Learning microstructure–property relationships in materials with robust features from vision transformers

Sheila E. Whitman
University of Arizona
Tucson, AZ 85721
sheilaw@arizona.edu

Guangyu Hu
University of Arizona
Tucson, AZ 85721
hug@arizona.edu

Marat I. Latypov
University of Arizona
Tucson, AZ 85721
latmarat@arizona.edu

Abstract

Machine learning of microstructure–property relationships from data is an emerging approach in computational materials science. Most existing machine learning efforts focus on the development of task-specific models for each microstructure–property relationship. We propose utilizing a pre-trained foundational vision model for the extraction of task-agnostic microstructure features and subsequent lightweight machine learning. We demonstrate our approach with a pre-trained DinoV2 model on unsupervised representation of an ensemble of two-phase microstructures and modeling of their overall elastic stiffness. Our results show the potential of foundational vision models for robust microstructure representation and efficient machine learning of microstructure–property relationships without the need for expensive task-specific training or fine-tuning.

1. Introduction

Structural alloys represent an important class of materials needed across all critical industries (energy, defense, transportation, infrastructure). Design of structural alloys relies on quantitative understanding of microstructure–property relationships. Computer models capable of capturing these relationships can significantly accelerate materials design endeavors. Machine learning is rapidly emerging as a powerful computational tool with numerous reports of models trained from experiments [16], physics-based simulations [14, 17, 31], or their combinations [23].

A challenge in enabling machine learning of microstructure–property relationships in structural materials is the need for a quantitative description of the microstructure. Robust description of microstructure is a non-trivial task because of a rich diversity of microstructures observable at different length scales and a variety of their aspects (spatial, geometric, statistical) relevant for different properties. Multiple strategies of quantitative microstructure description have been reported [3], including traditional geometric or statistical descriptions [16, 30], latent representations by convolutional neural networks (CNNs) [7, 14, 25, 31], and their combinations [7]. A limitation of machine learning with traditional microstructure descriptors is the need in selecting the most appropriate set of descriptors or distribution functions for each individual property-specific model (e.g., [30]). Similarly, training task-specific CNNs and designing their architectures for a variety of microstructure–property relationships is time consuming and computationally expensive.

While most machine learning studies on structural materials focus on task-specific models, research on language modeling and computer vision is undergoing a paradigm shift towards task-agnostic foundational models [11]. Foundational models advantageously learn representations of high-dimensional data (texts, images) that are universal for a spectrum of downstream tasks. Modeling with universal features can even yield better results than task-specific neural networks [4]. This progress has been possible with the advent of the transformer architecture [29] and strategies for unsupervised learning from large unlabeled datasets [5, 8, 12]. Materials research could benefit from the adoption and development of foundational models that facilitate learning relationships without task-specific reinvention of architectures, expensive training, or fine-tuning.

In this study, we demonstrate and evaluate a vision transformer (ViT) model DinoV2 [22] as a microstructure feature extractor for machine learning of microstructure–property relationships. We hypothesize that the general-purpose visual features that DinoV2 extracts from microstructures can serve as their robust representation for modeling properties without training or fine-tuning to any materials data. We test our hypothesis by low-dimensional representation of an ensemble of two-phase microstructures and training simple polynomial models of their elastic stiffness using features extracted by DinoV2 pre-trained on non-materials images.

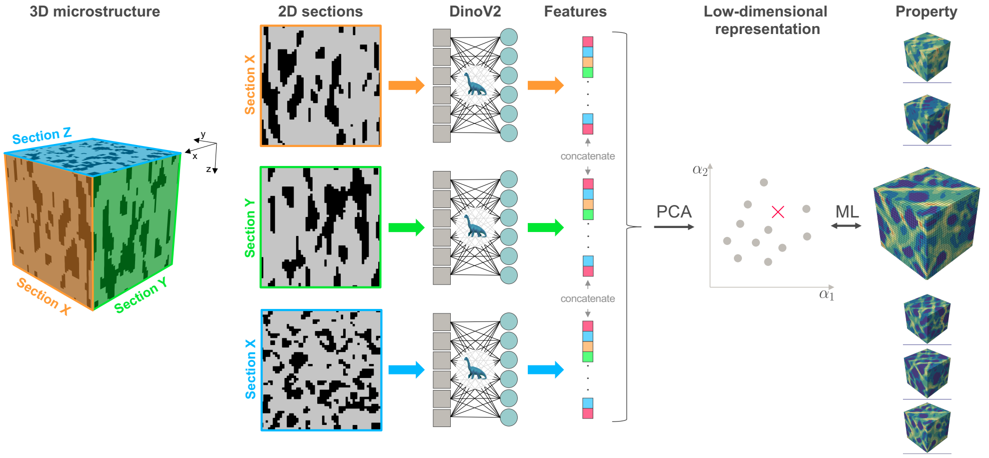

Figure 1. Overview of training a microstructure–property model based on features from DinoV2 model and data from simulations.

