Engineering with Computers (2026) 42:63
https://doi.org/10.1007/s00366-026-02290-9

ORIGINAL ARTICLE

Sim-to-real diffusion models for microstructure prediction in metal additive manufacturing

Ziyuan Xie[1] · Zichuan Fu[2] · Jingchi Zhang[2] · Kaihao Zhang[2] · Tianchen Zeng[3] · Yu Wu[3] · Shuheng Liao[4,5] · Xiang Li[6] · Tianju Xue[1]

Received: 4 October 2025 / Accepted: 9 February 2026

Abstract

Metal additive manufacturing is an efficient technique for producing metallic components with high flexibility. Understanding the mechanical properties of printed parts requires detailed analysis of their microstructures. While experiments are reliable for capturing as-printed microstructures, they are often costly and time-consuming. Alternatively, numerical simulations have become valuable tools to reduce the reliance on trial-and-error experimentation, powered by recent advances in high-performance computing such as GPUs for acceleration. However, a key question remains of effectively combining high-fidelity, yet expensive, experiment data with relatively lower-fidelity, yet massive, simulation data to achieve a coherent prediction of microstructures under various manufacturing process conditions. In this study, we leverage the power of deep generative models and propose a sim-to-real framework based on denoising diffusion probabilistic model (DDPM) to generate realistic microstructures for additively manufactured products. Specifically, we pre-train a DDPM on a large dataset of phase-field (PF) simulated microstructures. This model is then fine-tuned or distilled using a relatively small set of microstructure images obtained from electron backscatter diffraction (EBSD) experiments, with the goal of enhancing the authenticity of the generated images. The microstructures generated by our sim-to-real diffusion models show strong agreement with experimental results, evaluated using both machine learning and physics-based metrics. In particular, the sim-to-real diffusion models show accurate prediction under unseen experimental manufacturing process conditions (but covered by simulation data), demonstrating their excellent generalization ability.

Keywords Metal additive manufacturing · Microstructures · Deep generative models

1 Introduction

Metal additive manufacturing (AM) is fast growing due to its outstanding ability to manufacture products with complex geometries in a rapid and precise manner [1]. Additively manufactured products are now found in a wide range of industries, including aerospace, automotive, and healthcare [2, 3]. However, achieving consistent control over the properties of as-printed metal components remains a significant challenge, as variations in processing parameters, such as laser power and scanning speed, can lead to marked differences in the final product performance. In

---

✉ Kaihao Zhang
kaihaozhang@hkust-gz.edu.cn

✉ Yu Wu
wuyucs@whu.edu.cn

✉ Tianju Xue
cetxue@ust.hk

[1] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

[2] Smart Manufacturing Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453, China

[3] School of Computer Science, Wuhan University, Wuhan 430072, China

[4] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5] Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China

[6] China Iron and Steel Research Institute Group, Beijing 100081, China

the process–structure–property (PSP) relationship of metal AM, microstructures serve as a critical intermediary that bridges processing conditions and material properties. Consequently, understanding and controlling microstructures has become a key focus of research, with the goal of building up a practical PSP relationship and ultimately improving overall product quality [4–6].

Experimental methods, such as scanning electron microscopy (SEM) and electron backscatter diffraction (EBSD), are considered as the gold standard to characterize the microstructures of AM products. SEM enables the direct observation of grain size and morphology, while EBSD provides detailed crystallographic information such as grain orientation, texture, and phase distribution. These methods have been widely used to analyze AM-fabricated alloys, revealing microstructural characteristics such as epitaxial grain growth and process-dependent texture evolution [7–9]. Though experimental approaches provide high-fidelity data, they are often quite expensive. Hence, to reduce trial-and-error in optimizing manufacturing parameters, it is necessary to construct a relatively large reference database that covers printing cases under various conditions. Because of this need, rapid and cost-effective prediction of the microstructures of additively manufactured metallic products has become a major focus in the metal AM community. Physics-based numerical simulations and the emerging efforts on deep generative models are two promising solutions to this challenge.

Physics-based numerical simulation provides an auxiliary tool to reproduce the scenarios of experiments. The simulation approaches are indispensable due to its ability to reveal phenomena that are difficult to observe during experiments, such as grain development in the metal AM process [10]. The fast advancement of computing power, especially the improvement of high-performance GPUs in recent years, has significantly reduced simulation times, making it possible to generate massive data efficiently through computers. Cellular Automata (CA) [11, 12], Monte Carlo (MC) [13] and phase-field (PF) models [14, 15] are commonly adopted for prediction of metallic microstructures. These approaches enable the simulation of grain morphology and growth behavior during AM processing, with recent advancements in three-dimensional (3D) modeling significantly improving the fidelity of such predictions [13, 15–17]. Although these studies have demonstrated the effectiveness of numerical simulations, significant discrepancies still exist between simulations and experimental results, primarily because physical phenomena are often simplified during the simulation process.

The recent success of deep generative models shows their great potential to offer a powerful alternative to rapidly generate large volumes of realistic data. As a branch of machine learning algorithms, generative models excel at capturing complex data distributions and synthesizing realistic samples. Several frameworks have been developed for this purpose, including variational autoencoders [18] (VAEs), generative adversarial networks (GANs) [19], and normalizing flows [20]. While VAEs tend to produce samples of relatively lower quality, GANs are difficult to train, and flow models heavily rely on the expressive ability of their network architectures [21], diffusion models [22–24] have emerged as a mainstream generative approach due to their high generation quality and training stability. Recent studies have applied diffusion models to microstructure generation, indicating their ability to replicate structural characteristics of material. Training diffusion models directly on experimental data is a common practice. Azqadan et al. [25] applied denoising diffusion probabilistic models (DDPM) to generate microstructural images, with the training data characterized from 11 cast-forged magnesium alloy components. The generated images resembled the real ones and had reasonable particle distribution. Nevertheless, due to the difficulties of sample preparation, their training dataset contained 27 samples from these components, covering only 8 percent of the process parameters space, which limits the ability of their model. Training diffusion models on pure simulation data is another paradigm when experiment data is limited or costly to obtain. This approach has been adopted in several recent microstructure studies with diffusion models. Lyu et al. [26] successfully trained a conditional DDPM using synthesized microstructures of various composite materials and generated materials with specific permeability. Hoffman et al. [27] employed a diffusion model trained on microstructures simulated using the open-source software `SPPARKS` [28] to inpaint microstructures for various computer-aided design (CAD) objects. Even though these models trained solely on simulation data show good performance in generating microstructures, a typical issue is that the fidelity of the generated images is relatively lower than that achieved by models trained with experiment data, primarily due to the nature of the training data itself. Based on our description above, experiment data and simulation data exhibit complementary characteristics. Therefore, it is anticipated that using both forms of data in concert will produce superior outcomes.

Indeed, multi-fidelity approaches have recently been explored in diffusion models. Shu et al. [29] proposed a physics-informed diffusion model for high-fidelity flow field reconstruction, where their model is trained exclusively on high-fidelity simulation data and then applied to reconstruct flow fields from low-fidelity inputs. A similar multimodal generative frame work proposed by Qu et al. [30] was able to calibrate the global vertical temperature profiles by assimilating in-situ weather station data and ex-situ satellite imagery as the sampling conditions. In computer vision tasks for healthcare monitoring with limited real-world data, Gao et al. [31] proposed a diffusion-based sim-to-real transfer framework to train a diffusion model to generate in-bed human pose parameters from synthetic depth images, then generated poses conditioned on real depth images at inference. However, in the domain of metal AM, a comprehensive approach to utilize both the high-fidelity but expensive experiment data and the relatively low-fidelity but extensive simulation data to make a coherent prediction of microstructures is currently not available.

In this study, to fully leverage the advantages of both experiment and simulation data, we propose conditional sim-to-real diffusion models for generating high-fidelity microstructures within an applicable AM process window. The proposed model framework is conceptually illustrated in Fig. 1. A large amount of simulated microstructures and a relatively small amount of microstructural images are collected from the PF simulations and EBSD experiments, respectively. Analogous to the sim-to-real training in robotics [32, 33], to avoid the significant expense of collecting experiment data for training, our model is first trained on an extensive set of simulation data that spans a wide range of the process window for the typical metal AM process of 316 L stainless steel. Although these simulation data are relatively low-fidelity, they provide abundant and diverse training samples that capture the fundamental physical patterns of the system. This is particularly valuable because high-fidelity experimental data is often scarce and expensive to obtain. The pre-trained model is then fine-tuned [34] or distilled [35] using a limited set of experiment data to learn the realistic distribution and morphology of microstructures. These strategies allow the model to learn robust physical priors from large-scale simulations while adapting to the nuanced characteristics of real experimental observations, ultimately bridging the fidelity gap between simulation and reality. During the fine-tuning or distillation process, experimental images of certain manufacturing conditions are set aside as the testing set. Evaluation on the generated microstructures is made with both machine learning metric, which is the Fréchet inception distance (FID) score [36], and physics-based metrics, including grain size and morphology.

It should be noted that simulation itself or direct training of diffusion model on pure experiment data (referred to as the baseline diffusion model) has poor generalization ability and cannot predict accurate microstructures on the test dataset. To show the limitations of simulation and baseline diffusion model, we also evaluate the microstructures generated by these two methods. The evaluation results demonstrate that the microstructural images generated by our sim-to-real approach closely match the experimental ones, whereas the outputs from simulation and the baseline model each exhibit specific deficiencies. More importantly, the pretrained model based solely on simulation data is well-suited for transfer learning, enabling fast adaptation to other experimental datasets within the AM community. An example demonstrating this capability is presented in this work by considering experimental data from an independent study.

The paper is organized as follows: Sect. 2 describes the preparation of simulation and experiment data. Section 3

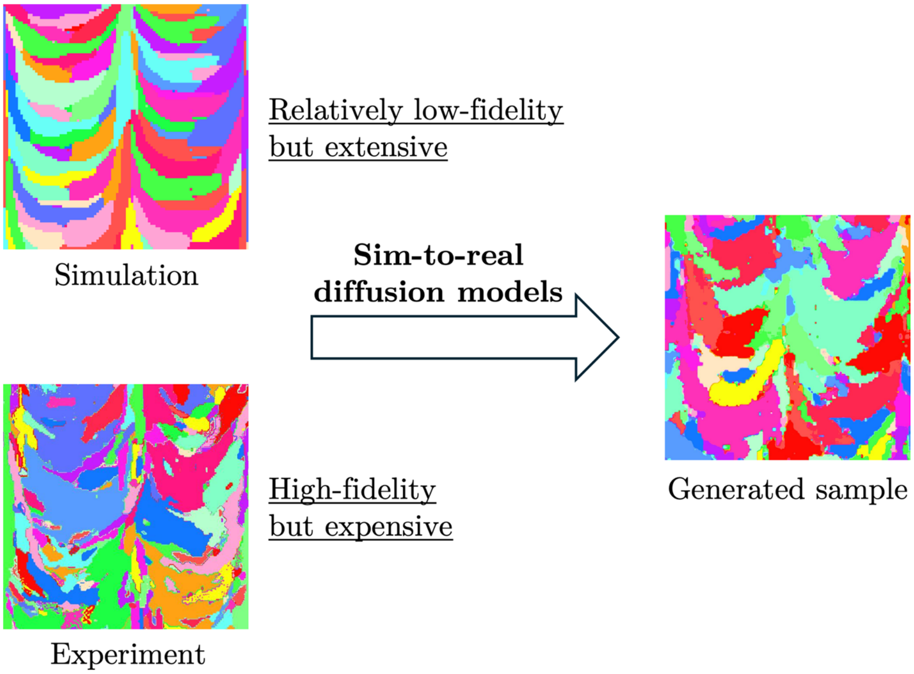


Fig. 1 Conceptual illustration of our proposed method

introduces the design of the sim-to-real diffusion models and the evaluation metrics for microstructural analysis. Section 4 presents a comparative evaluation of experimental, simulated, and generated microstructural images based on the proposed metrics. Finally, Sect. 5 summarizes the overall workflow and discusses the contributions of this study.

Our code is available at https://github.com/CMSL-HKUST/sim-to-real.

2 Data preparation

In this work, a laser powder bed fusion AM process is considered using 316 L stainless steel. We detail the preparation of simulation and experiment data used for model training. The simulation data, consisting of microstructural images, are generated using PF simulations. Prior to the PF simulations, a calibrated computational fluid dynamics (CFD) solver is employed to compute the temperature field, which serves as a prerequisite for modeling microstructure evolution. We also prepare experiment data in the form of EBSD images by a series of experiments, including sample printing, cutting, grinding, ion milling and EBSD characterization.

2.1 Numerical simulations

2.1.1 CFD model

Information on the temperature field is critical in the metal AM process. In this study, a CFD model is developed to capture the time-resolved temperature distribution influenced by a moving laser. The laser powder bed fusion AM simulation is illustrated in Fig. 2.

In the CFD model, the flow is assumed to be incompressible laminar flow and the influence of buoyancy and gravity is ignored. The simulated powder bed is considered as a continuum domain and powder is not explicitly modeled for simplicity. The governing equations are incompressibility condition, momentum conservation and energy conservation which are given as follows:

$$\nabla \cdot \mathrm{u} = 0 \tag{1}$$

$$\frac{\partial(\rho \mathrm{u})}{\partial t} + \nabla \cdot (\rho \mathrm{u}\mathrm{u}) = \nabla \cdot (\mu \nabla \mathrm{u}) - \nabla p + \mathrm{f} \tag{2}$$

$$\frac{\partial(\rho c_p T)}{\partial t} + \nabla \cdot (\rho c_p \mathrm{u} T) + \frac{\partial(\rho L f_L)}{\partial t} + \nabla \cdot (\rho L \mathrm{u} f_L) = \nabla \cdot (k \nabla T) + Q_{\text{laser}} \tag{3}$$

where ρ is the material density, μ is the viscosity, p is the pressure, T is the temperature, and L is the latent heat. The heat capacity c_p and the thermal conductivity k are temperature-dependent. The flow velocity vector is denoted as u. In Eq. (2), f is the Darcy drag force, given as $\mathrm{f} = \frac{180\mu}{\delta^2}\frac{(1-f_L)^2}{f_L^3}\mathrm{u}$, which represents the resistance imposed by the grain structure on the motion of the molten flow [37], and δ denotes the characteristic length of the mushy zone, which is taken to be the primary dendrite arm spacing. f_L is the liquid fraction describing the intermediate state (mushy zone) between liquid and solid, given as $f_L = \frac{T-T_s}{T_l-T_s}$, where T_s and T_l are the solidus and liquidus temperature, respectively. In Eq. (3), Q_{laser} is the volumetric heat source with a Gaussian distribution injected by the laser [38], which is:

