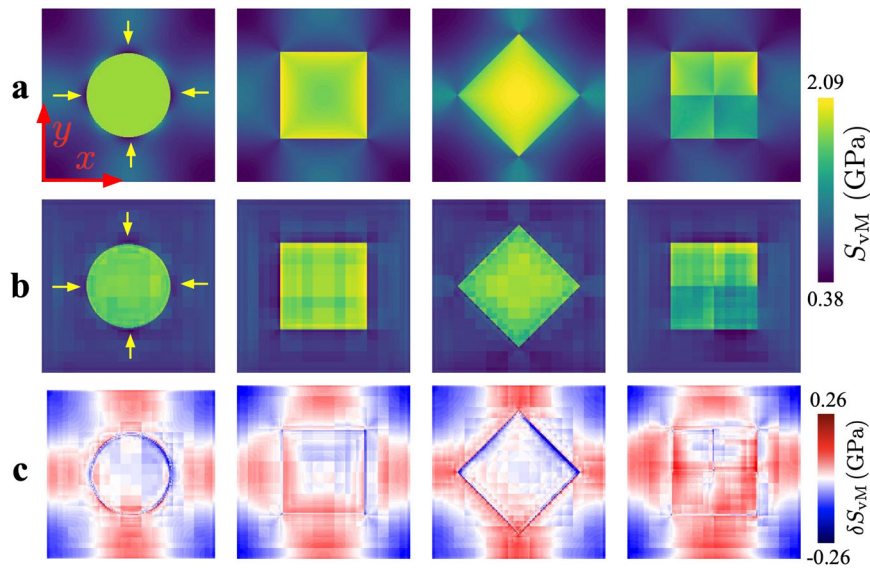


Fig. 3 Predicted stress distribution for selected geometries beyond the trained topologies. Simulation results (a) for isotropic elastic material response for topological cases far away from the trained set of Voronoi structures and the corresponding AI predicted (b) von Mises stress distribution for different hard inclusions in a soft matrix. The error between the AI prediction and the DAMASK solution is shown in (c).

one shown in Fig. 2a of all cases in the test dataset and dividing the sum by the number of cases (50 here). The resulting average error distribution map is shown in Fig. 2b. It indicates that these patches are indeed mostly located near the boundaries of the box. We speculate that this is due to the fact that our convolutional neural network, unlike the DAMASK solver, does not consider periodic boundary conditions. More specifically, when the kernel is applied to points that are located close to the outer borders of the simulation box, for the pixels beyond these borders their periodic counterpart values from the opposite borders should be used, like used in most similar set-ups where classical periodic boundary conditions are enforced. It is therefore likely that enforcing the periodicity of the input images in the neural network will help reducing the boundary effect errors. These aspects refer to work in progress which will be analyzed in more detail and be reported in ensuing publications. A positive aspect of this error close to the borders of the simulation domains, however, is that it seems to be merely a geometrical effect related to the simulation box and not a fundamental problem of the U-Net approach to identify surrogate solutions. Figure 2c shows the error distribution of all points in all cases of the test dataset. The analysis shows that most of the points (91% of all points) fall inside the red dashed lines indicating equal or less than 4% relative error.

Application of the U-Net to geometries far from the training data. Next, we look into geometries that could not be represented by Voronoi tessellation. These include classical elastic problems of circular and square shaped inclusions with larger Young's modulus embedded in a softer matrix. The von Mises stress distributions for such cases predicted by the AI and calculated by DAMASK are shown in Fig. 3.

Although the topologies of these high-mechanical contrast test cases are fundamentally different from the Voronoi structures that were used for training, the AI-based predictions still capture the correct stress partitioning between the soft and hard phases as shown in Fig. 3. The error between the reference patterns obtained by the full-field DAMASK simulations and the AI-predictions is shown in the third row. The maximum relative error is about 12%, however, as in the previous cases, the regions of largest error are limited to the domain boundaries. Otherwise, the stress distribution around and inside the hard inclusions follows

that observed in the reference pattern. For example, note that for the four deep blue regions around the circular inclusion at 0, 90, 180 and 270°, as marked by yellow arrows, the minimum von Mises stress based on the DAMASK-calculated and AI-predicted values are 802.5 and 849.4 MPa, respectively.

Effects of depth and kernel size of the U-Net. The Gibbs phenomenon is a classical numerical issue associated with solution algorithms of mechanical problems with sharp transition features in material properties (as shown in many works, e.g.,^{9,10}), especially when using the fast Fourier transform method as a solver. This undesired oscillating and overshooting effect is produced by Fourier solvers applied to piece-wise continuously differentiable functions at discontinuities (such as here when crossing interfaces). The oscillations in the solution fields due to such sharp transitions can be minimized using multiple filtering schemes⁹. Therefore, next we study in more detail the predicted stress profile and look into the effect of the depth of the network (i.e., the number of down- and upsampling steps) on such oscillations.

As shown in Fig. 4-right column the deeper network, i.e., the network with $N_s = 8$ (see “Setup and Training” section below for details about the network architecture) is capturing the wake of the stress field better than the shallower network with $N_s = 4$, however, the jump in the stress magnitude at the transition of the material property is larger. This seems analogous to the full Fourier approach used in spectral solvers where all frequencies are included (deep network) compared to the filtered approach (shallow network). In the spectral method, filtering the high frequencies is equivalent to approximating the derivatives with lower order finite differences (such as forward or backward differences). Since the convolution kernels used in the network could correspond to taking derivatives of the data, increasing the depth of the network, at least conceptually, will be analogous to work with higher order derivatives in conventional solvers, amplifying the fluctuations. In other words, the kernel weights, that are learned during the training, have the potential to signify differences between adjacent points. This is often the case in feature detection by U-Net where the edges are detected using larger differences (gradients) in image intensity or color. Here, we are speculating that since a deeper network has more layers

