Out of Domain Stress Prediction on a Dataset of Simulated 3D Polycrystalline Microstructures

Thomas Lu
Carnegie Mellon University
Pittsburgh, PA 15213
ttlu@cs.cmu.edu

Aarti Singh
Carnegie Mellon University
Pittsburgh, PA 15213
aarti@cs.cmu.edu

Abstract

Surrogate machine learning models for expensive material simulations can be an effective method to estimate relevant properties, which can help reduce the number of experiments needed. However, significant difficulties can occur when attempting to learn from small simulated datasets, particularly for samples out of the domain of the training data. This work provides an exploration on training deep learning models on a dataset of 36 synthetic 3D equiaxed polycrystalline microstructures with different cubic textures with a focus on out-of-domain accuracy, analyzing a number of transfer learning set ups, domain adaptation methods, model architectures, and featurizations across two formulations of the problem. We develop an evaluation set-up to validate our results, and report several methods that provide better results than our baseline of a simple U-Net architecture.

1 Introduction

Material simulations, conducted through density functional theory (DFT), molecular dynamics (MD), and other techniques have become indispensable tools in exploring the atomic and molecular interactions that govern material properties. These simulations offer remarkable precision, allowing scientists to predict properties, such as mechanical strength, electrical conductivity, and thermal behavior, at the atomic level. However, such experiments may be prohibitively expensive and time consuming. The main dataset for this work, for example, required weeks using a high power cluster to compute the elasto-viscoplastic fast Fourier transform (EVPFFT) simulations at the given resolution.

In order to alleviate this issue, surrogates can be learned to predict the properties that must otherwise be obtained from simulations. By doing so, machine learning methods hold the promise of enhancing the speed and generalization ability of material science research. However, researchers may encounter a number of issues when applying conventional machine learning methods. For example, sparse amounts of highly correlated data leads to overfitting and poor generalizability.

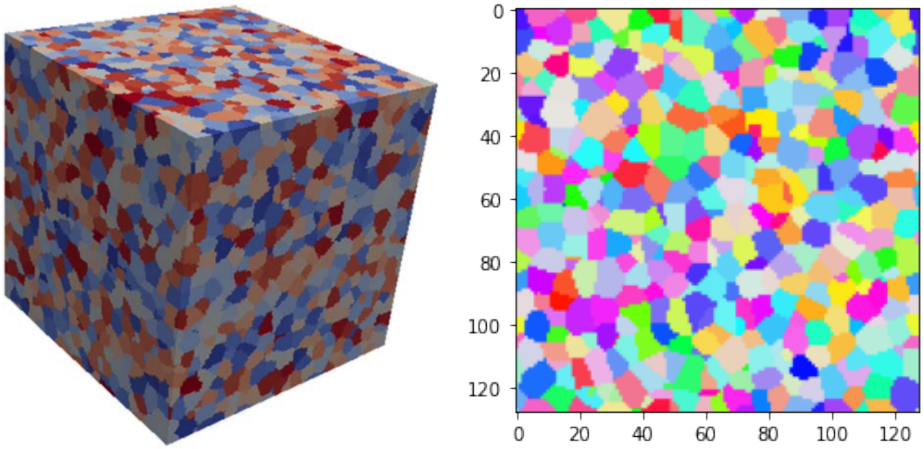
In light of these challenges, this paper explores practices to improve predictions in and out of domain for an example dataset of 36 simulated equiaxed polycrystalline microstructures across six different face centered cubic textures. We investigated the quality of several deep neural networks trained on images derived from these microstructures which are often data-hungry, and evaluated machine learning methods to improve the out-of-domain performance of these deep methods using a moderately small dataset.

2 Background and Related Work

Our work focuses on the dataset published by Mangal et al. as a use case [12] simulated using Dream 3D [6]. This dataset was accompanied by a work exploring other simple machine learning models to detect stress hotspots [11]. Key differences between this work and ours are the choice of featurization and predicted targets. Our work focuses on an image processing using deep nets on micrograph

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Figure 1: Left: Full 3D Microstructure from Dataset. Right: Single slice used as input to model.



images to extract features, and we attempt to predict a full stress field rather than the stress hotspots that would occur from applied stress.