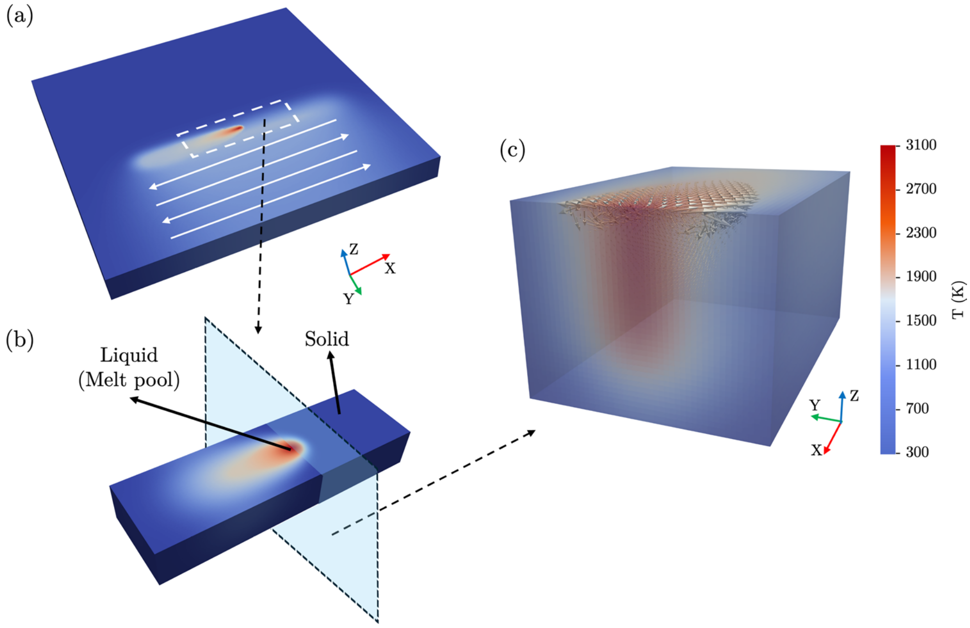


Fig. 2 Illustration of the CFD simulation: **a** Multi-track process; **b** One track showing regions of liquid and solid metal; **c** Fluid velocity field in the melt pool

$$Q_{\text{laser}} = \begin{cases} \dfrac{2P\eta}{\pi r_b^2 d} \exp(-\dfrac{2(x^2+y^2)}{r_b^2}), & z_{\text{top}} - z \le d \\ 0, & z_{\text{top}} - z > d \end{cases} \tag{4}$$

where P is the laser power, η is the absorption coefficient, r_b is the laser beam radius, and z_{top} denotes the top surface of the computational domain. For the thermal boundary conditions, we consider the heat loss due to heat convection and radiation [39]:

$$\begin{aligned} Q_{\text{conv}} &= h(T - T_0) \\ Q_{\text{rad}} &= \sigma e(T^4 - T_0^4) \\ -k\Delta T \cdot \text{n} &= Q_{\text{conv}} + Q_{\text{rad}} \end{aligned} \tag{5}$$

where h is the convection heat transfer coefficient, σ is the Stefan-Boltzmann constant, e is the emissivity of the material, and T_0 is the ambient temperature. While for the velocity boundary conditions, Marangoni shear stress is applied on the free surface of the melt pool, modeled via a temperature-dependent surface tension gradient [40, 41], following:

$$\tau_{\mathbf{M}} = \mu \frac{\partial \text{u}}{\partial z} = -\frac{\partial \gamma}{\partial T} \nabla_s T \tag{6}$$

where γ is the surface tension coefficient, and ∇_S is the surface gradient operator. To solve the transient CFD system, the convection terms are treated explicitly while the diffusion terms are treated implicitly:

$$\begin{aligned} &\rho \bar{c}^{n-1} \frac{T^n - T^{n-1}}{\Delta t} + \rho \bar{c}^{n-1} \nabla \cdot (\text{u}^{n-1} T^{n-1}) \\ &= \nabla \cdot (k^n \nabla T^n) + Q^n \end{aligned} \tag{7}$$

$$\begin{aligned} &\frac{\rho \left(\text{u}^* - \text{u}^{n-1}\right)}{\Delta t} + \rho \cdot \nabla \left(\text{u}^{n-1} \text{u}^{n-1}\right) \\ &= \nabla \cdot (\mu \nabla \text{u}^*) - \nabla p^{n-1} + \text{f}^n \end{aligned} \tag{8}$$

where $\bar{c}$ is the apparent heat capacity, given by $\bar{c} = c_p + L\frac{\partial f_L}{\partial T}$. The superscript n and $n-1$ refers to the value at the present and the last time step, respectively. Finite volume method is used for spacial discretization. Central difference scheme is used for the diffusion terms while the QUICK scheme is used for the convection terms. In Eq. (8), the incremental pressure-correction scheme is used [42]. At each time step, an estimated velocity field $\mathbf{u}^*$ is calculated based on the pressure gradient of the last time step. Then the present pressure is corrected by solving a Poisson's equation and the actual velocity field can be updated. The process is shown below:

$$\begin{aligned} p^n &= p^{n-1} + p' \\ \nabla^2 p' &= \frac{\rho}{\Delta t} \nabla \cdot \text{u}^* \\ \text{u}^n &= \text{u}^* - \frac{\Delta t}{\rho} \nabla p' \end{aligned} \tag{9}$$

The material focused in this study is 316 L stainless steel, for which the CFD model is further calibrated based on our experimental observations. We print five specimens with different energy density ρ_e, which is defined as [43]:

$$\rho_e = \frac{P}{vhl} \tag{10}$$

where v is the laser scanning speed, h is the hatch space, and l is the layer thickness. The values of h and l are set to 100 μm and 30 μm, respectively. For measurement of the melt pool dimensions, we cut cubic samples with size of $5 \times 5 \times 5\ \text{mm}^3$ from the specimens along the axial direction using electrical wire cutting. The preparation of these metallographic sample involves grinding in steps $\#180$ to $\#2500$ sandpaper, followed by polishing with alumina suspension of 1 μm to 50 nm. An optical microscope is used to observe and measure the morphology of the melt pool, where the cross-section area is chemically corroded using commercial ferrous chloride reagent for 10-20 s.

The melt pool dimensions obtained from our measurements are employed to calibrate the CFD model. Figure 3a presents the experimentally observed (the upper row) and numerically simulated (the lower row) melt pools, while Fig. 3b compares the melt pool dimensions (width and depth) obtained from both experiments and simulations. The overall errors in the predicted width and depth are 9.45% and 8.15%, respectively, demonstrating the reliability and accuracy of the proposed CFD model.

2.1.2 PF model

The phase-field (PF) method is effective in constructing an relatively accurate model of grain growth under a prescribed temperature field. A mesoscale PF model, which generally follows the descriptions in Ref. [15], is employed in this study to simulate the microstructures under the temperature profile from the CFD model mentioned above. The grain orientations are described by a parameter $\eta_i(\text{r}, t)$, where $i = 1, 2, ..., n$. If a grain takes the i^{th} orientation, then we have $\eta_j = 1, j = i$ and $\eta_j = 0, j \ne i$. A time-dependent Ginzburg-Landau equation is given as the governing equation:

$$\frac{\partial \eta_i}{\partial t} = -L_g \frac{\delta F(\zeta, \eta_i, T)}{\delta \eta_i(\text{r}, t)} \tag{11}$$

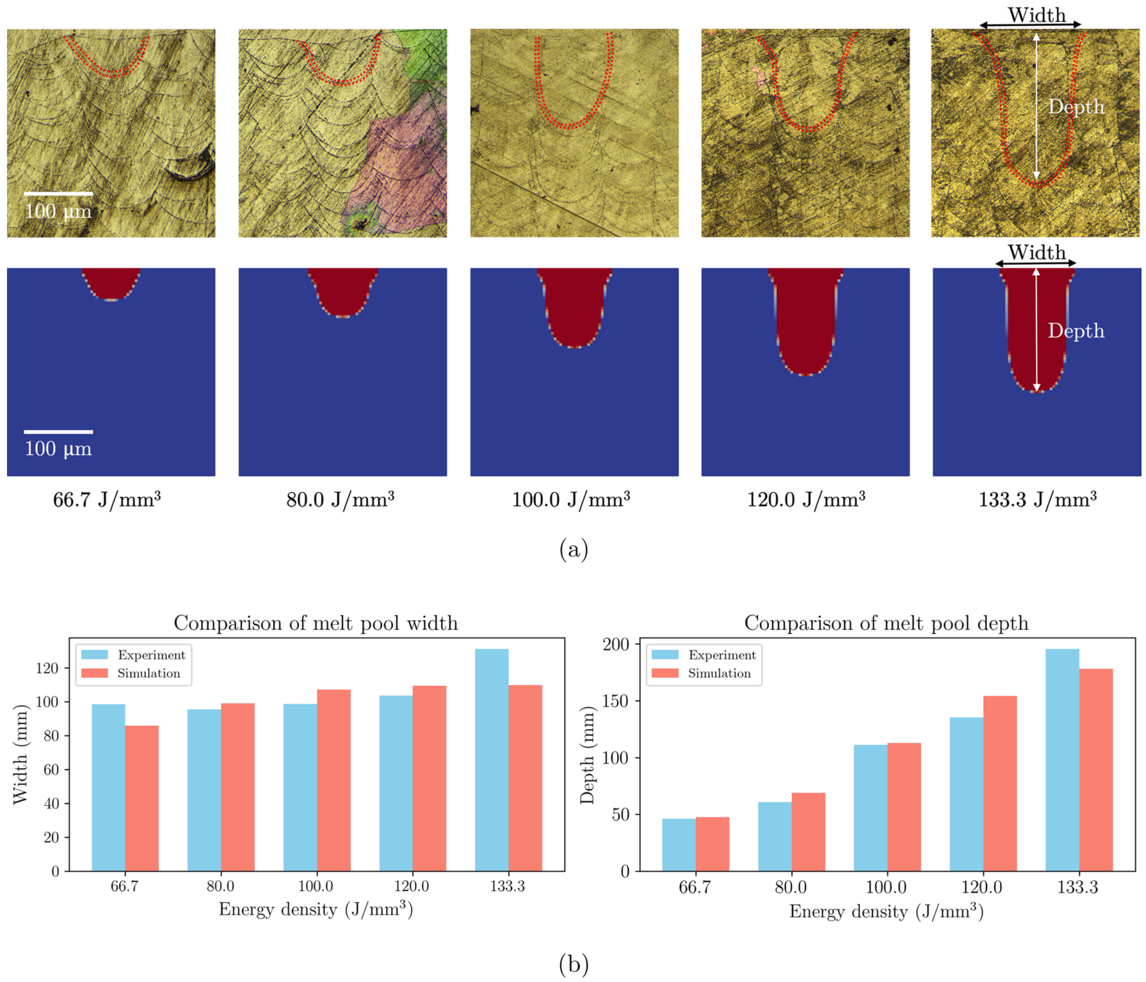


Fig. 3 Calibration of the CFD model: **a** Cross-section images; **b** Dimensions of the melt pools

where L_g is the kinetic coefficient for interfacial mobility at grain boundaries, defined as [17]:

$$L_g = L_0 \exp(-\frac{Q_g}{RT}) \tag{12}$$

where L_0 is a pre-exponential coefficient, Q_g is the activation energy, and R is the gas constant. The solid and liquid phases are represented by a parameter ζ, which is set to 1 in the solid phase, 0 in the liquid phase, and varies smoothly between 0 and 1 at the liquid–solid (L/S) interface, given as:

$$\zeta = \frac{1}{2}\left\{1 - \tanh\left[\vartheta \times (\frac{T}{T_l} - 1)\right]\right\} \tag{13}$$

where ϑ is a constant that drives ζ toward 0 in the liquid phase and 1 in the solid phase. In addition, the total free energy functional F in Eq. (11) is defined as:

$$F = \int_V (f_{\text{grain}} + f_{\text{grad}})dV \tag{14}$$

where f_{grain} and f_{grad} are the grain energy and gradient energy, respectively. The expressions of these energies are given below:

$$\begin{aligned} f_{\text{grain}} &= m_g \left[\sum_{i=1}^{n}\left(\frac{(\eta_i)^4}{4} - \frac{(\eta_i)^2}{2}\right) + \gamma \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} (\eta_i)^2 (\eta_j)^2 + \frac{1}{4} + (1-\zeta)^2 \sum_{i=1}^{n} (\eta_i)^2\right] \\ f_{\text{grad}} &= \frac{\kappa_g}{2} (\nabla \eta_i)^2 \end{aligned} \tag{15}$$

where m_g is the grain energy pre-coefficient, γ is the model parameter [44], and κ_g is the gradient term coefficient for the grain boundary. The grain energy f_{grain} reaches its minimum when all the grains are ordered in the system, which means $[\eta_1, \eta_2, \ldots, \eta_n] = [1, 0, \ldots, 0], [0, 1, \ldots, 0], \ldots, [0, 0, \ldots, 1]$. The grain texture introduced in Ref. [15] is not considered in this study, as our primary focus is on grain sizes and morphologies, while microstructural textures are excluded for the purpose of data augmentation. Moreover, Choi et al. [17] have shown that texture has a negligible effect on the final simulated grain structures. Equation (11) is treated with explicit Euler and finite difference methods for solving, and Dirichlet boundary condition is applied. The initial grain orientations are generated with the open-source software `Neper` [45].

The classical nucleation model [46, 47] is incorporated into the PF model to achieve the most faithful possible replication of grain nucleation and growth during molten metal solidification. The nucleation probability is implemented by the Poisson seed method [48] and given as:

$$P_n = 1 - \exp\left(-J\Delta t_p\left(\Delta x_p\right)^3\right) \tag{16}$$

where Δt_p and Δx_p denote the the PF time step and mesh size, respectively. J is the heterogeneous nucleation rate, which is given based on the nucleation theory considering the undercooling and elemental diffusion:

$$\begin{aligned} J &= A\exp\left(-\frac{\Delta G_V^{\mathrm{hom}} f(\theta)}{kT}\right) \\ A &= \frac{k_B T N_{\mathrm{atom}}}{h}\exp\left(-\frac{Q_g}{RT}\right) \\ f(\theta) &= 0.25\times(2+\cos\theta)\times(1-\cos\theta)^2 \\ \Delta G_V^{\mathrm{hom}} &= \frac{16\pi(\sigma_p)^3}{3(\Delta g_v)^2} \end{aligned} \tag{17}$$

where $\Delta G_V^{\mathrm{hom}}$ is the homogeneous nucleation barrier for triggering nucleus formation, k_B is the Boltzmann constant, h is the Planck's constant, N_{atom} is the number of atoms in the L/S interface, θ is the contact angle, and σ_p is the L/S interfacial energy. The difference between the solid and liquid Gibbs free energies is represented by Δg_v, which is approximately given as $\Delta g_v \approx L\times\frac{\Delta T}{T_l}$, where ΔT is the undercooling expressed as $\Delta T = T_l - T$. The nucleation rate reaches its maximum at a specific undercooling, which is set to 37 K, and nucleation rate at higher undercooling levels is not considered [15]. When the undercooling in the mushy zone reaches the appropriate level, a nucleation probability $P_n \in (0,1)$ is computed using Eq. (16). Simultaneously, a random number between 0 and 1 is generated. A nucleus is added to the corresponding mesh cell only if P_n exceeds the random number. The effect of incorporating the nucleation model is demonstrated through a single-layer, single-track PF simulation, as illustrated in Fig. 4.

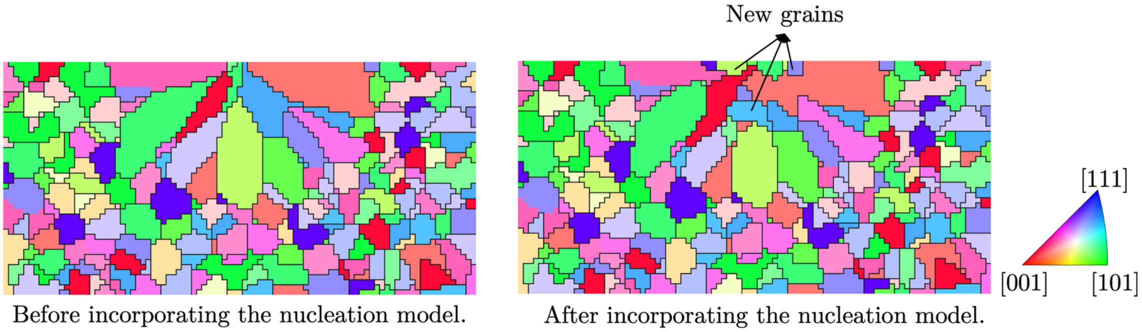
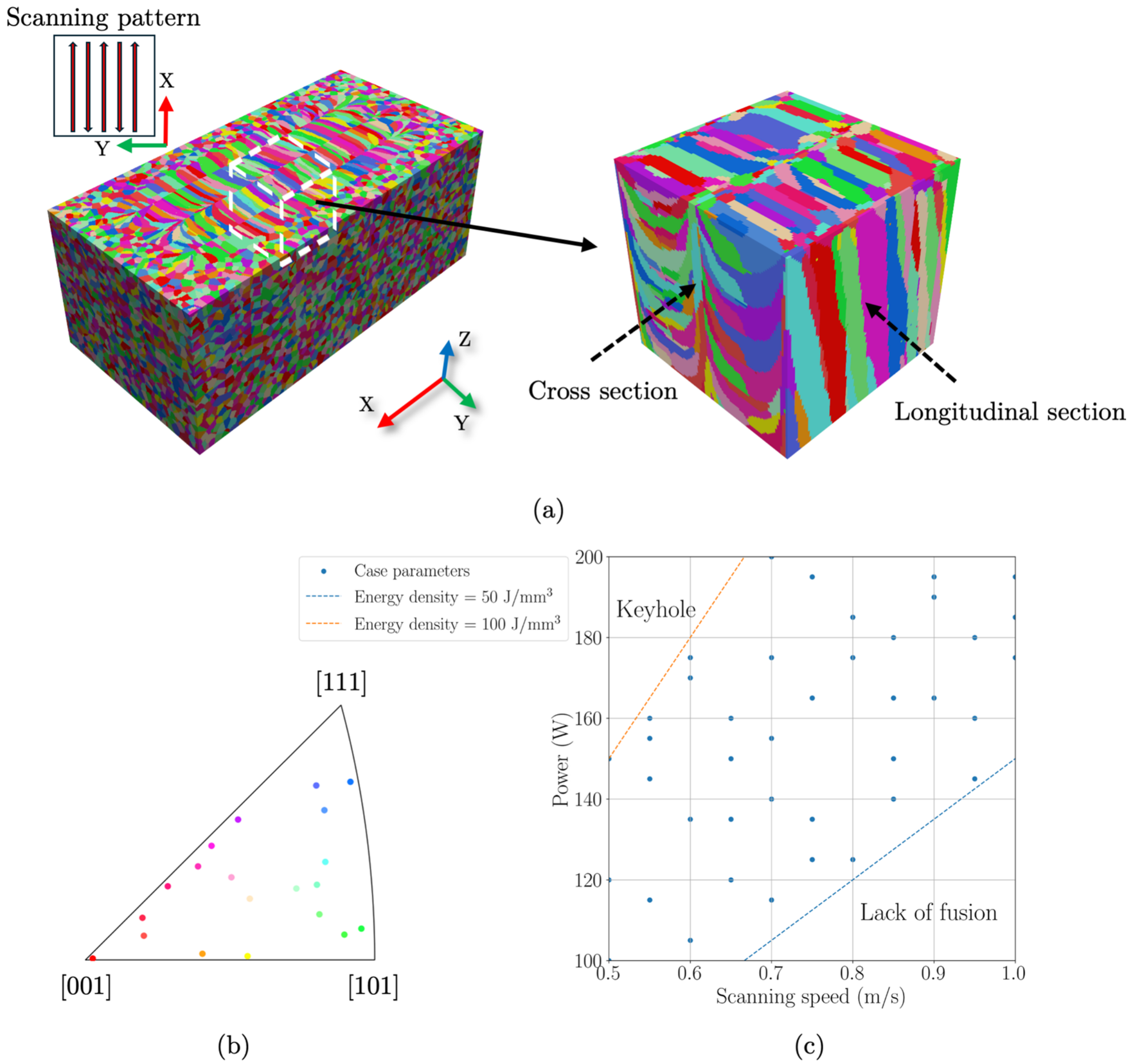


