

Generative design of high-fidelity microstructures using physics-aware machine learning

Corresponding Author: Professor Ruihao Yuan

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This paper proposes a physics-informed generative optimization strategy for achieving microstructure inverse design guided by competing performance objectives. The method integrates a variational autoencoder with a denoising diffusion probabilistic model, incorporating pixel similarity loss, edge detection loss, and structural similarity loss into the loss function to enhance the modeling capability of physical features such as grain boundaries and grain orientations. With only 25 sets of experimental data, it generates microstructures with bimodal grain size distributions and a synergistic improvement in strength and ductility, which is validated by crystal plasticity finite element simulations. However, several aspects still require refinement:

1. Although the introduction highlights the issues of insufficient physical fidelity and weak out-of-distribution generalization in traditional generative models, it does not specifically analyze the unique challenges in the microstructure design of nickel-based high-temperature alloys, mentioning them only briefly as typical aerospace materials. We suggest adding a discussion of the application scenarios and key material properties of nickel-based high-temperature alloys to improve the targetedness and practical relevance of the research.
2. Some of the figure legends and axis labels in the manuscript are missing units or use font sizes that are too small. We recommend unifying the image formats and improving clarity to ensure better visual consistency.
3. The article only uses bold formatting for headings at all levels, leading to an overly compact layout. We recommend clearly distinguishing headings from the main text and adjusting line breaks to enhance readability.
4. The study applied various data augmentation techniques, including cropping, rotation, and blurring, to the original images. Although these operations significantly expanded the dataset, some may introduce non-physical features. We recommend providing further clarification on whether such augmentations preserve key physical information (e.g., grain boundaries and orientations) and whether they influence subsequent physical modeling.
5. The overall language of the article is fluent, but some overly long sentences appear in the abstract and conclusion. We recommend splitting these sentences appropriately to clarify the logical thread and enhance readability.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors propose a generative optimization framework integrating VAEs, DDPMs, and NSGA-II for inverse microstructure design, incorporating a physics-aware loss based on Sobel edge detection and SSIM. While the methodology is clearly presented and relevant, substantial concerns remain regarding experimental validation, generalizability, and data treatment. I therefore recommend Major Revision, and the following issues must be addressed:

- 1) The study relies exclusively on CPFE simulations to validate out-of-distribution microstructures. For a high-impact journal, experimental synthesis and mechanical testing of at least one optimized candidate are strongly recommended to substantiate the central claims.
- 2) Although the optimization targets both yield strength and elongation, CPFE validation only predicts yield strength, omitting elongation due to insufficient calibration data. This directly weakens the claimed strength–ductility synergy and must be resolved through experimental validation or an extended CPFE damage model.

- 3) The model is trained on 25 microstructures expanded to 72,000 patches via cropping and aggressive augmentation (rotation, blurring, distortion, dilation). Such transformations may compromise the physical integrity of EBSD-derived crystallographic information. The authors must justify how these augmentations preserve orientation relationships and grain-boundary physics.
- 4) A CNN trained on the original patches predicts GND density for newly generated bimodal microstructures that lie outside the training distribution. Its predictive reliability under these conditions requires independent verification.
- 5) SSIM is applied to EBSD images encoding Euler angles, which reside in a periodic, symmetry-constrained orientation space. Standard pixel-wise SSIM does not inherently capture crystallographic misorientation. The authors must clarify how symmetry and periodicity were handled.
- 6) NSGA-II operates in a 6D latent space derived from PCA, yet the explained variance drops sharply after the first component. The cumulative retained variance and the impact of discarded components on reconstruction fidelity must be quantified.
- 7) Model parameters are calibrated using only three initial microstructures. Given the distinct bimodal grain morphologies generated, the validity of these parameters for novel microstructures must be reassessed.
- 8) The physics-aware loss is compared only to a pixel-wise baseline. Broader benchmarking against established generative approaches (e.g., GANs or unconstrained DDPMs) is needed to demonstrate methodological superiority.
- 9) Minor Issues:
- Correct "Grian area" to "Grain area" in Figure 3c.
 - Ensure Equations 4 and 8 are properly typeset with correct fraction formatting.
 - Improve axis readability in Supplementary Figure S1 (f–h).

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The author well address previous issues, and deserve a rapid publication in NC.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors have answered and revised the manuscript accordingly. I satisfy with the current version.

(Remarks on code availability)

I have no comments.

Open Access This Peer Review 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 cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review 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 <https://creativecommons.org/licenses/by/4.0/>

Generative design of high-fidelity microstructures using physics-aware machine learning

Weijie Liao,¹ Kaidi Li,^{1,2} Bin Tang,^{1,2} Jiangkun Fan,¹ Jun Wang,¹ Xiangyi Xue,¹ Jinshan Li,^{1,*} and Ruihao Yuan^{1,†}

¹State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an, 710072, China.

²Chongqing Innovation Center, Northwestern Polytechnical University, Chongqing, 401135, China.

We thank the reviewers for a careful reading of the manuscript and the constructive comments. Below is a point-by-point response to the reviewers' comments. The new submission has benefited significantly from the points raised by the reviewers. The changes made in the manuscript are given in text color blue.

REVIEWER #1:

Comment (1.1) *This paper proposes a physics-informed generative optimization strategy for achieving microstructure inverse design guided by competing performance objectives. The method integrates a variational autoencoder with a denoising diffusion probabilistic model, incorporating pixel similarity loss, edge detection loss, and structural similarity loss into the loss function to enhance the modeling capability of physical features such as grain boundaries and grain orientations. With only 25 sets of experimental data, it generates microstructures with bimodal grain size distributions and a synergistic improvement in strength and ductility, which is validated by crystal plasticity finite element simulations. However, several aspects still require refinement:*

Reply:

We thank the reviewer for this positive comment and have addressed the following issues.

