



## Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets



Zijiang Yang<sup>a</sup>, Yuksel C. Yabansu<sup>b</sup>, Reda Al-Bahrani<sup>a</sup>, Wei-keng Liao<sup>a</sup>, Alok N. Choudhary<sup>a</sup>, Surya R. Kalidindi<sup>b,c</sup>, Ankit Agrawal<sup>a,\*</sup>

<sup>a</sup> Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA

<sup>b</sup> George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

<sup>c</sup> School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

### ARTICLE INFO

**Keywords:**

Materials informatics  
Convolutional neural networks  
Deep learning  
Homogenization  
Structure-property linkages

### ABSTRACT

Data-driven methods are emerging as an important toolset in the studies of multiscale, multiphysics, materials phenomena. More specifically, data mining and machine learning methods offer an efficient toolset for extracting and curating the important correlations controlling these multiscale materials phenomena in high-value reduced-order forms called process-structure-property (PSP) linkages. Traditional machine learning methods usually depend on intensive feature engineering, and have enjoyed some success in establishing the desired PSP linkages. In contrast, deep learning approaches provide a feature-engineering-free framework with high learning capability. In this work, a deep learning approach is designed and implemented to model an elastic homogenization structure-property linkage in a high contrast composite material system. More specifically, the proposed deep learning model is employed to capture the nonlinear mapping between the three-dimensional material microstructure and its macroscale (effective) stiffness. It is demonstrated that this end-to-end framework can predict the effective stiffness of high contrast elastic composites with a wide range of microstructures, while exhibiting high accuracy and low computational cost for new evaluations.

### 1. Introduction

An important mission of the field of materials science is to design new/improved materials that can meet the stringent demands placed by emerging advanced technologies. The paradigm of process-structure-property (PSP) linkages plays a central role in capturing and curating the high value materials knowledge needed in this pursuit[1–11]. The extraction and deployment of these linkages has been hindered by the high dimensional representations needed for a rigorous description of the inherently heterogeneous material structure spanning multiple length or internal structure scales. Indeed, the precise physics-based connections between the material structure and its associated properties are very complex. However, from a practical viewpoint of materials design, it is imperative that we capture the high value information in these complex linkages in forms that allow computationally efficient explorations of the extremely large design spaces. Broadly speaking, PSP linkages can be cast in both directions of scale-bridging: (i) homogenization (going from smaller scales to larger scales) [12–14] and (ii) localization (going from larger scales to smaller scales)

[11,15–20]. Our focus here will be on homogenization, i.e., prediction of macroscale elastic properties of a high contrast composite given its microstructure information. Contrast in this context refers to the differences in the individual properties of the microscale constituents present in the material microstructure.

The conventional approaches for establishing structure-property linkages in composite materials have relied either on highly sophisticated analytical approaches based on statistical continuum theories [21–23] or on numerical approaches based on finite element (FE) models. Although the statistical continuum theories are very attractive because of their low computational cost (especially significant in exploring large design spaces), progress in this direction has been largely hindered by the need to establish accurately the Green's functions based kernels used in these theories, and the slow convergence of the series expansions for high contrast composites [24,25]. While the numerical approaches such as the finite element models circumvent these challenges effectively, they are not best suited for design explorations of the potentially very large materials space (i.e., solving inverse problems identifying the specific microstructures meeting a designer specified set

\* Corresponding author.

E-mail addresses: [zijiangyang2016@u.northwestern.edu](mailto:zijiangyang2016@u.northwestern.edu) (Z. Yang), [yabansu@gatech.edu](mailto:yabansu@gatech.edu) (Y.C. Yabansu), [redaalbahrani2012@u.northwestern.edu](mailto:redaalbahrani2012@u.northwestern.edu) (R. Al-Bahrani), [wkliao@eecs.northwestern.edu](mailto:wkliao@eecs.northwestern.edu) (W.-k. Liao), [choudhar@eecs.northwestern.edu](mailto:choudhar@eecs.northwestern.edu) (A.N. Choudhary), [surya.kalidindi@me.gatech.edu](mailto:surya.kalidindi@me.gatech.edu) (S.R. Kalidindi), [ankitag@eecs.northwestern.edu](mailto:ankitag@eecs.northwestern.edu) (A. Agrawal).

of desired property combinations [4,26,27]).

In recent years, data-driven approaches have attracted the attention of materials science researchers [28–35]. A new framework called Materials Knowledge Systems (MKS) [9,7,15,16,18,36,37] was formulated to take advantage of the relative merits of both the analytical and the numerical approaches described above in formulating structure-property linkages. In this data-driven approach, one first aggregates a sufficiently large ensemble of data points using the numerical approaches, where each data point includes information on the material microstructure (treated as input) and its effective property of interest (treated as output). After establishing the data set, one then calibrates the Green's function based kernels in the statistical continuum theory series expansions to the aggregated data set using suitable regression techniques. This approach has been shown to synergistically combine the respective merits of both the analytical and numerical approaches described earlier, and provide remarkably accurate, low computational cost, structure-property linkages for low to medium contrast composites [7,9,11,15,18,38]. Although the viability of the MKS approach has also been demonstrated for high contrast composites, there continue to be significant hurdles as the application to the high contrast composites requires feature engineering. In this regard, it is noted that feature engineering (i.e., selection of the important microstructure features influencing the effective property of interest) in the context of the MKS framework has been explored mainly using the framework of n-point spatial correlations and principal component analyses [6,37,39–43]. While the current feature engineering approach in the MKS framework was demonstrated to be highly successful in the consideration of the 2-point spatial correlations, its extension to include higher-order spatial correlations is nontrivial. This is mainly because of the explosion in the number of spatial correlations as one goes up systematically to the higher order spatial correlations.

In recent years, deep learning approaches have emerged as the methods of choice in addressing the problem of automated identification of features from an extremely large set of potential features. These methods have enjoyed successes in a broad range of application domains including computer vision (e.g., image segmentation, image classification and face recognition) [44–50]. This emerging new approach significantly outperforms traditional machine learning methods in its ability to learn the embedded model in an aggregated dataset. More specifically, deep learning approaches provide an end-to-end framework where an explicit feature design is not required. Consequently, the trained models usually exhibit higher generalization. Thus, deep learning approaches exhibit tremendous potential for addressing some of the main hurdles in materials research. In [51] Liu et al. applied deep convolutional neural networks to model a large image data collection of polycrystal electron patterns. Liu et al. [52] used deep neural networks to understand the relationship between the composition and the properties of materials. In [53], Li et al. implemented transfer learning approach to reconstruct material microstructures. Cang et al. [54] developed a convolutional deep belief network to automate conversion between microstructure and corresponding lower-dimensional feature representations. Later, Cang et al. [55] applied Variational Auto-Encoder to generate artificial material samples with same morphology distribution as the authentic ones. In [56,57], Yang et al. and Li et al. developed a Generative Adversarial Networks to identify the key microstructure representations and implemented it to design material microstructure with desired properties. Gopalakrishnan et al. [58] applied transfer learning technique to detect crack in pavement.

The target in this study is to establish structure-property linkages for homogenization of high contrast two-phase elastic composites. Homogenization in hierarchical multiscale modeling refers to transfer of information about the microstructure from a lower length scale to higher length scale. This information is usually expressed as an effective property of the material volume being studied and is calculated through various averaging techniques [12–14]. The main challenge in

calculating the effective stiffness is to solve the governing field equations formulated at the lower length scale. This is a computationally expensive task if one considers the large space of microstructures that needs to be explored. The proposed deep learning approach will address this task by building data-driven structure-property linkages (i.e., reduced-order models or surrogate models) between the 3-D microstructure and the effective elastic stiffness value.

