

Modeling Complex Polycrystalline Alloys Using a Generative Adversarial Network (GAN)-Enabled Computational Platform



Open Access This file is licensed under a Creative Commons Attribution 4.0

International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to

the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

In this work, the authors presented a highly efficient and powerful technique referred to as SliceGAN-Dream3D, to generate statically equivalent virtual microstructures (SEVM) of complex multi-scale polycrystalline materials, based on multi-modal imaging data in low-dimensions (e.g., 2D slices). The method nicely integrates GAN for feature learning and extraction from multi-modal imaging data and Dream3D, a powerful tool developed by the authors, for 3D virtual polycrystalline microstructure rendering. The utility of the proposed method was demonstrated by modeling two highly complex systems: cold spray formed Al 7050 alloys, containing prior particles with coarse grains (CG) and ultra-fine grains (UFGs), and additively manufactured (AM) Ti64 alloys, containing complex patterns of alpha lath variants in beta substrates. It was shown that the generated SEVM can be employed to accurately predict the nonlinear mechanical behavior of the material systems.

The paper is overall very-well written. The method presented is very impressive (especially the capability to generate realistic SEVM with 10^5 grains on multiple scales) and is of great technical significant and interest to the MSE community. I like the fact that the authors clearly demonstrated both the advantages and current limitations (mainly due to data inefficiency) of the method. I'm happy to recommend its publication in Nat Comm, and have some suggestions for optional minor revisions below.

It would be helpful to explicitly state in the caption of figure 1 that panels (a)-(e) are associated with Al alloy AA7050. Currently this information was only mentioned in the text.

The scale bar in Fig. 2a is very difficult to see.

Is it possible hard-constrain the generative model such that a better match of targeted simulated statistics (such as the aspect ratio distribution shown in Fig. 3e) can be achieved? This would be system-specific, but might be beneficial if there is particular interest in a common material system (e.g., Al 7050 alloy).

The authors might be interested in knowing a relevant work [S. Chen, et al., Stochastic Multi-Scale Reconstruction of 3D Microstructure Consisting of Polycrystalline Grains and Second-Phase Particles from 2D Micrographs, Metallurgical and Materials Transactions A 47, 1440 (2016)] in which a stochastic fusion via simulated annealing based on multi-modal 2D imaging data (EBSD, SEM etc) was employed to generate virtual 3D heterogeneous polycrystalline microstructure of Sn alloy containing secondary phase precipitates.

Reviewer #2 (Remarks to the Author):

Great to see the development of methods for tackling heterogeneous materials. It is also interesting to see the combination of traditional microstructural generation approaches with more modern ML methods. However, I don't think the paper currently makes a clear enough case that the method they present is sufficiently robust or general to be relevant to a wide section of the materials science community. The reported improvement in accuracy also seems to be open to interpretation. In addition, the following points should be addressed:

1. The importance of representativity is stressed, but later it simply says that the low-resolution image is assumed to be representative. Can you justify this.
2. Although the stress strain curves in figure 4 do seem to match up well, it's not clear how difficult this is. Would any random grain cluster basically have the same curve?
3. The pole plots in figure 5 are hard to compare. Please stack each above its equivalent and use the same colour bar scaling.
4. Figure 6 is hard to understand. A,b,c are presumably supposed to be comparable, but they've been rendered differently making this hard. It's not clear what (d) is, so perhaps add axis labels. In (e) shouldn't the green bars be the same within each row?
5. No information about the training process for the GAN. Was it trained successfully?
6. The code and data for this work are only available upon request, rather than shared in a well documented repository with the paper. This does not seem appropriate for a paper that cannot otherwise be validated.
7. Needs more discussion about the potential impact of the work. Specifically, given the particular improvements observed (i.e. in some metrics but not all), how would this impact certain common analysis reported in literature.

Reviewer #3 (Remarks to the Author):

In this work, the authors detail a process by which a combination of a Generative Adversarial Network for dimensional expansion (Kench and Cooper's sliceGAN) is integrated with microstructure synthesis packages on the Dream3D platform to create a combined synthetic microstructure generation approach that exceeds the capabilities of the constituent components individually. This approach is demonstrated on a cold spray formed aluminum alloy (AA7050) as well as an additively manufactured titanium alloy (LBPF Ti64), which together represent a range of complex microstructures containing both multimodal grain size distributions and anisotropic grain shapes. Microstructures generated with this new platform are compared to both individual components on a variety of grain size and shape metrics.

From a materials perspective, this paper presents novel contributions to the field of 3D microstructural synthesis and offers a promising platform for accessible generation of a variety of 3D microstructures that were previously challenging to realize. The constituent architectures themselves are not fundamentally new, but the integrated approach offers a degree of accessibility that would be of significant benefit to the materials community.

However, from a machine learning perspective, this reviewer feels that there are some critical points about the generative architecture and synthesis pipeline that should be addressed before this

manuscript can be published in Nature Communications. These critical points of issue are outlined first, followed by some additional minor comments for the authors to consider.

Major Points:

The points of greatest concern with this work center on clarity and transparency of results, especially as it pertains to the use of sliceGAN in within the pipeline. Despite the machine learning reference in the manuscript title, the current structure of the paper follows that of a material analysis with machine learning details scattered throughout, which makes the principles of the architecture implementation difficult to parse and creates opacity around the role of machine learning in the pipeline. Furthermore, the original work of Kench and Cooper presented many points of interest regarding material microstructure representation with sliceGAN, but it also had several points of concern around limitations of the associated approach that are specific to the sliceGAN architecture, and those points do not feel particularly well considered within this manuscript in its current form. Justifications for this argument are presented below:

1) Please include pixel and voxel counts alongside any associated image length scale representations. Length scales are essential to materials analysis, but from my understanding of Kench and Cooper's work, sliceGAN only interprets material length scale and information density in the context of pixels and voxels.

2) What are the output dimensions for the SEVM phase volumes produced with sliceGAN? Kench and Cooper's framework theoretically allowed for an arbitrary cubic volume of side length L , but I believe their actual architecture defaulted to L being 64 voxels. Was that same size used here? The only cube size reference I was able to find was 63.4 microns in Figure 6 b), but it is in microns, which again, from an information density perspective, is ambiguous.

3) There is a post-processing step applied to the sliceGAN output and the details of this are explained in the discussion section, which is a bit confusing. Discussion of its implications is fine, but information on input and output details, dataset structure, and processing steps that are necessary for GAN functionality should be grouped together in either the section discussing the implementation of the GAN or in the methods section. If length constraints are an issue, a supplemental section on architectural details is also fine, but either way, these details should be aggregated and accessible for anyone who wishes to replicate the process of this work.

4) Related to the prior point, review of the Section discussing the use of sliceGAN in the integrated platform and associated masking process for the CSF AA7050 also brings to light some confusion about sources of error in the pipeline process as they relate to the implementation of the GAN. The plots in Figure 9(g) clearly show some differences between sliceGAN, the true result, and the improved model, especially in the XY plane fractions, but it is not clear from the analysis whether those errors arise from the CG/UFG separation process or from sliceGAN. Were comparisons done with the associated area fractions after the separation and thresholding? Also, it is known that n -channel generative models like slice GAN have a tendency to add spurious connectivity to their distributed phases, and some evidence of this can be seen in the images shown in figure S2 across different scales (but again, it is difficult to evaluate because no pixel-based details are given). Was any sort of shape analysis (2D invariant/Hu moment) done on the associated phase outputs for sliceGAN as compared the original phase maps, both with and without the thresholding and erode/dilation? It feels like this would add some clarity about the sources of variability seen

between the original phase maps and those produced for the SEVMs.

5) One of the core motivations of this work was the use of GANs to generate microstructures with complexity that exceeds the capabilities of current synthetic pipelines, and one of the key demonstrators of this was CSF aluminum containing a widely varied grain size distribution with bimodal character. From a machine learning perspective, systems with this type of strong modal dependence are challenging for generative models that sample from a single latent space because that sampling is generally uniform (or at least unimodal), meaning that some aspects of latent sampling are overemphasized with respect to others. Karras et al. discusses this idea at length in their presentation of style-based generators (<https://arxiv.org/abs/1812.04948>).

When Kench and Cooper first published sliceGAN, they did not implement any sort of adjustments to account for multimodal/entangled features, so Chung and Ye released a Matters Arising Reusability Report (<https://doi.org/10.1038/s42256-021-00400-4>) outlining this issue and how it impacted the performance of the architecture. With this, Chung and Ye also released a publicly available modification of sliceGAN that includes the adaptive instance normalization (AdaIN) steps that mitigate this effect in place of the basic normalization protocols Kench and Cooper used. Further adding to relevance, they demonstrate this on systems of variable grain/phase feature size. Since the focus of this work is complex structures with strong modal dependence, was this AdaIN version of sliceGAN used, or just the original?

If this modified architecture was not used, it presents a major opportunity for performance improvement, as it not only helps disentangle correlations between phase feature size and distribution, but it also allows for a degree of tunability in that associated disentanglement, which would likely contribute to improved performance. Depending on the exact arrangement of the full pipeline (whose details should probably be detailed a bit more explicitly around Figure 8), implementing a changed architecture may prove to be a lot of work, but at the very least, the presence of this possible source of improvement should be acknowledged in the section on avenues for improvement.

Minor Points:

6) The title is a bit of a challenger to understand. Maybe something like “Generative Approaches for Statistically Equivalent Virtual Microstructures of Complex Polycrystalline Materials”? This change is entirely at the author’s discretion, but I had to read the title several times to parse it.

7) When stating the training details for the network (time, equipment, etc.), it would also be beneficial to include the approximate size (MB/GB) of the training set used, since learning convergence is dependent on passage of the entire set through the model.

8) Very minor, but some of the figure captions (1 and 6) look like they might bleed into the footer space. I appreciated the size of the figures for being able to visualize things, but don’t lose the captions in the final print.

9) In the avenues for improvement section, the authors mention they are looking at several ways of improving the approach to deal with data bottlenecks. In their reusability report, Chung and Ye mention that the approach would benefit from a progressive learning approach, which also seems highly salient here since sliceGAN is being used to produce maps across several different length

scales. Additionally, the authors mention the possibility of augmentation, which would be interesting to explore because to my understanding, Kench and Cooper's original model uses the rotational sensitivity of convolution to preserve the relative axis dependencies in sliceGAN. Augmentation would certainly provide more training information, but it would also disrupt these relationships. Creating an option for directionally-depending data augmentation would be interesting as a tool for encouraging isotropy along particular microstructural directions when relevant.

Reviewer #4 (Remarks to the Author):

The manuscript deals with the generation of multi-modal morphology of virtual microstructure based on generative adversarial network (GAN)-based methods. The authors integrated SliceGAN and Dream3D builders to generate virtual microstructures of cold spray formed (CSF) Al 7050 alloys and additively manufactured (AM) Ti64 alloys and compared generated virtual microstructures with ones from other conventional techniques (SliceGAN or Dream3D only). They also developed a multiscale model to examine mechanical properties of generated virtual microstructures.