Comment (1.2) *Although the introduction highlights the issues of insufficient physical fidelity and weak out-of-distribution generalization in traditional generative models, it does not specifically analyze the unique challenges in the microstructure design of nickel-based high-temperature alloys, mentioning them only briefly as typical aerospace materials. We suggest adding a discussion of the application scenarios and key material properties of nickel-based high-temperature alloys to improve the targetedness and practical relevance of the research.*

Reply:

We thank the reviewer for this suggestion. Accordingly, we have made a revision in the revised manuscript.

Page 1~2, lines 81~89

“Nickel-based superalloys with strength-ductility synergy are critical components used in aerospace scenarios, include turbine disks, blades, and combustors [32, 33]. The two properties, analogous to other materials such as high entropy alloys, are mainly optimized and explained by the microstructures. However, the design of microstructures mainly relies on experience and thereby we select it as a model case to demonstrate the efficacy of the proposed approach.”

Comment (1.3) *Some of the figure legends and axis labels in the manuscript are missing units or use font sizes that are too small. We recommend unifying the image formats and improving clarity to ensure better visual consistency.*

Reply:

We have cleaned up all these issues in the formats.

Comment (1.4) *The article only uses bold formatting for headings at all levels, leading to an overly compact layout. We recommend clearly distinguishing headings from the main text and adjusting line breaks to enhance readability.*

Reply:

We have accordingly revised these to improve readability.

Comment (1.5) *The study applied various data augmentation techniques, including cropping, rotation, and blurring, to the original images. Although these operations significantly expanded the dataset, some may introduce non-physical features. We recommend providing further clarification on whether such augmentations preserve key physical*

information (e.g., grain boundaries and orientations) and whether they influence subsequent physical modeling.

Reply:

This is an important point. We want to clarify that we have used two kinds of image augmentation strategies for model training. The first one, including cropping, rotation and flipping, preserves the original information of the microstructure and is therefore used directly for image reconstruction. The second one, including blurring and masking with black blocks, can lead to information loss such as the appearance of non-physical features. These images are used as input and the corresponding original images are used as output. That is, we train the model to restore the lost information, which can further improve the generalization ability and robustness of the generative model. To conclude, both strategies are designed to obtain a model that can generate high-fidelity microstructures.

We have accordingly made changes in the methods of revised manuscript.

Page 8~9, lines 470~490

“Microstructure database extension

To enrich the training data, a three-step image augmentation process is employed: (i) Each of the 25 original 800×1000 pixel images (Fig. S3a) is uniformly cropped into 80 small images of 100×100 pixels (Fig. S3b). This generates a total of 2,000 small images. (ii) Next, 10,000 new images are produced with flips (horizontal and vertical) and rotations (90° , 180° , 270°) based on the cropped ones. These geometric operations preserve original information, including grain-boundary morphology and lattice orientation (Fig. S3c). (iii) Then, Gaussian blurring, distortion, black blocks, erosion and dilation are applied to the 12,000 images to produce 60,000 images (Fig. S3d). These images (may have information loss) are used only as inputs and the corresponding original images are used as outputs for the image restoration training. In this stage, the model aims to restore the lost information and thereby further enhance the generalization ability of the model. The above steps result in a total of 72,000 small images for model training (12,000 for reconstruction and 60,000 for restoration).”

Comment (1.6) *The overall language of the article is fluent, but some overly long sentences appear in the abstract and conclusion. We recommend splitting these sentences appropriately to clarify the logical thread and enhance readability.*

Reply:

We have accordingly revised the manuscript, please see the updated abstract and conclusion.

REVIEWER #2:

Comment (2.1) *The authors propose a generative optimization framework integrating VAEs, DDPMs, and NSGA-II for inverse microstructure design, incorporating a physics-aware loss based on Sobel edge detection and SSIM. While the methodology is clearly presented and relevant, substantial concerns remain regarding experimental validation, generalizability, and data treatment. I therefore recommend Major Revision, and the following issues must be addressed:*

Reply:

We thank the reviewer for this constructive comment and have addressed the following issues.

Comment (2.2) *The study relies exclusively on CPFE simulations to validate out-of-distribution microstructures. For a high-impact journal, experimental synthesis and mechanical testing of at least one optimized candidate are strongly recommended to substantiate the central claims.*

Reply:

This is a very good point and we have accordingly validated the optimized candidate via experiments. We obtain a bimodal microstructures by tuning the processings based on comprehensive literature review and careful study of existing reports [e.g. Slone C E, Acta Materialia, 2019, 165: 496-507; Tung S Y, Acta Materialia, 2025: 121675]. Figure R 1e (Fig. 5 in the main text) shows the experimentally characterized microstructures, which have a clear bimodal grain size distribution (Figure R 1f). The alloy is processed by 60% hot rolling at 1000°C and then heat treatment at 850°C for 5 min ($1000^\circ\text{C}/60\%-850^\circ\text{C}/5$ min). Figure R 1g shows the corresponding tensile stress-strain curve by experiment (solid line), as well as that from an extended CPFE model (see detail in reply to comment 2.3 and Figure R 2). The two properties (yield strength and elongation) from experiments and CPFE model agree well with the machine learning-based predictions of point ① (Fig. 4f in the main text), as shown in Figure R 1h.

Thus, based on the reviewer's insightful suggestion, the optimized microstructure and associated properties have been confirmed by both experiments and the CPFE model. We envision that the strength-elongation synergy can be further improved if experimental results can augment the initial data and more iterations are conducted. Moreover, advances in processing techniques such as cryogenic treatment [Lee S W, Journal of Magnesium and Alloys, 2020, 8(2): 537-545] and additive manufacturing [Croteau J R, Acta Materialia, 2018, 153: 35-44] can further assist the realization of bimodal microstructure with enhanced properties.