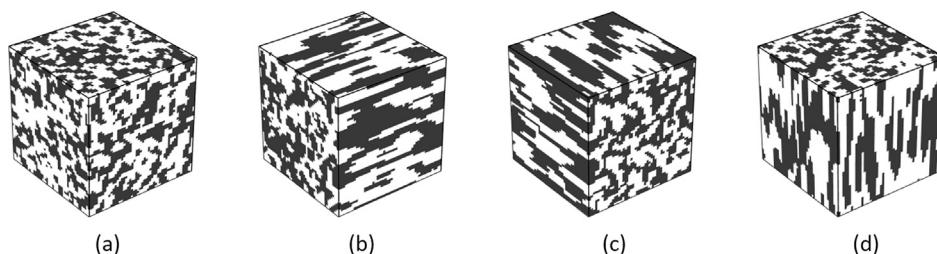
In this paper, we explore the benefits of using deep learning approaches in establishing high value structure-property homogenization linkages for high-contrast elastic 3-D composite microstructures. In a prior work [59], convolutional neural networks (CNN) were employed to build a model that converts the binary microstructure information into a set of filters that serve as higher-order microstructure information. However, this effort was not strongly explored as a completely feature-engineering free approach. In this study, a standalone CNN is built for the first time to establish structure-property linkages for high contrast elastic 3-D composites using a completely feature-engineering free approach. An extensive analysis of convolutional neural networks with different numbers of convolution and pooling layers was performed. The performance of the CNN is compared to structure-property linkages established with simple physics-based approaches and sophisticated physics-inspired approaches (these will be introduced in Section 3.2) employed in our prior work [7,9,60]. It will be shown through error metrics that CNN built in this study outperforms benchmark methods.

## 2. Datasets and methods

### 2.1. Generation of high contrast elastic 3-D datasets

In order to explore and evaluate the performance of CNN models in predicting the effective elastic properties of high contrast composites, we first need to generate a dataset that reflects the ground truth. In this work, because of the lack of a suitable experimental dataset, we assume that the ground truth is reasonably well captured by the results of micromechanical finite element models applied on digitally generated microstructures. Therefore, for this study, we generated 8550 3-D microstructures which are referred to as microscale volume elements (MVEs). The main purpose of these MVEs [7,11,15,18] is to produce the data needed to extract the desired structure-property linkages. They have to be large enough to capture the range of microscale interactions occurring naturally within the microstructural volume element, but small enough to allow for generation and aggregation of the needed data within reasonable computational cost.

The MVEs used in this study were generated by starting with a random assignment of numbers on a uniformly tessellated 3-D spatial (voxelized) grid, following by application of a 3-D Gaussian filter, and finally thresholding to obtain a targeted distribution of volume fractions in the ensemble (i.e., the collection of 8550 MVEs generated for this study). In the effort to generate a rich morphological diversity in the generated set of MVEs, 3-D Gaussian filters with different covariances were employed. The filters were selected in such a way that the MVEs had preferred directionality in three perpendicular directions. Different combinations of diagonal entries in covariance matrix were used to generate MVEs with different amounts of directionality. For this case study, the off-diagonal entries were always kept zero. However, a wider range of diversity in MVEs can be attained by using covariance matrices with non-zero entries in off-diagonal elements as well. Some examples of MVEs with different microstructural details are shown in Fig. 1. The MVE in part (a) is generated by a 3-D Gaussian filter with three identical diagonal entries in the covariance matrix. On the other hand, the microscale constituents of MVEs in part (b), (c) and (d) have clear directionality in x, y and z directions. The degree of directionality of the structural features are controlled with the values of the covariances used with the 3-D Gaussian filters. In total, 57 different 3-D Gaussian filters were employed and each filter is used to generate 150



**Fig. 1.** Visualization of selected MVEs generated by applying different 3-D Gaussian filters to a random number field.

MVEs with volume fractions ranging between 25 to 75 % for the hard phase. 100 MVEs among 150 generated MVEs associated with each Gaussian filter were randomly selected for training process and the rest 50 MVEs are set aside for testing process. Thus, out of the 8550 MVEs, 5700 MVEs were used for training process and 2850 MVEs for testing process. From these 5700 MVEs, 33 % were randomly selected as validation set and the rest as training set. In other words, 8550 MVEs are split into 3 sets, which are training set (3819 MVEs), validation set (1881 MVEs) and testing set (2850 MVEs).

Each MVE is populated with two potential microscale constituents (shown as white and black in Fig. 1, also referred as local states). Both local states are assumed to exhibit isotropic elastic responses. Because of our interest in high contrast composites for the present study, a ratio of 50 was employed between the Young's moduli of the local states. The Young's moduli of hard and soft phases were taken as  $E_1 = 120$  GPa and  $E_2 = 2.4$  GPa, while both phases were assigned the same Poisson ratio,  $\nu = 0.3$ .

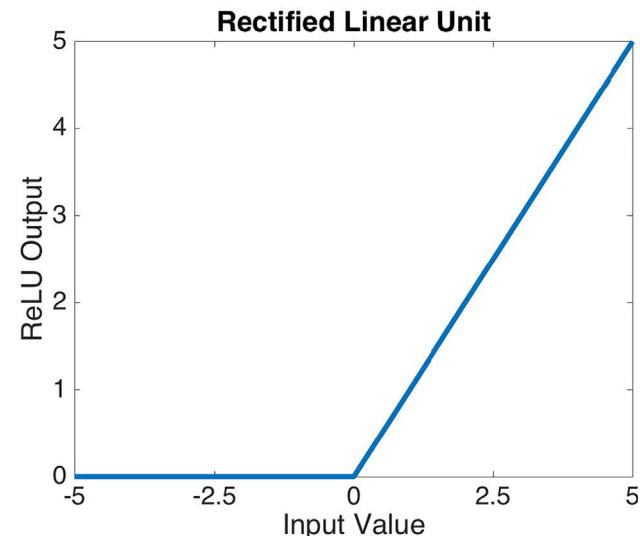
In this study, the effective property of interest was identified as the (1,1,1,1) component of the fourth-rank elastic stiffness tensor (usually denoted as  $C_{1111}$  in the tensorial notation or simply as  $C_{11}$  in a reduced notation). In order to estimate this property, finite element simulations were performed for each MVE. The simulations employed periodic boundary conditions [16] in such a way that the macroscale strain tensor had only one non-zero component,  $\langle \epsilon_{11} \rangle$  (the angled brackets represent the volume average). The overall approach described here can be extended to any other desired component of the effective elastic stiffness tensor (cf. [61]).

## 2.2. Artificial neural network

Artificial neural network is inspired by the biological neural networks. Multilayer perceptron neural network (MLP) [62] is a basic architecture of artificial neural networks. By stacking more layers to form a deep architecture, deep learning architecture has better learning capability. Convolutional neural networks (CNN) [63] is one of the deep learning architectures, which has been shown to be very powerful in solving computer vision problems [44–50,64,65]. In the following sections, further details of MLP and CNN are presented.

### 2.2.1. Multilayer perceptron neural network

An MLP usually consists of at least three layers, and each hidden layer (i.e., layers except input and output layers) consists of multiple neurons. A neuron takes the weighted sum of outputs of every neuron in the previous layer, and then pass the value through an activation function to produce the output. An activation function is generally used to exploit the nonlinear relationship between inputs and outputs. Rectified Linear Unit (ReLU) [66] activation function is one of the most commonly used activation functions. ReLU activation function can be formulated as  $f(z) = \max(0, z)$  and it is illustrated in Fig. 2. Fig. 3(a) illustrates an example of four-layer perceptron neural network. The hidden layers of MLP contain multiple neurons (blue<sup>1</sup> nodes in



**Fig. 2.** Plot illustrating the Rectified Linear Unit (ReLU).

Fig. 3(a)), and each neuron is connected to every neuron in the adjacent layers.