Although the topic is interesting and the novelty of the methodology is recognized, I am afraid that there is much ambiguity in the results and discussion showing the superiority of this study. For example, in page 21, the authors state "The SliceGAN-based SEVM in Figure 6(b) exhibits a poor representation of the grain size and aspect ratio distributions, but a good representation of the misorientation angle distribution. On the other hand, the Dream3D and SliceGAN-Dream3D platform-based SEVMs in Figures 6(a,c) match the three distributions closely, since grain-packing is achieved through Dream3D for both cases." Each method has its advantages and disadvantages depending on the physical properties of interest, and therefore the method developed in this study is not necessarily superior. In particular, CSF is characterized by a collision-distorted structure, and AM is characterized by a long and thin Widmanstatten structure. Therefore, it is important to reproduce the aspect ratio correctly. However, it is crucial to note that the method employed in this study may not adequately represent the aspect ratio. In fact, there is a concern that the aspect ratio in Figure 3(e) is misaligned between EBSD and SEVM.

Moreover, the primary selling point of this study is the generation of multi-modal morphology in the microstructure. However, if I understand correctly, it has generated the coarse grain (CG) and the regions of ultra-fine grains (UFGs), but not the details of the UFG microstructure itself. It seems to be only a phenomenological discussion of the mechanical properties, which are matched by parameters in the multi-scale model. I am concerned that the claims in this paper may mislead general readers into thinking that the methodology proposed in this study generates a multi-modal morphology of virtual microstructure that can reproduce the mechanical properties directly. Therefore, I do not recommend this manuscript for publication in Nature Communications in its current form.

Authors Response to Reviewers Comments and Suggestions

The authors take this opportunity to thank the editor and reviewers for their insightful and critical comments that have helped improve this manuscript. We have tried to incorporate suggested changes accordingly, wherever possible. Along with this authors' response, a revised manuscript with highlighted changes in red font has been attached.

Editors Comments and Response:

Given these mixed views, we are willing to reconsider the work if the referees' concerns are clearly address in the revised version. In particular, we ask you to openly share the code and provide a complete analysis of the performance of the platform, clearly showing the role of each Dream3D and SliceGAN in the prediction.

Authors' Response:

In this revised manuscript, we have made substantial changes in response to issues raised by the reviewers. We have created a public GitHub repository under a MIT license. All SliceGAN codes are available at <https://github.com/CMRL-JHU/sliceGAN>, and pre/postprocessing codes are available at <https://github.com/CMRL-JHU/csf-al7050-processing>.

In this repository, we have shared all necessary files to execute results in this paper. It includes our modification of the original SliceGAN code developed at Imperial College [1], creating a link between Dream3D and SliceGAN.

Editors Comments and Response:

Further, we ask you to address reviewer #3's comment on the alternative SliceGAN architecture. I do feel that in order for this manuscript to be editorially suitable for Nature Communications, these points should be significantly addressed. At the same time, the other substantial technical concerns must also be addressed.

Authors' Response:

We have tried to address all questions raised by the reviewers to our best ability. We sincerely hope that the editor and reviewers will be satisfied with our response and substantial revisions. Please find our response to individual reviewers next.

REVIEWER 1

General Comments:

In this work, the authors presented a highly efficient and powerful technique referred to as SliceGAN-Dream3D, to generate statically equivalent virtual microstructures (SEVM) of complex multi-scale polycrystalline materials, based on multi-modal imaging data in low-dimensions (e.g., 2D slices). The method nicely integrates GAN for feature learning and extraction from multi-modal imaging data and Dream3D, a powerful tool developed by the authors, for 3D virtual polycrystalline microstructure rendering. The utility of the proposed method was demonstrated by modeling two highly complex systems: cold spray formed Al 7050 alloys, containing prior particles with coarse grains (CG) and ultra-fine grains (UFGs), and additively manufactured (AM) Ti64 alloys, containing complex patterns of alpha lath variants in beta substrates. It was shown that the generated SEVM can be employed to accurately predict the nonlinear mechanical behavior of the material systems.

The paper is overall very-well written. The method presented is very impressive (especially the capability to generate realistic SEVM with 10^5 grains on multiple scales) and is of great technical significant and interest to the MSE community. I like the fact that the authors clearly demonstrated both

the advantages and current limitations (mainly due to data inefficiency) of the method. I'm happy to recommend its publication in Nat Comm, and have some suggestions for optional minor revisions below.

Authors' Response:

Thank you very much for the positive comments on our paper, as well as suggestions for improvement. We appreciate the encouragement and have addressed the issues raised.

Reviewer comment 1: It would be helpful to explicitly state in the caption of figure 1 that panels (a)-(e) are associated with Al alloy AA7050. Currently this information was only mentioned in the text.

Authors' Response:

The following text has been added to Figure 1:

“Microstructure characteristics of the (a-e) cold-sprayed Al7050 alloy and (f-g) Ti64 alloy manufactured by Laser Bed Powder Fusion (LBPF).”

Reviewer comment 2: The scale bar in Fig. 2a is very difficult to see.

Authors' Response:

We have added a new image with a larger scale bar.

Reviewer comment 3: Is it possible hard-constrain the generative model such that a better match of targeted simulated statistics (such as the aspect ratio distribution shown in Fig. 3e) can be achieved? This would be system-specific, but might be beneficial if there is particular interest in a common material system (e.g., Al 7050 alloy).

Authors' Response:

Thank you for this question. Indeed, more aspects of the microstructure can be considered in the computation of the loss functions, as proposed in [1]. Some of the microstructural features were added to the SliceGAN code. One of the attempts to create the cold-sprayed microstructure used some of these features, where we tried to obtain the Euler angles directly from the GAN model, However the results were not as good as we had expected (see Figure 6b and Supplementary Figure 1). This could be due to the fact that we did not have a high-quality microstructure and associated data that could be applied to the design of the neural network. The implementation of this idea will increase the level of complexity and will be explored in future work.

[1] Anindya Bhaduri, Ashwini Gupta, Audrey Olivier, Lori Graham-Brady. An efficient optimization-based microstructure reconstruction approach with multiple loss functions. Computational Materials Science 199 (2021). <https://doi.org/10.1016/j.commatsci.2021.110709>

Reviewer comment 4: The authors might be interested in knowing a relevant work [S. Chen, et al., Stochastic Multi-Scale Reconstruction of 3D Microstructure Consisting of Polycrystalline Grains and Second-Phase Particles from 2D Micrographs, Metallurgical and Materials Transactions A 47, 1440 (2016)] in which a stochastic fusion via simulated annealing based on multi-modal 2D imaging data (EBSD, SEM etc) was employed to generate virtual 3D heterogeneous polycrystalline microstructure of Sn alloy containing secondary phase precipitates.

Authors' Response:

We thank the reviewer for sharing this paper, which shows another good example where the classical microstructure generators fail to create complex microstructures. We have cited this paper (ref. 31) in the manuscript in the following line:

“For example, it is difficult to generate polycrystalline microstructures containing dispersed intragranular precipitates [8,31]...”

REVIEWER 2:

General Comments:

Great to see the development of methods for tackling heterogeneous materials. It is also interesting to see the combination of traditional microstructural generation approaches with more modern ML methods. However, I don't think the paper currently makes a clear enough case that the method they present is sufficiently robust or general to be relevant to a wide section of the materials science community. The reported improvement in accuracy also seems to be open to interpretation. In addition, the following points should be addressed:

Authors' Response:

Thank you for the comments on our paper. We have tried to address the issues raised next.

Reviewer comment 1: The importance of representativity is stressed, but later it simply says that the low-resolution image is assumed to be representative. Can you justify this.

Authors' Response:

The information needed to model cold sprayed aluminum spans multiple length scales. These include the prior particle length-scale $\sim 20\mu\text{m}$, the coarse grain length scale $\sim 1\mu\text{m}$, and the ultra-fine grain length scale $\sim 0.1\mu\text{m}$. This length scale range implies that in order to appropriately capture details required for a statistically significant number of prior particles with coarse grain representation and the inter-particle regions of ultra fine grains at a single resolution, a very large data-set would have to be generated and modeled. This is not deemed to be an efficient approach, both in terms of the SEVM generation, as well as modeling. Consequently, we have decided to pursue a multi-scale approach for both purposes.

The choice of the two EBSD maps with different scales was necessary for two reasons. The small EBSD provides more high-resolution data for the ultra-fine grains, while the larger EBSD map provides information on the coarse grains along with the morphology of prior particles. Representativity is defined in terms of the number of grains within the EBSD maps of each domain. While there is no fixed metric for this in general, in our analysis we have assumed a statistically sufficient representative EBSD map to contain at least 500 grains. The small EBSD map representing UFGs contains approximately 6000 grains after preprocessing (grain size filtering of grains containing less than 5 pixels), while the number of grains in the large EBSD map representing the coarse grains is around 1500.

Reviewer comment 2: Although the stress strain curves in figure 4 do seem to match up well, it's not clear how difficult this is. Would any random grain cluster basically have the same curve?

Authors' Response:

The optimization procedure undertaken for the bimodal microstructure simultaneously incorporates two models that individually capture the micromechanical behavior of the coarse grains and the ultra-fine grains respectively. Appropriate error measures are defined in terms of the overall discrepancy between the stress-strain behavior results obtained from simulations and experiments. The optimization process entails the following steps (provided in the Methods section):

- Selection of critical constitutive/crystal plasticity parameters,
- Computation of the bounds for each parameter for efficient evaluation
- Running an optimization procedure based on tree-structured Parzen (TPE) estimator using the python package Optuna [<https://optuna.org/>]
- Iteratively converge on constitutive parameters from simulations of the bimodal SEVM problem
- Check on the error function (objective function in Optuna) for convergence.

While this is not a difficult process in general, the steps outlined require careful sequential execution.

For different microstructural morphological and crystallographic characteristics, the overall stress-strain behavior is expected to be different. Even though the crystal plasticity parameters will remain

the same (as long as the material is the same), the upscaled constitutive model (UCM) coefficients will depend on the microstructural characteristics of the material. Consequently, the calibration process needs to be executed again for these multi-modal microstructures. We have added the following sentences to the Methods section (marked).

“It is important to note that the calibrated parameters depend on the morphological or crystallographic characteristics of the microstructure, as well as on the material properties. Hence, the calibration process needs to be repeated if any of these characteristics are different. For the bimodal CSF AA7050 alloy in this study, the number of outer loops needed for calibration is 38 with a maximum of 100 inner loops per outer loop. The UCM model simulations take approximately one minute on a single processor, while simulations with the CP model take approximately 2 hours on 24 processors of a Rockfish cluster. It consists of Intel Xeon Gold Cascade Lake 6248R processors with 4GBs of 2933MHz DDR4 memory per processor.”

Reviewer comment 3: The pole plots in figure 5 are hard to compare. Please stack each above its equivalent and use the same color bar scaling.

Authors’ Response:

Thank you for this suggestion. The images have now been stacked and bars added to Figure 5(f).