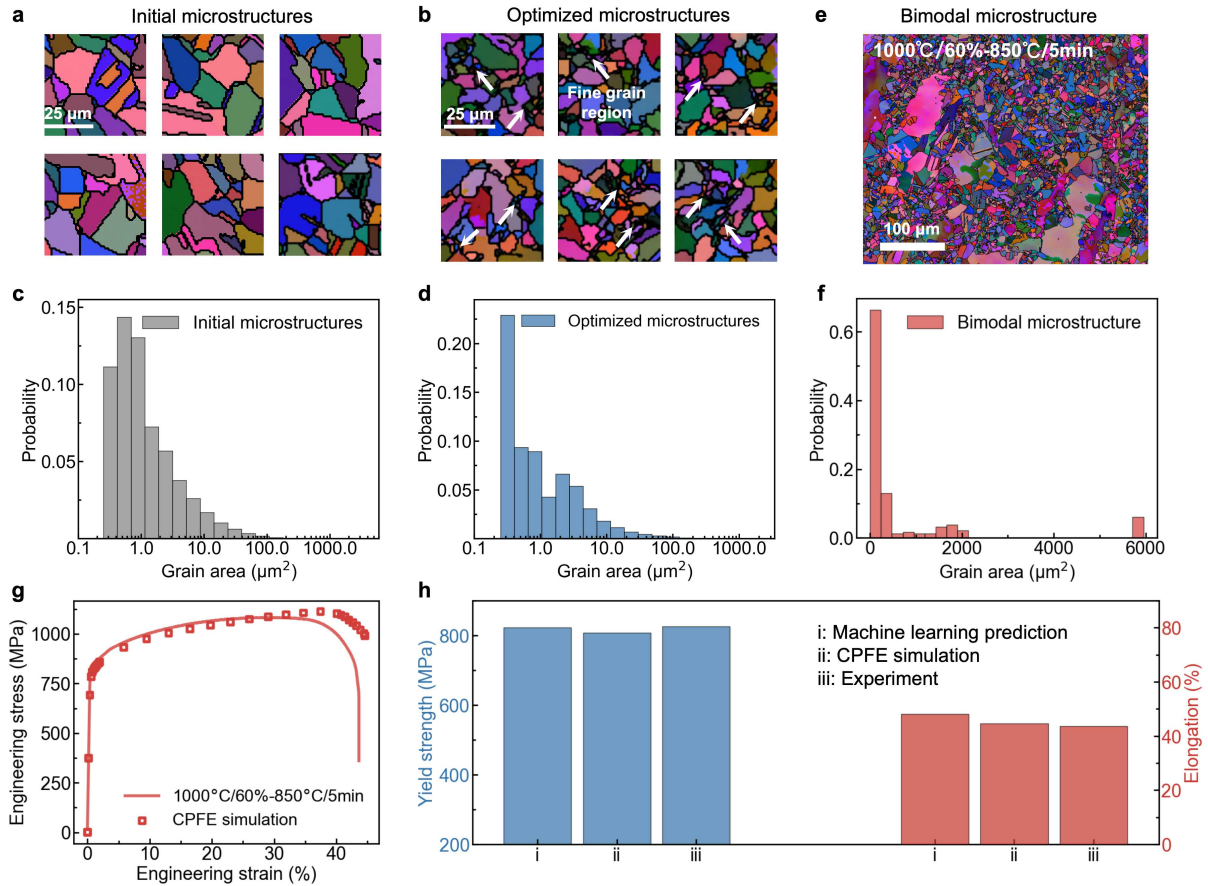


Figure R 1 | Analysis and validation of the optimized microstructure. a and b, Randomly selected samples from the cropped original and optimized microstructures, respectively. c and d, Grain area distribution for all original and optimized microstructures. e EBSD microstructure of the experimentally processed alloy obtained by 60% hot rolling at 1000°C followed by heat treatment at 850°C for 5 min (1000°C/60%-850°C/5 min). f Grain area distribution of e, showing a clear bimodal grain-size characteristic. g Experimental tensile response (red curve) and the corresponding CPFE simulation (red points) of the bimodal microstructure. h Comparison of the mechanical properties between the new experimental alloy and optimized generated alloy (points ① in Fig. 4f of the main text).

We have added the corresponding experimental results and discussion to the revised manuscript.

Page 7, lines 401~421

“Finally, we synthesize a bimodal microstructures by tuning the processings based on comprehensive literature review and careful study of existing reports [43, 44]. The alloy is processed by 60% hot rolling at 1000°C and then heat treatment at 850°C for 5 min (1000°C/60%-850°C/5 min). Fig. 5e shows the experimentally characterized microstructure, which has a clear bimodal grain size distribution (Fig. 5f). Fig. 5g shows the corresponding tensile stress-strain curve by experiment (solid line), as well as that from CPFE model. The two properties (yield strength and elongation) from experiments and CPFE model agree well with the machine learning-based predictions of point ① in Fig. 4f, as shown in Fig. 5h. Thus, the optimized out-of-distribution microstructure with strength-elongation synergy is successfully validated. We envision that the properties can be further improved if experimental results augment the initial data and more iterations are conducted. Moreover, advances in processing techniques such as

cryogenic treatment and additive manufacturing can further assist in building bimodal microstructure with enhanced properties [46, 47].”

Comment (2.3) Although the optimization targets both yield strength and elongation, CPFE validation only predicts yield strength, omitting elongation due to insufficient calibration data. This directly weakens the claimed strength-ductility synergy and must be resolved through experimental validation or an extended CPFE damage model.