Many deep learning methods were designed for the task of stress prediction in microstructures to supplant expensive Finite Element Method simulations on microstructures, which was typical for the task [18]. A number of deep neural approaches were applied across a number of different tasks and datasets along a process-structure-property breakdown [2]. 3D CNNs [14] and Physics-Informed Neural Networks [7] have also been developed to predict properties of composite microstructures.

In this work, several methods were used to further improve the data requirements and improve out-of-domain performance using data from other microstructure textures. Domain Adversarial Neural Networks and Multisource Domain Adaption Networks [5] [20] are aimed at improving source-target adaptation by an adversarial game against another trained classifier preventing a model from encoding texture differences in the latent representation. Deep CORAL [17] is another method that aims towards solving a similar problem within the loss functions of the model. Instance Reweighting is also a popular method to improve performance on a target dataset using labels from a source dataset [9].

3 Dataset

The dataset considered in this paper consists of 36 synthetic 3D equiaxed polycrystalline microstructures with different cubic textures, associated microstructural descriptors, and the corresponding results from applying a uniaxial tensile stress through EVPFFT simulations. These textures were selected from a number of common FCC rolling texture components such as brass, copper and Goss components, and all have different distributions of stress fields. For each texture, six microstructure instantiations are created, thus resulting in about 30000 grains per crystallographic texture. The texture intensity is characterized by multiples of random density (MRD) which is the intensity of a crystallite orientation with respect to it's intensity in a randomly textured material. For each texture studied here, the texture intensity is varied from weakly textured ($<$10 MRD) to strongly textured ($>$30 MRD) among the six instantiations. The microstructures are discretized on a $128 \times 128 \times 128$ grid, which allows the use of image based crystal plasticity models. Further details on the dataset are available in [12]. In our experiments, we use the 128 slices of strains along the z axis of each microstructure as images, and attempt to predict either the average stress tensor across the entire 3D microstructure, or the field along a given direction at each pixel.

4 Methods

4.1 Optimization Objective

When given a slice of a microstructure's Euler angle profile after appropriate transformations, which we use as an approximization of an experimental micrograph, our objective is to predict the stress profile in a number of directions. More precisely, two different targets are used in these experiments: one focused on a global property, and one focused on more local properties.

3D Stress Mean. The first is predicting the stress tensor when averaged across the entire microstructure as a measure of a global property across the whole 3D structure. Thus the target consists of a single numerical label for each 128x128 input slice. In this case, each microstructure has the same label for each slice, thus while the range of labels is large, the amount of unique labels is small.

Stress Field. The second is predicting all 128 x 128 values of each component of the stress field at each pixel. For the field prediction objective, in addition to trying out several out-of-domain prediction methods, we also compare the performance of models trained using a single simulated microstructure of the target texture along with the training dataset from the remaining textures. We compare this to two formulations: one where we only train on the non-target textures, and where we only train on the single available microstructure from the target texture.

For the average stress case, we optimize the mean squared error between the predicted stress in the 11 direction and the true label for each slice. For the stress field, we optimize on the mean squared error of the the 128x128 stress field in the 11 direction, and true 128x128 stress field. Results in other stress directions were similar, but we only report extensive results in the 11 direction.

4.2 Overview of Methods

In this work, a variety of conventional machine learning, domain adaptation and transfer learning methods are used to improve regression accuracy on our objectives. Domain adaptation techniques are a set of transfer learning strategies designed to address the problem of domain shift, where a model trained on data from one domain (source domain) may perform poorly when applied to a different domain (target domain) due to distributional differences. In this case, each texture had different attributes, and thus were treated as separate domains. However, a key challenge is that data is sparse in both source and target domains. As a baseline comparison, we use a standard U-Net model with an architecture provided by our colleague [1]. Hyperparameters were searched over a wide range of choices and computational complexities, and results reflect the best found hyperparameters for each model type. These hyperparameters are included in the supplementary material.

Key parameters which we varied in our experiments include model architectures, such as CNN [8], ViT with adapters [4] [19] [13], and U-Nets [15] [1]; pretrained networks such as VGG as a base featurizer [16] [10], instance reweighing, and deep adversarial domain adaptation methods; and auxiliary losses such as autoencoder reconstruction error. A more in-depth description of these methods are in the supplementary materials.