Fig. 4 Comparison between simulations with and without the nucleation model

2.1.3 Simulation cases and computation

The simulation substrate is modeled as a rectangular prism with dimensions of $1\times0.5\times0.16\ \mathrm{mm}^3$, within which approximately 30,000 grains are randomly distributed. The number of grains in the bulk is determined by calibrating the initial grain size based on our EBSD results. For computational efficiency, the time step and mesh size of the CFD model are twice as large as those used in the PF model. After computing the temperature field using the CFD solver, the results are linearly interpolated to match the time step and mesh resolution of the PF model. In this study, we consider two scenarios of microstructure prediction: (1) cross-section and (2) longitudinal-section microstructures. For clarity, we refer to the plane perpendicular to the X-axis as the cross-section, and the plane perpendicular to the Y-axis as the longitudinal section, as illustrated in Fig. 5a.

To generate simulated cross-section microstructures, we perform 8-layer, 3-track simulations. Each layer measures $1\times0.5\times0.03\ \mathrm{mm}^3$ and contains three scanning tracks, each with a length of $0.8\ \mathrm{mm}$. A reciprocating scanning strategy along the X-axis is adopted throughout the process. After the current layer is scanned, a new mesh layer is added to the top of the simulation domain, while the bottom layer of equal thickness is removed for acceleration. Since the powder-spreading process in the AM experiments usually takes at least 5 s, while the material cools to ambient temperature within milliseconds [49], we reset the temperature profile after each new layer is added. As shown in Fig. 5a, the final simulation result of the coupled CFD–PF model reveals grain nucleation, growth, re-melting, and resolidification processes occurring in the deposited layers and tracks. To visualize grain orientations, the inverse pole figure (IPF) coloring scheme is employed. In this study, 20 grain orientations are



(a)

(b)

(c)

Fig. 5 **a** An example of simulated microstructures; **b** 20 grain orientations and corresponding IPF colors; **c** Process window and sampled manufacturing parameters for 8-layer, 3-track simulations

randomly generated using a spherical sampling method. The corresponding Euler angles and IPF colors are computed using an open-source Python package `orix` [50], as presented in Table 1 and Fig. 5b, respectively.

The material properties of 316 L stainless steel and related modeling parameters are summarized in Table 2. Both the CFD and PF models are implemented using `JAX` [51] to enable GPU-based high-performance computing. A process window is defined for 316 L stainless steel to avoid keyhole mode and lack of fusion, thereby ensuring manufacturing stability and quality [43, 52, 53]. Additionally, at high parameter values (i.e., high laser power and scanning speed), the balling effect is likely to occur [53], while at low parameter values, poor wetting of the substrate or melt pool instability may arise [54]. Hence, guided by previous studies and empirical knowledge, the laser power and scanning speed are constrained to the ranges of 100–200 W and 0.5–1.0 $\mathrm{m/s}$, respectively. A total of 40 simulation cases are generated by uniformly sampling laser power and scanning speed values within the process window following a Sobol sequence [55]. The corresponding parameter combinations are shown in Fig. 5c. To construct the cross-section simulation dataset, we crop 250 images from each of the 40 simulated cases described earlier. The vertical centerline of each image lies on the XZ-plane and coincides with the middle of the three scanning tracks. The resolution of each image is 100×100 pixels, which corresponds to a physical area of $200 \times 200\ \mu\mathrm{m}^2$. As grain texture is not considered in this work, we apply a color-tuning data augmentation strategy that randomly exchanges the colors of grains to increase the diversity of training samples, with each image color-tuned once. In total, we obtain 20,000 simulated cross-section images.

To validate the adaptability of the proposed sim-to-real diffusion models to microstructures from an independent study

Table 1 Corresponding Euler angles (φ_1, Φ, φ_2) of the generated orientations in degrees

Orientation	φ_1	Φ	φ_2
1	288.268	99.552774	355.989992
2	348.574	63.854704	269.339636
3	112.833	178.774464	100.959837
4	249.236	113.286860	284.140558
5	315.500	134.949901	37.161362
6	322.058	144.619293	161.241669
7	30.616	128.864441	327.094381
8	14.060	107.991659	105.701093
9	61.139	101.915275	103.599122
10	316.131	85.547450	46.810286
11	35.405	99.300415	6.972105
12	151.599	68.257306	244.380792
13	344.840	126.234787	76.186122
14	191.939503	40.866612	95.596797
15	249.075761	160.948369	176.966337
16	113.585627	70.066149	19.210516
17	247.140334	99.519914	206.682338
18	300.465242	83.259101	52.822287
19	6.583780	135.990624	212.149993
20	270.051953	127.142326	251.913010

Table 2 Thermophysical and simulation parameters used in the coupled CFD-PF model, taken from [15, 17, 56]

Property and parameter	Value	Unit
Density (ρ)	8084	$\mathrm{kg\ m^{-3}}$
Viscosity (μ)	6×10^{-3}	Pa s
Latent heat (L)	2.7×10^{5}	$\mathrm{J\ kg^{-1}}$
Specific heat (c_p)	$462 + 0.134T$	$\mathrm{J\ kg^{-1} \cdot K^{-1}}$
Thermal conductivity (k)	$9.248 + 0.0157T$	$\mathrm{W\ m^{-1} \cdot K^{-1}}$
Solidus temperature (T_s)	1650	K
Liquidus temperature (T_l)	1723	K
Surface tension gradient ($\frac{d\gamma}{dT}$)	-4×10^{-4}	$\mathrm{N\ m^{-1} \cdot K^{-1}}$
Activation energy (Q_g)	1.4×10^{5}	$\mathrm{J\quad mol^{-1}}$
Pre-exponential coefficient (L_0)	3.5×10^{3}	$\mathrm{m^4 \cdot J^{-1} \cdot s^{-1}}$
Grain energy coefficient (m_g)	1.2×10^{6}	$\mathrm{J \cdot m^{-3}}$
Gradient coefficient for grain boundary (κ_g)	2.77×10^{-7}	$\mathrm{J \cdot m^{-1}}$
Number of atoms (N_{atom})	2.1×10^{10}	
Contact angle (θ)	4	$\circ$
L/S interface energy (σ_p)	0.385	$\mathrm{J \cdot m^{-2}}$
CFD mesh size (Δx)	4	μm
CFD time step (Δt)	4×10^{-7}	s
PF mesh size (Δx_p)	2	μm
PF time step (Δt_p)	2×10^{-7}	s

of Wang et al. [57], single-layer, single-track simulations are conducted to generate longitudinal-section data consistent with their experiments. We simulate 10 single-layer, single-track cases and extract 250 longitudinal-section images from each. The corresponding manufacturing parameters do not fully comply with our proposed process window, as the corresponding experiments are conducted externally. Table 3 lists the manufacturing parameters for these simulations. Each cropped image has its top edge aligned with the upper surface of the substrate. The resolution of each image is 64×64 pixels, corresponding to a physical area of 128×128 μm^2. To enhance data diversity, the same color-tuning augmentation scheme is applied. In total, 5,000 simulated longitudinal-section images are prepared. It is worth noting that these single-layer, single-track simulations are performed using the same settings and parameters as those used in the multi-layer cases.

Table 3 Manufacturing parameters of single-layer, single-track simulations

No.	1	2	3	4	5	6	7	8	9	10
Power (W)	100	120	160	140	180	150	150	200	200	200
Speed (m/s)	0.5	0.25	0.5	0.75	0.25	0.4	0.6	0.25	0.5	0.75

Table 4 Chemical composition of the 316 L stainless steel powder

Element	Cr	Ni	Mo	Mn	Si	Fe	Others
Content (wt%)	17.03	10.74	2.52	0.83	0.39	Bal	<0.1

Nevertheless, our proposed diffusion models demonstrate strong adaptive capability by effectively generating independent microstructures from the metal AM community.

The mesh size and computation time for the simulations are reported here. The 8-layer, 3-track simulation has approximately 10 million mesh cells used in the PF model, and 1.25 million in the CFD model. The total number of time steps, denoted as n_t, can be derived using the following equation:

$$n_t = \frac{\text{scanning distance}}{\text{scanning speed} \times \text{time step}} \tag{18}$$

For example, in a simulation with a scanning distance of 19.2 mm, a scanning speed of 0.8 m/s, and a time step of 2×10^{-7} s, the total number of time steps is calculated to be 120,000. Simulating an 8-layer, 3-track process requires 6 to 12 h on an NVIDIA L20 GPU, with the exact duration varying according to the laser scanning speed.

2.2 EBSD experiments

To obtain EBSD images, we conduct a series of in-house experiments. In addition, we consider external experiment data from an independent study [57] to demonstrate the adaptability of the proposed sim-to-real diffusion models.

A detailed description of our own experiments is presented below. The material is gas-atomized 316 L stainless steel powder supplied by Avimetal Powder Metallurgy Technology. The powder particles are nearly spherical, with a particle size distribution ranging from 15 to 53 μm. The detailed chemical composition of the powder is provided in Table 4. The AM system employed for the experiments is an AMC-M150B printer (Guangzhou Shinengine AM Technologies Co., Ltd.). The system is equipped with a fiber laser capable of delivering up to 500 W of power and features a laser beam spot size of 40–50 μm for melting and fusing the powder material.

Cross-section experimental microstructures are obtained through EBSD characterization conducted in-house. A total of ten specimens are fabricated using the same reciprocating scanning strategy as in the simulation. After electrical wire cutting and grinding, cubic samples from the specimens are further precisely polished using an ion milling instrument to obtain a strain-free surface, as shown in Fig. 6a and b. The ion milling parameters are as follows: acceleration voltage (Acc.V) of 6 kV, discharge voltage (Dis.V) of 1.5 kV, inclination angle of 20° and milling time of 30 min. EBSD characterizations are then performed using a Hitachi SU5000 field-emission SEM (Hitachi Consulting Co., Ltd), equipped with an Oxford Symmetry CMOS EBSD detector. Finally, due to surface damage during the polishing process, only eight samples are successfully characterized. The manufacturing parameters of these eight samples are listed in Table 5. Analysis of grain orientation is conducted along the building direction with the `AZtecCrystal` software. Figure 6c shows a grain orientation map colored with IPF scheme. The final images used for training are randomly cropped from various tracks of the original orientation maps in order to cover as many different grain morphologies that may arise during printing as possible. The region enclosed by the white square in Fig. 6c serves as an example. Each cropped image corresponds to a physical area of $200 \times 200\ \mu\text{m}^2$ and precisely covers the width of a single scanning track. A total of approximately 4,000 images are collected as the experimental cross-section dataset.

We also use the three longitudinal-section experimental microstructures reported in the study by Wang et al. [57] as part of our training dataset to demonstrate the adaptability of the proposed sim-to-real diffusion models to arbitrary microstructural data within the AM community. The manufacturing parameters of these three samples are also listed in Table 5. Cropped and processed experimental images based on their EBSD results are presented in Fig. 17. Each cropped image, aligned at the top edge of the original EBSD map, covers an area of $128 \times 128\ \mu\text{m}^2$. A total of 1,500 images are prepared as the experimental longitudinal-section dataset.

EBSD images of other alloys may also be used as experimental data for model refinement. However, to the best of our knowledge, no other alloy has been printed using the same scanning pattern as in our work and has had high-resolution cross-section or longitudinal-section EBSD images reported in the literature. As a result, we are unable to further validate our models across materials. Nevertheless, the 316 L stainless steel studied in Ref. [57] differs significantly from our in-house experiments in terms of chemical composition and powder size. The experimental setups, including the printing

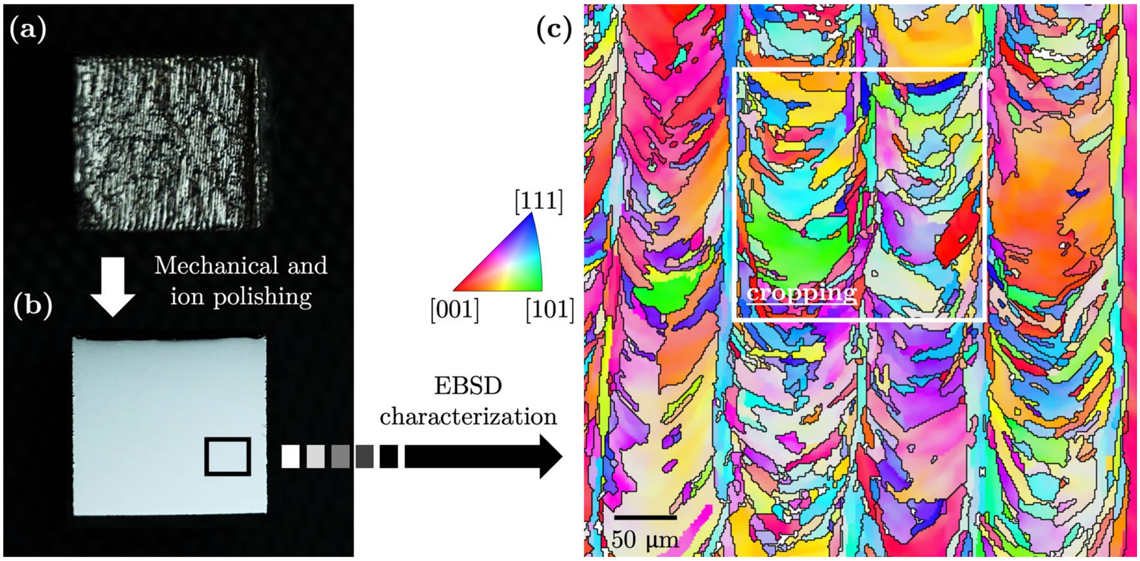


Fig. 6 Illustration of EBSD characterization: (a) Wire-cut sample; (b) Polished sample; (c) Grain orientation map with IPF colors

Table 5 Manufacturing parameters of experiments

Our experiments								
No.	1	2	3	4	5	6	7	8
Power (W)	100	120	120	150	150	180	180	200
Speed (m/s)	0.5	0.5	0.7	0.5	0.9	0.7	0.9	0.7
Experiments taken from Ref. [57]								
No.	1	2	3					
Power (W)	200	200	200					
Speed (m/s)	0.25	0.5	0.75					

machine and manufacturing parameters, also vary considerably. The promising results on this dataset still demonstrate the adaptability of our proposed models to some extent.