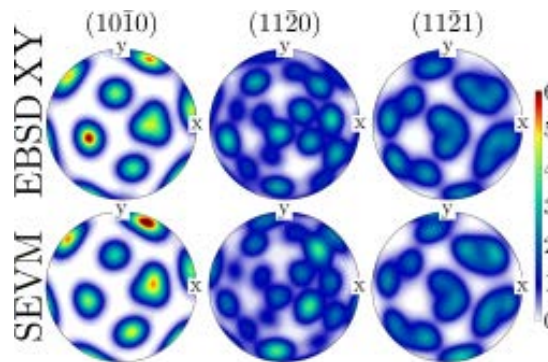


Figure 5(f). Comparison of the pole figures from EBSD map and SEVM .

Reviewer comment 4: Figure 6 is hard to understand. A,b,c are presumably supposed to be comparable, but they’ve been rendered differently making this hard. It’s not clear what (d) is, so perhaps add axis labels. In (e) shouldn’t the green bars be the same within each row?

Authors’ Response:

Thank you for pointing this. Figure 6 has been improved. We have replotted Figure 6(a-c) with the same scale for consistency. Figure (d) is a 2-point correlation function map of the prior particle domain. It is used to show patterns in the non-homogeneous distribution of CG and UFG regions. Essentially, the SEVM, whose 2-point correlation closely matches that of the EBSD has a more accurate distribution of the CG and UFG regions. We have added coordinates to this figure for clarity. Also Figure 6e) has been replotted to keep the same histogram bars width. We have changed the caption as follows:

“Comparing the 3D SEVM of the CSF AA7050 alloy generated by three alternative methods, viz. (a) Dream3D, (b) SliceGAN, and (c) integrated SliceGAN-Dream3D platform. Comparison of SEVM microstructural characteristics with those from EBSD maps: (d) Prior particle 2-point correlation function of a representative slice of each SEVM in each orthogonal plane, and the corresponding extremum of the central ellipse measured at $\$S_2=0.25\$$ across each axis, (e) PDFs of the major grain features viz. grain

size, aspect ratio, and misorientation for each SEVM measured as 2D slices in the YZ plane, and (f) texture (pole figures).”

Reviewer comment 5: No information about the training process for the GAN. Was it trained successfully?

Authors’ Response:

The training was performed successfully and without the presence of mode collapse. We have now added two figures to the supplementary material section VI, showing the loss function for training the GAN model to create (a) the cold spray formed microstructure and (b) the additively manufactured Ti64 alloy microstructure. The hyperparameters of the model have been slightly changed with respect to the original SliceGAN paper.

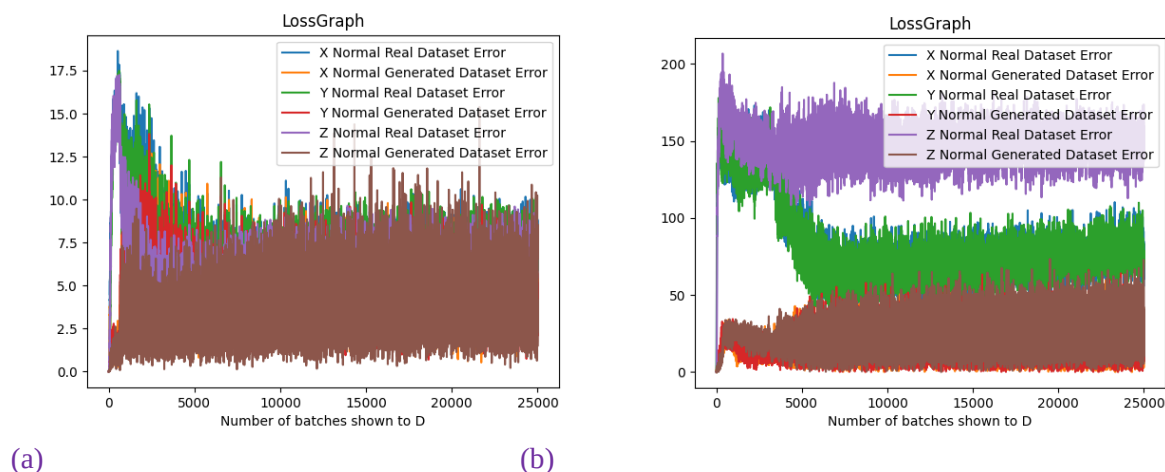


Figure S11. Loss function for training the GAN model to create (a) the cold spray formed AA7050 microstructure and (b) the additively manufactured Ti64 alloy microstructure.

Reviewer comment 6: The code and data for this work are only available upon request, rather than shared in a well documented repository with the paper. This does not seem appropriate for a paper that cannot otherwise be validated.

Authors’ Response:

We agree with the reviewer’s comment and acknowledge that this will help the materials science community to progress on the use of machine learning for microstructure generation.

We have created a public GitHub repository under a MIT license. All SliceGAN codes are available at <https://github.com/CMRL-JHU/sliceGAN>, and pre/postprocessing codes are available at <https://github.com/CMRL-JHU/csf-al7050-processing>.

In this repository, we have shared all necessary files to execute results in this paper. It includes our modification of the original SliceGAN code developed at Imperial College [1], creating a link between Dream3D and SliceGAN.

The following text has been added to the Data Availability section:

All datasets have been shared within the codes on the GitHub repository at <https://github.com/CMRL-JHU/sliceGAN>. The input data used to generate the cold-sprayed and Ti64 microstructures are in the input folder of SliceGAN.

The following text has been added to the Code availability section:

All codes for SliceGAN, and its pre/post-processing are available at <https://github.com/CMRL-JHU/sliceGAN> and <https://github.com/CMRL-JHU/csf-al7050-processing> respectively.

Reviewer comment 7: Needs more discussion about the potential impact of the work. Specifically, given the particular improvements observed (i.e. in some metrics but not all), how would this impact certain common analysis reported in literature.

Authors' Response:

Accurate representation of key microstructure features is an essential factor in the development of material models, as they can significantly affect material behavior. The methodology developed in this paper not only demonstrates a viable path for generating statistically equivalent material microstructures for complex multi-modal and heterogeneous microstructures, but also introduces a robust multiscale modeling technique that can overcome the challenges of accounting for material features at a range of length scales. Such complex heterogeneous microstructures include the cold-sprayed Al alloys and additively manufactured Ti64 alloys with Widmanstätten microstructures, among others. The proposed computational framework on GANs-based virtual microstructure generation, along with the multiscale modeling methods, is expected to open the door for addressing structure-property relations in complex materials that have thus far been difficult with conventional models, especially with respect to extreme properties like ductility and fatigue.

The following paragraph has been added to the Discussion Section:

Material properties, especially those corresponding to extreme values like fatigue life and ductility, are very sensitive to the morphological and crystallographic aspects of the microstructure [69]. Experimental studies of fatigue life in CSF Al alloys [68] have shown that cracks initiate in the prior particle substructure near the CG grain boundaries and have a strong influence on specimen failure. Studies in [57] have observed that cracks nucleate near the surface from particle-particle bonding defects, and subsequently evolve in the CG domain along the particle boundaries. The prior particle boundaries and the CG microstructure are major determinants of crack evolution in CSF alloys. Similarly, the fatigue crack behavior of additively manufactured Ti-6Al-4V alloy with a Widmanstätten microstructure has been reported to be highly sensitive to the crystallography and the morphology of the laths [58]. This underscores the importance of developing high-fidelity SEVMs, along with their implementation in a simulation tool for determining mechanical response.

The methodology developed in this paper not only demonstrates a viable path for generating statistically equivalent material microstructures for complex multi-modal and heterogeneous microstructures but also introduces a robust multiscale modeling technique that can overcome the challenges of accounting for material features at a range of length scales. The proposed computational framework on GANs-based virtual microstructure generation, along with the multiscale modeling methods, is expected to open the door for addressing structure-property relations in complex materials that have thus far been difficult with conventional models, especially with respect to extreme properties.

REVIEWER 3:

General Comments:

From a materials perspective, this paper presents novel contributions to the field of 3D microstructural synthesis and offers a promising platform for accessible generation of a variety of 3D microstructures that were previously challenging to realize. The constituent architectures themselves are not fundamentally new, but the integrated approach offers a degree of accessibility that would be of significant benefit to the materials community.

However, from a machine learning perspective, this reviewer feels that there are some critical points about the generative architecture and synthesis pipeline that should be addressed before this manuscript can be published in Nature Communications. These critical points of issue are outlined first, followed by some additional minor comments for the authors to consider.

Authors' Response:

We agree that this paper is not meant to be a significant advancement in machine learning, rather it is meant to provide an avenue for GANs to be used to overcome major challenges faced by conventional microstructure builders for complex materials. We have utilized advanced machine learning methods with interfaces to conventional synthetic builders and have made modifications to the existing code structure to address the class of problems considered. We believe that this is a significant development for computational modeling of complex material microstructures. In the following paragraphs, we try to address different questions raised by this reviewer.

Major Points:

The points of greatest concern with this work center on clarity and transparency of results, especially as it pertains to the use of sliceGAN in within the pipeline. Despite the machine learning reference in the manuscript title, the current structure of the paper follows that of a material analysis with machine learning details scattered throughout, which makes the principles of the architecture implementation difficult to parse and creates opacity around the role of machine learning in the pipeline. Furthermore, the original work of Kench and Cooper presented many points of interest regarding material microstructure representation with sliceGAN, but it also had several points of concern around limitations of the associated approach that are specific to the sliceGAN architecture, and those points do not feel particularly well considered within this manuscript in its current form. Justifications for this argument are presented below:

Reviewer comment 1: Please include pixel and voxel counts alongside any associated image length scale representations. Length scales are essential to materials analysis, but from my understanding of Kench and Cooper's work, sliceGAN only interprets material length scale and information density in the context of pixels and voxels.

Authors' Response:

We have now added the number of voxels in the images of figure 6. The number of pixels from the EBSD used for the generation of microstructures is 64. This value showed good quality results and the overlap of sub-images overlapped less than with larger values. The length of the microstructures is $128 \times \text{EBSD}$ resolution, where the latter is $0.5\mu\text{m}$. This can be checked for the CSF Al alloy microstructure where the length of the SEVM is $64\mu\text{m}$.

Moreover, to perform the grain packing we increase the resolution to describe grains in more detail and to avoid sharp interfaces between the grains which can give rise to meshing issues to perform finite element simulations. The number of voxels per direction is summarized in the following table and it has been added to Figure 6.

	EBSD	Dream3D	SliceGAN	SliceGAN+Dream3D
X	300	50	127	317
Y	300	50	127	317
Z	300	50	128	317

Reviewer comment 2: What is the output dimensions for the SEVM phase volumes produced with sliceGAN? Kench and Cooper's framework theoretically allowed for an arbitrary cubic volume of side length L , but I believe their actual architecture defaulted to L being 64 voxels. Was that same size used here? The only cube size reference I was able to find was 63.4 microns in Figure 6 b), but it is in microns, which again, from an information density perspective, is ambiguous.