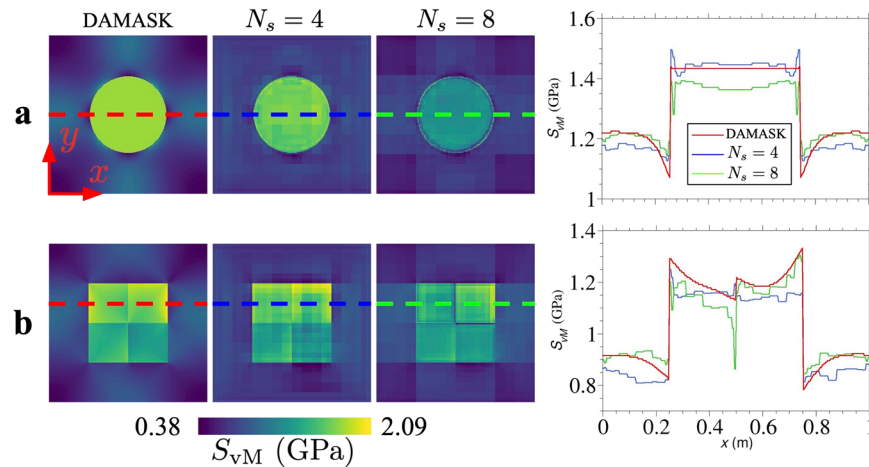


Fig. 4 Effect of network depth on predicted stress distribution. Von Mises stress distribution around circular (a) and square (b) inclusions calculated by DAMASK as well as two network architectures with depths of $N_s = 4$ and $N_s = 8$. The stress profiles are plotted along the dashed lines with red, blue and green corresponding to DAMASK, network with $N_s = 4$ and $N_s = 8$ based results, respectively.

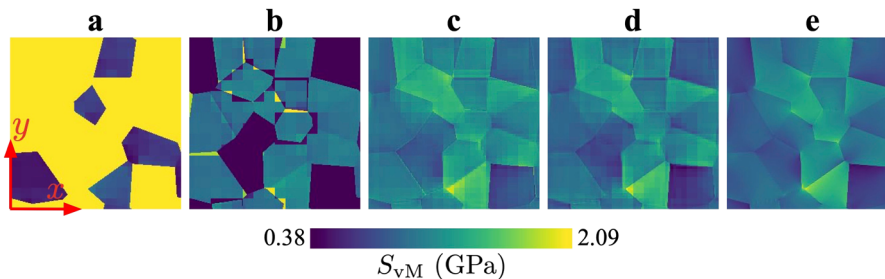


Fig. 5 Effect of kernel size on predicted stress distribution. The predicted von Mises equivalent stress distribution for kernels with size $k \times k$, $k = 3, 5, 7, 9$, shown in (a), (b), (c), and (d), respectively, compared to the results obtained from the full-field DAMASK simulations used here as reference (e). All models were trained under the same conditions.

calculating the differences from the previous layer, this will have an amplifying effect on oscillations.

Next we focus on the effect of the kernel size on the AI-predicted results. As seen in Fig. 5, smaller $k \times k$ kernels with $k = 3$ and $k = 5$ result in completely incorrect stress distributions. In fact, under the training conditions discussed in “Setup and Training” section below, it was not possible to reach the loss values corresponding to networks with larger k values (≈ 0.02). However, networks with $k \times k$ kernels of $k = 7$ and $k = 9$ were trained to the lower loss values and they perform much better. Note that the results discussed in the sections above were based on $k = 9$.

U-Net AI predictions for cases with elasto-plastic material response

As the next step to create a surrogate AI stress solver, we extend the U-Net solution to elasto-plastic problems. It is worth mentioning that in order to obtain the elasto-plastic response in the full-field DAMASK simulations used here as reference, the load must be applied incrementally (here we used 100 steps). This renders the conventional DAMASK simulation procedure computationally more demanding compared to the elastic problems where the load is applied instantaneously. In our AI approach, we can obviously omit the extra computational costs that are caused by this incremental approach. Again, similar to the elastic case, the AI-prediction is, once the system is trained, a single step forward calculation, i.e., one set of inputs, here (Y, ν, S_y) , representing Young’s modulus, Poisson’s ratio and Yield stress distributions, respectively, are fed to the neural network and the von Mises

stress is directly predicted, without any iterations. As we have in the current elastic-plastic material case more independent inputs and also a more complicated relation between the inputs and outputs, we first need to increase the predictive power of the neural network. Therefore, we increased the complexity of the neural network by doubling the number of the channels at each stage of the U-Net. Based on our observation of the effect of an increase in the depth of the U-Net, the increase in the number of channels appeared to be a suited approach for enhancing the complexity of the U-Net for predicting mechanical response in heterogeneous media.

Similar as done for the elastic case, we trained the U-Net using the DAMASK stress predictions to obtain reference values. Due to the nature of the material’s elasto-plastic behavior, the absolute value of the overall applied strain is important. For very small strains, all the domains remain elastic, and for very large strains all the domains enter the elastic-plastic regime. In the former case, we are actually solving a purely elastic problem and consequently the solution depends only on the elastic constants, and for the latter case the stresses could be estimated fairly well from the yield stress values and therefore the elastic constants do not play a significant role. To provide a generic training data-set for the U-Net, we choose the magnitude of the applied strain in a range so that some of the domains become plastic while the others remain elastic. In this regime, the stress is determined by a complicated interplay between elasticity and plasticity.

A typical example of the stress prediction using the trained U-Net for the case of elasto-plastic material behavior is depicted in Fig. 6.

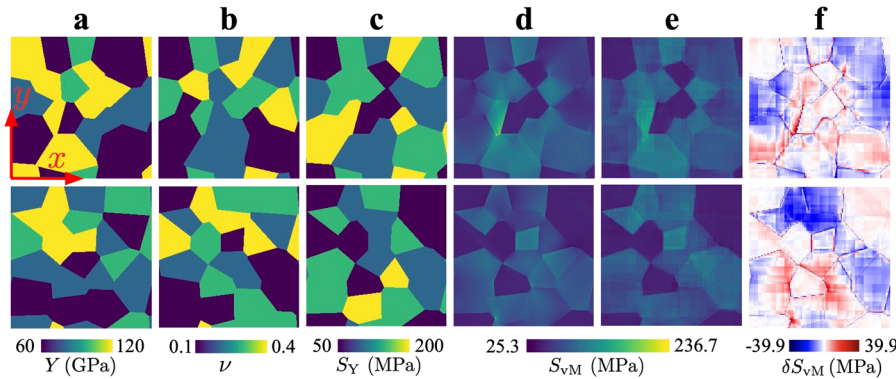


Fig. 6 Stress distribution in case of elasto-plastic material response. Comparison for two different topologies obtained by U-Net predictions (e) of the von Mises stress spatial distribution and the corresponding DAMASK calculations (d) for the shown input material property distributions of Young's modulus (a), Poisson ratio (b) and Yield stress (c). The error between U-Net and DAMASK predictions is shown in (f). The applied strain is chosen such that the von Mises stress depends on both elastic constants and the yield stresses, i.e., the spatial pattern of the von Mises stress is neither similar to the elastic constant distribution nor to the yield stress distribution.