3 Sim-to-real diffusion models

Effectively leveraging both high-fidelity but expensive experiment data and massive but relatively low-fidelity simulation data remains challenging. In this work, we propose sim-to-real diffusion models to generate realistic microstructures of metal AM products in a cost-effective way. The sim-to-real microstructure generation framework is built upon a conditional DDPM, which is pre-trained with a large amount of simulation data. To further refine the pre-trained model, we propose two parallel strategies: (1) fine-tuning and (2) distillation. A relatively small amount of experiment data is utilized during the refinement process. To quantitatively evaluate the quality of the microstructural images generated by the proposed sim-to-real approach, both machine learning and physics-based metrics are considered.

3.1 Conditional DDPM

Owing to the superior sample quality and stable training dynamics, DDPM is adopted in this study to achieve the microstructure generation. As the objective is to generate microstructures under different manufacturing conditions, the parameters of power and scanning speed are embedded into the model as conditional inputs. According to Ho et al. [22], a complete DDPM has two processes: the forward (diffusion) process and the reverse (denoising) process. The forward process is condition-independent, where Gaussian noise is progressively added to the original image x_0 at each diffusion step *t*, forming a Markov chain:

$$\begin{aligned} q(\mathrm{x}_{1:T} \mid \mathrm{x}_0) &:= \prod_{t=1}^{T} q(\mathrm{x}_t \mid \mathrm{x}_{t-1}), \\ q(\mathrm{x}_t \mid \mathrm{x}_{t-1}) &:= \mathcal{N}\left(\mathrm{x}_t; \sqrt{1-\beta_t}\,\mathrm{x}_{t-1}, \beta_t \mathrm{I}\right) \end{aligned} \tag{19}$$

where β_t is a constant variance, and I is the identity matrix. Using the notation $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$, we have:

$$q(\mathrm{x}_t \mid \mathrm{x}_0) = \mathcal{N}\left(\mathrm{x}_t;\ \sqrt{\bar{\alpha}_t}\,\mathrm{x}_0,\ (1-\bar{\alpha}_t)\mathrm{I}\right) \tag{20}$$

which can be further written as:

$$\mathrm{x}_t(\mathrm{x}_0, \boldsymbol{\epsilon}) = \sqrt{\bar{\alpha}_t}\,\mathrm{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon} \tag{21}$$

where the noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathrm{I})$. When conditioned on x_0, the forward process posteriors can be denoted as:

$$q(\mathrm{x}_{t-1} \mid \mathrm{x}_t, \mathrm{x}_0) = \mathcal{N}\left(\mathrm{x}_{t-1};\ \tilde{\boldsymbol{\mu}}_t(\mathrm{x}_t, \mathrm{x}_0),\ \tilde{\beta}_t\mathrm{I}\right) \tag{22}$$

$$\tilde{\boldsymbol{\mu}}_t(\mathrm{x}_t, \mathrm{x}_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathrm{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathrm{x}_t \tag{23}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{24}$$

The conditional reverse process $p_\theta(\mathrm{x}_{0:T} \mid \mathrm{y})$ is also a Markov chain, starting at the final noisy image $p(\mathrm{x}_T) = \mathcal{N}(\mathrm{x}_T;\ 0,\ \mathrm{I})$, which is defined as:

$$p_\theta(\mathrm{x}_{0:T} \mid \mathrm{y}) := p(\mathrm{x}_T)\prod_{t=1}^{T} p_\theta(\mathrm{x}_{t-1} \mid \mathrm{x}_t, \mathrm{y}) \tag{25}$$

$$p_\theta(\mathrm{x}_{t-1} \mid \mathrm{x}_t, \mathrm{y}) := \mathcal{N}\left(\mathrm{x}_{t-1};\ \boldsymbol{\mu}_\theta(\mathrm{x}_t, t, \mathrm{y}),\ \boldsymbol{\Sigma}_\theta(\mathrm{x}_t, t, \mathrm{y})\right) \tag{26}$$

where $\boldsymbol{\Sigma}_\theta(\mathrm{x}_t, t, \mathrm{y})$ is set to $\sigma_t^2\mathrm{I}$, and σ_t^2 is set to β_t, following the implementation in Ref. [22]. Reducing the Kullback–Leibler (KL) divergence between the parameterized probability distribution and the probability distribution of actual images as much as possible is essential in accordance with the core goal of generative models. The Evidence Lower Bound (ELBO), which is utilized in VAE [18], is extended into a chain representation for the diffusion model, which can be expressed as:

$$\mathcal{L} = \mathbb{E}_q\left[\underbrace{D_{\mathrm{KL}}\left(q(\mathrm{x}_T \mid \mathrm{x}_0) \,\|\, p(\mathrm{x}_T)\right)}_{L_T} + \sum_{t>1}\underbrace{D_{\mathrm{KL}}\left(q(\mathrm{x}_{t-1} \mid \mathrm{x}_t, \mathrm{x}_0) \,\|\, p_\theta(\mathrm{x}_{t-1} \mid \mathrm{x}_t, \mathrm{y})\right)}_{L_{t-1}} - \underbrace{\log p_\theta(\mathrm{x}_0 \mid \mathrm{x}_1, \mathrm{y})}_{L_0}\right] \tag{27}$$

Given that the term L_T is constant and L_0 is computable from Eq. (26), it is reasonable to directly compare Eq. (22) with Eq. (26) in the term L_{t-1}, leading to the following parameterization:

$$\boldsymbol{\mu}_\theta(\mathrm{x}_t, t, \mathrm{y}) = \tilde{\boldsymbol{\mu}}_t\left(\mathrm{x}_t,\ \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathrm{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_\theta(\mathrm{x}_t, t, \mathrm{y})\right)\right) = \frac{1}{\sqrt{\alpha_t}}\left(\mathrm{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta(\mathrm{x}_t, t, \mathrm{y})\right) \tag{28}$$

The final simplified loss function is given as:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t,\mathrm{x}_0,\boldsymbol{\epsilon},\mathrm{y}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\,\mathrm{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon},\ t, \mathrm{y}\right)\right\|^2\right] \tag{29}$$

where $\boldsymbol{\epsilon}_\theta$ represents the predicted noise, which is learned by a neural network. A U-Net [58] architecture is adopted for this purpose. The architecture of this U-Net adopts a symmetric encoder-decoder structure with 5 downsampling blocks and 5 upsampling blocks. The channel configuration is set to (32, 64, 128, 256, 512), progressively increasing feature dimensions to capture multi-scale information. Each resolution block contains 2 residual layers, ensuring sufficient feature extraction capacity. To balance computational efficiency and expressive power, attention layers are deployed only at the 2nd and 4th resolution levels (corresponding to 64 and 256 channels). The attention head dimension defaults to 8, with group normalization using 32 groups [59]. The conditional input is first mapped to a dimensional feature space via a random Fourier features (RFF) embedder [60], then concatenated with timestep embedded by a sinusoidal positional embedder [59] to form a 128 dimensional joint embedding vector. This serves as a global conditioning signal injected into various levels of the U-Net.

Given an initial Gaussian noise sample, a realistic microstructural image can be progressively reconstructed using the noise predicted by the U-Net. Specifically, conditioned on the manufacturing parameters and the diffusion step, the model estimates the noise that was added during the forward diffusion process, and iteratively removes it from the original Gaussian noise sample. Through this reverse denoising process, the noisy sample is gradually transformed into a microstructure that is consistent with the specified conditions. This reverse process is implemented using an implicit scheme [23], which enables deterministic sampling instead of the stochastic noise injection in DDPM, making it more suitable for high-fidelity microstructure generation. Furthermore, the implicit approach also leads to a significantly accelerated sampling process. The corresponding sampling formulation is given as:

$$\begin{aligned}\mathbf{x}_{t-1} &= \sqrt{\bar{\alpha}_{t-1}} \cdot \tfrac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\varepsilon_\theta(\mathbf{x}_t, t, \mathbf{y})\right) \\ &+ \sqrt{1-\bar{\alpha}_{t-1}}\varepsilon_\theta(\mathbf{x}_t, t, \mathbf{y})\end{aligned} \tag{30}$$

The conditional diffusion model is pretrained exclusively on a large-scale dataset of simulated microstructures (20,000 images, see Sect. 2.1.3), with 15% of the data reserved for validation. The overall workflow and model structure are shown in Fig. 7. The implementation is build on top of an open source library `Diffusers` [59].

3.2 Fine-tuning

Our first proposed strategy to incorporate experiment data to the pre-trained model is fine-tuning. Fine-tuning is a

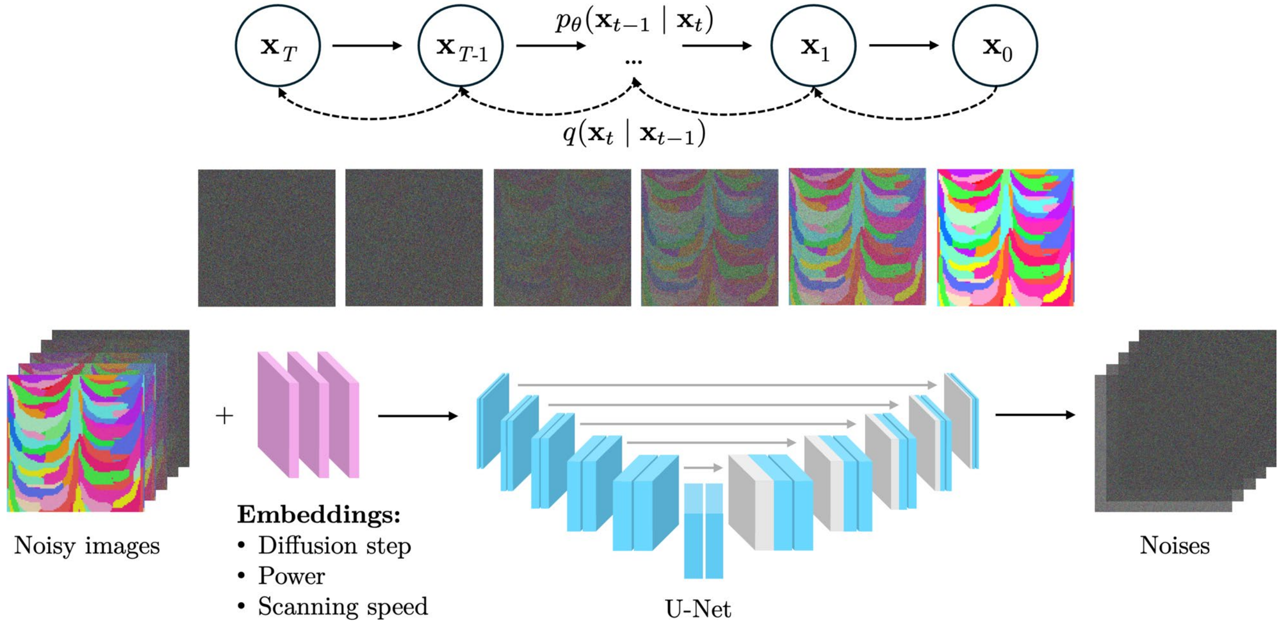
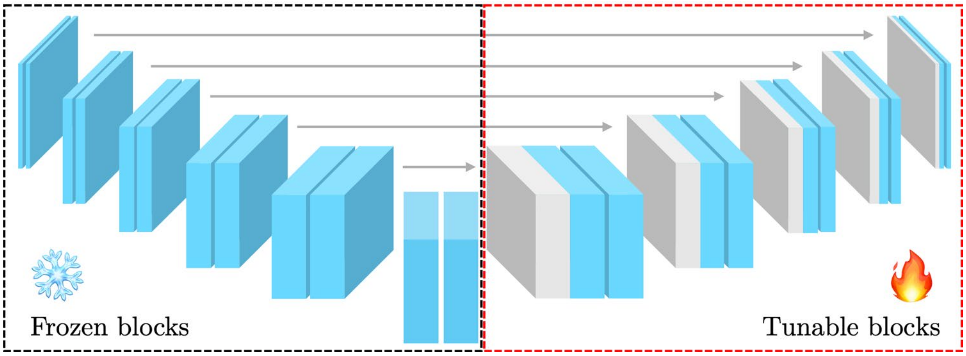


Fig. 7 Conditional DDPM

Fig. 8 Adjusting the U-Net for fine-tuning diffusion model



popular transfer learning strategy that entails refining a previously learned model on a particular target dataset in order to modify its parameters for a new task. Fine-tuning improves performance on specialized tasks with limited training data by utilizing the knowledge gained from pre-training to keep generalizable features while adjusting to the particulars of the target domain [34, 61].

In this study, the pre-trained DDPM is fine-tuned using a relatively small amount of experiment data to enable the generation of realistic microstructures. As illustrated in Fig. 8, we freeze all downsampling and middle blocks of the pre-trained U-Net to preserve the feature extraction capabilities and general representations learned from simulated microstructural images. This strategy helps prevent overfitting and catastrophic forgetting, which can possibly occur when fine-tuning on limited data [62]. During fine-tuning with experimental images, selected blocks in the upsampling path are left tunable to allow the model to adapt to the new data and capture the specific details and styles [58] present in the experimental microstructures. This approach not only accelerates training and reduces computational cost, but also ensures that the model maintains robust generalization while effectively learning realistic grain features. The number of tunable blocks in the upsampling path is task-dependent. Based on the results from models tuned with different number of tunable upsampling blocks, we set all upsampling blocks as tunable for the cross-section fine-tuning task, whereas only the last two upsampling blocks as tunable for the longitudinal-section fine-tuning task.

3.3 Distillation

Our second strategy adopted is distillation (also known as knowledge distillation). Distillation was first proposed by Hinton et al. [35] as a technique to compress a high-capacity teacher model into a smaller and more efficient student model. Based on the concept of distillation, we design a "teacher-student-reality" training framework and demonstrate its effectiveness for the sim-to-real microstructure generation task. The architecture of our sim-to-real distillation framework is illustrated in Fig. 9.

The teacher model is the pre-trained DDPM, whose parameters remain fixed during the distillation process. The noise predicted by the teacher model is referred to as teacher noise, denoted as ϵ_T. The student model, initialized from the teacher model with trainable parameters, is trained on experiment data. The noise predicted by the student model is denoted as ϵ_S. A third component, real noise ϵ, represents the actual Gaussian noise added to the clean image during the forward diffusion process. The goal of the student model is to accurately predict the real noise under the guidance of the teacher model, which possesses prior knowledge learned from simulation data. To this end, the overall loss function is formulated as the weighted sum of two mean squared error (MSE) terms:

$$\mathcal{L}_{\text{distill}} = \lambda_1 \left\|\boldsymbol{\epsilon}_S - \boldsymbol{\epsilon}\right\|^2 + \lambda_2 \left\|\boldsymbol{\epsilon}_S - \boldsymbol{\epsilon}_T\right\|^2 \tag{31}$$

where λ_1 and λ_2 are two coefficients introduced to balance the two objectives. A common setting for hyperparameters λ_1 and λ_2 is to keep $\lambda_1 + \lambda_2 = 1$. As we aim to generate high-fidelity realistic microstructures, we need choose to use a larger λ_1 in this work, which controls how much the model can learn from experiment data. Based on the training results with different hyperparamter settings, we empirically set λ_1 and λ_2 to 0.8 and 0.2, respectively.

For both the two refining strategies, there are 4,000 experimental images for the cross-section refining task and 1,500 for the longitudinal-section refining task, as mentioned in Sect. 2.2. Still, 15% of the data reserved for validation. It is worth noting that, since the pre-trained model has already learned the simulated microstructure distributions across a wide range of manufacturing conditions within the process window, the fine-tuned or distilled model may retain the ability to generate realistic microstructures even when the corresponding experimental conditions are not included in the sim-to-real training. To evaluate this generalizability under unseen conditions, we set aside a subset of experimental cases for testing. Specifically, for the cross-section refining task, cases (150 W, 0.5 m/s) and (200 W, 0.7 m/s) are excluded from training and used solely for testing. For the longitudinal-section refining task, case (150 W, 0.5 m/s) is reserved for testing.

3.4 Evaluation metrics

To quantitatively evaluate the authenticity of the generated microstructures, we employ both machine learning and physics-based metrics for analysis. Before analysis, generated grains with a real width or length smaller than 4 μm, which corresponds to the size of two mesh cells in our simulation, are considered as noisy pixels and removed to minimize the influence of noise. The total area of the removed grains accounts for less than 10% of the entire image area. In addition, to facilitate grain counting (i.e., ensuring that each grain has a distinct and uniform color), we fix the grain colors to the 20 colors shown in Fig. 5b, and the color of each grain is mapped to the closest color among these 20 predefined colors. These image processing techniques will not have significant influence on the generated results, because the overall grain size and aspect ratio are kept.