Authors' Response:

The SliceGAN code enables the creation of microstructures with lengths of the power of 2. The microstructures shown in the present work are 64 voxels. The following sentences have been added to the Methods section, subsection iii, step 2:

For the examples shown in this work, the sampling window has a size of 64 pixels x 64 pixels. This size can be changed but this value provides enough sub-images from the masks created in step 1. They also contain the principal features of the microstructure.

Reviewer comment 3: There is a post-processing step applied to the sliceGAN output and the details of this are explained in the discussion section, which is a bit confusing. Discussion of its implications is fine, but information on input and output details, dataset structure, and processing steps that are necessary for GAN functionality should be grouped together in either the section discussing the implementation of the GAN or in the methods section. If length constraints are an issue, a supplemental section on architectural details is also fine, but either way, these details should be aggregated and accessible for anyone who wishes to replicate the process of this work.

Authors' Response:

Thank you for your suggestions. First of all, we have created a public GitHub repository under a MIT license. All SliceGAN codes are available at <https://github.com/CMRL-JHU/sliceGAN>, and pre/postprocessing codes are available at <https://github.com/CMRL-JHU/csf-al7050-processing>. This repository can be used by users to verify the results.

We have also added more details on the postprocessing of the SliceGAN output to the supplementary section: II. Additional information on the SEVM generation by the SliceGAN-Dream3d platform. The creation of the new mask needs information of the grain morphology around the mask, and a dilation/erosion process is performed based on a volume threshold. This threshold value is chosen such that the area fractions were closer to the area fractions of the EBSD maps. The following paragraph and figure S2 have been added to the paper.

“The schematic in figure S2 shows the process of creation of an improved mask similar to that seen in figure 9(f). The original mask generated by SliceGAN (approximated as a sphere just for the schematic representation), is used as an initial guide to create the actual region containing the coarse grains. All coarse grains, matching the criteria given in Equation (6) are retained, and hence the original mask is expanded to contain all voxels belonging to those grains. All grains not matching the criteria in Equation (6) are eliminated, such that the original mask is dilated to remove all voxels belonging to those grains. Using a volume fraction $V_{\text{frac}}=0.5$ should result in a mask that is similar in volume fraction to the original, whereas a $V_{\text{frac}}<0.5$ should result in a mask with a higher volume fraction.”

Reviewer comment 4: Related to the prior point, review of the Section discussing the use of sliceGAN in the integrated platform and associated masking process for the CSF AA7050 also brings to light some confusion about sources of error in the pipeline process as they relate to the implementation of the GAN. The plots in Figure 9(g) clearly show some differences between sliceGAN, the true result, and the improved model, especially in the XY plane fractions, but it is not clear from the analysis whether those errors arise from the CG/UFG separation process or from sliceGAN. Were comparisons done with the associated area fractions after the separation and thresholding? Also, it is known that n-channel generative models like slice GAN have a tendency to add spurious connectivity to their distributed phases, and some evidence of this can be seen in the images shown in figure S2 (now S3) across different scales (but again, it is difficult to evaluate because no pixel-based details are given). Was any sort of shape analysis (2D invariant/Hu moment) done on the associated phase outputs for sliceGAN as compared the original phase maps, both with and without the thresholding and erode/dilation? It feels like this would add some clarity about the sources of variability seen between the original phase maps and those produced for the SEVMs.

Authors' Response:

- *it is not clear from the analysis whether those errors arise from the CG/UFG separation process or from sliceGAN:*

The errors most likely originated from SliceGAN for two main reasons. First, the resolution of the large EBSD maps that are used in the reconstruction is low. Second, the amount of training data for the network is limited. The effect of these limitations in the data are reinforced through the new Figure S5. However, despite the sparsity of training data, SliceGAN compared favorably to the Dream3D only method in all our testing. The preprocessing of the binary EBSD maps (Figure 9 a-c) was a choice where the noise (small blue points within the red phase, see Figure 9 a-b) was decreased but as a matter of fact the preprocessing altered the binary map. We realize that an increase in the amount and fidelity of the input data to SliceGAN would certainly increase the fidelity of the results. However, in this paper we are just presenting the computational platform for digitally representing multi-modal microstructures that generally works even with sparse data. We have added the following text to Supplementary section II.

“Figure S5 shows box plot expressions of much of the same data contained in the PDF plots of Figure 9(g) with a few key additions. To see the effect of the grain cluster thresholding on the distributions, it includes the area fractions of the EBSD data after thresholding out the UFGs, but before thresholding out the small pockets of CG clusters for SliceGAN training. It is observed that the grain cluster thresholding operation primarily stretches and shifts the distributions upwards. Additionally, the figure shows dashed lines denoting the singular area fractions of the full EBSD maps. Because SliceGAN requires sub-sampling rectangular windows of the EBSD images for training, there exists a sampling bias such that there is a decreasing gradient from the center to the edges of the EBSD images of the probability of data being included in a given sample. This sampling bias can be seen clearly in the deviation from the mean of the EBSD-Cleaned distribution versus the EBSD-Cleaned True value in the XY plane. While the true value is the most important to match the results, it is also important to understand the bias that went into training the GAN. It can also be seen clearly in Figure S5 that the mask improvement process does not affect the shape of the area fraction distributions in a major way, but instead shifts the distributions downwards. It can be seen here that the shift has the effect of closely aligning the mean of the SEVM mask area fraction distributions with the true area fractions of the cleaned EBSD data.”

- *Were comparisons done with the associated area fractions after the separation and thresholding?*

Yes, the comparison was made after separation and thresholding. A new figure S5 has been added including all the same information as Figure 9(g), as well as the area fractions just after the GAKAM thresholding in order to compare the effects of the CG cluster size filter. A new section has been added to the supplementary material around what is now Figure S5, giving more details about the effects of the grain cluster thresholding operation, the mask improvement operation, and some sources of error in between.

- *it is known that n-channel generative models like slice GAN have a tendency to add spurious connectivity to their distributed phases, and some evidence of this can be seen in the images shown in figure S2 (now S3)*

While we did not notice any direct effect of spurious connectivity, such occurrences could be attributed to size discrepancy of the SliceGAN training set. For example, the dimensions of prior particles are 10 μ m-35 μ m in the XY plane. and 20 μ m-50 μ m in the YX plane. This corresponds to an average prior particle size of ~20 μ m, which translates to 40 pixels. While this average size is well within the characteristic length of sub-images in the SliceGAN training set (64x64 pixels), as well as the SEVM size of 64 μ m, there could be rare occurrences where this is violated, if the particle is not nearly centered or larger. Increasing the sample training size could be a remedy to this occurrence. The prior particles in our study seem to be interconnected in the YZ plane and create a complex network. We have noticed that in general the prior particle phase generated by SliceGAN is separated in the XY but the phase is interconnected in the YZ and XZ planes, which corresponds to the general features of the experimental EBSD maps.

- *Was any sort of shape analysis (2D invariant/Hu moment) done on the associated phase outputs for sliceGAN as compared the original phase maps, both with and without the thresholding and erode/dilation?*

In addition to Figure S5, where the area fraction of 2D slices of the SEVM are compared with the values from the EBSD map, we performed a comparison of the 2-point correlation (which has implications of shape) of the prior particle phase before and after performing the thresholding and erode/dilation. The new Figure S6 shows the 2-point correlation of the EBSD (Raw data) and EBSD (Cleaned data) for the XY and XZ planes. From these results we can confirm that the morphology of the prior particle domain was not altered significantly as the general shape of the 2-point correlation function remains almost unchanged. This can also be seen in Figures 9(b and c) where the prior particle shape didn't change. The following text has been added to the Supplementary section II.

“A comparison of the 2-point correlation S_2 of the prior particle phase before and after performing the thresholding and erode/dilation is depicted in Figures S6 for the EBSD (Raw data) and EBSD (Cleaned data) in the XY and XZ planes. From these results, it can be inferred that the morphology of the prior particle domain is not altered significantly as the general shape of the S_2 function remains almost unchanged. This can also be seen in Figures 9(b, c), where the prior particle shape did not change.”

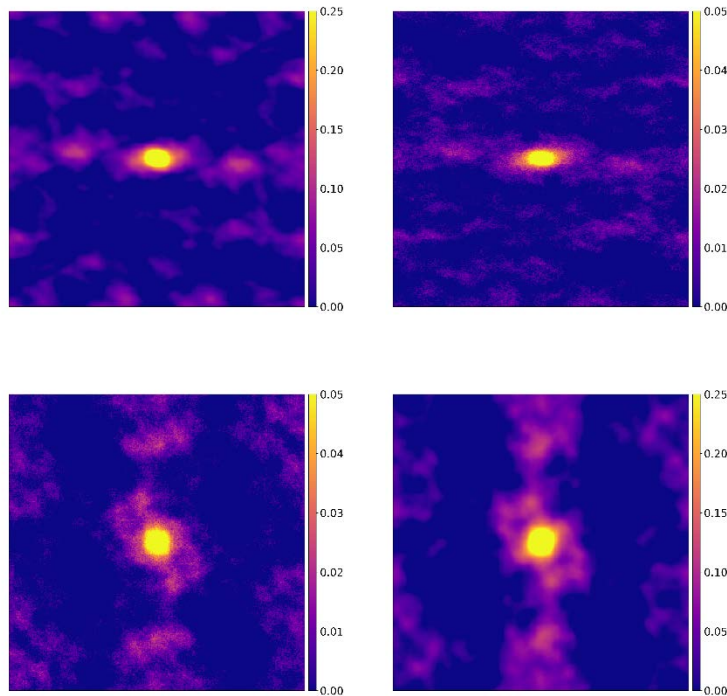


Figure S6. Comparison of the 2-point correlation function S_2 maps of the EBSD data just after the GAKAM threshold to isolate coarse grains as shown in Figure 2 (Raw), and just after the grain cluster size threshold as shown in Figure 9(c,d) (Cleaned).

Reviewer comment 5: One of the core motivations of this work was the use of GANs to generate microstructures with complexity that exceeds the capabilities of current synthetic pipelines, and one of the key demonstrators of this was CSF aluminum containing a widely varied grain size distribution with bimodal character. From a machine learning perspective, systems with this type of strong modal dependence are challenging for generative models that sample from a single latent space because that sampling is generally uniform (or at least unimodal), meaning that some aspects of latent sampling are overemphasized with respect to others. Karras et al. discusses this idea at length in their presentation of style-based generators (<https://arxiv.org/abs/1812.04948>).

When Kench and Cooper first published sliceGAN, they did not implement any sort of adjustments to account for multimodal/entangled features, so Chung and Ye released a Matters Arising Reusability Report (<https://doi.org/10.1038/s42256-021-00400-4>) outlining this issue and how it impacted the performance of the architecture. With this, Chung and Ye also released a publicly available modification of sliceGAN that includes the adaptive instance normalization (AdaIN) steps that mitigate this effect in place of the basic normalization protocols Kench and Cooper used. Further adding to relevance, they demonstrate this on systems of variable grain/phase feature size. Since the focus of this work is complex structures with strong modal dependence, was this AdaIN version of sliceGAN used, or just the original?