Reply:

We have further introduced a ductile damage model into the CPFE framework so that both strength and elongation could be evaluated. The model uses the cumulative plastic work density W_p as the damage-history variable and incorporates stress triaxiality η to characterize the effect of the local stress state on damage initiation and propagation [Liu Y, Materials Science and Engineering: A, 2018, 724: 298-323; Kiran R, Engineering Fracture Mechanics, 2013, 102: 101-117]. The detailed formulation has been added to the Supplementary materials.

Based on this extended CPFE framework, we predicted the tensile responses of both the initial and optimized microstructures, including their strength and ductility. After fitting the experimental stress-strain data of three samples processed under different conditions (Figure R 2a), the required material parameters are identified and subsequently used to initialize the CPFE simulations. As shown in Figure R 2b and c, the simulated yield strengths and elongations of the optimized microstructures agree well with the machine learning predictions.

We have added these results to the Supplementary materials (Page 17~18 and 23, Fig. S11 and S14) and mentioned them in the revised manuscript.

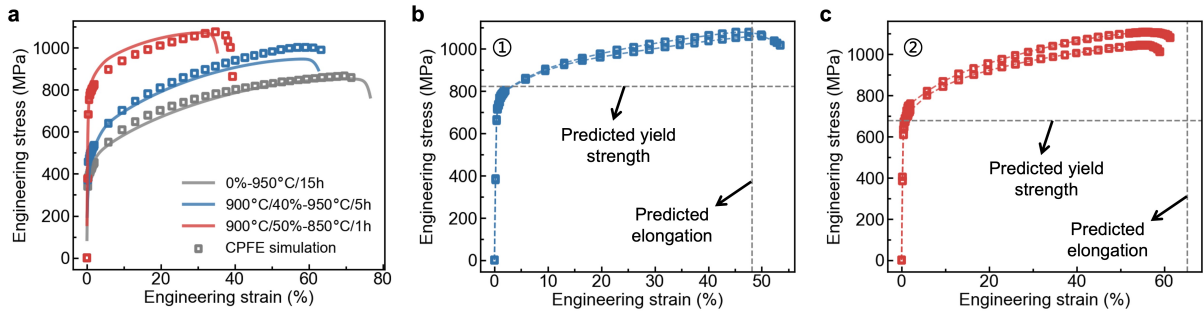


Figure R 2 | CPFE simulation of initial and generated optimized microstructures. a Calibration of the CPFE model using experimental stress-strain curves from representative processed alloys. As an example, 900°C/40%-950°C/5 h indicates 40% hot rolling at 900°C followed by heat treatment at 950°C for 5 h. The solid lines represent the experimental curves, and the square points represent the CPFE simulation results. b,c Simulated tensile responses of the optimized microstructures (point ① and ② in Fig. 4f of the main text). The horizontal and vertical gray dashed lines indicate the predicted yield strength and elongation of machine learning models, respectively.

Comment (2.4) The model is trained on 25 microstructures expanded to 72,000 patches via cropping and aggressive augmentation (rotation, blurring, distortion, dilation). Such transformations may compromise the physical integrity of EBSD-derived crystallographic information. The authors must justify how these augmentations preserve orientation relationships and grain-boundary physics.

Reply:

This is an important point. We want to clarify that we have used two kinds of image augmentation strategies for model training. The first one, including cropping, rotation and flipping, preserves the original information of the microstructure and is therefore used directly for image reconstruction. The second one, including blurring and masking with black blocks, can lead to information loss such as appearance of non-physical features. These images are used as input and the corresponding original images are used as output. That is, we train the model to restore the lost information, this can further improve the generalization ability and robustness of the generative model. To conclude, both two strategies are designed to obtain a model that can generate high-fidelity microstructures.

We have accordingly made changes in the methods of the revised manuscript.

“Microstructure database extension

To enrich the training data, a three-step image augmentation process is employed: (i) Each of the 25 original 800×1000 pixel images (Fig. S3a) is uniformly cropped into 80 small images of 100×100 pixels (Fig. S3b). This generates a total of 2,000 small images. (ii) Next, 10,000 new images are produced with flips (horizontal and vertical) and rotations (90° , 180° , 270°) based on the cropped ones. These geometric operations preserve original information, including grain-boundary morphology and lattice orientation (Fig. S3c). (iii) Then, Gaussian blurring, distortion, black blocks, erosion and dilation are applied to the 12,000 images to produce 60,000 images (Fig. S3d). These images (may have information loss) are used only as inputs and the corresponding original images are used as outputs for the image restoration training. In this stage, the model aims to restore the lost information and thereby further enhance the generalization ability of the model. The above steps result in a total of 72,000 small images for model training (12,000 for reconstruction and 60,000 for restoration).”

Comment (2.5) A CNN trained on the original patches predicts GND density for newly generated bimodal microstructures that lie outside the training distribution. Its predictive reliability under these conditions requires independent verification.

Reply:

This is an important point. We validate the performance of the CNN on out-of-distribution microstructures, that is, the experimentally obtained bimodal microstructure in Figure R 1e. Following the same procedure described in the main text (methods, microstructure database extension), a 1000×800 -pixel original EBSD image is cropped into 80 100×100 -pixel patches (Figure R 3a). These patches are then fed into the trained CNN to predict the GND density, and the predictions are compared with the corresponding experimentally measured GND density. As shown in Figure R 3b, the model achieves a coefficient of determination, $R^2 = 0.73$, indicating the reasonable predictive ability of the CNN model on out-of-distribution microstructures.

We have added Fig. R3 to the Supplementary materials (Page 24, Fig. S15).