Regarding the machine learning metric, we adopt the FID score [36], a standard evaluation metric for generative models. FID compares the distribution of generated images to that of real images in the feature space of a pre-trained `Inception-v3` neural network [63, 64]. Specifically, FID measures the difference in means and covariances of these distributions, and is defined as:

$$\text{FID} = \left\|\mu_r - \mu_g\right\|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right) \tag{32}$$

where μ_r and Σ_r denote the mean and covariance of the real images, and μ_g and Σ_g represent the mean and covariance of the generated images. A lower FID score suggests a smaller statistical distance between the generated and real image distributions, which reflects better generative quality. To evaluate the FID scores of the generated microstructures, each well-trained diffusion model is used to generate 100 images under the testing manufacturing conditions. The FID scores are then computed by comparing these generated images with the corresponding experimental images under the same conditions.

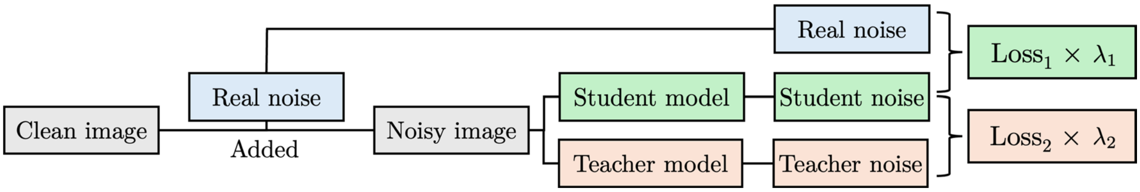
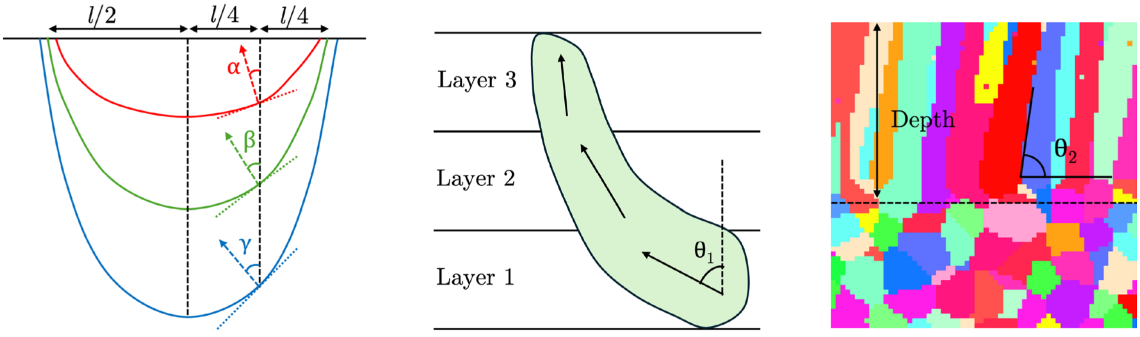


Fig. 9 Architecture of the sim-to-real distillation diffusion model



(a) (b) (c)

Fig. 10 Physics-based evaluation metrics: **a**, **b** Grain inclination angle in cross section; **c** Grain inclination angle and melt pool depth in longitudinal section

Regarding physics-based metrics, we consider grain size and morphology, as they are important features closely associated with the mechanical properties of the as-built samples. The first physics-based metric considered is the grain inclination angle. Previous studies have shown that the initial grain growth direction tends to be perpendicular to the melt pool boundary at the onset of solidification [15, 57]. To standardize the measurement of grain inclination in the cross section, a vertical reference line is drawn at one-quarter of the melt pool width. As the melt pool moves upward during multi-layer scanning, the angle between the grain growth direction and the vertical reference line gradually decreases, as illustrated in Fig. 10a. The initial growth angle in the first layer, denoted as θ_1, is selected as the target metric, as shown in Fig. 10b. In the longitudinal section, the grain inclination angle is defined as the angle between the grain growth direction and the horizontal reference line, denoted as θ_2 in Fig. 10c. Additionally, since the longitudinal-section case is single-layer, single-track scanning, the melt pool depth is also considered as a physics-based metric specific to this configuration.

The second physics-based metric considered is grain size, which is further characterized by three sub-metrics: grain length, grain width, and grain area. To quantify grain length and width, we adopt the linear grain boundary intercept method, as proposed in the AM Bench 2022 [65]. In this approach, horizontal and vertical lines are overlaid on the microstructural images. The number of pixels between the intersection points of these lines and the grain boundaries is counted, representing the grain length and width, respectively. The grain area is defined as the total number of pixels enclosed within each grain region. As illustrated in Fig. 11a, two representative horizontal and vertical lines are shown in white, with their intercepts on grain boundaries marked by white crosses. For statistical analysis, a total of 100 horizontal and 100 vertical lines are used.

The third physics-based metric is the grain aspect ratio, which characterizes the elongation of individual grains. To estimate this quantity, an ellipse is fitted to the contour of each grain with principle component analysis (PCA) method [66], as illustrated in Fig. 11b. The aspect ratio is then approximated by the ratio of the lengths of the major and minor axes of the fitted ellipse. The distributions of grain size and aspect ratio are quantified by counting their occurrences within predefined bins and are visualized as histograms, as shown in Sect. 4.

4 Results and Discussion

The results and analysis of the generated microstructural images in both the cross section (our in-house experiments) and longitudinal section (from literature) are presented in this section. A general analysis based on visual inspection is first provided. Quantitative evaluations are conducted using the metrics including the FID score, grain inclination angle, melt pool depth (applied only to the longitudinal section), grain size and aspect ratio. We also discuss the influence of sim-to-real data proportion and distillation hyperparameters on the model performance.

4.1 Cross-section prediction

Figure 12 shows the cross-section microstructures from experiment, simulation and three diffusion models trained with different strategies. The baseline diffusion model is defined as a model directly trained with experiment data, and no sim-to-real transition is incorporated. The details of fine-tuning and distillation can be found in Sect. 3.2 and 3.3. Due to the generative nature of diffusion models, the images generated by a single diffusion model may exhibit some variation, even under identical input conditions. However, our diffusion models are conditioned with the

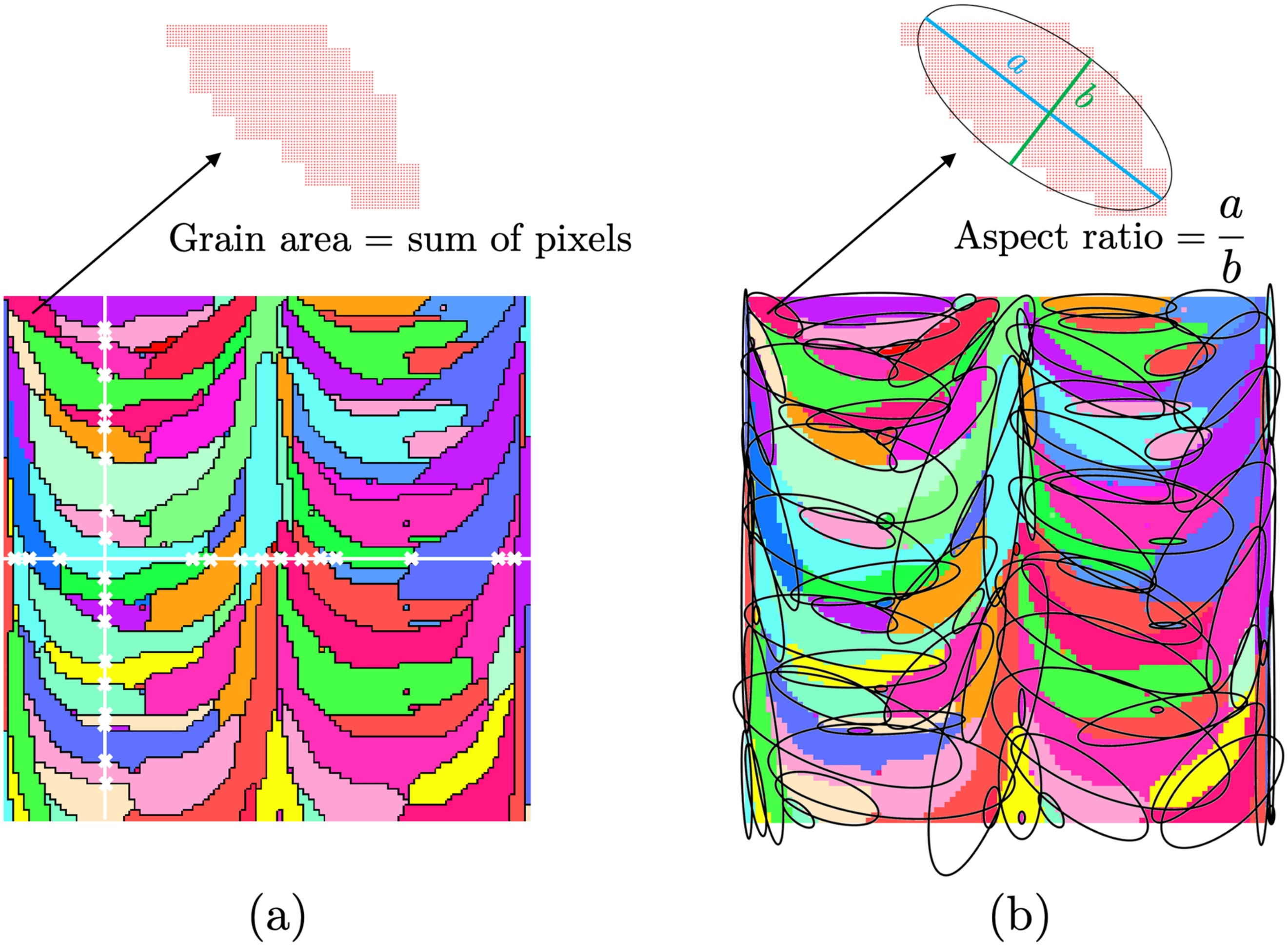


Fig. 11 Physics-based evaluation metrics: **a** Grain size; **b** Grain aspect ratio

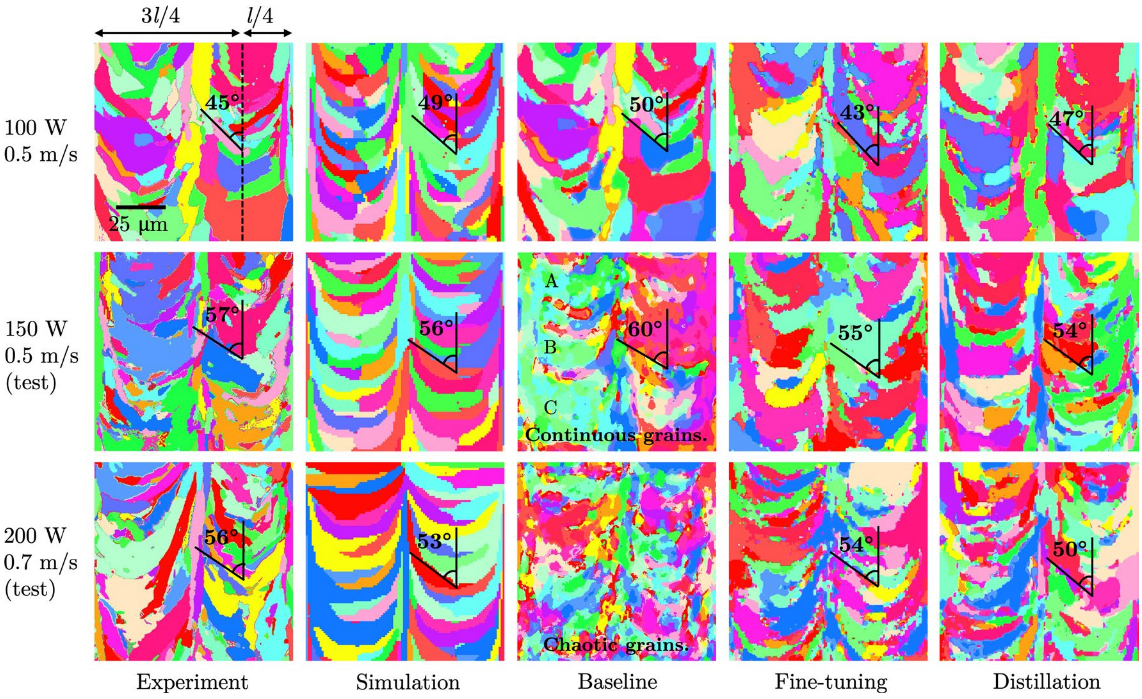


Fig. 12 Cross-section microstructures obtained from experiment, simulation, and diffusion-based models (baseline, fine-tuning, and distillation)

manufacturing conditions, so the generated images should have similar characteristic. In this work, we present only one representative image for each model under a specific set of conditions. Nevertheless, the selected image adequately reflects the overall generation quality. To clearly indicate the location of each image in the figure, we use a coordinate format of [Method, (Power, Scanning speed)] to describe the position of the corresponding generated image.

4.1.1 General analysis

An overview of the microstructures shown in Fig. 12 is first provided. Compared with the experimental grains shown in the first column, the simulated grains exhibit a similar growth trend, primarily oriented upwards along the central axis and both sides. However, the simulated grains appear clearer and more regular, with fewer nucleated grains accumulating along the central axis. This discrepancy can be attributed to the fact that the thermal history and nucleation phenomena in the experiments are more complicated and less controlled [57].

The generated images from diffusion models with fine-tuning and distillation demonstrate organized microstructures under all the conditions presented. In contrast, although the baseline diffusion model successfully predict the microstructure under seen condition, it fails to generate realistic grain patterns under testing conditions. For example, in the image [Baseline, (150 W, 0.5 m/s)], grains A, B, and C are fused into a single grain, which deviates significantly from the experimental microstructure. Under another unseen condition (200 W, 0.7 m/s), the generated grains appear chaotic, with no discernible growth trend. These failures are mainly because the baseline diffusion model has no memory of the real microstructures under the testing conditions. As a result, it lacks prior knowledge of how the microstructures should appear in such scenarios and fails to produce reliable predictions. This poor performance is primarily attributed to the limited generalizability of diffusion models trained on small experiment data.

In fact, improving the generalization ability of diffusion models remains a challenging problem in the field of computer vision. Recent studies [67–70] have demonstrated that diffusion models tend to reconstruct images based on memorized patterns from the training data, rather than generating novel content. This issue becomes even more pronounced when the training dataset is limited or when multimodal learning is involved.

Interesting phenomena regarding microstructure distribution can be observed. The image [Baseline, (100 W, 0.5 m/s)] exhibits a microstructure highly similar to the experimental result under the same manufacturing conditions, apart from some minor differences in position and color. This replication behavior serves as evidence of the limited generalization ability of the baseline diffusion model. After fine-tuning or distillation, the generalization performance of our diffusion models improves significantly. Additionally, the two images generated under unseen conditions by the distilled diffusion model display distinct pattern styles. The image [Distillation, (150 W, 0.5 m/s)] resembles the simulation style, whereas the image [Distillation, (200 W, 0.7 m/s)] appears more experiment-like. This stylistic variation may result from the value of two coefficients included in Eq. (31).

4.1.2 FID score

The FID scores of the generated microstructures from the baseline, fine-tuning, and distillation diffusion models are compared in Table 6. Only the images generated under the testing manufacturing conditions, along with their corresponding experimental images, are used for the FID calculation, specifically for the cases of (150 W, 0.5 m/s) and (200 W, 0.7 m/s). As shown in the table, the FID score of the baseline model exceeds 230. However, after applying fine-tuning and distillation, the score decreases significantly. This improvement can be attributed to the fact that, during the fine-tuning and distillation processes, the pre-trained model has learned the microstructure distributions from the simulation data. Although the simulation data is of lower fidelity compared to the experiment data, it still provides the model with basic grain characteristics under the given testing conditions. Subsequent fine-tuning and distillation with experiment data enable the model to further capture real microstructural features, thereby enhancing its ability to generate more accurate predictions of unseen microstructures.