If this modified architecture was not used, it presents a major opportunity for performance improvement, as it not only helps disentangle correlations between phase feature size and distribution, but it also allows for a degree of tunability in that associated disentanglement, which would likely contribute to improved performance. Depending on the exact arrangement of the full pipeline (whose details should probably be

detailed a bit more explicitly around Figure 8), implementing a changed architecture may prove to be a lot of work, but at the very least, the presence of this possible source of improvement should be acknowledged in the section on avenues for improvement.

Authors' Response:

The authors thank the comment of the reviewer as it has an important potential in the future of generative models for microstructure generation. Disentanglement is highly important when the model must be trained using a richer and bigger data set, in the context of microstructure generation this represents different microstructures with different characteristics inherent from manufacturing or thermomechanical processes. In the present work the original SliceGAN code was used, the main reason for this is that in the present work only uses one data point (one microstructure), it means that the latent space is enough to represent this data set and disentanglement is not necessary. However, as stated by the reviewer, disentanglement is an important feature and will improve the performance of the model as shown in the reusability report. While we appreciate the reviewer's reference to the latest updates in the SliceGAN techniques, we would like to emphasize that our objective in this paper is not necessarily to use the best ML methods proposed in this field, but to use the general ML strategy to create a computational platform to meet the stated objective.

We understand that the field of machine learning is rapidly evolving and while we would like to keep up with the advances, we may sometimes miss the latest advances that are taking place in the field. The advances by Karras, Laine and Aila, and Chung and Ye are of high importance to the literature and have now been added to the references. The following text has been added to the "Avenues for improvement ." part of the Discussions section .

"For the case of richer datasets or microstructures with different characteristics, disentanglement could potentially be an issue, and style-based generators can be used to improve the variation of generated images and disentangle the latent space [71,72]. These techniques are currently under investigation."

Minor Points:

Reviewer comment 6: The title is a bit of a challenger to understand. Maybe something like "Generative Approaches for Statistically Equivalent Virtual Microstructures of Complex Polycrystalline Materials"? This change is entirely at the author's discretion, but I had to read the title several times to parse it.

Authors' Response:

To improve the clarity of the title of the paper has been changed to the following:

Statistically Equivalent Virtual Microstructures for Modeling of Complex Polycrystalline Alloys Using a Generative Adversarial Network (GAN)-Enabled Computational Platform.

Reviewer comment 7: When stating the training details for the network (time, equipment, etc.), it would also be beneficial to include the approximate size (MB/GB) of the training set used, since learning convergence is dependent on passage of the entire set through the model.

Authors' Response:

The GPUs used have a memory of 40GB. The following text has been added to the Methods section ii:

"The training runs of the binary CSF AA7050 and additively manufactured Ti64 microstructures were performed on 1 GPU Nvidia A100 with 40GB of memory for a period of 2 hours."

Reviewer comment 8: Very minor, but some of the figure captions (1 and 6) look like they might bleed into the footer space. I appreciated the size of the figures for being able to visualize things, but don't lose the captions in the final print.

Authors' Response:

Thanks for catching this. We have fixed it now.

Reviewer comment 9: In the avenues for improvement section, the authors mention they are looking at several ways of improving the approach to deal with data bottlenecks. In their reusability report, Chung and Ye mention that the approach would benefit from a progressive learning approach, which also seems highly salient here since sliceGAN is being used to produce maps across several different length scales. Additionally, the authors mention the possibility of augmentation, which would be interesting to explore because to my understanding, Kench and Cooper's original model uses the rotational sensitivity of convolution to preserve the relative axis dependencies in sliceGAN. Augmentation would certainly provide more training information, but it would also disrupt these relationships. Creating an option for directionally-dependent data augmentation would be interesting as a tool for encouraging isotropy along particular microstructural directions when relevant.

Authors' Response:

Thank you very much for these insights. We are currently working on implementing data augmentation in the code and comparing results with those from the baseline SliceGAN code. The main objective is to generate polycrystalline microstructure with a high geometrical anisotropy created by additive manufacturing.

REVIEWER 4:

The manuscript deals with the generation of multi-modal morphology of virtual microstructure based on generative adversarial network (GAN)-based methods. The authors integrated SliceGAN and Dream3D builders to generate virtual microstructures of cold spray formed (CSF) Al 7050 alloys and additively manufactured (AM) Ti64 alloys and compared generated virtual microstructures with ones from other conventional techniques (SliceGAN or Dream3D only). They also developed a multiscale model to examine mechanical properties of generated virtual microstructures.

Although the topic is interesting and the novelty of the methodology is recognized, I am afraid that there is much ambiguity in the results and discussion showing the superiority of this study.

Reviewer comment 1: For example, in page 21, the authors state "The SliceGAN-based SEVM in Figure 6(b) exhibits a poor representation of the grain size and aspect ratio distributions, but a good representation of the misorientation angle distribution. On the other hand, the Dream3D and SliceGAN-Dream3D platform-based SEVMs in Figures 6(a,c) match the three distributions closely, since grain-packing is achieved through Dream3D for both cases." Each method has its advantages and disadvantages depending on the physical properties of interest, and therefore the method developed in this study is not necessarily superior. In particular, CSF is characterized by a collision-distorted structure, and AM is characterized by a long and thin Widmanstätten structure. Therefore, it is important to reproduce the aspect ratio correctly. However, it is crucial to note that the method employed in this study may not adequately represent the aspect ratio. In fact, there is a concern that the aspect ratio in Figure 3(e) is misaligned between EBSD and SEVM.

Authors' Response:

Coupling Dream3D and SliceGAN enables us to bring the main advantages of the two individual methodologies. In Figure 3(e) the EBSD and SEVM are misaligned because the 2D data of the EBSD and the 3D data of the SEVM are compared, even though, the data is relatively close. This kind of misfit is a common occurrence with stereology techniques when microstructural characteristics are mapped from 2D to 3D. A better plot to understand the accuracy of Dream3D is shown in Figure 6(e), where 2D slices from

the 3D SEVM are compared to the 2D EBSD data. The following text has been added to the sub-section “Integrating SliceGAN with the synthetic builder DREAM3D for creating SEVMs” in the Results section II:

“Note that this kind of misfit is a common occurrence with stereology techniques when microstructural characteristics are mapped from 2D to 3D. A better plot to understand the accuracy is shown in Figure 6(e), where 2D slices of the 3D SEVM, shown in Figure 3(a), are compared with the 2D EBSD data.”

Reviewer comment 2: Moreover, the primary selling point of this study is the generation of multi-modal morphology in the microstructure. However, if I understand correctly, it has generated the coarse grain (CG) and the regions of ultra-fine grains (UFGs), but not the details of the UFG microstructure itself. It seems to be only a phenomenological discussion of the mechanical properties, which are matched by parameters in the multi-scale model. I am concerned that the claims in this paper may mislead general readers into thinking that the methodology proposed in this study generates a multi-modal morphology of virtual microstructure that can reproduce the mechanical properties directly. Therefore, I do not recommend this manuscript for publication in Nature Communications in its current form.

Authors’ Response:

We are sorry for the misunderstanding. Indeed, we have generated the UFG microstructures, as shown in Figure S8 in the Supplementary Material section III to calibrate the UCM. The UFG characteristics are obtained from the high resolution EBSD map shown in Figure 1(c), which has statistically sufficient UFGs needed to sample the grain size, aspect ratio, misorientation angle distributions and texture. A micromechanical model of the UFGs is created from these statistics, as shown in Figure S8. The parameter calibration process for the CP and UCM models requires this detailed UFG SERVE construction to derive the UCMs.

We have tried to clarify the main goals and processes in this revised manuscript. The goal of this paper is two-fold. First, it uses a combination of GANs and conventional synthetic microstructure builders to create statistically equivalent virtual microstructures (SEVMs) for complex heterogeneous (including multimodal) materials that are difficult to model otherwise. The second objective is to use this virtual microstructure to develop a concurrent multiscale model for simulating these microstructures that inherently have multiple scales in their morphology and crystallography. We have now added this to the introduction for clarity.

“The goal of this paper is two-fold. First, it uses a combination of GANs and conventional synthetic microstructure builders to create statistically equivalent virtual microstructures (SEVMs) for complex heterogeneous (including multimodal) materials that are difficult to model otherwise. The second objective is to use this virtual microstructure to develop a concurrent multiscale model for simulating these microstructures that inherently have multiple scales in their morphology and crystallography.”

We hope that the revised manuscript will satisfy the reviewer with respect to these concerns.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have constructively addressed all of my comments and suggestions. I appreciate the authors' clarification concerning possible addition of hard-constraints to obtain better matching of certain targeted statistics. The paper can be accepted for publication in the current form.

Reviewer #1 (Remarks on code availability):

The codes are accessible and seems to be compile correctly.

Reviewer #2 (Remarks to the Author):

Thank you for engaging with the question I posed. My responses below follow the same numbering scheme as my original review.

1. Although I agree that representativity is a tricky topic, I don't think it's enough to simply say "we have assumed a statistically sufficient representative EBSD map to contain at least 500 grains". What is this based on and how is it validated? 500 crystals is just a 22x22 grain image and I don't think it is at clear from your images that this would be representative.
2. The additional information about how difficult the method is is interesting, but I don't feel that it actually answers my question. Perhaps another way to ask is: considering all of the steps and compute required, how different is the stress/strain behaviour of your structures from that of random/arbitrary grain structures?
3. Thank you, I think they is now much easier to compare.
4. I think this is much clearer, thank you.
5. The new figure seems to suggest highly unstable training. Can you explain the shape of these training graphs? How did you decided when to terminate training?
6. This is great to see, thank you.
7. I think the paragraph you've added is nice, but it doesn't quite answer the question I was asking: Considering your approach only improves on some metrics, but not on others, where will this be acceptable? For example, will it be good for mechanical simulations, but not for percolation? Or, do you think it's more appropriate for alloys than porous composites? When should a read apply your method and when should they avoid it?

Overall, although I appreciate the authors efforts to make the paper clearer, I still don't believe that this would be of interest to a wide audience as neither robustness nor generality have been established convincingly.

Reviewer #2 (Remarks on code availability):

Didn't review the code, but the repos looks sensibly structured and well documented.

Reviewer #3 (Remarks to the Author):

This work presents a platform based approach for GAN-based generation of microstructures with directionally-dependent and spatially correlated microstructures. The approach presented is well-documented and accessible, and provides clear avenues both for recreation of the work, as well as model training for additional microstructural configurations beyond those discussed. The benefits of the method are directly demonstrated on a microstructure with anisotropic and spatially-dependent morphologies, and the strengths and weaknesses of the approach are adequately discussed from a materials perspective. From a machine learning perspective, there are still avenues of improvement available, but the information included in the revision is sufficient to identify what these possible avenues are and how they could be pursued with scientific rigor. Based on this, the reviewer feels that this work meets the standards of scientific contribution expected by Nature Communications.