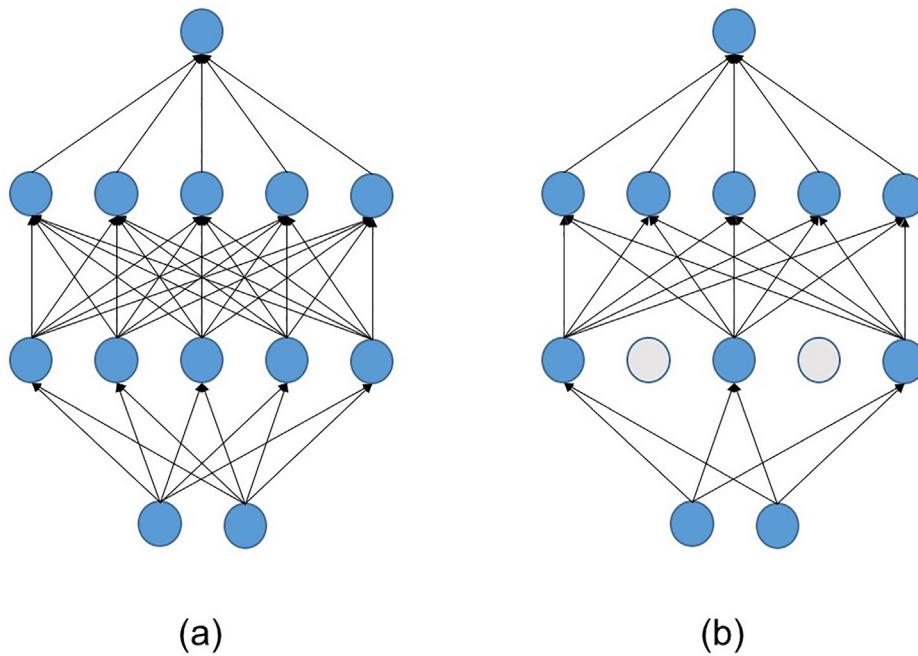
### 2.2.2. Convolutional neural network

A CNN model is usually comprised of three basic operations, with these operations being repeated multiple times. These three unit operations are referred to as a convolution layer, a pooling layer and a fully connected layer in the CNN terminology. Fig. 4 provides a schematic illustration for an example 3-D CNN configuration.

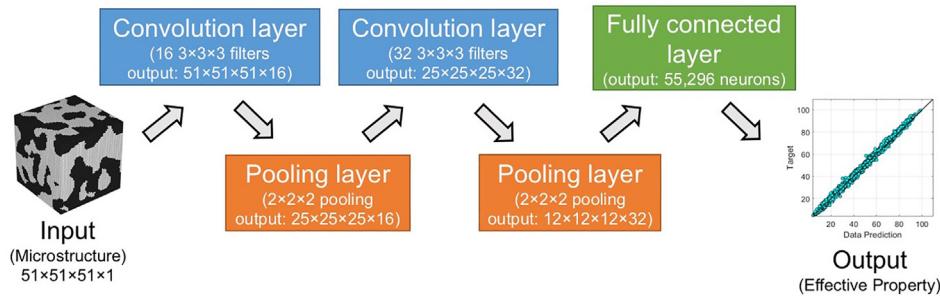
The convolution layer forms the core layer of CNN models, and its goal is to extract important features from the input images objectively. This is accomplished using a set of convolution filters [64] in each convolution layer that are to be systematically learned from the available data. Fig. 5 provides a simple illustration of the computations involved in the application of a convolution filter. In this illustration, a very simple  $3 \times 3$  image (colored brown) is first padded with zeros (colored gray) and then convolved with a  $3 \times 3$  filter (colored blue). Finally, a bias (colored yellow) is added to produce a feature map. Further computational details involved in the convolution layer can be found in many references [45,47,49]. Essentially, the application of each filter involves computing a dot product of the filter weights with an equal sized subsection of the input image centered around the voxel of interest in the input image. Putting together the results of the dot products following the same spatial sequence as the individual voxels considered in the input image produces the output of the convolution. Note that padding the image as described in Fig. 5 allows the output image size to be the same as the size of the input image. It is also important to recognize that a single convolution layer involves the application of multiple filters and bias, each producing a different feature map. The general algorithm described above is equally applicable to 2-D or 3-D images.

Pooling layer is usually employed after one or several stacked

<sup>1</sup> For interpretation of color in Fig. 3, the reader is referred to the web version of this article.



**Fig. 3.** Dropout (a) Conventional neural network. (b) After applying dropout.



**Fig. 4.** Example architecture of 3-D convolutional neural network.

convolution layers. The purpose of the pooling layers is to reduce the dimensionality of feature maps. Fig. 6 shows an example of a commonly used  $2 \times 2$  max-pooling. In this type of pooling, the values in the outputs are taken as the maximum of each  $2 \times 2$  block in the input feature map. This is depicted in Fig. 6 via suitably colored blocks. The concept described above is easily extended to 3-D using  $2 \times 2 \times 2$  pooling.

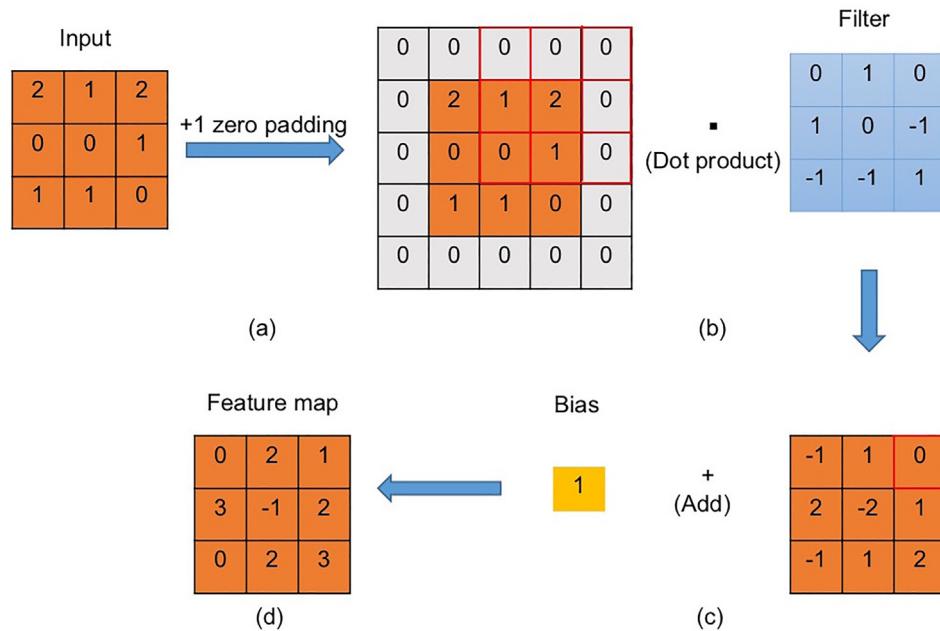
The outputs obtained after employing stacks of convolution and pooling layers (see Fig. 7 for an example) are flattened to a vector. This vector is then fed into a fully connected layer, which is the same as a hidden layer in the conventional multilayer perceptron neural network. Thus, each input of this layer is a voxel from the outputs obtained after a set of convolution and pooling layers. The final layer in CNN is called output layer and the values in it depend on whether the network is established for classification or regression. For classification, the output layer contains as many neurons as the number of classes in the ensemble of datasets and the value of each neuron in the output layer changes between 0 to 1 to reflect the probability of a selected data point belonging to the selected class. On the other hand, in regression, if there is only one target, then the output layer contains one neuron with a continuous value reflecting the predicted output of CNN.