Speed-up and efficiency evaluation

In the previous sections, the network training and evaluation were performed on a machine powered by an Intel Xeon CPU clocked at 2.30GHz with a NVIDIA Tesla P4 GPU. However to benchmark the performance of the AI-based solution in comparison with DAMASK, we restrict here both to use only one core of an AMD EPYC 7702 64-Core Processor, clocked at 3.34 GHz. Limiting the overhead of both calculations as much as possible and running both on the same CPU for a large number of evaluations provides a good measure of the total number of floating point (FL) operations necessary to reach the final result. Of course, working with GPU for AI evaluation could lead to better scalability compared to parallel processing conducted in the typical spectral simulations in DAMASK. However, here we restrict the discussion to only single core comparison to evaluate the methods with respect to their basic FL operation costs.

Under these conditions, elastic calculations took in average 12.13 s (measured over 100 calculations) while the AI-based approach only took 0.12 s (measured over 2000 calculations), revealing a speed-up of about 103 times, roughly corresponding to the reduction in the number of required FL operations.

The calculation for elastic-plastic constitutive material response required in DAMASK to reach the applied load in 100 increments took about 22 min, an average per time increment of 13.2 s, while the AI-based calculation took only 0.158 s (measured as an average over 2000 evaluations). This corresponds to a speed-up by a factor of about 8300 times, corresponding to the same factor in reduced required total FL operations when using the AI-based solver for the application of the full load. Note that the simulation time for the non-linear solver in DAMASK will depend on the step size and numbers of increments applied. The current selection of 100 increments is not the optimum choice, however, based on the measurements, even calculation of a single increment for elasto-plastic material behavior is about 84 times slower than the AI-based solver for a scenario where the total external load that brings the material into the elastic-plastic regime is applied instantaneously. It should be considered, however, that also conventional solvers for materials with non-linear constitutive response could further be optimized for the specific problem at hand to reach converged results faster. This means that the details of the speed-up will highly depend on the specific problem at hand. The values reported above represent a typical simulation with DAMASK and the speed-up may vary between different solver configurations. The important point to note here is the conceptual difference in nature of the AI-based solver as opposed to conventional non-linear solvers, since the former will always work with a single-step evaluation of the network regardless of

the non-linearity of the task posed while the latter is always iterative.

DISCUSSION

Advanced materials and the products consisting of them have become immensely complex in their internal structure and chemistry as well as in the mechanical loading conditions they get exposed to, ranging from cell phones to space ships. Particularly structural materials, i.e., materials that are primarily meant to carry mechanical loads, nowadays contain multiple crystals, defects and phases with high mechanical contrast. Mapping such complex microstructures and parts in the form of digital twins and subjecting them to mechanical calculations is a prerequisite to their microstructural design, improved processing, further development and safe application. Similar aspects apply to other mechanically heterogeneous media as encountered in such diverse fields as soil-, structure-, building-, construction-, earthquake-, ice-, or colloidal mechanics.

Conducting such mechanical simulations for materials with large internal phase contrast and complex constitutive response (e.g., elasto-plastic behavior), together with non-linearities arising from large deformations, lead to immense computational costs when using conventional finite-element or spectral solver methods. The high computational costs are primarily due to the iterative nature of the solution algorithms used by these solvers. They limit the range, size and complexity of problems that are accessible to simulation-based investigations using current computers.

Therefore, an alternative surrogate machine-learning-based solver for mechanically heterogeneous and non-linear fields (here stress distribution) is introduced in this work. The accuracy and computational costs of this U-Net based solver are compared with a high-performance spectral solver, both for elastic and elasto-plastic constitutive scenarios. The proposed DNN can predict the local stress distribution with 3.8 and 6.4% mean absolute percentage error (MAPE) for heterogeneous elastic and elasto-plastic material response, respectively. The performance tests show a reduction in computation time of about 103 and 8300 times for elastic and elasto-plastic materials, respectively. Besides the acceptable accuracy and the substantially reduced computational costs, the trained DNN also shows a great generalization capability by predicting stress distributions in geometries far different from those used in the training data-sets.

Although the observed MAPE shows a reasonably accurate stress prediction by the ML method, the errors are currently admittedly in some cases larger than those of a typical (e.g., FEM)

solver. How these errors affect and act on further downstream calculations will depend on the specific type of these calculations. For example, if the absolute level of stress is in focus, for example as required for damage growth modeling, the ML solution must be improved substantially. However, in applications such as shape, microstructure, loading and topology optimization; augmentation and efficiency increase of conventional solvers; or the identification of regions of high stress and strain concentrations, the current accuracy is sufficient. A detailed error analysis and network architecture study reveals several pathways for its further improvement in the future. Incorporating filtering schemes to remove the high frequency noise and enforcing periodicity in kernel operations of the DNN represent specific items along these lines that are currently in progress. In addition, predicting full stress and strain tensor components is straightforward based on the current framework and require no conceptual change.

The huge speed-up in calculation of the stress distribution in such highly non-linear mechanical systems paves the way for many important applications, specifically addressing more complicated simulations, materials, loading scenarios and optimization problems, not previously computationally accessible. Furthermore, combining the proposed surrogate model with conventional solution methods can be an approach towards higher accuracy of conventional solvers at considerably lower computational costs. This can be for instance achieved by employing the U-Net based stress prediction as an initial stress guess for conventional iterative solvers. This could greatly reduce the number of iterations required, thus substantially saving computation time that is usually required for convergence. In general, problems with an inherent instability or solution bifurcation or, in particular, with severe localized deformation and failure are speculated to be harder to solve with the AI-based method. Reaching close to machine precision accuracy in predicting solutions to more complex and highly heterogeneous fields will probably require the use of exponentially more training data with diminishing return in accuracy. Coupling the U-Net with the governing PDEs in a concept known as physics-informed neural network (PINN)¹¹ will be helpful in improving the accuracy of predictions as well as quantifying their quality. This aspect as well as extending the model to multiple stress / strain components and general boundary conditions are work in progress and will be reported in subsequent publications.