2. Related work

Given a large body of work on data-driven modeling of microstructure–property relationships, here we provide a brief snapshot of most relevant research. As mentioned above, machine learning models primarily differ in the approach to quantitative description of the microstructure. One strategy is to use geometric descriptors of microstructures (e.g., phase volume fraction, grain size) that are intuitive and familiar from traditional models (e.g., Voigt/Reuss bounds, Hall-Petch relation [10, 24]). Another strategy is to describe microstructures with distribution functions: n-point correlations [15], lineal path functions [19], or chord length distributions [28]. This strategy was shown successful for modeling a range of properties based on data from both simulations (e.g, [18]) and experiments (e.g., [16]). Given the high dimensionality of statistical distributions, principal component analysis (PCA) is often used for dimensionality reduction, which provides features ultimately used for machine learning. With low-rank representation of statistical distributions used as features, numerous machine learning methods have been reported: from simple polynomial regression [18] to more sophisticated statistical learning (e.g., [20]). Besides geometric and statistical microstructure descriptions inspired by theories, purely data-driven approaches (e.g., CNNs) have been also explored. CNNs for modeling microstructure–property relationships are typically designed and trained from scratch for each specific property of interest [7, 14, 25, 31].

While task-specific machine learning models prevail in computational materials research, foundational models are taking over in natural language processing and computer vision. Specifically, there is an active development of large models with the transformer architecture that are entirely unsupervised and task-agnostic [6, 12, 27]. Researchers seek to develop models that produce features generalizable to numerous applications and tasks without the need for fine-tuning. DinoV2 is an example of such a foundational model [22]. By introducing self-supervised learning at both image level and masked patch level [32], DinoV2 achieves task-agnostic features with a semantic meaning.

3. New ViT approach

We propose to model microstructure–property relationships with foundational vision models as feature extractors for microstructures. In this study, we adopt DinoV2 as a foundational ViT model of choice and focus on machine learning of microstructure–property relationships based on high-fidelity finite-element simulation data. We further propose to model 3D properties based on 2D sections of 3D microstructures. First, 2D microstructure data is much more widely accessible than full 3D data given the high cost and requirements of highly specialized and expensive 3D characterization equipment [9]. Second, 2D microstructure images are readily compatible with most foundational vision models pre-trained on vast amounts of 2D images, including DinoV2. Finally, machine learning of microstructure–property relationships based on three orthogonal 2D sections of microstructure was shown to achieve a reasonable accuracy with traditional microstructure descriptions [13]. With these ideas, our approach (illustrated in Figure 1) involves the following steps for developing a model based on DinoV2 features and data from numerical simulations:

1. collect training data: generate 3D microstructures and obtain their properties from simulations;
2. get 2D sections from 3D microstructures;
3. extract image-level features with a pre-trained ViT (DinoV2) for each 2D section;
4. aggregate features (e.g., mean pooling or concatenation) from all 2D sections of a microstructure;
5. reduce the dimensionality (using PCA or similar);

6. train a lightweight regression-type statistical learning model that captures the relationship between microstructure features and the property of interest.

Upon training, we only need 2D sections of any new 3D microstructure to extract its features with the same (DinoV2) ViT model for inference of the property. Below, we predict the elastic stiffness from three orthogonal sections of two-phase microstructures using this approach.

4. Experiments

For demonstration of the proposed approach, we use a published dataset of 5900 two-phase microstructures and their corresponding overall Young's modulus values calculated with finite element simulations [7]. The microstructures consist of a stiff phase and a compliant phase with a stiffness ratio of 50 – a relatively high property contrast, which is generally challenging for traditional models [17, 31].

The raw microstructure data in the dataset is represented by $51 \times 51 \times 51$ arrays of phase labels, which we sliced along the three axes to obtain three orthogonal 2D sections of size 51×51. DinoV2 is pre-trained to process RGB images in patches of size 14×14 pixels. To make our 2D sections consistent with the input expected by DinoV2, we first converted our binary 2D arrays with phase labels into RGB images. We then resized our 51×51 by either (i) cropping to a compatible size of 42×42 pixels; or (ii) interpolating to the size of 224×224 native to DinoV2. The number of features extracted for each image depends on the size of the pre-trained DinoV2 model with the base version yielding 768 features for each image [22]. We first analyze the representation of microstructures with DinoV2 features in a low-dimensional space and then evaluate a simple polynomial model of the Young's modulus using these features.