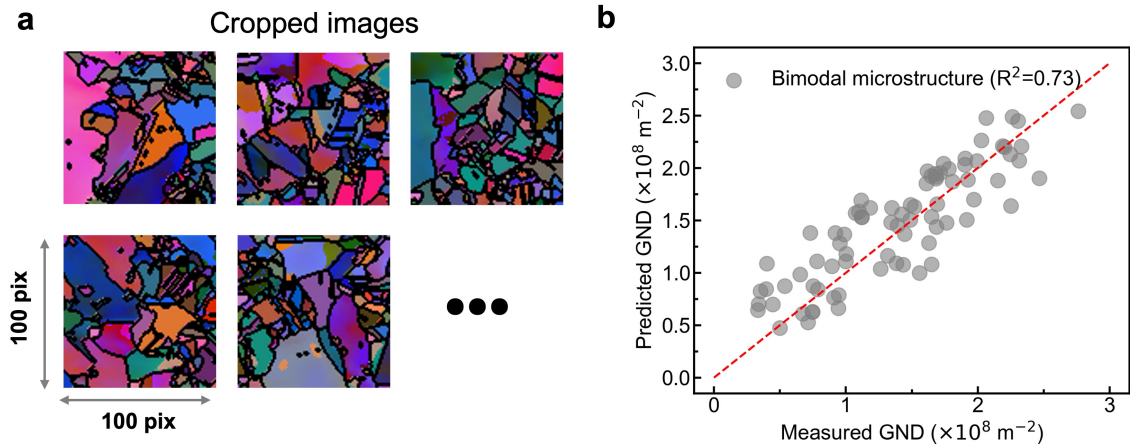


Figure R 3 | Validation of the CNN model’s predictive performance on out-of-distribution data. a The experimentally obtained bimodal microstructure image is cropped into 100×100 -pixel patches. b The predicted GND density from the CNN model is compared with the measured GND density.

Comment (2.6) SSIM is applied to EBSD images encoding Euler angles, which reside in a periodic, symmetry-constrained orientation space. Standard pixel-wise SSIM does not inherently capture crystallographic misorientation. The authors must clarify how symmetry and periodicity were handled.

Reply:

This is a very good point and we really appreciate this comment. We admit that we have not directly used the symmetry and periodicity during the modeling process. However, we would like to claim that the information is not destructed and overlooked in the study. Figure R 4 illustrates how the Euler angles are mapped to the pixel values for training image generation models, from which the output image pixels are mapped back to Euler angles for

misorientation calculations. Figure R 4a shows that three Euler angles are mapped to the red, green, and blue channels, respectively, which constitute an image. Taking Euler1 as an example (Figure R 4b), each angle is first transformed to a pixel value using $\text{Pixel value} = \text{Euler1 value}/360 \times 255$. These pixel values are used to train a model to generate pixels for a certain image. Then the pixels of the generated images are transformed back to Euler1 angles using the same method. That is, such a one-to-one mapping allows us to preserve the original information in the Euler angles and thereby the SSIM based on pixel values will not cause information loss in symmetry and periodicity. There have been several studies that use pixels or Euler angle to train surrogate models for image generation or reconstruction etc [e.g. Labady S, Acta Materialia, 2026: 122240; Jung J, npj Computational Materials, 2021, 7(1): 96]. The symmetry and periodicity may be more rigorously considered when advanced 5D or 6D rotation representation methods are available [e.g. Zhou Y, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5745-5753], which deserves further study in future work.

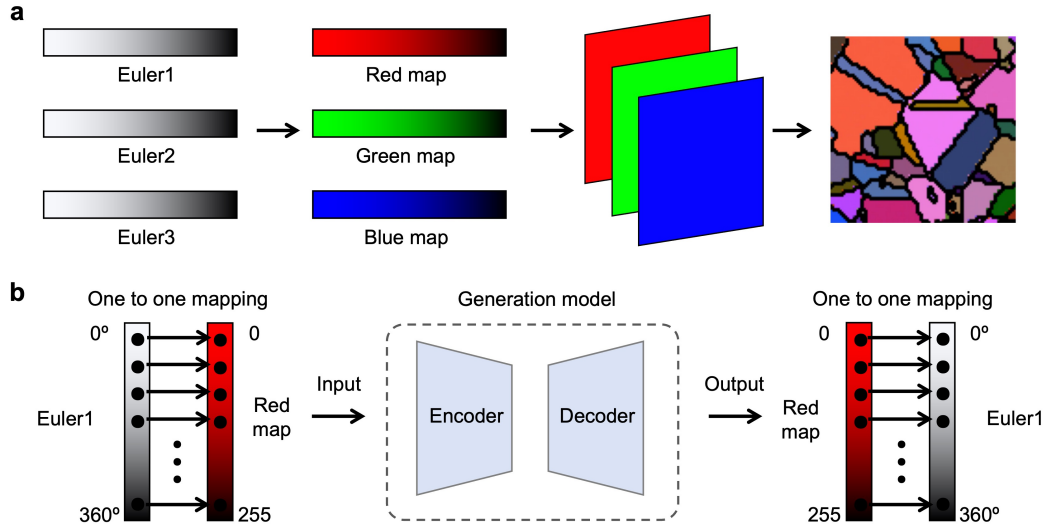


Figure R 4 | One-to-one mapping from Euler angle to image pixel. a Schematic illustration of the conversion from three Euler angles to the RGB image, where Euler1, Euler2, and Euler3 are mapped to the red, green, and blue channels, respectively. b One-to-one mapping from each value in Euler1 to each pixel in the red channel before model training, and the reverse transformation after image generation.