METHODS

Short summary of large deformation elasto-plastic mechanics

Here we present a concise summary of large-deformation elasto-plastic solid mechanics, as implemented in the Düsseldorf Advanced Material Simulation Kit (DAMASK)⁸. For a complete description as well as the details of the different numerical implementation, parameter identification and solution algorithms, the reader is referred to the original papers^{8,12}. Assuming large deformations, neglecting inertial and body forces, the strong form of the mechanical equilibrium in a continuous domain is

$$\begin{aligned} \text{Div } \mathbf{P} &= \mathbf{0} \text{ in } \Omega_0 \\ \mathbf{P} &= \mathbf{P}_{BC} \text{ on } \Gamma_P \\ \mathbf{u} &= \mathbf{u}_{BC} \text{ on } \Gamma_u \end{aligned} \quad (1)$$

where \mathbf{P} is the first Piola–Kirchhoff (PK) stress tensor and Ω_0 is the volumetric domain. The above partial differential equation (PDE), together with the boundary conditions (prescribed either as traction or displacement on non-overlapping surfaces Γ_P and Γ_u of Ω_0), describe the strong form of the mechanical equilibrium. Under external or internal loads, the domain will deform and the material points will move from their reference positions \mathbf{X} to their current positions \mathbf{x} through the deformation field χ . In the context of large-strain solid mechanics, the deformation gradient $\mathbf{F} = \text{Grad } \chi$ is multiplicatively decomposed into elastic (\mathbf{F}_e) and plastic (\mathbf{F}_p) parts

$$\mathbf{F} = \mathbf{F}_e \mathbf{F}_p \quad (2)$$

with the plastic flow rule

$$\dot{\mathbf{F}}_p = \mathbf{L}_p \mathbf{F}_p. \quad (3)$$

In a first approach, we assume that the material undergoes isotropic plastic deformation when loaded beyond its yielding point. This means that the inelastic deformation is governed by the second invariant of the deviatoric stress tensor, referred to as J_2 . In other words, tensorial directionality is reduced to a scalar stress value (which is typically calculated through the von Mises equivalent stress measure) to which the strength of the material (plastic resistance against inelastic deformation) is compared. Under the assumption of such an isotropic J_2 plastic material response model, the velocity gradient \mathbf{L}_p is formulated as

$$\mathbf{L}_p = \dot{\gamma}_p \frac{\mathbf{S}^{\text{dev}}}{\|\mathbf{S}^{\text{dev}}\|} \quad (4)$$

with $\mathbf{S}^{\text{dev}} = \mathbf{S} - \text{tr } \mathbf{S} / 3$ the deviatoric part of the second PK stress (\mathbf{S}) and

$$\dot{\gamma}_p = \dot{\gamma}_0 \left(\frac{S_{VM}}{S_y} \right)^n \quad (5)$$

the plastic strain rate. $\dot{\gamma}_0$ and S_y in Eq. (5) are the initial strain rate and yield stress, respectively. n is the plastic yielding exponent that describes the rate sensitivity and

$$S_{VM} = \sqrt{\frac{3}{2}} \|\mathbf{S}^{\text{dev}}\| \quad (6)$$

the von Mises stress equivalent measure of \mathbf{S} . Once the total deformation is decomposed into its plastic and elastic parts through an iterative solution of Eqs. (2)–(5) as explained in⁸, the stress is calculated by the generalized Hooke's law

$$\mathbf{S} = \mathbf{C} : \mathbf{E} \quad (7)$$

where

$$\mathbf{E} = \frac{1}{2} (\mathbf{F}_e^T \mathbf{F}_e - \mathbf{I}) \quad (8)$$

is the Green strain tensor and \mathbf{S} is the second PK stress, related to \mathbf{P} as

$$\mathbf{P} = \det \mathbf{F}_p \mathbf{F}_e \mathbf{S} \mathbf{F}_p^{-T}. \quad (9)$$

The first PK stress introduced above, should satisfy the mechanical equilibrium in Eq. (1), which is solved numerically in DAMASK with the help of a spectral solver^{13,14}. The non-linear systems of equations outlined above are solved in iterations until convergence is achieved. These solutions serve here as a reference for training as well as for assessing the quality and the predictive capability of the machine learning method.

In this work, we investigate two constitutive test cases of purely elastic and elasto-plastic materials with perfect plastic (zero strain hardening) response. For simplicity, only isotropic elasticity is considered here. These constitutive assumptions reduce the material properties to two (three) parameters of Young's modulus Y , Poisson's ratio ν (and yield stress S_y) in case of elastic (elasto-plastic) material, respectively. The mechanical heterogeneity of the material is then mapped as a topological aggregate (mimicking a polycrystal) where each domain assumes a set of different material parameter values in the ranges of [60, 120] GPa, [0.1, 0.4] and [50, 200] MPa for Y , ν and S_y , respectively. Note that although these property values are not based on any specific material, they roughly correspond to ductile polycrystalline metals.

A machine learning approach based on U-Net

Using machine learning and specifically deep neural networks has become ubiquitous in material science (see Refs. 15–26 for a review). Most of the current ML related innovations in materials science and engineering successfully aim at accelerated material discovery^{27–30}, efficient interatomic potential development^{31–34} or feature identification in complex pattern that have relevance for materials performance^{35–39}. However, we show here that ML can also help to fundamentally change the way how we solve (non-linear) partial differential equation systems in conjunction with advanced constitutive laws that describe complex material microstructures, much faster than via classical finite element or spectral solvers^{11,40}. This tenet change has the potential to revolutionize continuum-based simulations of materials, allowing a substantial enhancement in the accessibility of materials and topologies of high complexity, size and speed, to quantitative predictions.