4.1. Microstructure map with ViT features

Figure 2 shows a low-dimensional representation of the 5900 microstructures obtained by PCA of the 768 features from the base DinoV2 model and by PCA of the two-point auto-correlation function of the stiff phase (for 2D sections normal to the x axis) – approach from literature [21] used as a baseline here. The low-dimensional representation ("microstructure map") is visualized in terms of the first two principal component scores. We observe that PCA of both descriptions lead to a dense low-dimensional representation without any clusters for two-point correlation and with more pronounced clustering in the case of the DinoV2 features. The most striking difference is that the first principal component of the two-point correlation function is highly correlated with the volume fraction of the stiff phase, which is not the case for the principal components of the DinoV2 features. Indeed, the volume fraction steadily increases along the horizontal axis from zero to one (see color). The first principal component of the two-point correlation function (highly correlated with the volume fraction of the stiff phase) is also a significantly dominant one capturing 92% variance in the dataset as seen in the scree plot (Figure 2b). At the same time, the principal components are more balanced in terms of explained variance in the case of the DinoV2 features (Figure 2a).

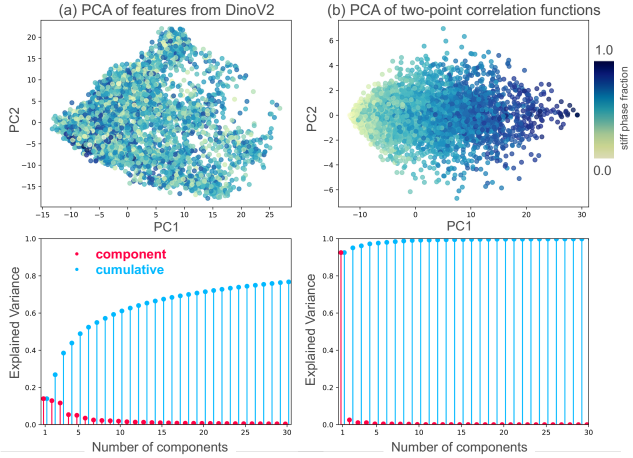

Figure 2. PCA of features from 5900 2D sections of two-phase microstructures obtained (a) by DinoV2 (base), (b) as two-point correlation function; the explained variance is also shown for the first 30 principal components (cumulative and per component)

4.2. Property modeling with ViT features

After extracting features from microstructures with DinoV2 and obtaining their low-dimensional representation with PCA, we used simple polynomial models for modeling their Young's modulus. We tested all available sizes of DinoV2 (small, base, large, giant), various numbers of principal components, and two aggregation strategies for features extracted from the three 2D sections: (i) element-wise mean of three feature vectors or (ii) concatenation of the vectors.

We evaluated these regression settings in terms of the mean absolute percentage error (MAPE) for a subset of the training data. We considered polynomials of only second order to avoid large numbers of terms in the regression model and overfitting. Table 1 summarizes the results, including the best number of principal components identified for each DinoV2 size and two-point auto-correlation function for both of the aggregation strategies (mean and concatenation). Table 1 also lists the corresponding MAPE values for the test set unseen by the models during the search of the best regression settings. Among models based on DinoV2 features, the polynomial model of the second order with 16 principal components of concatenated $(3 \cdot 768)$ features from the base DinoV2 model gives the best accuracy in predicting the Young's modulus for two-phase microstructures unseen during training. The small DinoV2

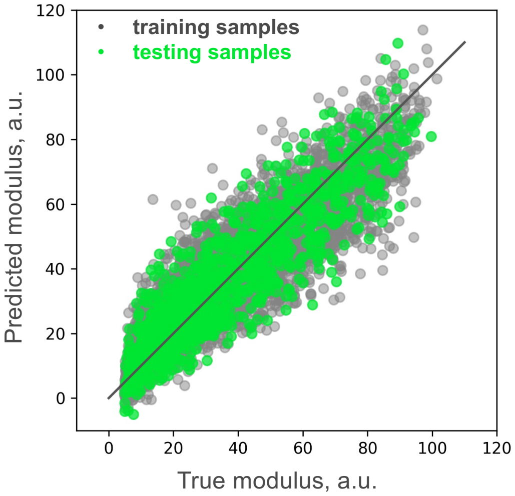

Figure 3. Parity plot showing the prediction of the Young's modulus for microstructures unseen during training by the best polynomial model of second degree with 16 principal components of the concatenated features from the base DinoV2 model.