Reviewer #3 (Remarks on code availability):

Code on both repositories is complete (essential file structures present) and provides README instructions for implementation. Implementation was not directly tested by the reviewer, but instructions are clear and email contacts are provided for questions and extenuating circumstances.

Reviewer #4 (Remarks to the Author):

I appreciate the authors' revisions to the manuscript in accordance with the referees' comments. The authors adeptly addressed all referee comments, resulting in significant improvements to the manuscript. The study's merit, demonstrating the efficacy of combining GANs and conventional synthetic microstructure builders to create statistically equivalent virtual microstructures for complex heterogeneous materials, is now more apparent. This conclusion is supported by concrete results and discussion.

Therefore, I now recommend the manuscript for publication in Nature Communications

Authors Response to Reviewers Comments and Suggestions

The authors thank the editor and reviewers for positive comments on our manuscript. With respect to the issues raised by the second reviewer, we have tried to address them to our best ability. A revised manuscript with highlighted changes in red font is attached.

REVIEWER 2:

General Comments:

Thank you for engaging with the question I posed. My responses below follow the same numbering scheme as my original review.

Reviewer Comment 1: *Although I agree that representativity is a tricky topic, I don't think it's enough to simply say "we have assumed a statistically sufficient representative EBSD map to contain at least 500 grains". What is this based on and how is it validated? 500 crystals is just a 22x22 grain image and I don't think it is at clear from your images that this would be representative.*

Authors' Response:

The number of grains we have assumed is an estimate, that was established through some parametric studies. An example of the results of this study is shown below in figure 1. In this figure, the grain size probability density is plotted with reconstructed histograms using samples containing 50, 100, 200, 300, 400, and 500 grains and compared with the pdf of the original larger EBSD (approximately 6000 grains). It is concluded from this figure that 500 grains are optimally sufficient to reconstruct the original pdf with adequate accuracy.

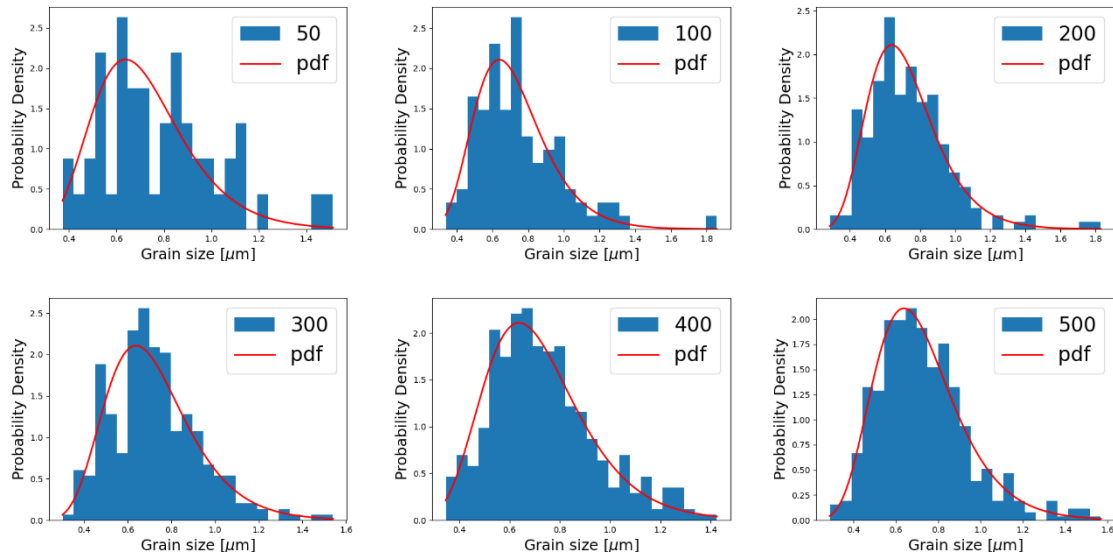


Figure 1: Plots of the grain size probability density with reconstructed histograms using samples containing. 50, 100, 200, 300, 400, and 500 grains, compared with the pdf of the original EBSD containing approximately 6000 grains.

The minimum number of grains is also assessed using other characteristics, viz. misorientation angle or aspect ratio as shown in figure 2. The Kolmogorov-Smirnov (KS) test is used to analyze the goodness of fit between the distributions from the reconstructed SEVM and the EBSD. The KS test quantifies the difference in the cumulative distribution functions (CDF). Figure 2 shows the KS values as a function of the number of grains for aspect ratio, grain size and misorientation angle. This figure also shows that the reconstructed microstructure is more representative and accurate with 500 grains.

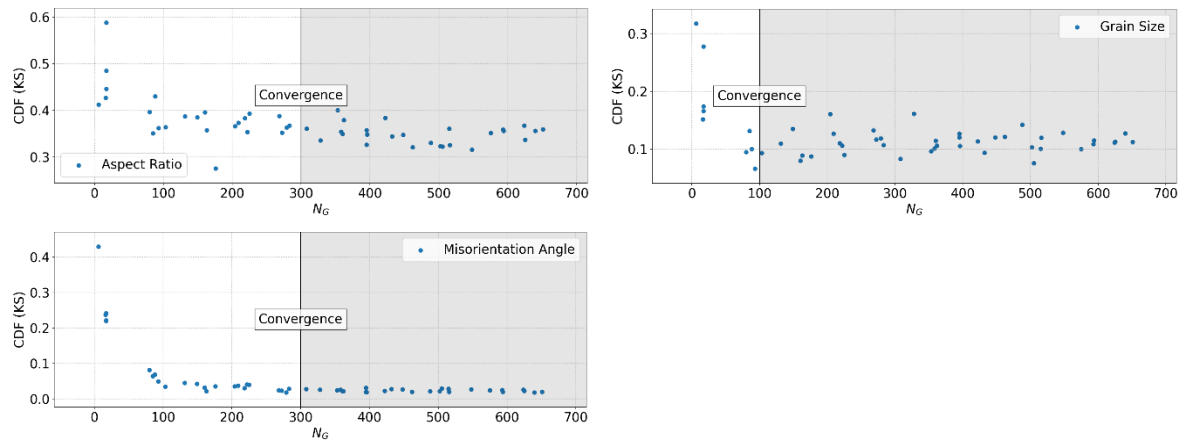


Figure 2: KS value for (a) aspect Ratio, (b) grain size and (c) misorientation angle, as a function of the number of grains.

Reviewer Comment 2: The additional information about how difficult the method is interesting, but I don't feel that it actually answers my question. Perhaps another way to ask is: considering all of the steps and compute required, how different is the stress/strain behavior of your structures from that of random/arbitrary grain structures?

Authors' Response:

We are sorry that this aspect was not clear from our previous discussion. The mechanical behavior can vary significantly with changes in microstructural descriptors. In terms of the overall stress-strain behavior, the grain size distribution has a significant effect on the yield stress, while misorientation angles, aspect ratios and grain size distribution can significantly affect the post-yield plastic behavior. All of these characteristics will also have a significant effect on the local stresses and strains affecting extreme properties like ductility and fatigue. The effects can be very pronounced with anisotropic local texture. A number of papers in the literature have explicitly shown these effects. Prior work in the corresponding author's group, e.g. in [1,2], has also explicitly shown the effect of detailed microstructural characteristics on the overall material behavior (stress-strain relations).

1. S. Kotha, D. Ozturk and S. Ghosh, "Parametrically homogenized constitutive models (PHCMs) from micromechanical crystal plasticity FE simulations, Part I: Sensitivity analysis and parameter identification for Titanium alloys", *International Journal of Plasticity*, Vol. 120, pp. 296-319, 2019. <https://doi.org/10.1016/j.ijplas.2019.05.007>

2. S. Kotha, D. Ozturk and **S. Ghosh**, “Parametrically homogenized constitutive models (PHCMs) from micromechanical crystal plasticity FE simulations, Part II: Thermo-elasto-plastic model with experimental validation for Titanium alloys”, *International Journal of Plasticity*, Vol. 120, pp. 320-339, 2019. <https://doi.org/10.1016/j.ijplas.2019.05.008>

Reviewer Comment 5: The new figure seems to suggest highly unstable training. Can you explain the shape of these training graphs? How did you decided when to terminate training?

Authors’ Response:

The instabilities of the loss function are not due to unstable training, but they are due to the difference of the sub-images taken to train with every batch shown to the Discriminator. In general, GANs keep learning as long as the loss in the discriminator approaches a constant value. At some point in the training, when the discriminator already knows all the features of the real data, the loss function for the real data surges, and the model stops learning. The number of iterations, while based on the original work of Kench and Cooper [ref. 44], were slightly modified for this dataset. The most common issue with GAN training is mode collapse. In this case the discriminator loss knows all the features from the real data, and the feedback to the generator produces images that do not process the real data quality anymore. An example of mode collapse and bad training is shown in the following figure 3 , where the discriminator loss on the real data continues to increase.

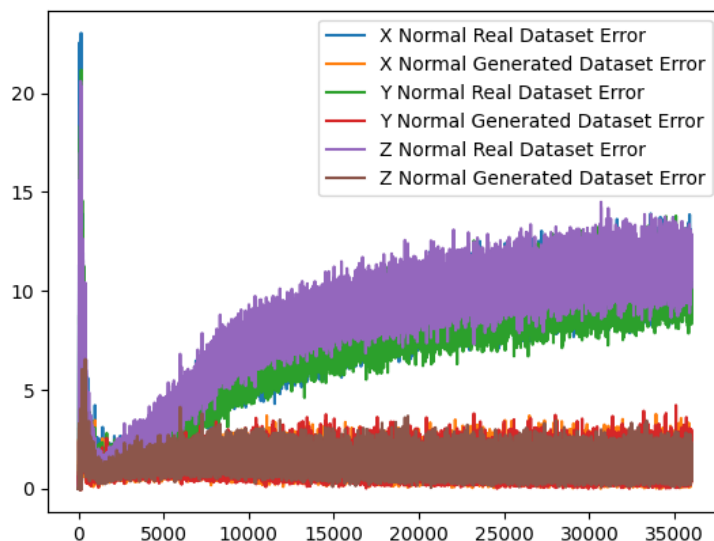


Figure 3: Plot of the discriminator loss evolution depicting mode collapse.

We have added the following sentences to the Supplementary Section VI

“Supplementary Figure 11 shows the discriminator loss evolution during training, calculated using Equation 4. The loss evolution approaches a constant value. The number of batches shown to the discriminator is selected such that the evolution of the loss is stable. The number of batches is also chosen to avoid mode collapse and an abrupt increase in the loss function. The evolution of the loss function can be improved by using more data or other changes in the model as explained in the Discussion section.”

Reviewer comment 7: I think the paragraph you've added is nice, but it doesn't quite answer the question I was asking: Considering your approach only improves on some metrics, but not on others, where will this be acceptable? For example, will it be good for mechanical simulations, but not for percolation? Or, do you think it's more appropriate for alloys than porous composites? When should a read apply your method and when should they avoid it?