Introducing a deep learning alternative for a physics-based solution scheme should be carried out with stringent assessment of the performance in quality of the solutions (see Refs. ^{41,42} for a review on challenges of introducing AI as a tool in material science). The necessity of a quantitative quality assessment of the predictions, among other factors, is motivated by a specific common challenge associated with all deep learning approaches: unlike in conventional solver schemes, where the solution is directly built upon the fundamental governing equations (e.g., Eq. (1)), in deep learning, the network only learns to *reproduce* the correct output based the training data. The training data-set has the correct value of output which are calculated by the physics-based conventional solver (hence, theory-trained AI⁴⁰), and this establishes an indirect way for the network to be exposed to the physics of the problem. In other words, the neural network learns a mapping between the inputs and their corresponding outputs which are calculated by a physics-based solver. Although the network does not directly incorporate the physical laws, it can mimic the outputs which are based on these physical laws. This means that deep learning is for such tasks well equipped for digesting and reproducing pattern features very efficiently and relate patterns to topologies, but this does not per se include gaining insights into the underlying physics-related origin of certain features of such patterns (except for the topological ones). Thus, when using deep learning primarily as an efficient solver rather than as a physics tool seems to exploit its biggest strength in the current context. This lack of direct inclusion of the physical laws (and in general human knowledge) motivated invention of procedures to shed light on how the neural networks make their predictions; these methods are commonly referred to as explainable AI methods (for example see Refs. ^{43–46}). In this article, unlike the common approaches in explainable AI (e.g., Ref. ^{47,48}), we develop an understanding about the relationship between the inner structure of the proposed neural network and its capability in finding an approximate solution surrogate to the posed boundary value problem by using a number of tests. Utilizing these findings, we make connections between details of the machine learning solution to the well defined solution algorithms used for solving PDE with conventional spectral solvers.

Neural networks vary in their basic neural units, the arrangement of these units in the layers and their connectivity, the character of the loss functions (e.g. terms of the quantification of the deviations of the predicted values relative to the reference values in the training data), and the “reductionist” spirit of the network design (e.g., see Refs. ^{49–51}). This diversity in possible architecture details is reflective of the diversity of the applications. For applications with specific characteristics (and constraints) one might require to develop a network topology and network workflow that matches the problem solving at hand more adequately. Examples of such architectures which have been tailor-made for specific physics- and materials science applications can be found in Refs. ^{52,53}.

In this article, our goal is to develop and to use a neural network architecture which is capable of estimating the local stress fields in heterogeneous media exposed to external loads, using only the local material properties and the microstructure topology (e.g., the domain, grain or phase structure) as input information. This requires the network to be able to (i) capture local features efficiently, and also (ii) have the same dimensions for inputs and outputs. These are also requirements for many computer vision problems like segmentation and object detection. Therefore, as a first step, we implement one of the most common neural network architectures in computer vision which is known as the “U-Net”⁵⁰. U-Net was originally developed for biological image segmentation and has since been one of the most commonly used architectures in solving computer vision problems.

Another consideration, which favors using the U-Net architecture, arises from the main equation to be solved, i.e., Eq. (1). The operator in this equation is a derivative, which could be captured effectively using the convolutional layers of the U-Net. With all the above mentioned considerations, we find adapting this architecture particularly promising for the application of solving the mechanical equilibrium and local material response in solid mechanics.

Generating simulation data

We used Voronoi tessellation to create 1000 two-dimensional random geometries with 20 different domains in each simulation box. Note that the difference between each domain is not the crystallographic orientation in our current training set-up, but the elastic and plastic material properties are different in each domain. For simplicity, an isotropic plasticity model (and not more advanced models such as crystal plasticity) is employed in

the elasto-plastic case. The size of the simulation box is $256 \times 256 \times 1$ grid points. Two cases are considered, one with isotropic elasticity and a second one with elastic-ideal-plastic response (i.e., no strain hardening). The material properties of each domain are assigned randomly from the sets listed in Table 1.

A random choice of Young’s modulus Y and Poisson’s ratio ν from Table 1 (elastic column) is assigned to the domains (which mimic grains) in the 1000 different geometries that were generated for conducting the full-field simulations in DAMASK. The DAMASK simulations then serve as training data for the U-Net. For the case of elasto-plastic material response, in addition to Y and ν taken from the elasto-plastic column in Table 1, we assigned different yield stress values, S_y , to each domain of the simulation box as well. Furthermore, in case of elasto-plastic material response, $n = 20$ and $\dot{\gamma}_0 = 10^{-3} \text{ s}^{-1}$ are adopted as parameter values in the plastic strain rate equation, i.e., Eq. (5). These parameters result in a perfect plastic model (zero strain hardening) with different yield stresses in adjacent domains (i.e., grains).

Once the geometry and the material properties have been assigned, all simulation boxes (i.e., the polycrystal aggregates) were loaded to a tensile strain of $E_{xx} = 0.01$ or $E_{xx} = 0.001$ for elastic or elasto-plastic cases, respectively (see Fig. 7 for coordinates). The elasto-plastic load was set to a modest value that fell in the middle of the range of the yield strength values in order to enhance complexity, i.e., to obtain a mixture of domains in elastic and elasto-plastic regimes. Furthermore, note that the minimum yield stress in the training data is only 50 MPa, therefore the applied external load in the elasto-plastic case is lower compared to the purely elastic case. Stress components in all other normal directions (i.e., yy and zz) as well as deformation in the off-diagonal components were set to zero. In case of elastic material behavior, the load was applied instantaneously in one step, while in the elasto-plastic case, it was applied sequentially over 100 steps. The calculations were performed using the open-source micromechanical simulation software package DAMASK⁸. Note that even in case of the linear elastic model, the PDEs are not linear due to the non-linear nature of the underlying strain model presented in Eq. (8) that is capable of modeling both large deformations and rigid rotations. Although the applied strain is small (1%), the local strains are higher due to stress

Table 1. Sets of the material properties used for generating the training data.