### 2.3. Proposed deep learning architecture for homogenization

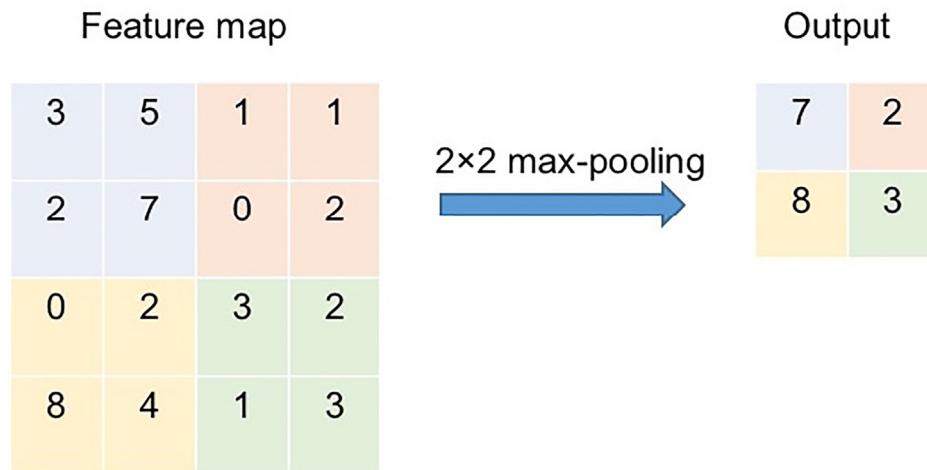
In order to establish homogenization linkages, it is necessary to extract higher-order neighborhood information within three-dimensional microstructures and model the relationship between the

neighborhood information and the effective stiffness of MVEs. Thus, CNN is adapted here to build such homogenization linkages.

In this work, the inputs to CNN are assumed to be  $51 \times 51 \times 51$  three-dimensional microstructures where each voxel is assigned a value of either zero (i.e., hard phase) or one (i.e., soft phase). Since the original data is comprised of binary images with 0s and 1s, applying element-wise convolution operations with the large amount of 0s in the input might significantly deteriorate the functionality of the filters. Therefore, the input data is rescaled from  $[0, 1]$  to  $[-0.5, 0.5]$  before training the models. After rescaling, 3-D CNNs were used to establish the desired homogenization linkage. In this approach, the structure-property linkage is therefore directly trained on the input 3-D microstructure. Among various activation functions, such as ReLU, Leaky ReLU and Softplus, we found that ReLU preserves the gradients in the microstructures better. Thus, in each layer of the CNN, ReLU function is used as activation function to explore the nonlinearity, and weights are initialized by normalized initialization [67]. This initialization method samples a  $U[-r, r]$  with  $r = \sqrt{\frac{6}{fan.in + fan.out}}$  where  $fan.in$  and  $fan.out$  are the number of inputs and outputs of the layer, respectively. CNNs usually have millions of parameters, which makes overfitting a common problem. In order to avoid overfitting,  $L2$  regularization (regularization strength is 0.001) is used in each convolution layer and fully connected layer and/or dropout is added after first fully connected layers. Dropout randomly drops neurons of neural network during training, and the percentage of dropped neurons in the total number of neurons is called



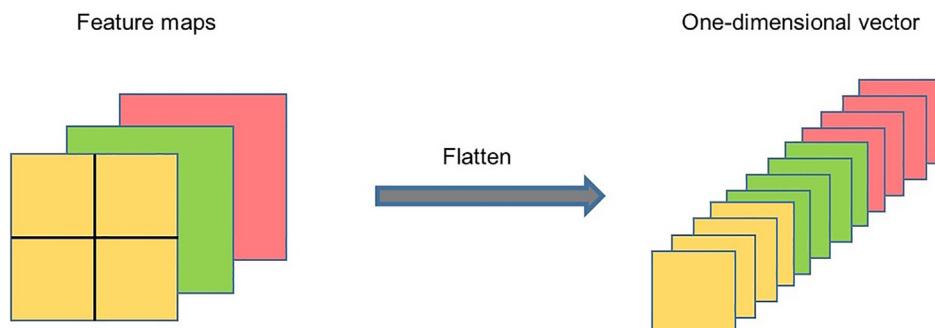
**Fig. 5.** Computation process of an example  $3 \times 3$  convolution filter. (a) Input is zero padded; this allows us to keep the size of the output the same as the size of the input image. (b) Convolution filter and computation. As an example, the value in top right corner in (c) is computed by dot product between red window and filter. (c) Results after convolution operation are offset by a bias. (d) Output is called a feature map.



**Fig. 6.**  $2 \times 2$  max-pooling process. Each  $2 \times 2$  block is replaced by the maximum value in the block.

dropout rate. The comparison of neural network with and without applying dropout is shown in Fig. 3. Since this is a regression problem, there is only one neuron in the output layer to produce continuous values of the effective stiffness.

Mean square error is used as the loss function, and Adam optimizer [68] with setting the learning rate as 0.001,  $\beta_1$  value as 0.9 and  $\beta_2$  value as 0.999 is used as the optimizer. Adam optimizer is an advanced optimization algorithm compared to conventional stochastic gradient



**Fig. 7.** Example of flattening process. Three  $2 \times 2$  feature maps are flattened to a one-dimensional vector with 12 elements.

descent algorithm, and produces faster convergence. During each iteration of the training process, the loss function calculates the error between predicted values and ground truth values. Then, the optimizer is used to propagate the error back through the entire neural network so that the weights can be adjusted accordingly. Thus, the loss function can be minimized through such iterative training. Early stopping technique is applied to monitor the learning process of CNN, and it will terminate the training process when the value of model's loss on the validation set does not improve for 10 epochs. We train several CNN models with different architectures from shallow to deep, and subsequently explore the hyperparameters space, i.e., number of filters in each convolution layer, batch size (i.e. number of MVEs that will be propagated through the network in a training iteration), and dropout rate.

### 3. Results and discussion

#### 3.1. CNN performance analysis

All the experiments were carried out on a NVIDIA DIGITS DevBox with 4 TITAN X GPUs with 12 GB memory for each GPU and Core i7-5930 K 6 Core 3.5 GHz CPU with 64 GB DDR4 RAM. All the programs were implemented in Python 2.7 and Keras [69], a high-level neural networks library build on top of TensorFlow [70] used to build deep learning models. As mentioned before, 5700 MVEs were used for training the CNN (i.e. training set and validation set) and the remaining 2850 MVEs were solely used as testing set. The model performance is evaluated by calculating mean absolute stiffness error (MASE) and mean absolute error (MAE) for both training (i.e. 5700 MVEs including validation set) and testing data. MASE for a selected set of data reflects the average error between the predicted values and ground truth values calculated from FE models. This error metric is defined as

$$MASE = \frac{1}{N} \sum_{i=1}^N \left| \frac{S_i - \hat{S}_i}{S_{\text{average}}} \right| \times 100\% \quad (1)$$

where  $N$  denotes the total number of MVEs in the selected set,  $S_i$  and  $\hat{S}_i$  represent ground truth effective stiffness and predicted effective stiffness for the  $i^{\text{th}}$  MVE, respectively.  $S_{\text{average}}$  denotes the average effective stiffness of all the MVEs in the dataset. In other words, MASE can be calculated by dividing mean absolute error between ground truth values and predicted values by the average of effective stiffness in the dataset.