model extracting 384 features from each image shows comparable accuracy, whereas features from the giant model lead to the highest MAPE. The mean aggregation of features from three 2D sections gives lower accuracy for all DinoV2 sizes. Interpolating the 2D sections to the image size 224 × 224 native to DinoV2 prior to feature extraction does not improve the results (see "224B" column) compared to images cropped to 42 × 42 size, despite some loss in the microstructure information from the crop. The polynomial models based on features from DinoV2 exhibit test errors comparable to that of concatenated two-point correlation functions (baseline) as the microstructure descriptions of three orthogonal sections (1681 "features" per section).

5. Discussion and conclusion

We utilized the foundational ViT model, DinoV2, as a feature extractor for quantitative microstructure representation for both unsupervised exploration and supervised learning of microstructure–property relationships. Traditionally, statistical descriptors like two-point correlation functions have been widely used to this end. For multi-phase materials, these functions tend to heavily prioritize the volume fraction of the phase being analyzed, often dominating the first principal component (Figure 2, [17]). This dominance may overshadow more subtle details of the microstructure (e.g., morphology, directionality) both in visualization of microstructure ensembles and learning of relationships. We found that the principal components of features from the DinoV2 model did not have a strong correlation with the phase volume fractions and were more balanced in terms of the variance they explain (Figure 2a).

	TP	S	B	224B	L	G
# features	1681	384	768	768	1024	1536
Test MAPE	0.31	0.33	0.32	0.36	0.35	0.38
(mean, #PCs)	(25)	(16)	(22)	(30)	(26)	(22)
Test MAPE	**0.27**	0.29	0.29	0.35	0.33	0.38
(cat, #PCs)	(37)	(30)	(17)	(51)	(32)	(37)

Table 1. The best polynomial regression settings with features from DinoV2 of different sizes; for the base model, we include an additional result with interpolated 224 × 224 images (base 224); numbers of principal components (PCs) are shown in parentheses.

In this study on digital two-phase microstructures, the best polynomial model based on DinoV2 features had comparable accuracy to the baseline model using two-point correlation functions. The two-point auto-correlation functions and other statistical or geometric descriptions are most straightforward to compute for binary microstructures considered here. However, their calculations are not as trivial for real-world images obtained in experiments (e.g., grayscale) which require an additional step of image segmentation. The fidelity of geometric or statistical descriptions therefore relies on the quality of segmentation prone to errors [2]. In contrast, ViT foundational models are not limited to binary images as input and can extract features from raw experimental images without the need for segmentation. We thus anticipate that the foundational ViT models such as DinoV2 have the potential for extracting features for superior models of process–structure–property relationships based on real-world microstructure data without segmentation. While we considered microstructure–property relationships, the approach can be also used for process–microstructure models that relate microstructure features from ViT models to process variables (e.g., [26]).

In conclusion, we demonstrated the use of a pre-trained foundational ViT model, DinoV2, for feature extraction from microstructure images for (i) unsupervised learning and visualization of an ensemble of two-phase microstructures and (ii) supervised learning of microstructure–property relationships. We showed that (i) ViT features do not have strong correlation with the volume fraction and lead to more balanced visualization of a microstructure ensemble; and (ii) concatenation of DinoV2 features from orthogonal 2D sections followed by dimensionality reduction and simple polynomial regression leads to a reasonably good prediction of the overall elastic stiffness of two-phase microstructures even with limited 2D microstructure input. The code for this paper is made available on GitHub [1].

Acknowledgments. SEW acknowledges the support by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2137419.

References

[1] Elastic DinoV2 code: https://github.com/materials-informatics-az/Elastic-DinoV2. 4

[2] Ben Bales, Tresa Pollock, and Linda Petzold. Segmentation-free image processing and analysis of precipitate shapes in 2d and 3d. *Modelling and Simulation in Materials Science and Engineering*, 25(4):045009, 2017. 4

[3] Ramin Bostanabad, Yichi Zhang, Xiaolin Li, Tucker Kearney, L Catherine Brinson, Daniel W Apley, Wing Kam Liu, and Wei Chen. Computational microstructure characterization and reconstruction: Review of the state-of-the-art techniques. *Progress in Materials Science*, 95:1–41, 2018. 1