4.3 Evaluation Set-Up

We train and evaluate all models using a 5-fold validation set-up, with 4 sets of held-out data ranging from most out of texture, to least out of texture.

1. Out of Texture: In each fold, all slices of all microstructures in one entire texture was held-out to evaluate out-of-texture performance. This consists of 768 total samples.
2. Different Microstructures: Of the 5 remaining textures, one microstructure was held out from each texture, for which all slices were used to evaluate performance within texture, but in different microstructures. This consists of 640 total samples.
3. Distant Slices: We have found that slices within the same microstructure but a distance of at least 10 slices away tend to vary considerably from other slices within the same microstructure. To that end, one validation set consists of a chunk of 28 slices from each training microstructure, which are 10 slices from the training samples. This buffer of 10 slices on either side is unused for training or validation. This consists of 700 total points.
4. Same Distribution: Of the remaining samples, 20% is randomly selected as an additional validation set. As these points are directly adjacent to the training data, this fold often has similar error to the training set. This consists of around 400 points.

The remaining 1600 samples are kept as the training set. This entails around 320 training slices for each texture.However, because this data was produced from a single structure, adjacent slices were very highly correlated, thus many samples were nearly identical, though these samples were not removed.

[1]Thanks to the illustrious George Pasparakis for his TensorFlow implementation of the U-Net!

Table 1: **Mean squared error predicting microstructure-averaged stress across 11 direction.**

Mean Pred Model	Train Set	Train Dist	Distant Slice	Unseen Mcrst, Seen Text	Unseen Text
Mean	1.00	1.00	1.00	1.24	16.26
CNN	0.06	0.08	1.01	1.34	16.32
CNN + Rod. Vec	0.07	0.08	1.03	1.43	16.70
CNN + Reconstruct loss	0.08	0.09	1.04	1.36	16.98
CNN + MDAN	0.37	0.29	1.04	1.38	17.45
VGG	0.04	0.11	0.92	1.49	16.80
VGG + Reconstruct Loss	0.04	0.13	0.94	1.46	15.99
ViT	**0.02**	**0.01**	0.06	0.48	20.53
ViT + Reconstruct Loss	**0.02**	**0.01**	0.06	0.46	19.99
ViT + Random Crop	0.04	0.02	**0.03**	**0.39**	19.415
ViT + Binning	-	3.61	3.48	4.96	**10.79**

Our input data was featurized using the Euler angles of each microstructure layed out in a 2D image. We also experiment with additional transforms on these features such as the Log transform, found to be effective experimentally, or by converting to other popular angular representations, such as Rodrigues vectors [3]. For clarity, we only report results for stress along the 11 direction.

5 Results and Analysis

5.1 3D Mean Objective

Our strongest model for most out of domain validation used a Visual Transformer as a featurizer and the Log of the euler angles as features, and predicted bins rather than direct regressions. The mean of training values in each bin was used as a numerical prediction when comparing error. We suspect the benefit from binning came from preventing the predictions on images that were too far out of domain from being too far out of the typical range. Results are shown in table 1.

We find that adversarial domain adaptation methods were somewhat ineffective in our experiments. Our experiments with transfer learning revealed that pretrained models did not perform significantly better than freshly initiated models with similar architectures, though training on other textures was beneficial. For this objective, transformer-based models outperformed convolution based models, possibly due to the attention operation that captures non-local similarity being more conducive to global relations. ViT models were able to perform especially well when using the log of the Euler angle as a feature. A reconstruction loss was marginally beneficial to VGG and CNN models in some cases, but not in the ViT model. Augmenting the dataset with random crops was beneficial to the ViT model, possibly to combat overfitting.

5.2 Field Objective

For predicting the field objective, our best model used a randomly initialized CNN, where no downsampling occurred. The use of a CNN over a U-Net, which is more typically used in the literature, provided a noticeable improvement to out of domain accuracy, though for training error and in-distribution error) it performed worse than U-Net. ViT, which produced the strongest results predicting a global property out of domain, performs the poorest here. Results are shown in table 2.