Different architectures of varying depths and different number of filters in each convolution layer of 3-D CNNs are explored to examine which architectures produce the best model for the present application. The depths of 3-D CNNs are gradually extended from 10 layers to 16 layers where multiple stacked convolution layers and pooling layers are added gradually. In other trials, the number of filters in each convolution layer was gradually increased for a fixed depth. Table 1 presents the results of six of the 3-D CNNs explored in this study. In the description of the architecture, "ConvBlock" denotes a combination of a convolutional layer and a pooling layer, and "FC" a fully connected layer. As an example, the architecture ConvBlock(16–32×2)-FC(2048–1024) means that there is one convolution layer with 16 filters, followed by a pooling layer, followed by two other convolution layers (each with 32 filters), followed by a pooling layer, followed by two fully connected layers with 2048 neurons and 1024 neurons, respectively. For all the experiments,  $3 \times 3 \times 3$  filter is applied for each 3-D convolution layer.  $L_2$  regularization (regularization strength is 0.001) is applied in every convolution layer and fully connected layer, and the batch size is 32 MVEs. Since the model's performance is evaluated by its performance on new data (i.e., testing set), we can observe that the third 3-D CNN architecture with 14 layers in Table 1 achieves the best performance (good accuracy without over-fitting). Indeed, when the model's complexity reaches a certain point, further increasing the depth

**Table 1**  
Results comparison of different 3-D CNN architectures.

Architecture	Training MASE/MAE (GPa)	Testing MASE/MAE (GPa)
input-ConvBlock(16-32-64-128)-FC(2048-1024)-output	2.13%/0.71	3.12%/1.05
input-ConvBlock(8-16-32-64-128)-FC(2048-1024)-output	3.19%/1.07	3.72%/1.25
input-ConvBlock(16-32-64-128-256)-FC(2048-1024)-output	<b>2.10%/0.70</b>	<b>3.10%/1.04</b>
input-ConvBlock(32-64-128-256-512)-FC(2048-1024)-output	2.65%/0.89	3.47%/1.17
input-ConvBlock(16-32-64-128-256×2)-FC(2048-1024)-output	4.01%/1.34	4.55%/1.53
input-ConvBlock(16-32-64-128×2-256×2)-FC(2048-1024)-output	4.37%/1.46	4.95%/1.66

of model or the number of filters in convolution layers does not improve the model performance. More specially, the results of first, third, fifth and sixth CNNs in Table 1 suggest that further increasing depth of CNN does not increase performance. Meanwhile, the results of second, third and fourth CNNs in Table 1 show that further increasing the number of filters in convolution layer might deteriorate performance for the selected depth. We thus conclude that the third architecture is the best choice for the present study, and carry out further exploration and optimization of the hyperparameters space for this architecture.

Trials were carried out to search for the best combination of CNN hyperparameters for the third model architecture in Table 1. Because a deep learning model usually has many hyperparameters, and training a deep learning model is time consuming, in practice one uses a greedy approach to search for the best combination of hyperparameters. More specially, we choose one hyperparameter, find its optimum value and fix it for later trials. We repeat this process until all the hyperparameters are tuned. First, the effect of different batch sizes on model performance is explored. Table 2 shows batch size with 32 MVEs gives the best performance. Then with the batch size as 32 MVEs, the effect of dropout rate on model's performance is explored and summarized in Table 3. One can observe that the model without dropout achieves the best performance.

#### 3.2. Comparison with currently employed method

We compare the results obtained with deep learning approach to the results obtained with the simple physics-based approaches (i.e. rule of mixtures methods) and the sophisticated physics-inspired approaches [7,9,60] (i.e., two-point statistics methods). To demonstrate the difficulty of the problem, three different simple physics-based approaches (i.e. upper (Voigt) bound, lower (Reuss) bound, average (Voigt-Reuss-Hill)) were employed as benchmark in addition to sophisticated physics-inspired approaches. Simple physics-based approach is a direct approach of predicting the effective stiffness of a composite material based on the volume fraction and material property of microscale constituents. The bounds specified with simple physics-based approaches either assumes isostrain or isostress conditions (i.e. uniform strain or stress throughout the entire composite volume). The workflows used for the comparisons of simple physics-based approaches, sophisticated physics-inspired approaches and deep learning approach are depicted in Fig. 8.

**Table 2**  
Effect of batch size on CNN performance.

Batch size	Training MASE/MAE (GPa)	Testing MASE/MAE (GPa)
16	3.13%/1.05	3.85%/1.29
32	<b>2.10%/0.70</b>	<b>3.10%/1.04</b>
64	4.56%/1.53	4.88%/1.64

**Table 3**

Effect of dropout rate on CNN performance.

Dropout rate	Training MASE/MAE (GPa)	Testing MASE/MAE (GPa)
No dropout	<b>2.10%/0.70</b>	<b>3.10%/1.04</b>
0.2	3.39%/1.13	3.90%/1.31
0.4	3.70%/1.24	4.29%/1.44
0.6	5.40%/1.81	5.78%/1.94
0.8	3.51%/1.17	4.21%/1.41

In the sophisticated physics-inspired approaches, the reduced-order representation of spatial correlations in the microstructure are fit to the property values through standard regression methods [7,9,60]. More specifically, this approach computed two-point autocorrelations of a selected microscale constituent in each MVE, and then projected that statistical representation of microstructure to a reduced-order space by employing PCA [7,9,60].

From the results shown in Table 4, it is clear that all three simple physics-based approaches have inferior performance compared to sophisticated physics-inspired and deep learning approaches. This demonstrates the high level of difficulty of the problem due to the fact that volume fraction solely is not an adequate measure to predict the effective stiffness of high contrast 3-D elastic composites. On the other hand, we can observe that the 3-D CNN achieves the best performance with 2.10 % and 3.10 % for training and testing MASE, and 0.70 (GPa) and 1.04 (GPa) for training and testing MAE. Sophisticated physics-inspired approaches get 6.81% and 6.79 % for training and testing MASE, and 2.28 (GPa) for both training and testing MAE. More specifically, the proposed 3-D CNN improves the model performance of the sophisticated physics-inspired approaches by as much as 54% ( $1 - 3.10/6.79$ ) in terms of testing MASE. The parity plots of two methods for both training and testing sets are shown in Fig. 9. The top and bottom rows depict the results of both methods for training and testing sets, respectively. The left and right columns correspond to CNN and sophisticated physics-inspired approaches, respectively. From Fig. 9, it is seen that the parity plots of 3-D CNN for both training and testing sets exhibit high accuracy and low variance. In particular, it is noted that the deep learning approach performs better in very low and very high effective stiffness values, where the sophisticated physics-inspired approaches produce significantly higher errors in its predictions.

Sophisticated physics-inspired approaches provide an approach to extracting the salient features in the microstructure. However, in the present study, only 2-point correlations are used. Clearly, higher-order

**Table 4**

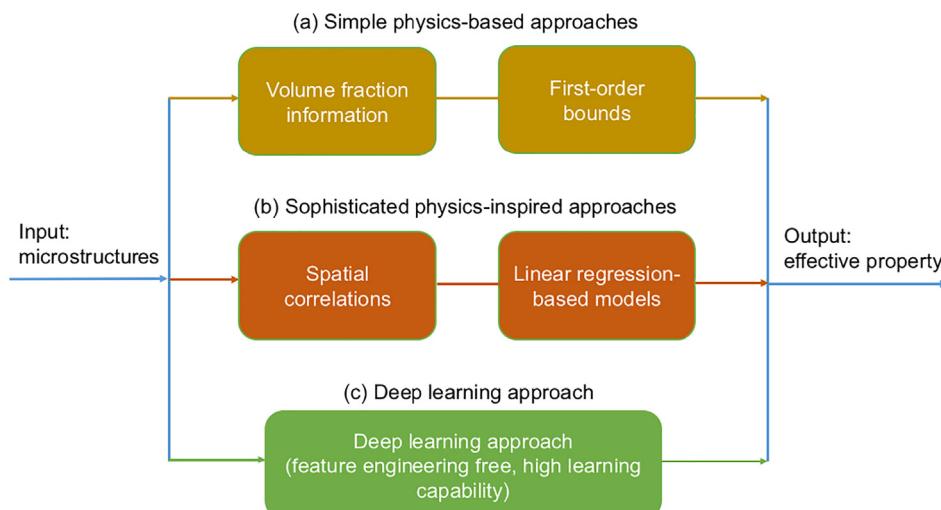
Performance comparison of predictive models.