	Elastic	Elasto-plastic
Y (GPa)	{80, 90, 100, 110, 120}	{60, 80, 100, 120}
ν	{0.1, 0.2, 0.3, 0.4}	{0.1, 0.2, 0.3, 0.4}
S_y (MPa)	–	{50, 100, 150, 200}

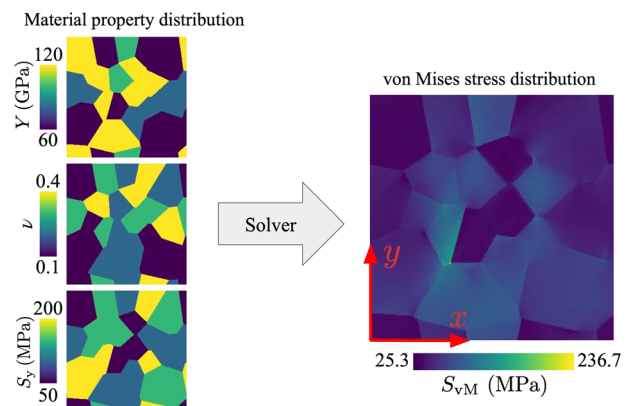


Fig. 7 Schematics of the material property-response relationship. Example of a typical set of maps of an input material property distribution (left) and corresponding output in terms of von Mises equivalent stress distribution (right), calculated with the open-source micromechanical simulation software package DAMASK⁸. The uniaxial tensile load is applied in x direction. The full-field forward simulations serve as training data for the U-Net-based machine learning solver.

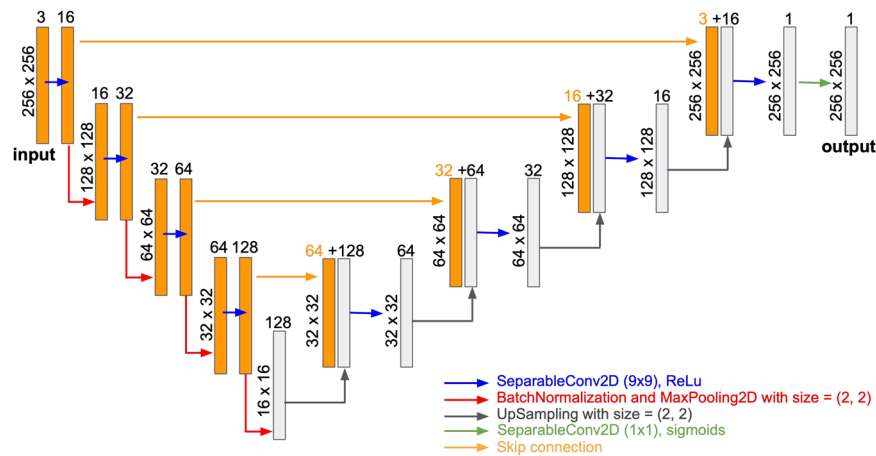


Fig. 8 Architecture of the U-Net used in this work. Here, the notation of Tensorflow⁵⁵ is adopted for naming the layers. The layers consist of separable convolutional layers (SeparableConv2D) with either a rectified linear unit (ReLU) or sigmoid activation functions to extract the features and apply the non-linearities, batch normalization (BatchNormalization) to transform the layers' outputs to a mean value of zero and a standard deviation of 1, max pooling (MaxPooling2D) for coarse-graining, and up sampling (UpSampling2D) for going from the coarse grained image to a high resolution one. The skip connections send the images from each contracting step to its expanding counterpart.

concentration at the domain junctions and related topological features of high mechanical contrast. However, here we restrict the investigation to the domain-to-domain differences in a near-linear heterogeneous system of an elastic material exposed to small deformations and for the non-linear case of elasto-plastic behavior.

The inputs to the machine learning algorithm are the distributions of the material properties (Y, ν in case of elastic and Y, ν, S_y in case of elasto-plastic response, respectively), represented as color coded images. The output or respectively target of the ML training lies in the prediction of the von Mises equivalent values of the second PK stress, S_{VM} as defined in Eq. (6), which is plotted at the end of each simulation as a color coded image. Note that each grid point in the simulation is represented by one pixel in the input/output images (i.e., all images are of size 256×256). Consequently, the simulation domain is restricted to this size. An example of a set of such input and output images are shown in Fig. 7.

For better visualization for readers, the simulated maps are in this paper color coded using the viridis or seismic color maps. However, note that the neural network is trained and working with monochrome images for each of the scalar fields.

Network architecture

The network architecture used here is an adaption of the U-Net⁵⁰, see Fig. 8. The network has two distinct parts, a contracting and an expansive part. The contracting part consists of $N_s = 4$ steps where each step is a repeated stacking of contracting modules where each module consists of a convolutional layer, a non-linear activation function, and a batch normalization layer, which is followed by a downsampling layer. The convolutional layer increases the number of channels and the downsampling decreases the number of rows and columns of the data (i.e. the width and height of the image). After these N_s steps conducted in the contracting part, the expansive part of the network starts, which also consists of N_s steps. Every step in the expansive part is similar to the contracting part except that in this case the convolutional layers decrease the number of channels and the downsampling is replaced by upsampling which increases the number of rows and columns. Following the U-Net architecture, in each downsampling step, the width and the height of the image are reduced by a factor of two. Corresponding to each downsampling step, an upsampling step is implemented where the scaling factor is again two. As the number of downsampling steps equals that of the upsampling steps, the output data have the same dimensions as the input. This architecture also enables us to pass the information of each step of the downsampling sequence directly to the corresponding upsampling part, as it is done in the original version of U-Net as well⁵⁰. The last contraction step is followed by a convolutional layer which reduces the number of channels to 1 and applies a sigmoid function which maps the values into the range $[0, 1]$. In spite of its overall similarity to the original U-Net structure, the variant of the neural network that we use here deviates from the original design, as explained in the following.

First and foremost, the kernel size of the convolutional layer in our model ($k = 9$) is much larger than the kernel size used regularly in U-Net (i.e., $k = 3$). Here, a kernel of size k refers to a $k \times k$ matrix whose elements are to be learned during the training of the network. The kernel, i.e., specifically its matrix size, is a key feature of convolutional neural networks that is applied step-wise in the form of a sliding array across pattern data. This process has the aim to amplify, identify and extract certain features from an input image. It is usually coded in the form of a simple matrix (of much lower rank compared to the image size), that is sequentially slid across the image and multiplied at each sequential position with a subset of the input array such that the output enhances certain topological pattern features such as edges, corners, gradients, etc. The rationale behind our choice for increasing the kernel size from 3 to 9 is that a larger kernel size can lead to a more accurate derivative estimate. In fact, as shown in the results section, we observe that for small kernel sizes the network's capability to predict the results is significantly reduced. The analysis shows that the kernel matrix size is indeed an essential parameter for rendering this approach a viable solver alternative, capable for instance of picking up local stress peak and stress gradient features. These local stress features are related to the mechanical contrast variations which are characteristic of heterogeneous materials.