Overall, we find transfer learning to not be particularly effective, as our VGG and ViT based models produced similar error to our models with the same architecture but freshly initialized weights.

A comparison of models differentiating U-Net and CNN is shown in a section of table 2, where the key differences between the U-Net and CNN structures are the downsampling of the intermediate representations, and the U-Net style residual connections. In these results, the CNN model does not reduce representations to a lower resolution. The CNN with Compression model reduces representations to a lower resolution in the same way as our U-Net. The CNN with Residuals model connects layers in a U shape much like U-Net, but does not lower resolution. In these experiments, the models with compression (U-Net and compressed CNN) perform similarly with lower train and in-distribution error than the uncompressed CNNs, but higher out-of-domain error, while the uncompressed models perform better on the unseen microstructures and textures. This is typically emblematic of overfitting on the training set. An analysis of the predicted values for U-Net and CNN architectures reveals that the U-Net produces predictions more closely fitting the true distribution for the slices adjacent to training slices than the CNN, which produces sharp predictions near the mean. However, for a distant slice, this relation is reversed, and the CNN produces predictions that more

Table 2: **Mean squared error predicting stress field across 11 direction.**

Field Pred Model	Train Set	Train Dist	Distant Slice	Unseen Mcrst, Seen Text	Unseen Text
Mean	178.54	176.13	179.39	146.91	148.95
CNN	105.55	108.25	**125.28**	**101.90**	**115.53**
U-Net	**37.51**	**64.32**	171.81	141.60	147.81
VGG	118.75	133.99	193.25	158.06	159.95
ViT Patch Size 4	96.14	114.73	172.43	141.61	146.89
CNN	105.55	108.25	125.28	101.90	**115.53**
CNN + Residuals	105.92	107.87	**124.82**	**101.78**	116.80
CNN + Compress	47.19	80.63	194.47	160.54	164.24
U-Net	**37.51**	**64.32**	171.81	141.60	147.81
ViT Pretrained	63.74	112.91	209.98	172.42	173.63
ViT Patch Size 4	96.14	114.73	**172.43**	**141.61**	**146.89**
ViT Patch Size 8	72.43	105.38	214.15	177.89	178.95
ViT Patch Size 16	50.62	80.85	213.64	178.02	179.35
ViT Patch Size 32	**44.81**	**70.45**	207.69	172.44	174.06
U-Net	**37.51**	**64.32**	171.81	141.60	147.81
U-Net with MDAN	45.50	65.08	**161.60**	**132.86**	**140.69**
U-Net with CORAL	57.82	66.12	165.46	148.95	148.76
U-Net with Reweighting	44.88	64.33	164.03	135.38	145.81

closely fit the true distribution, though neither fits it especially well. This suggests that downsampling leads to lesser performance when predicting on unseen samples compared to the U-Net. We believe that the compressions dilute the local features to some degree, thus a U-Net's low resolution features do not provide enough signal, and instead cause overfitting to the training set. Thus, a CNN with compression, which does not have the U-Net's residual connections from the high resolution features performs the worst on the out-of-distribution test set, though it is able to fit to the training data about as well as the U-Net.

Another analysis was performed to investigate our hypothesis that ViT performs more poorly on field models due to a lesser ability to discover relationships between local features. To that end, we experimented with multiple patch sizes on freshly initiated instances of ViT. These results are shown in a section of table 2. We find that out of texture performance improved with the lowest patch size, suggesting locality is an important factor to improve test accuracy on this task. However, a larger patch size led to improved performance within the training distribution.

We find that deep adversarial and alignment based methods did not provide as drastic of an improvement as architectural choice. U-Net with CORAL and MDAN did not perform especially better than the U-Net on its own, and, in some cases, caused a much larger error. We hypothesize this is due to a lack of diverse training data for these complex targets, thus these auxiliary losses were unable to provide a meaningful signal.

6 Conclusion

In this paper, we have explored numerous techniques to learn from limited simulated data spread across multiple dissimilar sources. By developing models that can learn simulated properties, we can have more accurate predictions prior to running more computationally-heavy simulations, which can direct a more efficient allocation of resources for expensive simulations and experiments.