Method	Training MASE/MAE (GPa)	Testing MASE/MAE (GPa)
3-D CNN	<b>2.10%/0.70</b>	<b>3.10%/1.04</b>
Sophisticated physics-inspired approaches	6.81%/2.28	6.79%/2.28
Simple physics-based approaches (Upper bound)	146.25%/48.93	145.15%/48.78
Simple physics-based approaches (Lower bound)	79.33%/26.54	79.38%/26.68
Simple physics-based approaches (Average)	47.08%/15.75	46.66%/15.68

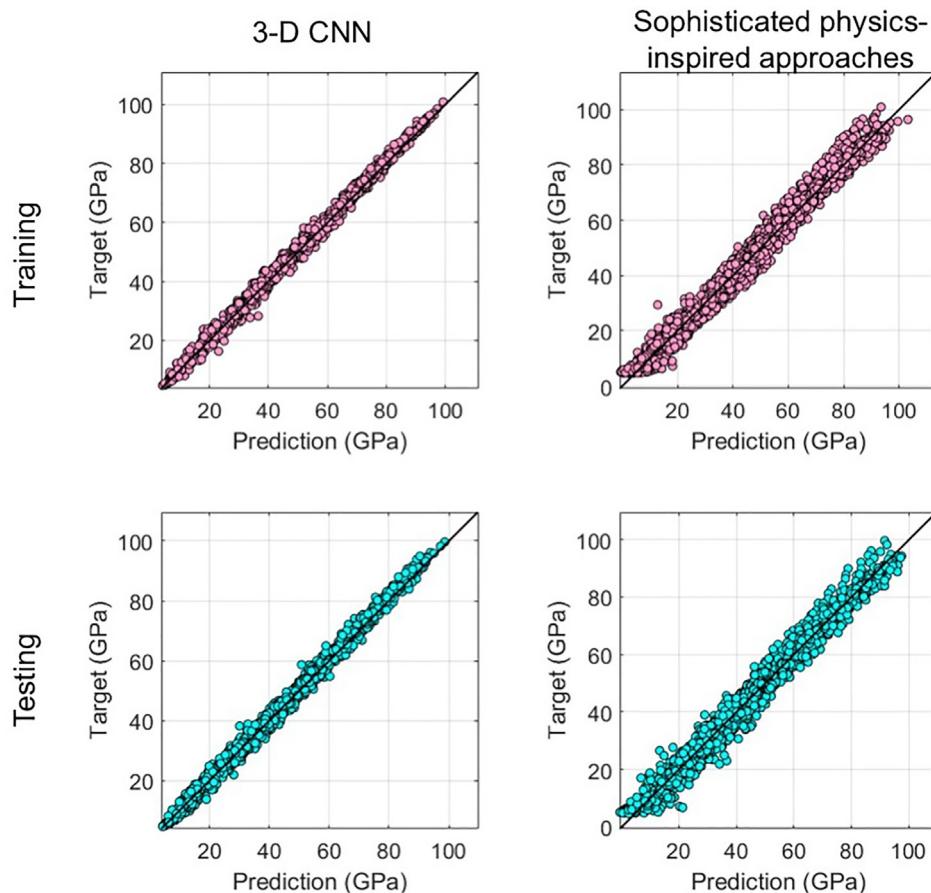
spatial correlations would improve the model [71–73], but do add significantly to the cost of building the models. In this regard, 3-D CNN provides a good alternative. The CNN filters are essentially capturing selected higher-order spatial correlations in an automated manner that is essentially feature-engineering-free (from the user's perspective). Therefore, with careful design of architecture and tuning of hyperparameters, deep learning approach can produce a highly reliable and robust prediction of the effective property associated with a given microstructure.

To the best knowledge of authors, this is the first time that a standalone 3-D CNN is implemented to establish structure-property linkages for high contrast elastic 3-D composites. Thereby, the extensive exploration of CNN architecture and hyperparameters provides important insight and guidance, and the proposed model in this work can serve as a pre-trained model to accelerate the research of structure-property linkages for high contrast elastic 3-D composites. In addition, the deep learning strategies described in this work outline a new data-driven framework that (i) allows practical inverse solutions to materials design through the use of computationally inexpensive surrogate models calibrated to the expensive physics based numerical simulations, and (ii) allow a more objective calibration of the parameters in the models using the limited amount of experimental data. In fact, these data-driven surrogate models offer the only avenue available today to mediate between the multiscale experiments and multiscale simulations in ways that can effectively inform each other.

However, there are still some limitations of current work. Firstly, because experimental data is limited and they currently lack the desired accuracy, resolution (i.e. spatial and temporal) and throughput, the proposed model is trained on simulation dataset, while its effectiveness on experimental dataset still needs to be evaluated. If a good



**Fig. 8.** Framework of the comparisons between the conventional machine learning method and deep learning method. (a) Simple physics-based approaches (i.e. rule of mixtures methods). (b) Sophisticated physics-inspired approaches (i.e. two-point statistics methods). (c) Deep learning approach.



**Fig. 9.** Parity plots of CNN and sophisticated physics-inspired approaches. The top and bottom rows depict the results for training and testing data, respectively. The columns represent the CNN and sophisticated physics-inspired approaches, respectively.

performance on experimental dataset can be achieved, it is capable to replace numerical approaches, such as FE model. Moreover, experiments or numerical approaches are usually expensive and time consuming. In some tasks like materials design, we can use deep learning model to do prescreening to select candidates that are more likely to achieve desired properties, and then use either experiment or numerical approaches to evaluate them, which can significantly reduce the time and cost to discovery new materials. Secondly, though the proposed model can achieve very accurate predictions, it still works like a black box. Thus, how to explain deep learning with domain knowledge is an interesting research direction.

#### 4. Conclusions

In this paper, 3-D CNN is implemented to model elastic homogenization linkages for three-dimensional high-contrast composite material system. The results show that the proposed 3-D CNN outperforms the sophisticated physics-inspired approaches by as much as 54% in terms of testing MASE. The deep learning approach thus demonstrates its superior capability for building a model with high accuracy, low computational cost, and higher learning capability.

Although intensive search of the space of hyperparameters and CNN architectures has been carried out to find the best model, there are still additional possibilities to further improve the model performance. First, we can observe from Fig. 9 that deep learning approach does not perform very well for some mid-range values of effective stiffness. It might be possible to further improve the model accuracy by implementing some advanced techniques, such as batch normalization [74], Residual structure [75], and Inception structure [76]. In addition, applying dropout does not improve the performance of proposed model, so other

advanced regularization techniques, such as DropConnect [77] and stochastic pooling [78] can be explored in the future studies. Second, deep learning is only beginning to be applied actively in materials science problems. We believe the proposed approach is sufficiently generalized and can be applied to other multiscale materials phenomena. This can potentially lead to a transformational change in the field of materials science and engineering, especially in the area of multiscale materials design.