Authors' Response:

In the present work we focused on properties that are relevant to the mechanical behavior of polycrystalline alloys. We believe that most properties considered were improved for the problems studied (It is not clear which metrics are being referred to).

However, we agree that there could be other physical properties and metrics beyond those considered in this study that will need further consideration. Such things are beyond the scope of this paper. As mentioned by the reviewer, percolation properties may need additional metrics that have not been considered in this work. The study of pores is possible with this method, where the pore/void can be added as an additional phase. However, we have not studied that here and would refrain from making claims in this paper.

We have added the following sentence to the Discussion Section

“Other physical properties and metrics like percolation are beyond the scope of this paper and may need additional considerations.”

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

1. I don't see how figure 1 supports your claims that "500 grains are optimally sufficient to reconstruct the original pdf with adequate accuracy."? Accuracy of what? What is it you are measuring, what was the threshold of "adequate accuracy", and how was this threshold chosen? I'm also struggling to interpret figure 2. Are these graphs suggesting you measured the KS of the CDF for samples as small as $N_G=10$ grains? Is each point the average of many samples of this size or a single sample? Why do some values of N_G have multiple repeated points? A "convergence" value was added to each graph, but how was this threshold chosen? It seems to be quite arbitrary. Presumably there is a metric that you care about and you can state that you were willing to be within a tolerance of this metric with N_G grains. You can then extract this metric from many samples and state a confidence interval (e.g. "this data suggest that for a sample with 500 grains we would be 95% confident of metric X being within 1% of the true value").
2. Can you show a comparison between the stress/strain graphs from the grain structure resulting from the method you described in this paper and a few random grain structures? I think this would allow the reader to judge the importance/utility of using your method.
5. Thank you for adding a few words on this topic.
7. Table 1 seems to suggest that your method is good for some metrics but not for others, or perhaps I am misunderstanding it (apologies if so!)? If this is the case, then can you explain when to use which method?

Authors Response to Reviewers Comments and Suggestions

The authors thank the reviewer for comments on this manuscript. In this rebuttal, we have tried to address the issues raised by Reviewer 2. The manuscript was revised as necessary with highlighted changes in red font attached.

REVIEWER 2:

General Comments:

The authors have constructively addressed all of my comments and suggestions. I appreciate the authors' clarification concerning possible addition of hard-constraints to obtain better matching of certain targeted statistics. The paper can be accepted for publication in the current form.

Authors' Response:

Thank you for the comments on our paper. We have tried to address the issues raised next.

Comment #1: I don't see how figure 1 supports your claims that "500 grains are optimally sufficient to reconstruct the original pdf with adequate accuracy."? Accuracy of what? What is it you are measuring, what was the threshold of "adequate accuracy", and how was this threshold chosen?

Authors' Response:

Thank you for bringing up this point. The 500-grain decision was made from a combination of results in Figs. 1 and 2 (shown below). In Figure 1 the distributions were compared with respect to the difference in the area under the histograms with additional grains, i.e. ($L1 = \Sigma(A_{\text{hist1}} - A_{\text{hist2}})$). For example, the difference in the histogram area between the 50 and 100 grain is calculated to be approximately 18%, whereas the same difference between the 400 and 500 grain models is approximately 4%. Consequently, it was inferred that the difference between the 400 and 500 grains is sufficiently small to accept the 500-grain model as the reference.

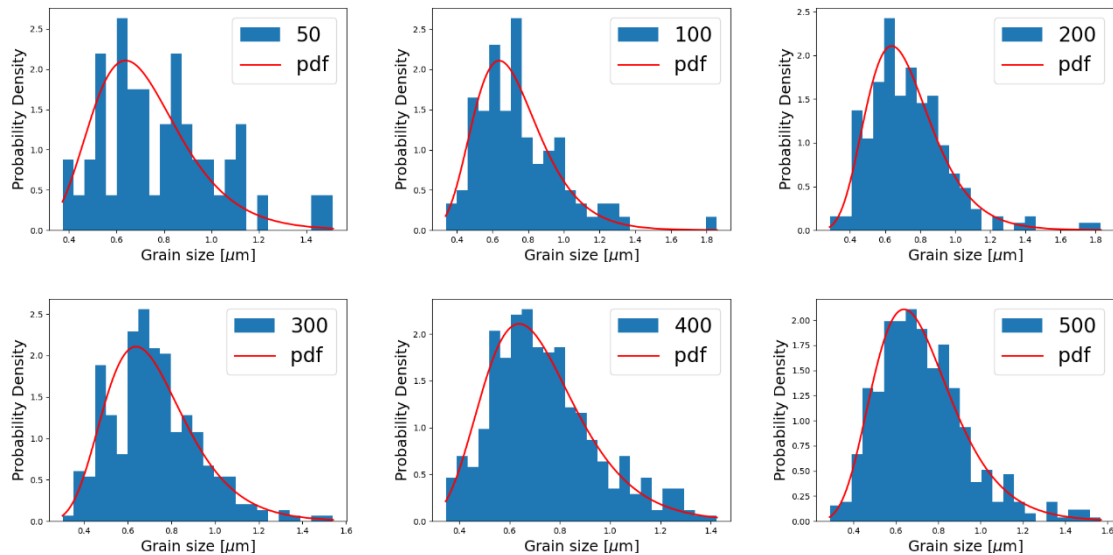


Figure 1: Plots of the grain size probability density with reconstructed histograms using samples containing. 50, 100, 200, 300, 400, and 500 grains, compared with the pdf of the original EBSD containing approximately 6000 grains.

Are these graphs suggesting you measured the KS of the CDF for samples as small as $N_G=10$ grains? Yes, the grain packing algorithms do not limit the minimum number of grains. You can generate microstructures with one grain to a limit imposed by the memory of the machines used.

Is each point the average of many samples of this size or a single sample? No, each point corresponds to a unique microstructure.

Why do some values of N_G have multiple repeated points? The multiple repeated points are because to generate this data, we fixed the dimension of the microstructure domain and packed the grains. For a given dimension, this process was repeated 4 to 5 times to check the variance between the microstructures. The number of grains can vary and, in some cases it can be close. This explains the repeated points which are in fact different microstructures.

A “convergence” value was added to each graph, but how was this threshold chosen? It seems to be quite arbitrary. Presumably there is a metric that you care about and you can state that you were willing to be within a tolerance of this metric with N_G grains. You can then extract this metric from many samples and state a confidence interval (e.g. "this data suggest that for a sample with 500 grains we would be 95% confident of metric X being within 1% of the true value").

The convergence value is selected to yield the smallest possible number of grains for which the KS value is stable (with a small variance) for characteristics like misorientation angle or grain size. We chose a sufficiently small threshold between 5% and 10% for the misorientation angle and grain size distribution in the KS test. It can be seen that for the misorientation angle, the KS value stabilizes faster than that for the grain size. For this data, the converged KS value is 0.03 and 0.1 for the misorientation angle and grain size respectively, and thus the convergence value.

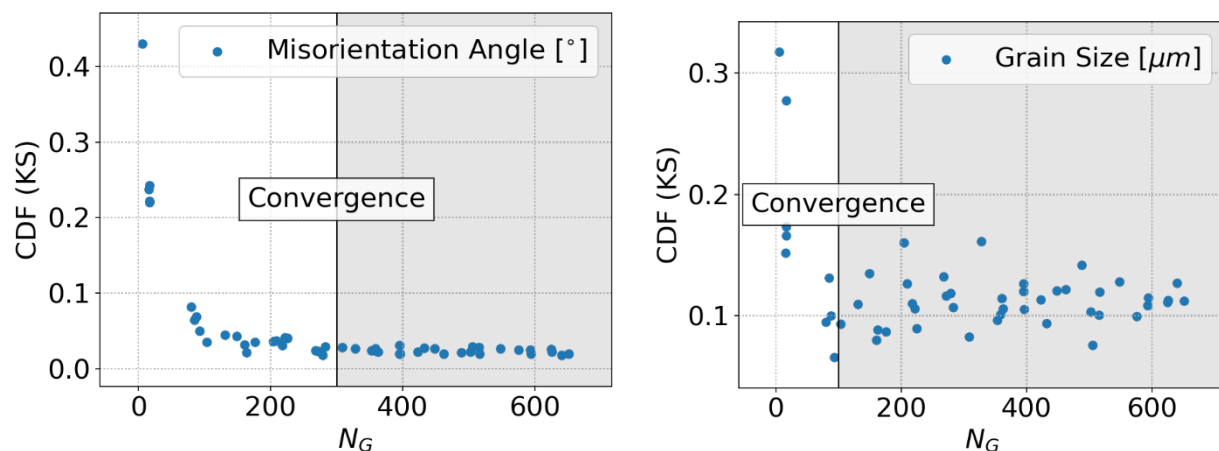


Figure 2: KS test on misorientation and grain size distribution to establish convergence.

Comment #2: Can you show a comparison between the stress/strain graphs from the grain structure resulting from the method you described in this paper and a few random grain structures? I think this would allow the reader to judge the importance/utility of using your method.

Authors' Response:

To show the need for the mask geometry and distributions in the material simulation, as opposed to a randomly created microstructure, results of additional tests are performed and shown in Figure 4. The random microstructure 1 has the same UFG volume fraction of 0.6 as the SERVE, while the random microstructure 2 has a UFG volume fraction of 0.75. Random aspect ratio and misorientation distributions, viz. the Mackenzie distribution, are assigned to the volumes as shown in Figure 3(a,b). For the random volumes 1 and 2, the mask is generated by selecting grains randomly and assigning CP and UFG properties. The overall stress-strain response of the three different simulated microstructures (Simulation: (SERVE), (Random Volume 1) and (Random Volume 2)) are compared with the experimental data in Figure 4. It is observed that the randomness of the mask can significantly change the response of the material and hence a more rigorous method such as that in this paper is desirable.