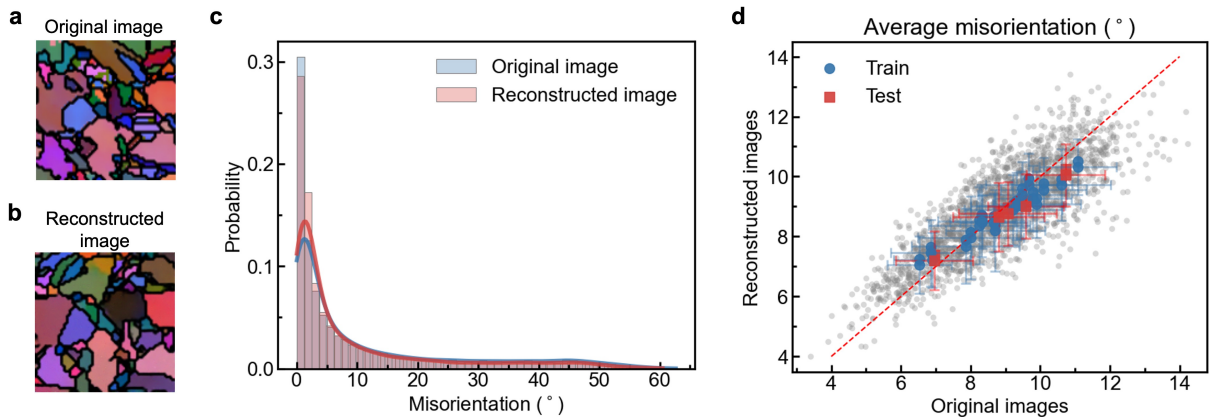


Figure R 5 | The analysis of orientation difference distributions between original and reconstructed images. a The original image. b The reconstructed image. c The misorientation distributions of the original and reconstructed images closely align. d Comparison of average misorientation for all original and reconstructed images. The grey points represent the cropped patches. The blue and red points with error bars indicate the mean and standard deviation of the physical features for the original large images in the training and test sets, respectively.

To confirm the possible information loss in symmetry and periodicity, we calculate the misorientation distribution of the reconstructed images and the original images, as shown in Figure R 5a-c. For the two images, the distributions show a good consistency (Figure R 5c). To make the comparison more reliable, we extend the analysis to all 2,000 cropped image patches and their reconstructions (Figure R 5d). The misorientation values of the original and reconstructed images agree well in both training and test data (details please see the calculation for Fig. 3 in the main text).

We have added these results to the Supplementary materials (Page 6~8, Fig. S4 and S5) and revised manuscript.

Page 5, lines 239~247

“The crystallographic misorientation distributions are also analyzed (Fig. S4), which show a similar tendency to those in Fig. 3. It should be noted that the symmetry and periodicity of the lattice are not directly considered during modeling. However, we adopt a one-to-one mapping from Euler angle to image pixel that can avoid possible crystallographic information loss, as supported by both Fig. 3 and Fig. S4-5.”

Comment (2.7) *NSGA-II operates in a 6D latent space derived from PCA, yet the explained variance drops sharply after the first component. The cumulative retained variance and the impact of discarded components on reconstruction fidelity must be quantified.*

Reply:

This is an insightful comment. We have analyzed both the cumulative retained variance and the reconstruction fidelity as functions of the number of principal components (PCs). Figure R 6 a and b show that the first PC explains more than ~30% of the variance, and the cumulative retained variance increases slowly and reaches 90% at about 20 PCs. However, the following analysis suggests that using more PCs is not necessarily to improve the microstructure fidelity. This is due to the fact that PCA only guarantees optimality for second-order statistics, and PCs ranked lower may introduce redundant information or noise that have no influence on image fidelity.

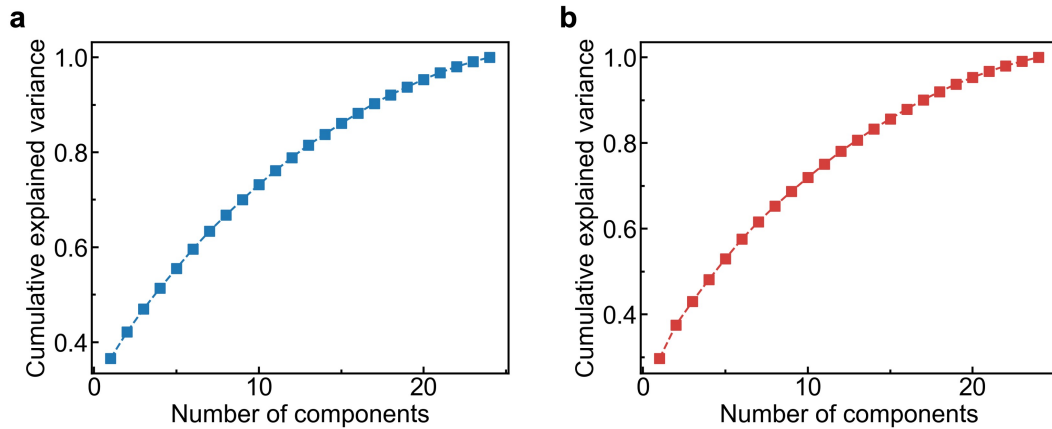


Figure R 6 | The cumulative explained variance plot for the μ and σ obtained via Principal component analysis (PCA). **a** The PCA applied to the μ vector. **b** The PCA applied to the σ vector.

We next systematically evaluate the effect of the number of PCs on the representation of original microstructure and the fidelity of the generated microstructure. For simplicity, we randomly select an image from the test data to examine the effect, as shown in Figure R 7a. Following the same image cropping and encoding method in the main text (Page 5, first paragraph in subsection ‘Multi-objective driven microstructure design’), we get the PCs for the original image. In the main text, we use 3 PCs to represent the original image and guide the optimization of microstructures. Here, we will change the number of PCs. In Figure R 7b and c, the effect of the number of PCs on the similarity between original and generated images is checked. We use a quantitative metric to measure the similarity, which is based on the encoded latent variables of original and generated images. The metric, Fréchet distance (FD) is given by, $FD = FD(P, Q)$, where P and Q represent the overall latent-variable distributions for the original and generated images, respectively (details in Supplementary materials, Page 13). A lower FD means a higher similarity. Figure R 7b shows that the FD decreases monotonically with increasing PCs. Note that after the third one, the FD becomes almost stable. This suggests that adding more PCs has little positive effect on the fidelity of the generated microstructures. To obtain a simple model that can benefit the inverse optimization process, we thus select 3 PCs. The fidelity is