#### Acknowledgements

This work is supported in part by the following grants: AFOSR award FA9550-12-1-0458; NIST award 70NANB14H012; NSF award CCF-1409601; DOE awards DESC0007456, DE-SC0014330; and Northwestern Data Science Initiative.

#### References

- [1] A. Agrawal, A. Choudhary, Perspective: materials informatics and big data: realization of the fourth paradigm of science in materials science, *APL Mater.* 4 (5) (2016) 053208.
- [2] S.R. Kalidindi, A.J. Medford, D.L. McDowell, Vision for data and informatics in the future materials innovation ecosystem, *JOM* 68 (8) (2016) 2126–2137.
- [3] J.H. Panchal, S.R. Kalidindi, D.L. McDowell, Key computational modeling issues in integrated computational materials engineering, *Comput. Aided Des.* 45 (1) (2013) 4–25.
- [4] G.B. Olson, Computational design of hierarchically structured materials, *Science* 277 (5330) (1997) 1237–1242.
- [5] O. Wodo, J. Zola, B.S.S. Pokuri, P. Du, B. Ganapathysubramanian, Automated, high throughput exploration of process-structure-property relationships using the mapreduce paradigm, *Mater. Discovery* 1 (2015) 21–28.
- [6] Y.C. Yabansu, P. Steinmetz, J. Hötzer, S.R. Kalidindi, B. Nestler, Extraction of reduced-order process-structure linkages from phase-field simulations, *Acta Mater.* 124 (2017) 182–194.

- [7] N.H. Paulson, M.W. Priddy, D.L. McDowell, S.R. Kalidindi, Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics, *Acta Mater.* 129 (2017) 428–438.
- [8] J. Yan, M.A. Sutton, A.P. Reynolds, Process–structure–property relationships for nugget and heat affected zone regions of aa2524-t351 friction stir welds, *Sci. Technol. Weld. Joining* 10 (6) (2005) 725–736.
- [9] M.I. Latypov, S.R. Kalidindi, Data-driven reduced order models for effective yield strength and partitioning of strain in multiphase materials, *J. Comput. Phys.* 346 (2017) 242–261.
- [10] J. Smith, W. Xiong, W. Yan, S. Lin, P. Cheng, O.L. Kafka, G.J. Wagner, J. Cao, W.K. Liu, Linking process, structure, property, and performance for metal-based additive manufacturing: computational approaches with experimental support, *Comput. Mech.* 57 (4) (2016) 583–610.
- [11] Y.C. Yabansu, S.R. Kalidindi, Representation and calibration of elastic localization kernels for a broad class of cubic polycrystals, *Acta Mater.* 94 (2015) 26–35.
- [12] S. Nguyen, A. Tran-Le, M. Vu, Q. To, O. Douzane, T. Langlet, Modeling thermal conductivity of hemp insulation material: a multi-scale homogenization approach, *Build. Environ.* 107 (2016) 127–134.
- [13] X.-Y. Zhou, P. Gosling, C. Pearce, Z. Ullah, et al., Perturbation-based stochastic multi-scale computational homogenization method for the determination of the effective properties of composite materials with random properties, *Comput. Methods Appl. Mech. Eng.* 300 (2016) 84–105.
- [14] A. Cruzado, B. Gan, M. Jiménez, D. Barba, K. Ostolaza, A. Linaza, J. Molina-Aldareguia, J. Llorca, J. Segurado, Multiscale modeling of the mechanical behavior of in718 superalloy based on micropillar compression and computational homogenization, *Acta Mater.* 98 (2015) 242–253.
- [15] T. Fast, S.R. Kalidindi, Formulation and calibration of higher-order elastic localization relationships using the mks approach, *Acta Mater.* 59 (11) (2011) 4595–4605.
- [16] G. Landi, S.R. Niezgoda, S.R. Kalidindi, Multi-scale modeling of elastic response of three-dimensional voxel-based microstructure datasets using novel dft-based knowledge systems, *Acta Mater.* 58 (7) (2010) 2716–2725.
- [17] G. Landi, S.R. Kalidindi, Thermo-elastic localization relationships for multi-phase composites, *Comput. Mater. & Continua* 16 (3) (2010) 273–293.
- [18] Y.C. Yabansu, D.K. Patel, S.R. Kalidindi, Calibrated localization relationships for elastic response of polycrystalline aggregates, *Acta Mater.* 81 (2014) 151–160.
- [19] R. Liu, Y.C. Yabansu, A. Agrawal, S.R. Kalidindi, A.N. Choudhary, Machine learning approaches for elastic localization linkages in high-contrast composite materials, *Integrating Mater. Manuf. Innovation* 4 (1) (2015) 13.
- [20] R. Liu, Y.C. Yabansu, Z. Yang, A.N. Choudhary, S.R. Kalidindi, A. Agrawal, Context aware machine learning approaches for modeling elastic localization in three-dimensional composite microstructures, *Integrating Mater. Manuf. Innovation* (2017) 1–12.
- [21] H. Garmestani, S. Lin, B. Adams, S. Ahzi, Statistical continuum theory for large plastic deformation of polycrystalline materials, *J. Mech. Phys. Solids* 49 (3) (2001) 589–607.
- [22] E. Kröner, Bounds for effective elastic moduli of disordered materials, *J. Mech. Phys. Solids* 25 (2) (1977) 137–155.
- [23] E. Kröner, Statistical modelling, *Modelling Small Deformations of Polycrystals*, Springer, 1986, pp. 229–291.
- [24] D.T. Fullwood, B.L. Adams, S.R. Kalidindi, A strong contrast homogenization formulation for multi-phase anisotropic materials, *J. Mech. Phys. Solids* 56 (6) (2008) 2287–2297.
- [25] J. Michel, H. Moulinec, P. Suquet, A computational method based on augmented lagrangians and fast fourier transforms for composites with high contrast, *CMES (Comput. Modell. Eng. Sci.)* 1 (2) (2000) 79–88.
- [26] B.L. Adams, S. Kalidindi, D.T. Fullwood, *Microstructure-sensitive Design for Performance Optimization*, Butterworth-Heinemann, 2013.
- [27] A. Jain, J.A. Bollinger, T.M. Truskett, Inverse methods for material design, *AIChE J.* 60 (8) (2014) 2732–2740.
- [28] A.G. Gagorik, B. Savoie, N. Jackson, A. Agrawal, A. Choudhary, M.A. Ratner, G.C. Schatz, K.L. Kohlstedt, Improved scaling of molecular network calculations: the emergence of molecular domains, *J. Phys. Chem. Lett.* 8 (2) (2017) 415–421.
- [29] L. Ward, R. Liu, A. Krishna, V.I. Hegde, A. Agrawal, A. Choudhary, C. Wolverton, Including crystal structure attributes in machine learning models of formation energies via voronoi tessellations, *Phys. Rev. B* 96 (2) (2017) 024104.
- [30] A. Furmanchuk, A. Agrawal, A. Choudhary, Predictive analytics for crystalline materials: bulk modulus, *RSC Adv.* 6 (97) (2016) 95246–95251.
- [31] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials. Available from: <[1606.09551](https://doi.org/10.4236/jmd.201609551)>.
- [32] R. Liu, A. Kumar, Z. Chen, A. Agrawal, V. Sundararaghavan, A. Choudhary, A predictive machine learning approach for microstructure optimization and materials design, *Sci. Rep.* 5 (2015) 11551.
- [33] A. Agrawal, P.D. Deshpande, A. Cecen, G.P. Basavarsu, A.N. Choudhary, S.R. Kalidindi, Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters, *Integrating Mater. Manuf. Innovation* 3 (1) (2014) 1–19.
- [34] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B* 89 (9) (2014) 094104.
- [35] K. Gopalakrishnan, A. Agrawal, H. Ceylan, S. Kim, A. Choudhary, Knowledge discovery and data mining in pavement inverse analysis, *Transport* 28 (1) (2013) 1–10.
- [36] S.R. Kalidindi, S.R. Niezgoda, G. Landi, S. Vachhani, T. Fast, A novel framework for building materials knowledge systems, *Comput. Mater. & Continua* 17 (2) (2010) 103–125.
- [37] S.R. Kalidindi, *Hierarchical Materials Informatics: Novel Analytics for Materials Data*, Elsevier, 2015.
- [38] T. Fast, S.R. Niezgoda, S.R. Kalidindi, A new framework for computationally efficient structure–structure evolution linkages to facilitate high-fidelity scale bridging in multi-scale materials models, *Acta Mater.* 59 (2) (2011) 699–707.
- [39] S.R. Niezgoda, A.K. Kanjrala, S.R. Kalidindi, Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data, *Integrating Mater. Manuf. Innovation* 2 (1) (2013) 3.
- [40] S.R. Kalidindi, J.A. Gomberg, Z.T. Trautt, C.A. Becker, Application of data science tools to quantify and distinguish between structures and models in molecular dynamics datasets, *Nanotechnology* 26 (34) (2015) 344006.
- [41] P. Altschuh, Y.C. Yabansu, J. Hötzler, M. Selzer, B. Nestler, S.R. Kalidindi, Data science approaches for microstructure quantification and feature identification in porous membranes, *J. Membr. Sci.* 540 (1) (2017) 88–97.
- [42] A. Choudhury, Y.C. Yabansu, S.R. Kalidindi, A. Dennstedt, Quantification and classification of microstructures in ternary eutectic alloys using 2-point spatial correlations and principal component analyses, *Acta Mater.* 110 (2016) 131–141.
- [43] A. Iskakov, Y.C. Yabansu, S. Rajagopalan, A. Kapustina, S.R. Kalidindi, Application of spherical indentation and the materials knowledge system framework to establishing microstructure–yield strength linkages from carbon steel scoops excised from high-temperature exposed components, *Acta Mater.* 144 (2018) 758–767.
- [44] H. Schulz, S. Behnke, Learning object-class segmentation with convolutional neural networks, in: *ESANN*, 2012.
- [45] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, P.E. Barbano, Toward automatic phenotyping of developing embryos from videos, *IEEE Trans. Image Process.* 14 (9) (2005) 1360–1371.
- [46] P. Sermanet, Y. LeCun, Traffic sign recognition with multi-scale convolutional networks, *Neural Networks (IJCNN), The 2011 International Joint Conference on*, IEEE, 2011, pp. 2809–2813.
- [47] D.C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, J. Schmidhuber, Flexible, high performance convolutional neural networks for image classification, in: *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, Barcelona, Spain, 2011, pp. 1237.
- [48] T. Wang, D.J. Wu, A. Coates, A.Y. Ng, End-to-end text recognition with convolutional neural networks, *Pattern Recognition (ICPR), 2012 21st International Conference on*, IEEE, 2012, pp. 3304–3308.
- [49] S. Ji, W. Xu, M. Yang, K. Yu, 3d convolutional neural networks for human action recognition, *IEEE Trans. Pattern Anal. Machine Intelligence* 35 (1) (2013) 221–231.
- [50] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. Available from: <[1409.1556](https://doi.org/10.4236/jmd.201404009)> .
- [51] R. Liu, A. Agrawal, W.-k. Liao, M.D. Graef, A. Choudhary, Materials discovery: understanding polycrystals from large-scale electron patterns, in: *Proceedings of IEEE BigData Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH)*, 2016, pp. 2261–2269.
- [52] R. Liu, L. Ward, C. Wolverton, A. Agrawal, W. Liao, A. Choudhary, Deep learning for chemical compound stability prediction, in: *Proceedings of ACM SIGKDD Workshop on Large-scale Deep Learning for Data Mining (DL-KDD)*, 2016, pp. 1–7.
- [53] X. Li, Y. Zhang, H. Zhao, C. Burkhardt, L.C. Brinson, W. Chen, A transfer learning approach for microstructure reconstruction and structure–property predictions, *Sci. Rep.* (2018) arXiv preprint arXiv:1805.02784.
- [54] R. Cang, Y. Xu, S. Chen, Y. Liu, Y. Jiao, M.Y. Ren, Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design, *J. Mech. Des.* 139 (7) (2017) 071404.
- [55] R. Cang, H. Li, H. Yao, Y. Jiao, Y. Ren, Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model, Available from: <[1712.03811](https://doi.org/10.4236/jmd.201703111)> .
- [56] Z. Yang, X. Li, L.C. Brinson, A.N. Choudhary, W. Chen, A. Agrawal, Microstructural materials design via deep adversarial learning methodology, *J. Mech. Des.* (2018) arXiv preprint arXiv:1805.02791.
- [57] X. Li, Z. Yang, L.C. Brinson, A.N. Choudhary, A. Agrawal, W. Chen, A deep adversarial learning methodology for designing microstructural material systems, in: *Processing of the ASME 2018 International Design Engineering Technical Conferences*. (accepted).
- [58] K. Gopalakrishnan, S.K. Khatan, A. Choudhary, A. Agrawal, Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection, *Constr. Build. Mater.* 157 (2017) 322–330.
- [59] A. Cecen, H. Dai, Y.C. Yabansu, S.R. Kalidindi, Material structure–property linkages using three-dimensional convolutional neural networks, *Acta Materialia* (accepted).
- [60] A. Gupta, A. Cecen, S. Goyal, A.K. Singh, S.R. Kalidindi, Structure–property linkages using a data science approach: application to a non-metallic inclusion/steel composite system, *Acta Mater.* 91 (2015) 239–254.
- [61] S.R. Kalidindi, G. Landi, D.T. Fullwood, Spectral representation of higher-order localization relationships for elastic behavior of polycrystalline cubic materials, *Acta Mater.* 56 (15) (2008) 3843–3853.
- [62] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation. Tech. rep. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [63] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [64] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* 25, 2012, pp.