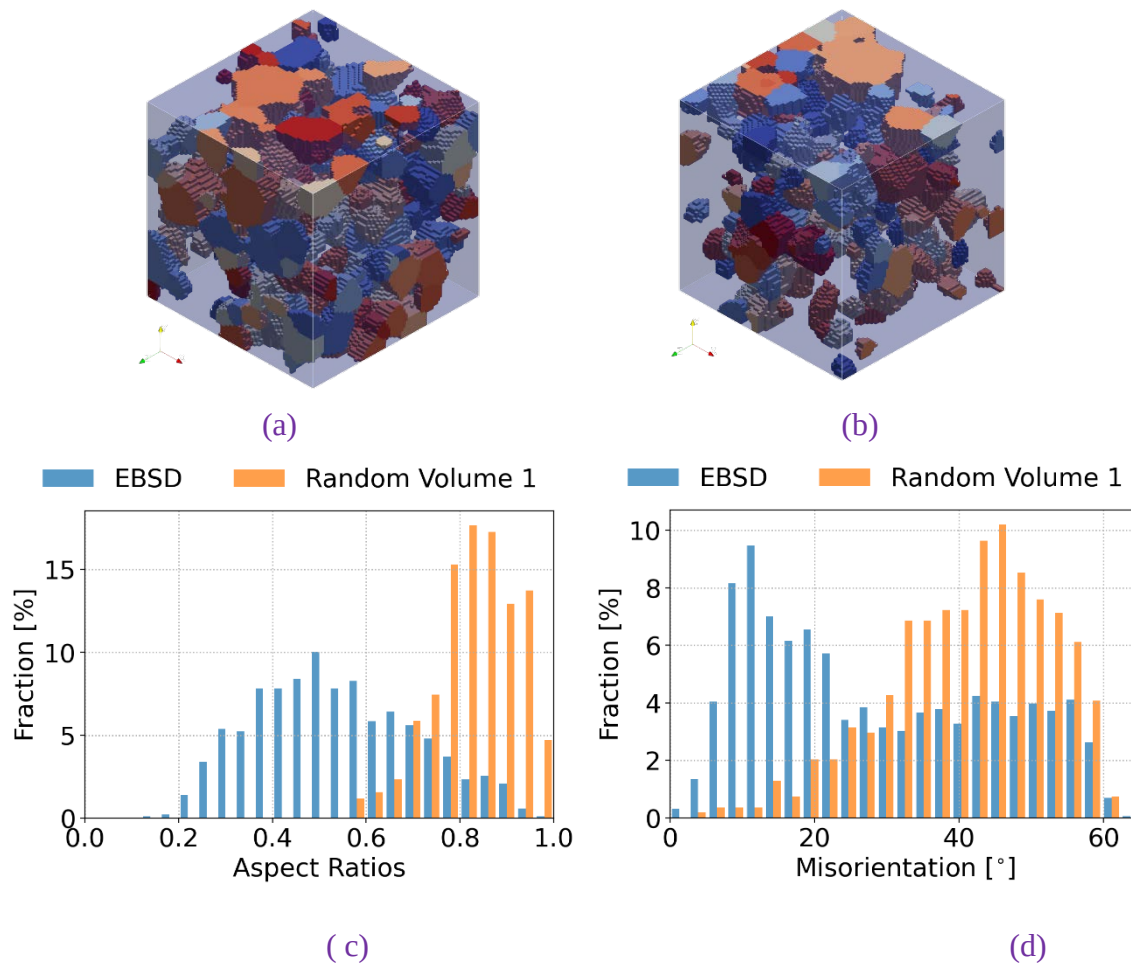


Figure 3. Microstructure features generated randomly: (a) Random volume 1 with a UFG volume fraction of 0.6, (b) Random volume 2 with a UFG volume fraction of 0.75, (c) Aspect ratio distribution, (d) Misorientation angle distribution.

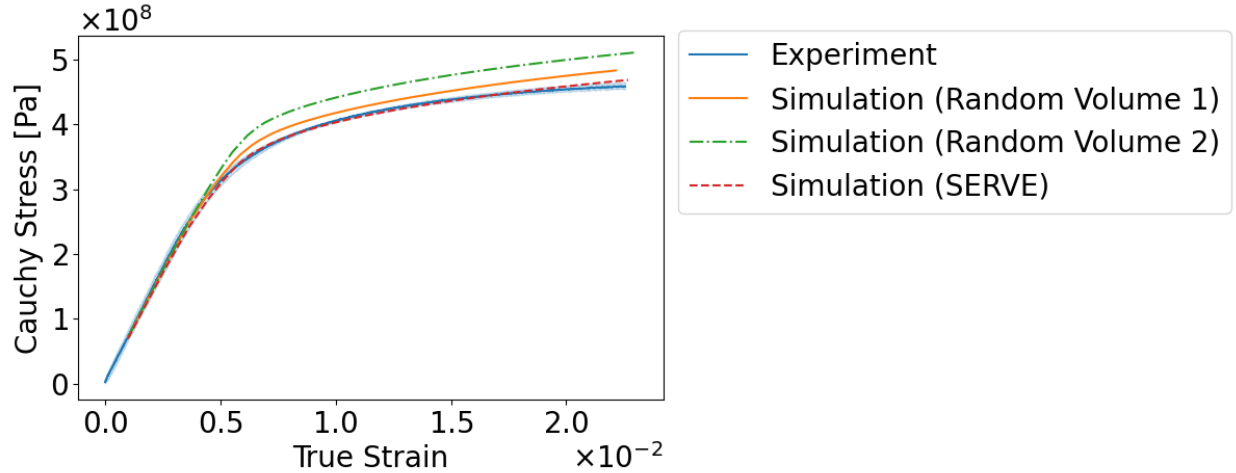


Figure 4. Stress-Strain response of the different SERVEs and the experimental data. Experiment (mean and range), Simulation (Random volume 1), Simulation (Random volume 2) and Simulation (SERVE).

We have added the following text to the Supplementary Material Section IV

Comparing the macroscopic response of the created SERVE with other randomly generated microstructures

To show the need for the mask geometry and distributions in the material simulation, as opposed to a randomly created microstructure, results of additional tests are performed and shown in Figure \ref{fig:ervesrandom}(b). The random microstructure 1 has the same UFG volume fraction of 0.6 as the SERVE, while the random microstructure 2 has a UFG volume fraction of 0.75. Random aspect ratio and misorientation distributions, viz. the Mackenzie distribution, are assigned to the volumes as shown in Figure \ref{fig:ervesrandom}(a). For the random volumes 1 and 2, the mask is generated by selecting grains randomly and assigning CP and UFG properties. The overall stress-strain response of the three different simulated microstructures (Simulation: (SERVE), (Random Volume 1) and (Random Volume 2)) are compared with the experimental data in Figure \ref{fig:ervesrandom}(b). It is observed that the randomness of the mask can significantly change the response of the material and hence a more rigorous method such as that in this paper is desirable.

\begin{figure}[H]

\centering

\subfigure[]{\includegraphics[height=0.3\textwidth]{sfig10-1.png}}

```

\subfigure[]{}

\includegraphics[width=1.0\textwidth]{sfig10-2.pdf}}

\caption{ (a) Microstructures Random volume 1 with a UFG volume fraction of 0.6 and Random volume 2 with a UFG volume fraction of 0.75 with randomly assigned aspect ratio and misorientation angle distribution, and (b) comparison of the stress-strain plots for the different microstructural volumes with experimental results.}

\label{fig:servesrandom}

\end{figure}

```

Comment #3: Table 1 seems to suggest that your method is good for some metrics but not for others, or perhaps I am misunderstanding it (apologies if so!)? If this is the case, then can you explain when to use which method?

Authors' Response:

Thank you for the clarification on the specific metrics. We have reviewed our presentation in the paper and found that some metrics were not presented appropriately. Correspondingly we have changed a part of table 1 and only include the length ratio. We have chosen to represent the second part using three different plots of the KS that represent an ensemble effect of 200 slices as opposed to only 5 slices before. As you can see the results reflect the uniformly good performance of the SliceGAN+ Dream3D model for all metrics considered. So based on these results we prefer this method. As mentioned in the paper, the Dream3D model can be used to generate microstructures with unimodal distributions. However, when the microstructures are more difficult to generate and present additional complexities the present method, SliceGAN+Dream3D, offers an additional feature to generate more complex grains or phases.

We have added the following figure and text to the manuscript.

Figures 6(d,e), Table 1 and Figure 7 compare the two-point correlation function and the distributions of major microstructural descriptors for the EBSD maps with two orthogonal sections taken from SEVMs generated by Dream3D, SliceGAN and the SliceGAN+Dream3D platform.

Table~\ref{tab:sevm_compare_s2} shows the ratio of the maximum to minimum lengths of the ellipsoid representing the peak aspect ratio, with the SliceGAN-generated SEVM being the closest to the characteristics of the EBSD map.

Figure~\ref{fig:sevm_compare_ks} compares the KS value for each distribution in both planes of the EBSD results with orthogonal slices of synthetic microstructures created by Dream3D, SliceGAN, and SliceGAN+Dream3D for descriptors viz. (a) the equivalent grain size, (b) aspect ratio, and (c) misorientation distributions.

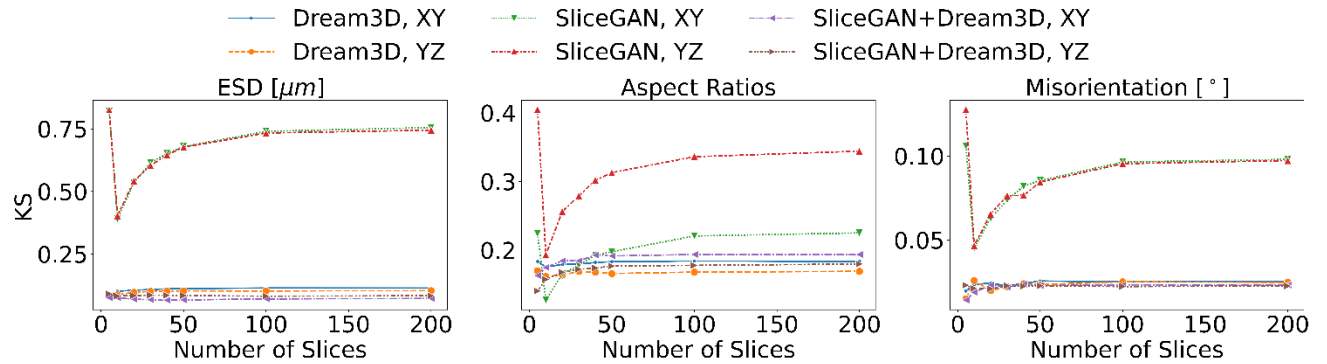


Figure 4. Comparison of microstructural features of the EBSD with synthetic microstructures created by Dream3D, SliceGAN, and SliceGAN+Dream3D, in terms of the KS values for descriptors viz. (a) the equivalent grain size, (b) aspect ratio, and (c) misorientation distributions.

REFERENCES

- [1] S. Kotha, D. Ozturk, and S. Ghosh. Parametrically homogenized constitutive models (PHCMs) from micromechanical crystal plasticity FE simulations, part I: Sensitivity analysis and parameter identification for Titanium alloys. *Int J Plast* 120, (2019).
- [2] J. Shen, S. Kotha, R. Noraas, V. Venkatesh and S. Ghosh, Developing parametrically upscaled constitutive and crack nucleation models for the α/β Ti64 alloy. *Int J Plast* 151, (2022).
- [3] M. Pinz, G. Weber, J. C. Stinville, T. M. Pollock, and S. Ghosh, Data-driven Bayesian model-based prediction of fatigue crack nucleation in Ni-based superalloys. *npj Comput Mater* 8, 39 (2022).

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

Once again, I am pleased by the constructive engagement from the authors. I still feel like the representativity analysis is somewhat arbitrary, but I appreciate that it's a poorly defined concept and your results looks sensible. I am also still unsure of whether many people will be able to make use of this intricate workflow, but I hope to be proven wrong! I am happy for the article to be published in its present form.