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 1

[5] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 132–149, 2018. 1

[6] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 2

[7] Ahmet Cecen, Hanjun Dai, Yuksel C Yabansu, Surya R Kalidindi, and Le Song. Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Materialia*, 146:76–84, 2018. 1, 2, 3

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 1

[9] McLean P Echlin, Alessandro Mottura, Christopher J Torbet, and Tresa M Pollock. A new tribeam system for three-dimensional multimodal materials analysis. *Review of Scientific Instruments*, 83(2):023701, 2012. 2

[10] EO Hall. Variation of hardness of metals with grain size. *Nature*, 173(4411):948–949, 1954. 2

[11] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 1

[12] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022. 1, 2

[13] Guangyu Hu and Marat I Latypov. Learning from 2d: machine learning of 3d effective properties of heterogeneous materials based on 2d microstructure sections. *Frontiers in Metals and Alloys*, 1:1100571, 2022. 2

[14] Olga Ibragimova, Abhijit Brahme, Waqas Muhammad, Daniel Connolly, Julie Lévesque, and Kaan Inal. A convolutional neural network based crystal plasticity finite element framework to predict localised deformation in metals. *International Journal of Plasticity*, 157:103374, 2022. 1, 2

[15] Yang Jiao, FH Stillinger, and S Torquato. Modeling heterogeneous materials via two-point correlation functions: Basic principles. *Physical review E*, 76(3):031110, 2007. 2

[16] Nikhil Khatavkar, Sucheta Swetlana, and Abhishek Kumar Singh. Accelerated prediction of vickers hardness of co-and ni-based superalloys from microstructure and composition using advanced image processing techniques and machine learning. *Acta Materialia*, 196:295–303, 2020. 1, 2

[17] Marat I Latypov and Surya R Kalidindi. Data-driven reduced order models for effective yield strength and partitioning of strain in multiphase materials. *Journal of Computational Physics*, 346:242–261, 2017. 1, 3, 4

[18] Marat I Latypov, Laszlo S Toth, and Surya R Kalidindi. Materials knowledge system for nonlinear composites. *Computer Methods in Applied Mechanics and Engineering*, 346: 180–196, 2019. 2

[19] Binglin Lu and Salvatore Torquato. Lineal-path function for random heterogeneous materials. *Physical Review A*, 45(2): 922, 1992. 2

[20] Andrew Marshall and Surya R Kalidindi. Autonomous development of a machine-learning model for the plastic response of two-phase composites from micromechanical finite element models. *JOM*, 73(7):2085–2095, 2021. 2

[21] Stephen R Niezgoda, Anand K Kanjarla, and Surya R Kalidindi. Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integrating Materials and Manufacturing Innovation*, 2:54–80, 2013. 3

[22] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 1, 2, 3

[23] Darren C Pagan, Calvin R Pash, Austin R Benson, and Matthew P Kasemer. Graph neural network modeling of grain-scale anisotropic elastic behavior using simulated and measured microscale data. *npj Computational Materials*, 8 (1):259, 2022. 1

[24] NJ Petch. Xvi. the ductile fracture of polycrystalline α-iron. *Philosophical Magazine*, 1(2):186–190, 1956. 2

[25] Reeju Pokharel, Anup Pandey, and Alexander Scheinker. Physics-informed data-driven surrogate modeling for full-field 3d microstructure and micromechanical field evolution of polycrystalline materials. *JOM*, 73(11):3371–3382, 2021. 1, 2

[26] Evdokia Popova, Theron M Rodgers, Xinyi Gong, Ahmet Cecen, Jonathan D Madison, and Surya R Kalidindi. Process-structure linkages using a data science approach: application to simulated additive manufacturing data. *Integrating materials and manufacturing innovation*, 6:54–68, 2017. 4

[27] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning

transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 2

[28] Salvatore Torquato and B Lu. Chord-length distribution function for two-phase random media. *Physical Review E*, 47(4):2950, 1993. 2

[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[30] Hongyi Xu, Ruoqian Liu, Alok Choudhary, and Wei Chen. A machine learning-based design representation method for designing heterogeneous microstructures. *Journal of Mechanical Design*, 137(5):051403, 2015. 1

[31] Zijiang Yang, Yuksel C Yabansu, Reda Al-Bahrani, Weikeng Liao, Alok N Choudhary, Surya R Kalidindi, and Ankit Agrawal. Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. *Computational Materials Science*, 151:278–287, 2018. 1, 2, 3

[32] Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Cihang Xie, Alan Yuille, and Tao Kong. ibot: Image bert pre-training with online tokenizer, 2022. 2