- 1097–1105.
- [65] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, Face recognition: a convolutional neural-network approach, *IEEE Trans. Neural Networks* 8 (1) (1997) 98–113.
- [66] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.
- [67] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Aistats, Vol. 9, 2010, pp. 249–256.
- [68] D. Kingma, J. Ba, Adam: A method for stochastic optimization, Available from: <[1412.6980](https://arxiv.org/abs/1412.6980)> .
- [69] F. Chollet, Keras, <<https://github.com/fchollet/keras>> , 2015.
- [70] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: large-scale machine learning on heterogeneous systems, software available from tensorflow.org, 2015. URL <<http://tensorflow.org/>> .
- [71] A. G. Gray, A. W. Moore, N-body problems in statistical learning, in: Advances in neural information processing systems, 2001, pp. 521–527.
- [72] S. Torquato, G. Stell, Microstructure of two-phase random media. i. the n-point probability functions, *J. Chem. Phys.* 77 (4) (1982) 2071–2077.
- [73] B.L. Adams, P. Etinghof, D.D. Sam, Coordinate free tensorial representation of n-point correlation functions for microstructure by harmonic polynomials, *Materials Science Forum*, Vol. 157 Trans Tech Publ, 1994, pp. 287–294.
- [74] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, 2015, pp. 448–456.
- [75] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [76] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
- [77] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, R. Fergus, Regularization of neural networks using dropconnect, in: International Conference on Machine Learning, 2013, pp. 1058–1066.
- [78] M.D. Zeiler, R. Fergus, Stochastic pooling for regularization of deep convolutional neural networks, Available from: <[1301.3557](https://arxiv.org/abs/1301.3557)> .