Another point worth mentioning is the magnitude of the FID scores. It seems even the fine-tuned and distilled models do not achieve very low FID values. We speculate that there may be two main reasons for this. First, the commonly used `Inception-v3` network [63, 64], which is employed to extract features in the latent space, is pre-trained on the large-scale natural image dataset `ImageNet` [71]. In contrast, our dataset consists of microstructural images, which may not be fully compatible with the feature representations learned from natural images. Second, residual noise in the generated images cannot be completely eliminated. Despite these limitations, our sim-to-real approach still demonstrates satisfactory performance in generating

Table 6 FID scores of cross-section microstructural images generated by different diffusion models. The lower the FID score is, the higher quality the generated samples have

	Baseline	Fine-tuning	Distillation
FID	236.76	167.87	178.32

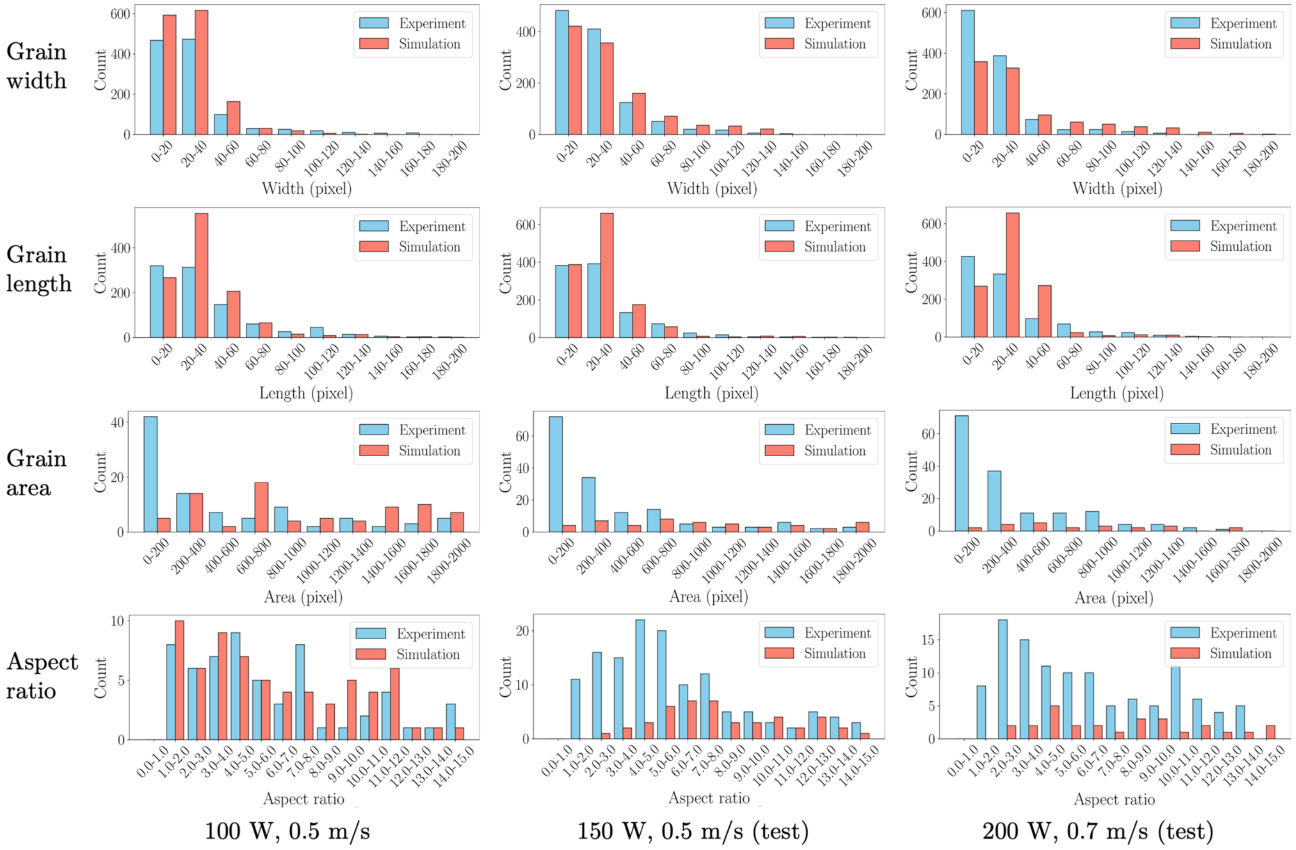


Fig. 13 Grain-level comparison of cross-section microstructures between experiment and simulation

microstructures with reasonable grain size and morphology, as discussed in the next subsection.

4.1.3 Grain size and morphology

The grain inclination angle shown in Fig. 12 is first analyzed in this subsection. From the experimental images, it can be observed that the inclination angle in the case of (100 W, 0.5 m/s) is smaller than those in the other two cases. This is attributed to the shallower melt pool depth resulting from the lower laser energy density. The energy densities of the (150 W, 0.5 m/s) and (200 W, 0.7 m/s) cases are relatively similar, leading to comparable inclination angles observed in the experiments. Our simulation results successfully capture this trend in grain inclination angle. The image [Baseline, (100 W, 0.5 m/s)] shows good agreement with the experimental result, as the baseline diffusion model has been trained under this condition. The inclination angle in [Baseline, (150 W, 0.5 m/s)] is slightly overestimated, even though the overall microstructure prediction in this case is inaccurate. For the image [Baseline, (200 W, 0.7 m/s)], no clear microstructural pattern is observed, making it impossible to measure the grain inclination angle. Among the microstructures generated by fine-tuned and distilled diffusion models, the maximum absolute error of inclination angle is 6° occurred at the image [Distillation, (200 W, 0.7 m/s)]. The error may stem from the inherent uncertainty of the generation process or the diversity of the training data. Despite this, the fine-tuned and distilled diffusion models are still capable of accurately predicting grain inclination angles under both seen and unseen conditions.

The following presents a quantitative analysis of grain width, length, area, and aspect ratio. To perform this analysis, simulated or generated microstructural images are randomly selected and compared with a randomly chosen image from the experimental dataset under the same manufacturing conditions. Due to inherent fluctuations in the experiment data, slight variations in grain count may be observed for each metric. Nevertheless, the overall grain distribution remains consistent and stable across the dataset.

The comparisons between experimental and simulated grains are shown in Fig. 13. Since the initial grain size was adjusted in the simulation, the resulting grain width distributions exhibit relatively good agreement with the experiment data. However, despite the satisfactory results for grain width, the correspondence deteriorates when examining grain length. Even larger discrepancies are observed in the distributions of grain area and aspect ratio, where

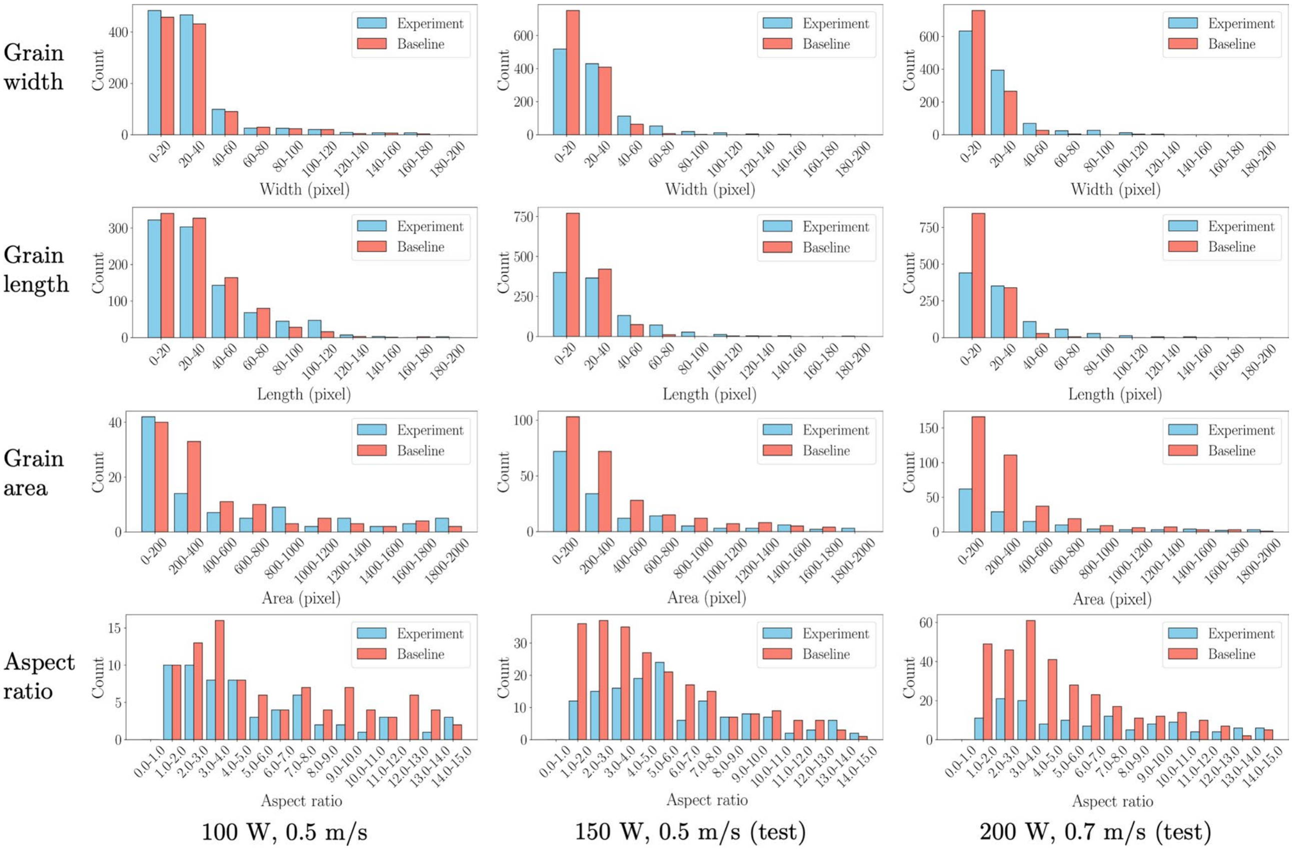
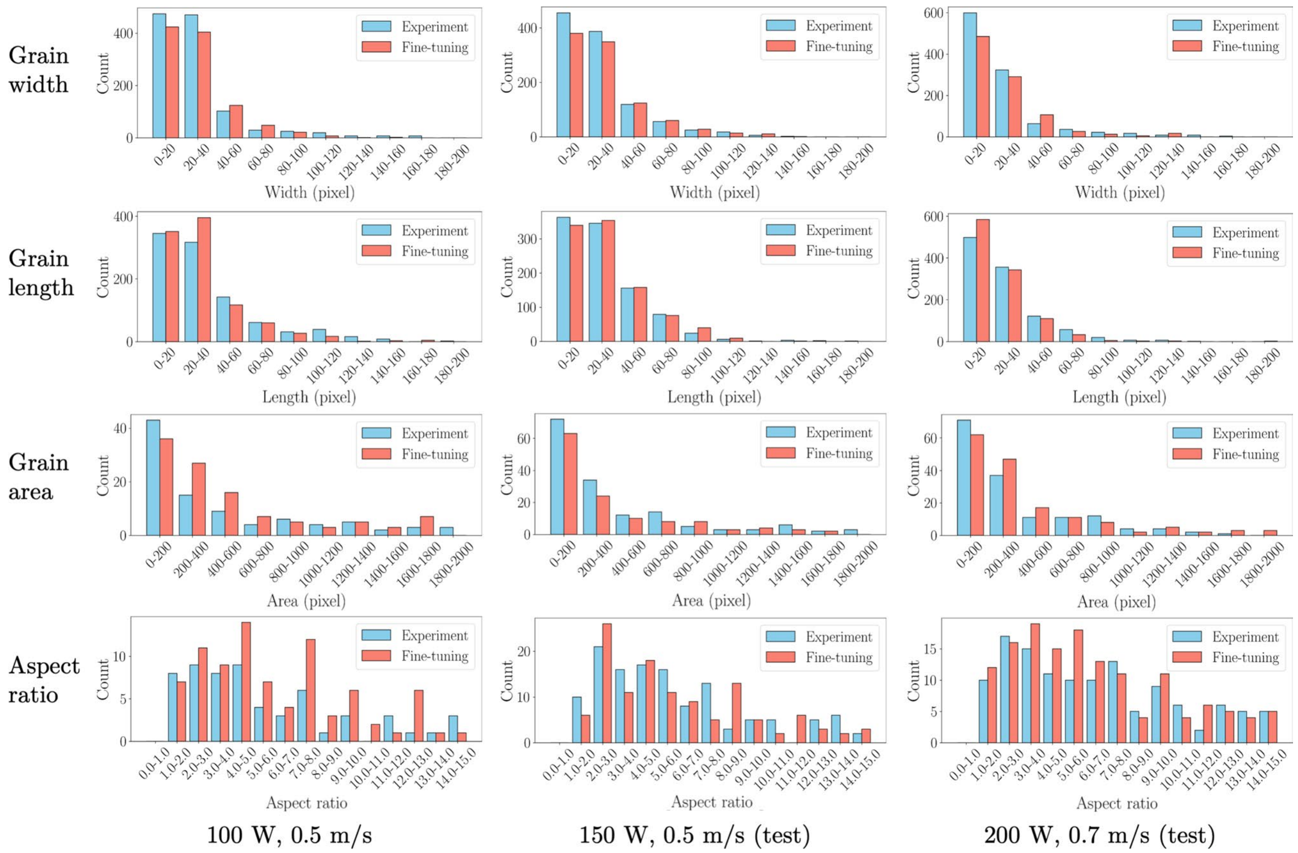


Fig. 14 Grain-level comparison of cross-section microstructures between experiment and baseline diffusion model

significant mismatches exist between the experimental and simulated results. This can be attributed to the complexity of metal AM, which involves physical interactions across macroscale, mesoscale, and microscale regimes [72]. Current simulations are limited in their ability to fully capture all phenomena occurring during the printing process. By summing the occurrences of binned grains from the experimental and simulated datasets separately, it is evident that the total number of grains in the experimental images exceeds that in the simulated ones. This is probably due to a higher nucleation density in the printing experiments. Additionally, noise from EBSD characterization may contribute to the observed difference.

The comparisons between the experimental results and the baseline predictions are presented in Fig. 14. Under the seen condition (100 W, 0.5 m/s), the generated microstructure exhibits good consistency with the experimental result in terms of grain width and length, and demonstrates reasonably good agreement in grain area and aspect ratio. This indicates that the baseline model is capable of capturing essential morphological features of grains when trained and evaluated under similar conditions. In contrast, under unseen conditions, the total number of generated grains exceeds that observed in the experiment data across all evaluated metrics. This discrepancy is primarily attributed to the chaotic and less controlled distribution of grains produced by the model when extrapolating beyond its training domain. The overgeneration of grains and noise under unseen conditions indicates reduced predictive reliability, likely due to the inability of the baseline model to generalize to microstructural variations not encountered during training.

Figure 15 presents the comparative results between the microstructures generated by the fine-tuned model and those obtained from experiments. The results indicate that the model is capable of generating grains that align well with experimental features across all evaluated metrics, not only under seen conditions but also under unseen conditions. Similar trends observed in Fig. 16 further demonstrate the effectiveness of the distillation training strategy in generating realistic microstructures. However, for the metric of grain area, the number of grains with areas ranging from 200 to 400 pixels exceeds that observed in the experiment data under all conditions. We speculate that this discrepancy is due to the presence of noisy, medium-sized grains generated during the distillation process. Despite this deviation, the distilled model still performs well across the other metrics, indicating its overall robustness and generalization ability.



100 W, 0.5 m/s 150 W, 0.5 m/s (test) 200 W, 0.7 m/s (test)

Fig. 15 Grain-level comparison of cross-section microstructures between experiment and fine-tuned diffusion model

Table 7 summarizes the mean absolute errors (MAE) and mean absolute percentage errors (MAPE) of the cross-section grain analysis results. The unit for MAE is the count of grains. The errors are calculated for each bin and then averaged. As the counts for the larger pixel ranges (i.e., the latter bins) are very low for both the real and generated data shown in these histograms, we calculate the errors using only the first five bins for grain size, while the first 10 bins for grain aspect ratio. From the table, we can observe a substantial decrease in MAEs after applying the proposed training strategies. Due to the inherently small counts in some bins, the calculated MAPEs are relatively larger. Nevertheless, a significant improvement in performance can still be observed.