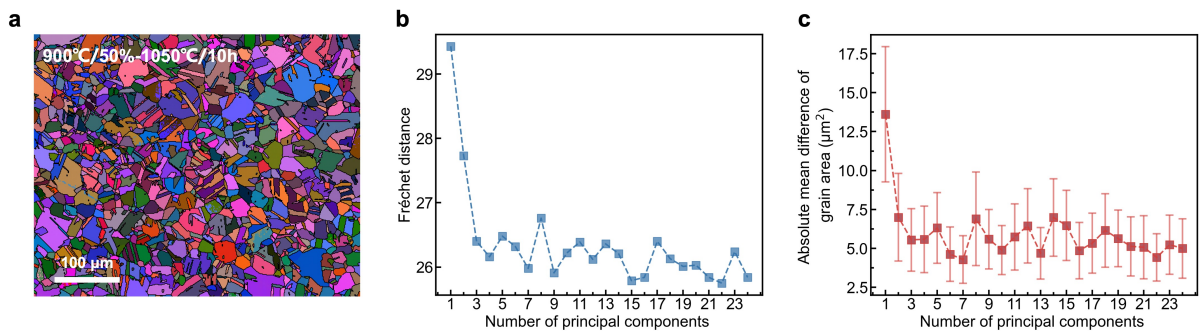


Figure R 7 | The fidelity analysis between original and generated images as the number of principal components (PCs) increases. **a** A randomly selected initial microstructure from the test set. **b** The Fréchet distance (FD) between the overall latent-variable distributions of all original cropped images and generated images as a function of the number of PCs. **c** The absolute mean difference of grain area between generated images and the original image **a**, with error bars indicating the standard deviation.

further confirmed by examining the grain area of generated microstructures with varying PCs, as shown in Figure R 7c. The tendency is similar to that in Figure R 7b, that is, the deviation in grain area between original and generated microstructures becomes close when 3 or more PCs are used.

We have added the corresponding results and details to the Supplementary materials (Page 13~15, Fig. S9 and S10) and accordingly revised the main manuscript.

Page 6, lines 304~310

“The original dimension of PCs for μ and σ is 24, which is so high for the next inverse optimization step. Thus, we analyze how the number of PCs affects the fidelity of the generated microstructures (Fig. S10). Finally, for each descriptor, the first three principal components are retained, resulting in a 6-dimensional representation for each microstructure.”

Comment (2.8) Model parameters are calibrated using only three initial microstructures. Given the distinct bimodal grain morphologies generated, the validity of these parameters for novel microstructures must be reassessed.

Reply:

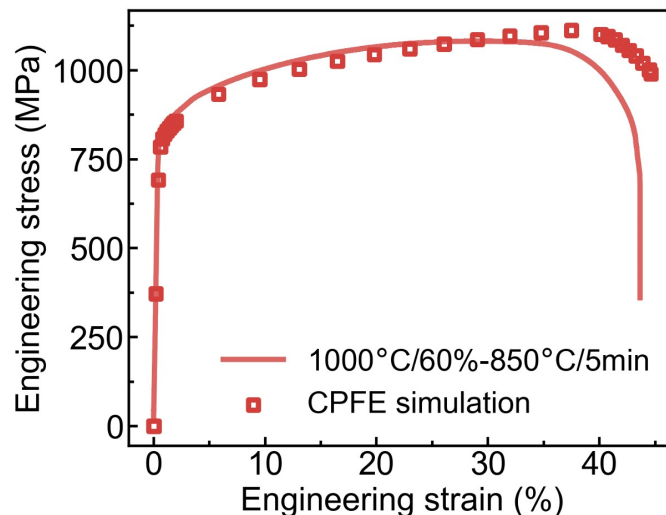


Figure R 8 | Experimental tensile response (red curve) and the corresponding CPFE simulation (red points) of the bimodal microstructure. The simulated curve shows good agreement with the experimental result, supporting the applicability of the calibrated CPFE parameters to bimodal microstructures.

The CPFE model used in this study explicitly incorporates the strengthening effect of the grain-boundary influence zone and has been shown to capture the Hall-Petch-type response associated with grain-size variation. Therefore, its applicability is not restricted to the three initial microstructures used for calibration, but extends to microstructures with different grain-size distributions, including bimodal ones. We would like to refer the reviewer to our previous work if the reviewer is interested in the CPFE model [Li K, Materials Science and Engineering: A, 2024, 907: 146747]. To further verify this point, we construct CPFE models directly from the experimentally synthesized bimodal microstructures. The simulated tensile response is compared to the corresponding experimental curve, as shown in Figure R 8. A good agreement is achieved, supporting the transferability of the calibrated parameters to the bimodal microstructures.

We have accordingly revised this in the main manuscript.