Another important modification which we implemented in our version of the U-Net lies in replacing the conventional convolutional layers by separable convolutional layers. In contrast to the convolutional layer used in the original U-Net design, where the outputs obtained by applying the kernel to each channel are summed up for all the channels, in our approach the outputs are added together with weights that are learned. Using an unweighted sum is a more suitable choice for segmentation applications as one could expect high correlation between different channels of the input, e.g., between the red, green, and blue colors of an image. For our application, however, we are not expecting such a high correlation between the different material properties (i.e., between the Young's modulus, Poisson's ratio, and the yield stress). The depth of the U-Net (i.e., the number of contraction/expansion steps) and also the number of channels are modified in our implementation. Effects that stem from an increase in the depth of the U-Net are studied systematically and discussed in the results section.

The input data: random microstructures

The input data include in the present case the spatial distribution of the material properties, including the local Young's modulus Y , the local Poisson ratio ν , and in the case of the elasto-plastic behavior, the local yield strength values S_y . We arrange this information by stacking (w, h) arrays of Y and ν to form a $(w, h, 2)$ array, with $w = 256$ and $h = 256$. For the elasto-plastic problems, an extra channel is added which contains the yield stress values S_y , as an isotropic measure of the material's resistance against inelastic shape change. We also introduce an additional channel for Y/ν . The choice of this additional feature is motivated by the common form of

the solutions emerging from elasticity theory, where Y and ν appear in mixed terms of Y/ν or $Y/(1+\nu)$. The values in the channels are shifted and scaled such that they all fall into the range [0, 255].

The output data: von Mises stress distribution

Using the input data, i.e., the scaled material properties, we aim to predict the distribution of the von Mises stress introduced in Eq. (6), again in scaled form. The von Mises stress value is an equivalent stress measure which reduces a tensorial form to a scalar surrogate for isotropic cases. In general, we are here focusing on the von Mises stress measure as it is a first order parameter that allows to reveal the most important key features of the mechanical heterogeneity in isotropic media subjected to elastic-plastic loads.

Note that in our AI-approach, we do not calculate the von Mises stress by separate prediction of the individual stress components, but we calculate it directly in an end-to-end approach, i.e., using the inputs mentioned above. We speculate that calculation of the individual stress tensor components would be more straightforward for the neural network, but this would require the use of more output channels. For highly mechanically anisotropic cases it would be however pertinent to calculate the components of the stress tensor individually (and extract from these also secondary measures such as equivalent stress values).

Training the network

The neural network is implemented, trained, and tested using Keras⁵⁴ which is an application programming interface (API) written in Python, running on top of the machine learning platform TensorFlow⁵⁵. The 1000 samples are split randomly into the training and test sets of size 950 and 50 samples, respectively. To train the network, the mean absolute error is used as the loss function (which should be minimized). The minimization of the loss function is done using the ADAM optimizer which is a stochastic first-order gradient-based method⁵⁶. We use random samples in batches of size 32 for the gradient estimation, and continue training for 400 epochs, i.e., 400 complete passes through the whole training dataset. We set the hyperparameters of the ADAM optimizer to $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$, and a learning rate within the range $[1, 5] \times 10^5$. The training is carried out for 400–800 epochs. During the training the mean absolute error of ≈ 0.02 is achieved without any significant sign of over-fitting.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The code used for machine learning in this study is open-source and accessible in github.com/tensorflow. The Düsseldorf Advanced Material Simulation Kit (DAMASK) is also open-source and accessible at damask.mpie.de.

Received: 20 March 2021; Accepted: 3 June 2021;

Published online: 01 July 2021

REFERENCES

- Aydin, R. C., Braeu, F. A. & Cyron, C. J. General multi-fidelity framework for training artificial neural networks with computational models. *Front. Mater.* **6**, 61 (2019).
- Cecen, A., Dai, H., Yabansu, Y. C., Kalidindi, S. R. & Song, L. Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Mater.* **146**, 76–84 (2018).
- Fernández, M., Rezaei, S., Mianroodi, J. R., Fritzen, F. & Reese, S. 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 (2020).
- Li, K., Tang, K., Li, J., Wu, T. & Liao, Q. A hierarchical neural hybrid method for failure probability estimation. *IEEE Access* **7**, 112087–112096 (2019).
- Yang, Z., Yu, C.-H. & Buehler, M. J. Deep learning model to predict complex stress and strain fields in hierarchical composites. *Sci. Adv.* **7**, eabd7416. <https://doi.org/10.1126/sciadv.abd7416> (2021).
- Wang, H., Planas, R., Chandramowlishwaran, A. & Bostanabad, R. Train once and use forever: Solving boundary value problems in unseen domains with pre-trained deep learning models. Preprint at <https://arxiv.org/abs/2104.10873> (2021).
- Pandey, A. & Pokharel, R. Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions. *Scr. Mater.* **193**, 1–5 (2021).
- Roters, F. et al. DAMASK - the düsseldorf advanced material simulation kit for modeling multi-physics crystal plasticity, thermal, and damage phenomena from the single crystal up to the component scale. *Comput. Mater. Sci.* **158**, 420–478 (2019).
- Willot, F. Fourier-based schemes for computing the mechanical response of composites with accurate local fields. *C.R. Mecanique* **343**, 232–245 (2015).
- Khorrami, M., Mianroodi, J. R., Shanthraj, P. & Svendsen, B. Development and comparison of spectral algorithms for numerical modeling of the quasi-static mechanical behavior of inhomogeneous materials. Preprint at <https://arxiv.org/abs/2009.03762> (2020).
- Raissi, M., Perdikaris, P. & Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
- Roters, F. et al. Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. *Acta Mater.* **58**, 1152–1211 (2010).
- Shanthraj, P., Eisenlohr, P., Diehl, M. & Roters, F. Numerically robust spectral methods for crystal plasticity simulations of heterogeneous materials. *Int. J. Plast.* **66**, 31–45 (2015).
- Eisenlohr, P., Diehl, M., Lebensohn, R. & Roters, F. A spectral method solution to crystal elasto-viscoplasticity at finite strains. *Int. J. Plast.* **46**, 37–53 (2013).
- Sha, W. et al. Artificial intelligence to power the future of materials science and engineering. *Adv. Intell. Syst. Comput.* **2**, 1900143 (2020).
- Ye, W., Chen, C., Wang, Z., Chu, I. H. & Ong, S. P. Deep neural networks for accurate predictions of crystal stability. *Nat. Commun.* **9**, 1–6 (2018).
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
- Schmidt, J., Marques, M. R., Botti, S. & Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019). <https://doi.org/10.1038/s41524-019-0221-0>
- Ramprasad, R., Batra, R., Piliñia, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: Recent applications and prospects. *Npj Comput. Mater.* **3**, 54 (2017). <https://doi.org/10.1038/s41524-017-0056-5>
- Niezgoda, S. R., Kanjarla, A. K. & Kalidindi, S. R. Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integr. Mater. Manuf. Innov.* **2**, 54–80 (2013).
- McDowell, D. L. & Lesar, R. A. The need for microstructure informatics in process-structure-property relations. *MRS Bull.* **41**, 587–593 (2016).
- Bereau, T., Andrienko, D. & Kremer, K. Research update: computational materials discovery in soft matter. *APL Mater.* **4**, 053101 (2016).
- Wodo, O., Broderick, S. & Rajan, K. Microstructural informatics for accelerating the discovery of processing-microstructure-property relationships. *MRS Bull.* **41**, 603–609 (2016).
- Agrawal, A. & Choudhary, A. Perspective: materials informatics and big data: Realization of the fourth paradigm of science in materials science. *APL Mater.* **4**, 053208 (2016).
- Dimiduk, D. M., Holm, E. A. & Niezgoda, S. R. Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr. Mater. Manuf. Innov.* **7**, 157–172 (2018).
- Masi, F., Stefanou, I., Vannucci, P. & Maffi-Berthier, V. Thermodynamics-based artificial neural networks for constitutive modeling. *J. Mech. Phys. Solids* **147**, 104277 (2021).
- Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
- Meftahi, N. et al. Machine learning property prediction for organic photovoltaic devices. *Npj Comput. Mater.* **6**, 1–8 (2020).
- Sun, W. et al. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. *Sci. Adv.* **5**, 1–9 (2019).
- Dai, D. et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. *Comput. Mater. Sci.* **175**, 109618 (2020).
- Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
- Kolb, B., Lentz, L. C. & Kolpak, A. M. Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods. *Sci. Rep.* **7**, 1192 (2017).
- Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).