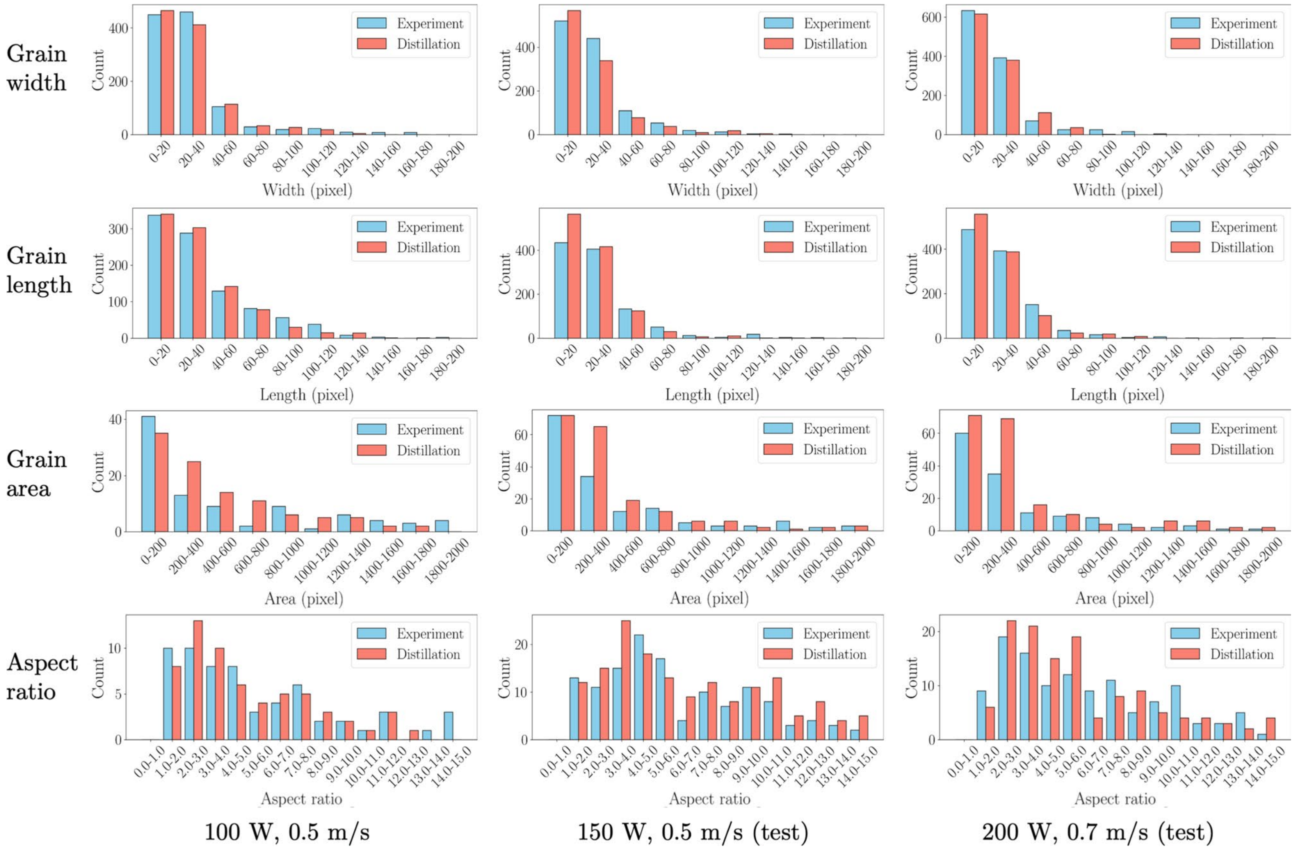
From the quantitative evaluations of the cross-section microstructures, we conclude that the simulations are not yet capable of fully capturing the complexity of real experimental data, due to the absence of features such as porosity, spatter, vapor depression, melt-pool instabilities, and other phenomena present in actual printing processes, while the baseline model struggles to generalize microstructural features under unseen conditions. In contrast, microstructures generated by both the fine-tuned and distilled models demonstrate significantly better alignment with experimental results across all evaluated metrics. These findings underscore the effectiveness of incorporating strategies such as fine-tuning and distillation with experiment data into diffusion models to provide physical priors and enhance microstructural predictions.

4.2 Longitudinal-section prediction

Figure 17 shows the longitudinal-section microstructures from experiment, simulation and three diffusion models trained with three different strategies. Similar to the cross-section images shown in Fig. 12, only one representative image is displayed for each method under a specific set of manufacturing conditions.

4.2.1 General analysis

The longitudinal-section microstructures exhibit different patterns compared to their cross-section counterparts. In this view, grains in the experimental images tend to tilt in the direction of laser scanning, which reflects the influence of thermal gradients and solidification dynamics. Since the longitudinal-section images in this study are obtained from single-layer, single-track scanning cases, the melt



100 W, 0.5 m/s 150 W, 0.5 m/s (test) 200 W, 0.7 m/s (test)

Fig. 16 Grain-level comparison of cross-section microstructures between experiment and distilled diffusion model

Table 7 Errors of the cross-section grain analysis results under test conditions

Grain metric	Width	Length	Area	Aspect ratio
MAE				
Simulation	96	95.4	16.8	8.1
Baseline	74.3	128.3	30.3	15.2
Fine-tuning	32.1	28.2	8.1	3.7
Distillation	19.9	44.2	10.7	3.3
MAPE				
Simulation	51.6%	54.4%	64.4%	66.7%
Baseline	58.4%	74.4%	149.0%	146.3%
Fine-tuning	15.8%	28.6%	39.4%	51.7%
Distillation	26.5%	31.1%	32.1%	36.9%

The unit for MAE is the count of grains

pool depth can be clearly identified by distinguishing the boundary between the grown grains and the initial substrate grains. A lower scanning speed leads to stronger grain tilting and a deeper melt pool. This growth trend is successfully captured in our simulations. However, the simulated grain structures generally appear coarser than those observed in the experiments.

For the baseline model, the images under seen conditions, such as [Baseline, (200 W, 0.75 m/s)] and [Baseline, (200 W, 0.25 m/s)], show relatively good agreement with experimental results. This demonstrates the ability of diffusion models to replicate microstructures under conditions they have encountered during training. Nonetheless, considerable noise appears in the training results, such as the region labeled D in the image [Baseline, (200 W, 0.25 m/s)]. By comparison, the image under the unseen condition [Baseline, (200 W, 0.5 m/s)] is dominated by noise and lacks recognizable grain morphology, which reflects the poor generalization capability of the baseline model. After applying fine-tuning or distillation, the diffusion models exhibit improved performance under both seen and unseen conditions, generating microstructures with clearer grain boundaries and more realistic morphology.

4.2.2 FID score

Only the images predicted under the testing condition of (200 W, 0.5 m/s) are used to calculate the FID scores, in comparison with their corresponding experimental images. The results are summarized in Table 8. Similar to the cross-section case, the FID score of the baseline prediction exceeds 220, but is reduced to approximately 190 after applying our proposed approaches. The improved FID scores of the longitudinal-section training are slightly higher

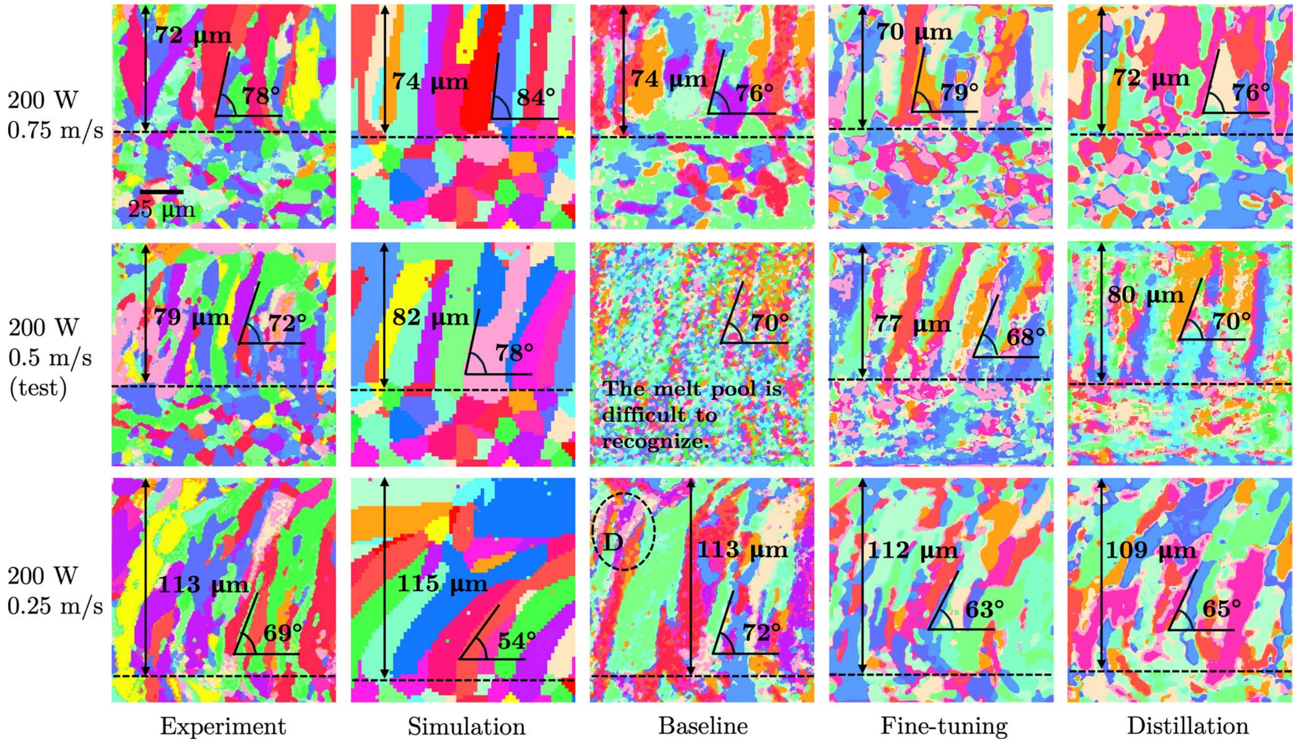


Fig. 17 Longitudinal-section microstructures obtained from experiment, simulation, and diffusion-based models (baseline, fine-tuning, and distillation). The experimental images here are cropped and further processed based on the results of Ref. [57], with permission from Elsevier

Table 8 FID scores of longitudinal-section microstructural images generated by different diffusion models

	Baseline	Fine-tuning	Distillation
FID	221.96	185.52	189.09

when compared to the cross-section results in Table 6. This difference may be attributed to minor degradation during the processing of experimental images, as well as the limited amount of experimental data available for refining the pre-trained diffusion model in this scenario.

4.2.3 Grain size and morphology

The grain inclination angles and melt pool depths of the microstructures shown in Fig. 17 are quantitatively analyzed first. By examining the experimental images from top to bottom, it can be observed that the inclination angle of the grains gradually decreases. This trend may be attributed to the increased curvature of the melt pool at lower scanning speed [57]. Meanwhile, the melt pool becomes deeper, which can be explained by the higher energy density associated with the reduced scanning speed. Our simulations exhibit high predictive accuracy for melt pool depth across various processing conditions. However, their performance in capturing grain inclination angles remains limited, particularly at very low scanning speeds. When trained solely on experimental images, the baseline model can generate well-distributed grains with accurate inclination angles and melt pool depths under seen conditions. In contrast, under the testing condition, the predicted grain structures become disordered, as shown in the image [Baseline, (200 W, 0.5 m/s)].

In the following analysis, the grain width, length, area, and aspect ratio are quantitatively evaluated. In the same manner as described in Sect. 4.1.3, representative experimental microstructural images and their corresponding simulated or generated images are randomly selected for comparison under each set of conditions.

Figure 18 shows comparisons between experimental and simulated results using the proposed grain-level metrics. It is evident that the number of grains in the experimental images is significantly higher than that in the simulated ones. As described in Sect. 2.1.3, the initial settings in the longitudinal-section simulations is set to be the same as that in the cross-section simulations. This configuration is intended to demonstrate the adaptability of the fine-tuned and distilled diffusion model to different experimental setups.

Figure 19 presents comparisons between the baseline model predictions and experimental microstructural images. The generated microstructure under a seen condition of (200 W, 0.75 m/s) shows relatively good agreement with the experiment data. However, under another seen condition of (200 W, 0.25 m/s), the generated microstructure displays noticeable mismatches. This unexpected degradation

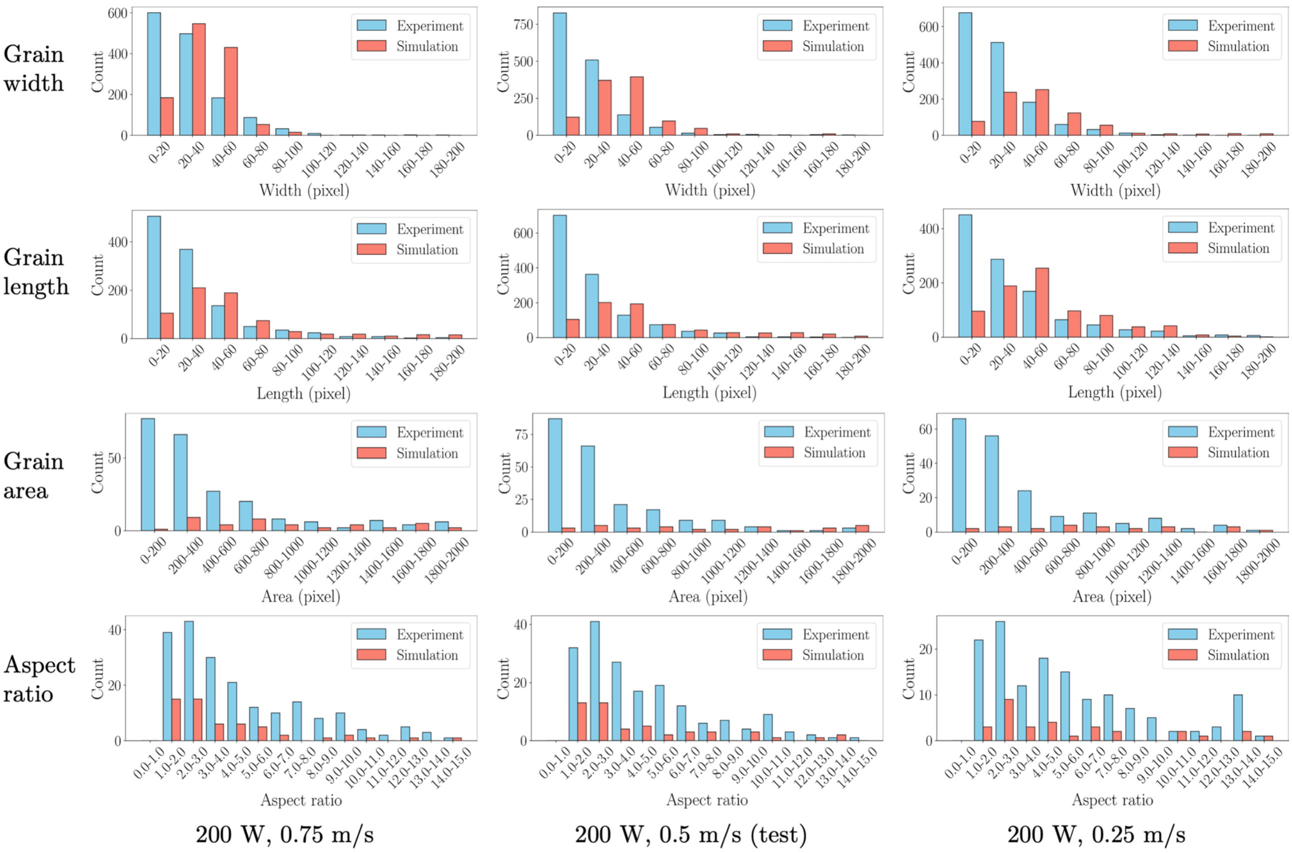


Fig. 18 Grain-level comparison of longitudinal-section microstructures between experiment and simulation

in performance is likely a consequence of the limited quantity of experimental training data, which may have led to increased noise and reduced fidelity in the output. Furthermore, the generated microstructure under an unseen condition of (200 W, 0.5 m/s) performs even worse, which can be attributed to the poor ability of diffusion models to generate data beyond its training memory.

The comparative results after fine-tuning and distillation are presented in Fig. 20 and Fig. 21, respectively. Under seen conditions, the histogram distributions of both models closely match the experiment data. For the unseen condition, both approaches demonstrate substantial improvements over the baseline model. Despite the improvements, the generated microstructures still exhibit noisy grains, which may lead to slightly higher grain counts compared to those observed in the experiment data across all metrics. Notably, in this longitudinal-section training scenario, the fine-tuned model yields better performance than the distilled model, implying that the distillation process may be more dependent on the size and diversity of the training dataset.

Table 9 summarizes the mean MAEs and MAPEs of the longitudinal-section grain analysis results. As shown in the table, the refined models demonstrate improved performance, with the fine-tuned model outperforming the distilled model. This observation is consistent with our previous discussion.

These comparisons reveal the limitations of simulation and baseline training methods, while demonstrating that our proposed approaches can effectively generate realistic microstructures, even when fine-tuning and distillation data are sourced from an independent study [57] within the AM community.