Comment (2.9) *The physics-aware loss is compared only to a pixel-wise baseline. Broader benchmarking against established generative approaches (e.g., GANs or unconstrained DDPMs) is needed to demonstrate methodological superiority.*

Reply:

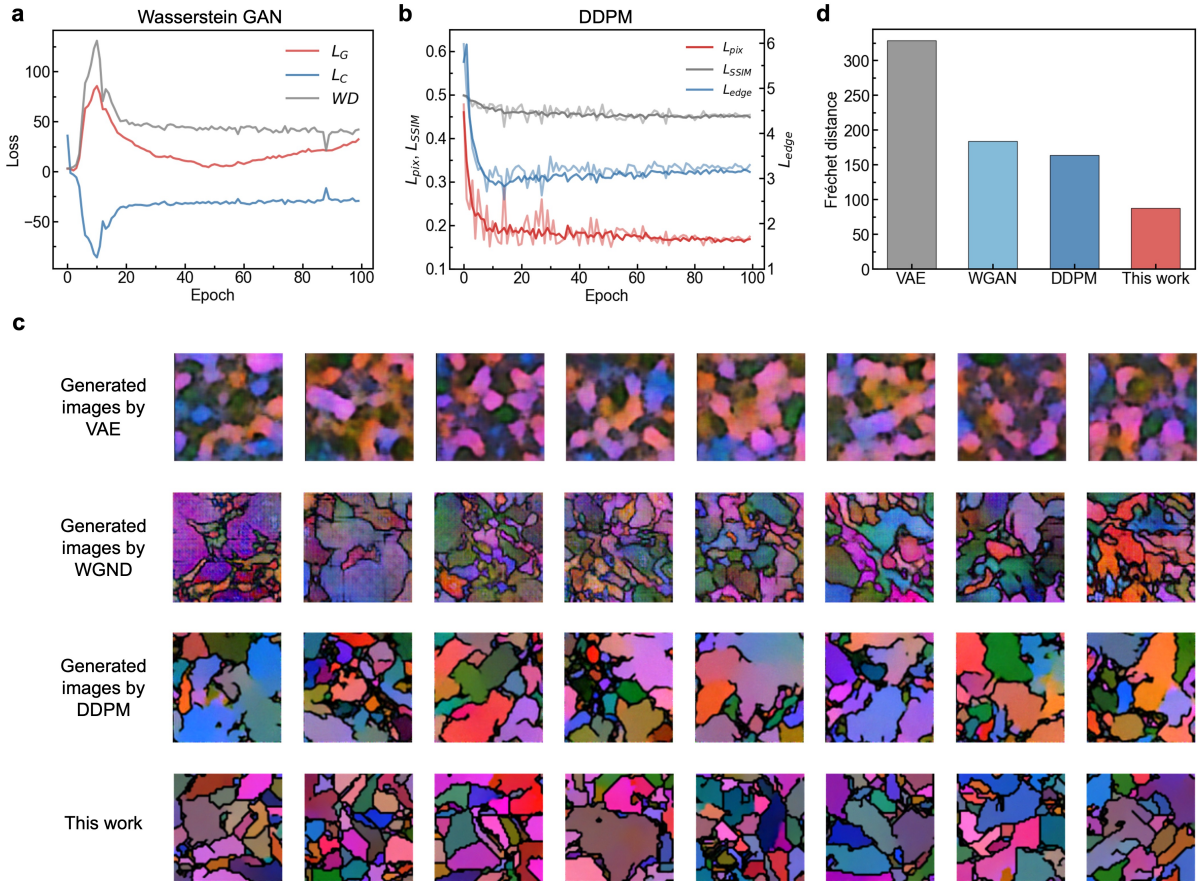


Figure R 9 | The performance comparison of four generative models. a The loss functions evolution of the discriminator and generator in WGANs. b The loss functions evolution for standard DDPMs. c Randomly generated microstructures from the standard VAE (first row), WGAN (second row), standard DDPM (third row), and this work (fourth row). d The FD evaluation results for different generative models.

We have extended the benchmarking study by including three additional baseline models: a Wasserstein GAN (WGAN), a standard VAE, and a standard DDPM. The results show that the proposed framework outperforms all three models. The WGAN is selected because it is a representative GANs variant with improved training stability and reduced mode collapse [Gulrajani I, Advances in neural information processing systems, 2017, 30]. The detailed formulation of the WGAN has been added to the Supplementary materials. We also test the standard VAE model

which has the same parameters to that used in the main text. In addition, the standard DDPM model uses the same architecture to that in the main text except for employing the VAE output as condition layer. Figure R 9a and b show the loss as a function of epoch for WGAN and DDPM. The losses for WGAN fluctuate in the first ~ 20 epochs because of adversarial optimization and then gradually become stable. The losses for DDPM gradually become stable after ~ 40 epochs. Figure R 9c compares the generated images by different models, those from our model is also included as comparison. The VAE generates the most blurry microstructures than the others, while WGAN and DDPM generated microstructures exhibit substantial noise and hallucination in crystallographic orientations and interfaces.

We further quantitatively compare the generations from the four models, in total 2,000 microstructures (the same number to the cropped original images) are generated by each model. The Fréchet Distance (FD) is used as the metric to assess the similarity between generated and original images, which is a widely used metric for evaluating generative model performance at the distribution level [Heusel M, Advances in neural information processing systems, 2017, 30]. Specifically, the generated and original images are first input into an Inception-v3 network to extract the feature representations. Then the FD between the two feature distributions is calculated by $FD = FD(P, Q)$, where P and Q represent the extracted feature representations for the original and generated images, respectively (details in Supplementary materials, Page 10). A lower FD indicates a better image fidelity of generated microstructures with original ones. Figure R 9d compares FD of the four models and indicates that the images generated by our model has the highest fidelity.

We have added these results to the Supplementary materials (Page 10~11, Fig. S7) and revised manuscript.

Page 5, lines 250~254

“Additionally, we benchmark the proposed framework against three baseline models, namely a standard VAE, a Wasserstein GAN (WGAN), and a standard DDPM. The comparisons show that the proposed method outperform all other models (Fig. S7).”

Comment (2.10) *Minor Issues:*

- Correct “Grian area” to “Grain area” in Figure 3c.
- Ensure Equations 4 and 8 are properly typeset with correct fraction formatting.
- Improve axis readability in Supplementary Figure S1 (f-h).

Reply:

We have cleaned up all these issues.

* ljsh@nwpu.edu.cn (Jinshan Li)
† rhyuan@nwpu.edu.cn (Ruihao Yuan)