Results from this paper may be limited to this dataset, but may extend to similar works on other simulated bodies. Compared to the U-Net baseline, specialized deep methods specifically geared towards improving transfer (MDAN, CORAL etc.) provided a minor benefit ($\approx 5\%$ error reduction) compared to architectural choices (U-Net and CNN) and data concatenation ($\approx 22\%$ error reduction). The choice of model architecture based on prior knowledge of the target leads to significant improvements for learning the correct relationships. Labels where local features are more relevant, such as field prediction, make better use of CNN models which preserve the high resolution features. For global properties, models that enable interactions across the whole image such as ViT led to more accurate predictions ($\approx 97\%$ error reduction in domain, $\approx 37\%$ out of domain). However, overfitting must be carefully controlled using regularization methods such as binning to improve predictions out of texture.

Acknowledgments

This project was funded in part by DEVCOM ARL through the HTMDEC project.

References

[1] Anindya Bhaduri, Ashwini Gupta, and Lori Graham-Brady. Stress field prediction in fiber-reinforced composite materials using a deep learning approach. *Composites Part B: Engineering*, 238:109879, 2022.

[2] Frederic E. Bock, Roland C. Aydin, Christian J. Cyron, Norbert Huber, Surya R. Kalidindi, and Benjamin Klusemann. A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Frontiers in Materials*, 6, 2019.

[3] Jian S. Dai. Euler–rodrigues formula variations, quaternion conjugation and intrinsic connections. *Mechanism and Machine Theory*, 92:144–152, 2015.

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv*, abs/2010.11929, 2020.

[5] Yaroslav Ganin, E. Ustinova, Hana Ajakan, Pascal Germain, H. Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. Domain-adversarial training of neural networks. *ArXiv*, abs/1505.07818, 2015.

[6] Michael Groeber and Michael Jackson. Dream.3d: A digital representation environment for the analysis of microstructure in 3d. *Integrating Materials and Manufacturing Innovation*, 3:5, 02 2014.

[7] Ehsan Haghighat, Danial Amini, and Ruben Juanes. Physics-informed neural network simulation of multiphase poroelasticity using stress-split sequential training. *Computer Methods in Applied Mechanics and Engineering*, 397:115141, 2022.

[8] Yann Lecun, Patrick Haffner, and Y. Bengio. Object recognition with gradient-based learning. 08 2000.

[9] Shuang Li, Shiji Song, and Gao Huang. Prediction reweighting for domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 28(7):1682–1695, 2017.

[10] Julia Ling, Maxwell Hutchinson, Erin Antono, Brian L. DeCost, Elizabeth A. Holm, and Bryce Meredig. Building data-driven models with microstructural images: Generalization and interpretability. *ArXiv*, abs/1711.00404, 2017.

[11] Ankita Mangal and Elizabeth A Holm. Applied machine learning to predict stress hotspots i: Face centered cubic materials. *International Journal of Plasticity*, 111:122–134, 2018.

[12] Ankita Mangal and Elizabeth A. Holm. A dataset of synthetic face centered cubic 3d polycrystalline microstructures, grain-wise microstructural descriptors and grain averaged stress fields under uniaxial tensile deformation. *Data in Brief*, 19:2029–2036, 2018.

[13] Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. Adapterhub: A framework for adapting transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020): Systems Demonstrations*, pages 46–54, Online, 2020. Association for Computational Linguistics.

[14] Chengping Rao and Yang Liu. Three-dimensional convolutional neural network (3d-cnn) for heterogeneous material homogenization. *ArXiv*, abs/2002.07600, 2020.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015.

[16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[17] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *ECCV Workshops*, 2016.

[18] Taj, B Prasanth Kumar, and K. Vijayasree. Microstructure based finite element analysis for deformation behavior of mg based composites. *Materials Today: Proceedings*, 4(8):9095–9099, 2017. International Conference on Advancements in Aeromechanical Materials for Manufacturing (ICAAMM-2016): Organized by MLR Institute of Technology, Hyderabad, Telangana, India.

[19] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics.

[20] Han Zhao, Shanghang Zhang, Guanhang Wu, José M. F. Moura, Joao P Costeira, and Geoffrey J Gordon. Adversarial multiple source domain adaptation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.