34. Dragoni, D., Daff, T. D., Csányi, G. & Marzari, N. Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron. *Phys. Rev. Mater.* **2**, 013808 (2018).
35. Jiang, Y. et al. Topological representations of crystalline compounds for the machine-learning prediction of materials properties. *Npj Comput. Mater.* **7**, 1–8 (2021).
36. Adachi, Y., Taguchi, M. & Hirokawa, S. Microstructure recognition by deep learning. *Tetsu-to-Hagane* **102**, 722–729 (2016).
37. DeCost, B. L., Francis, T. & Holm, E. A. Exploring the microstructure manifold: Image texture representations applied to ultrahigh carbon steel microstructures. *Acta Mater.* **133**, 30–40 (2017).
38. DeCost, B. L. & Holm, E. A. A computer vision approach for automated analysis and classification of microstructural image data. *Comput. Mater. Sci.* **110**, 126–133 (2015).
39. DeCost, B. L. & Holm, E. A. Characterizing powder materials using keypoint-based computer vision methods. *Comput. Mater. Sci.* **126**, 438–445 (2017).
40. Rad, M. T., Viardin, A. & Apel, M. On theory-training neural networks to infer the solution of highly coupled differential equations. Preprint at <https://arxiv.org/abs/2102.04890> (2021).
41. DeCost, B. L. et al. Scientific AI in materials science: a path to a sustainable and scalable paradigm. *Mach. Learn.: Sci. Technol.* **1**, 033001 (2020).
42. Vasudevan, R. K., Ziatdinov, M., Vlcek, L. & Kalinin, S. V. Off-the-shelf deep learning is not enough, and requires parsimony, bayesianity, and causality. *Npj Comput. Mater.* **7**, 1–6 (2021).
43. Doran, D., Schulz, S. & Besold, T. R. What does explainable AI really mean? A new conceptualization of perspectives. Preprint at <https://arxiv.org/abs/1710.00794> (2017).
44. Miller, T., Howe, P. & Sonenberg, L. Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences. Preprint at <https://arxiv.org/abs/1712.00547> (2017).
45. Preece, A., Harborne, D., Braines, D., Tomsett, R. & Chakraborty, S. Stakeholders in explainable AI. Preprint at <https://arxiv.org/abs/1810.00184> (2018).
46. Arrieta, A. B. et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).
47. Bach, S. et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **10**, e0130140 (2015).
48. Arras, L., Horn, F., Montavon, G., Müller, K.-R. & Samek, W. “What is relevant in a text document?”: An interpretable machine learning approach. *PLoS One* **12**, e0181142 (2017).
49. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
50. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Med. Image. Comput. Comput. Assist. Interv.*, 234–241 (2015).
51. Lechner, M. et al. Neural circuit policies enabling auditable autonomy. *Nat. Mach. Intell.* **2**, 642–652 (2020).
52. Linka, K. et al. Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. *J. Comput. Phys.* 110010 (2020).
53. Mohan, A. T., Lubbers, N., Livescu, D. & Chertkov, M. Embedding hard physical constraints in neural network coarse-graining of 3d turbulence. Preprint at <https://arxiv.org/abs/2002.00021> (2020).
54. Chollet, F. et al. Keras. https://keras.io/getting_started/faq/#how-should-i-cite-keras (2015).
55. Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems (2015). <https://www.tensorflow.org/>. Software available from tensorflow.org.
56. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).

ACKNOWLEDGEMENTS

We thank Martin Diehl for constructive discussions about the evaluation of the speed-up and comparison with conventional solvers.

AUTHOR CONTRIBUTIONS

J.R.M. and N.H.S. developed the initial concept and model formulation. J.R.M. carried out the solid mechanics calculations to generate the training and test dataset. N.H.S. designed the neural network and carried out the training. J.R.M. and N.H.S. wrote the initial draft. D.R. supervised the work and helped in interpretation of the results. All authors discussed and modified the paper together.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to J.R.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021