4.3 Discussion

In this section, we will discuss the influence of sim-to-real data proportion and distillation hyperparameters on the performance of the refined diffusion models.

4.3.1 Sim-to-real data proportion

In the proposed sim-to-real framework, balancing the proportion of simulation data and experiment data is an important issue. By comparing Table 7 and Table 9, it appears that the performance on the cross-section task is better than that on the longitudinal-section task. The numbers of simulation images used for pre-training are 20,000 and 5000,

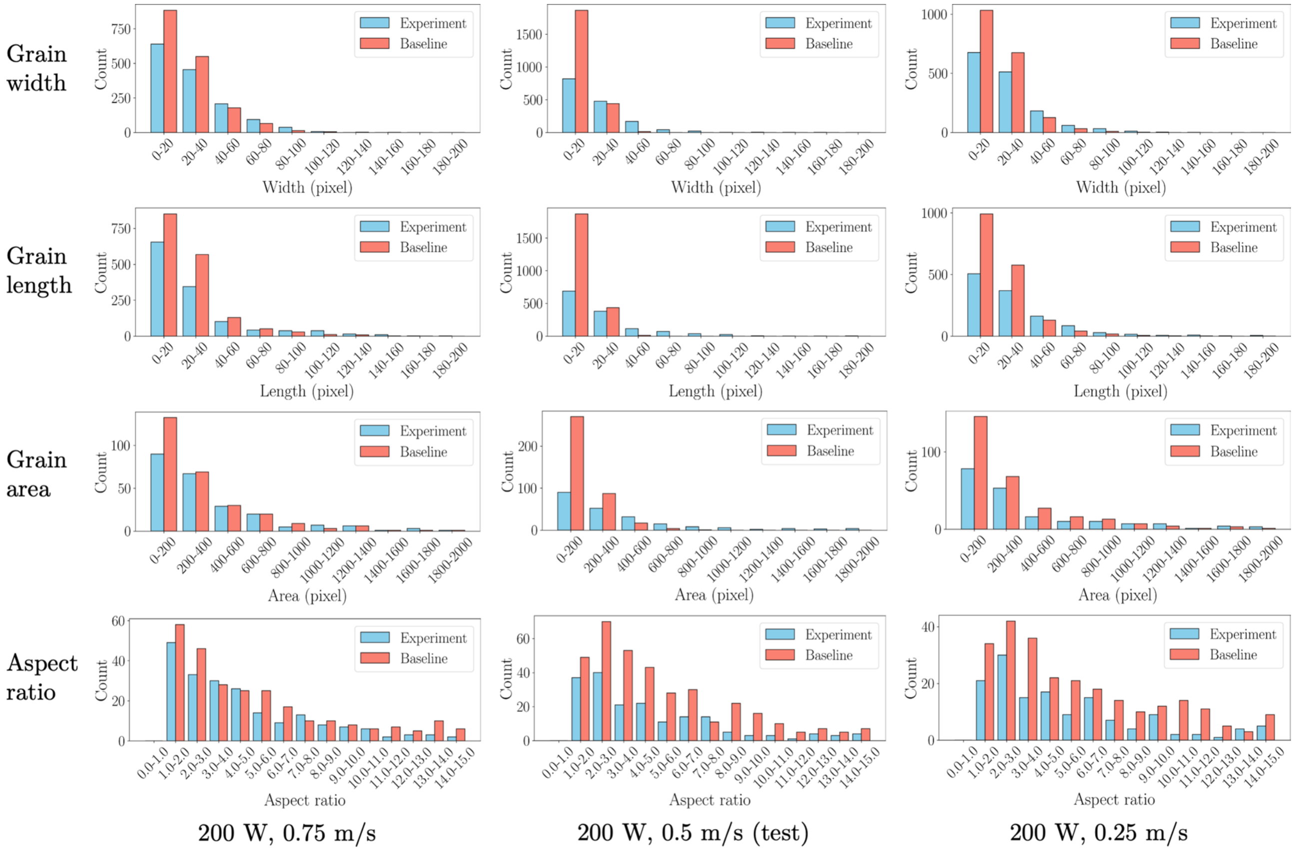


Fig. 19 Grain-level comparison of longitudinal-section microstructures between experiment and baseline diffusion model

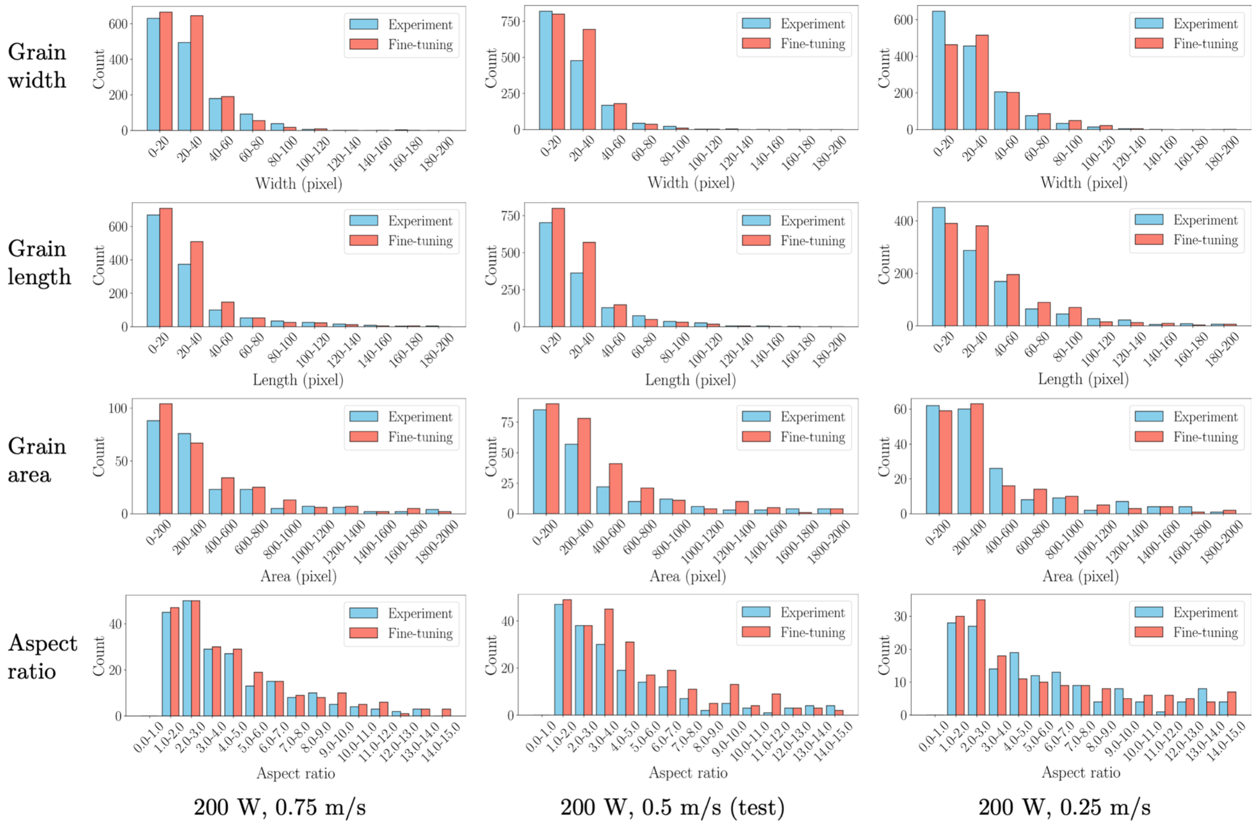


Fig. 20 Grain-level comparison of longitudinal-section microstructures between experiment and fine-tuned diffusion model

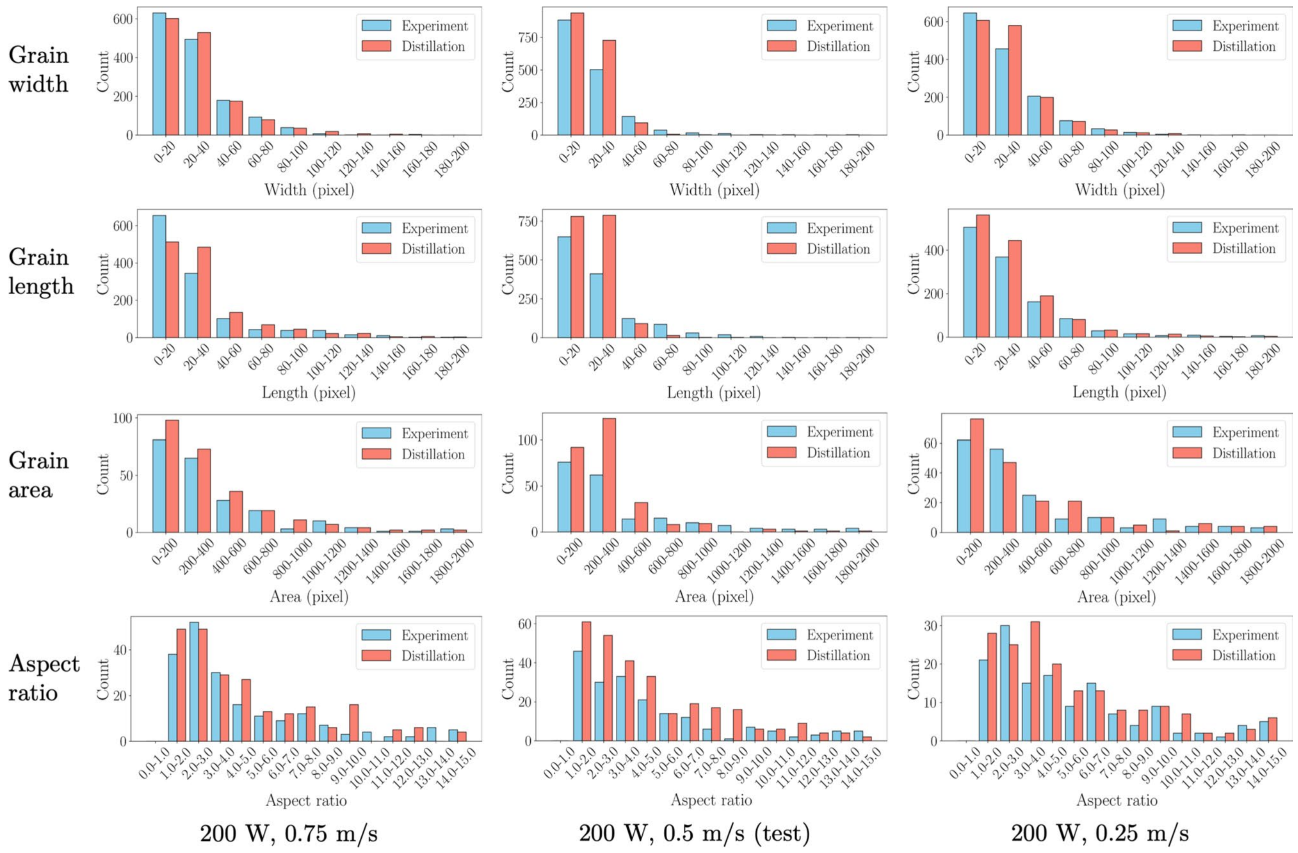


Fig. 21 Grain-level comparison of longitudinal-section microstructures between experiment and distilled diffusion model

Table 9 Errors of the longitudinal-section grain analysis results under test conditions

Grain metric	Width	Length	Area	Aspect ratio
MAE				
Simulation	79.2	61.6	18.6	12.5
Baseline	101.8	93.0	14.8	15.1
Fine-tuning	23.0	13.0	8.8	5.8
Distillation	27.4	36.0	10.8	11.2
MAPE				
Simulation	160.0%	40.4%	79.3%	72.6%
Baseline	90.5%	89.1%	74.8%	131.0%
Fine-tuning	34.0%	14.8%	42.2%	76.6%
Distillation	54.1%	51.2%	53.9%	112.4%

The unit for MAE is the count of grains

respectively, indicating that pre-trained models with more simulation data tend to be more stable.

To figure out the suitable data ratio in the microstructure prediction task, we further use the 8 cross-section experimental cases to conduct several fine-tuning trainings with different number of cases set aside for testing. The generated microstructures conditioned on (200 W, 0.7 m/s) from these models are shown as examples in Fig. 22, and corresponding FID scores are listed in Table 10. It can be observed that models refined with a larger amount of experimental data are able to generate higher-quality microstructures, whereas the performance drops significantly when experimental data are very limited. From the image generated by the model fine-tuned with only 2 cases, we can conclude that the limited experimental data cannot offer adequate real information, and may even confuse the physical prior learned from the simulation data. Since experimental data are difficult to obtain, the amount used may need to be adjusted according to the desired quality of the results.

4.3.2 Distillation hyperparameters

As introduced in Sect. 3.3, there are two hyperparameters, λ_1 and λ_2, that control the performance of the distilled diffusion models. To further investigate the influence of the hyperparameter settings on the generated microstructures, we use cross-section experiment data to distill the pre-trained diffusion model using different combinations of λ_1 and λ_2, and the generated results under test condition (150 W, 0.5 m/s) are shown in Fig. 23. The FID scores are also calculated to quantitatively analyze the results, which are listed in Table 11. It can be observed that a relatively larger value of λ_1 enables the generated microstructures to capture more experimental details, as the influence of

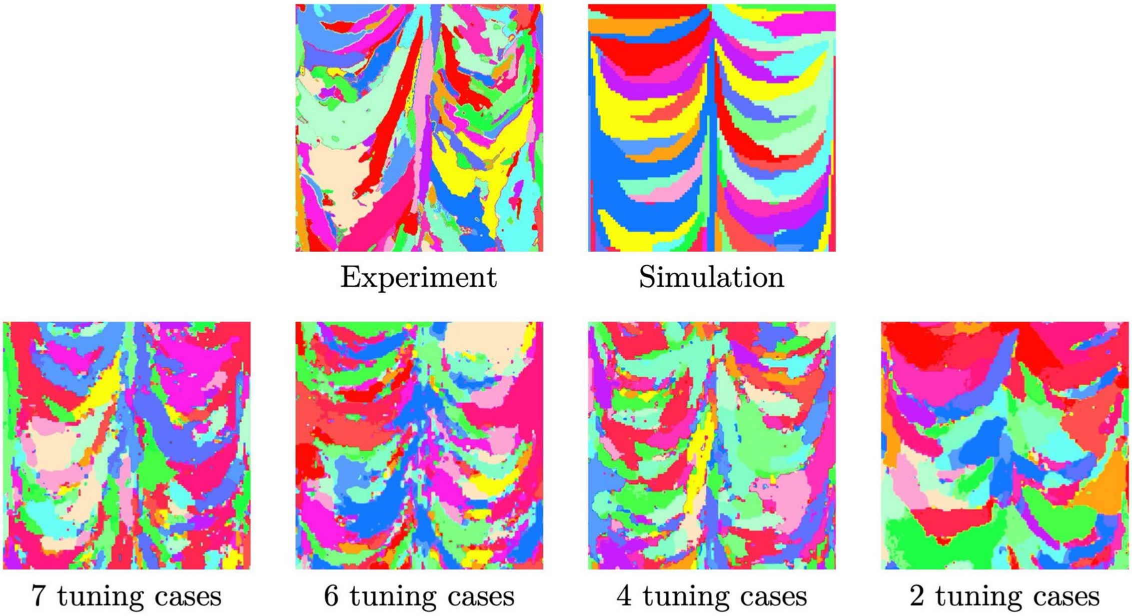

Experiment
Simulation
7 tuning cases
6 tuning cases
4 tuning cases
2 tuning cases

Fig. 22 Cross-section generated results under test condition (200 W, 0.7 m/s) with different settings of experiment data proportion in fine-tuning

Table 10 FID scores of the cross-section generated results under test conditions with different settings of experiment data proportion in fine-tuning

Number of fine-tuning cases	7	6	4	2
FID	143.65	167.87	162.41	215.86

Table 11 FID scores of the cross-section generated results under test conditions with different settings of hyperparameters in distillation

λ_1 - λ_2	0.5 - 0.5	0.6 - 0.4	0.7 - 0.3	0.8 - 0.2	0.9 - 0.1
FID	251.82	212.96	208.39	178.32	189.95

experimental data on the student model is weighted by λ_1. However, an extremely large value of λ_1 may prevent the model from learning effectively from the simulation data. In this work, we set λ_1 and λ_2 to 0.8 and 0.2, respectively, as this combination yields the lowest FID score, as shown in Table 11.

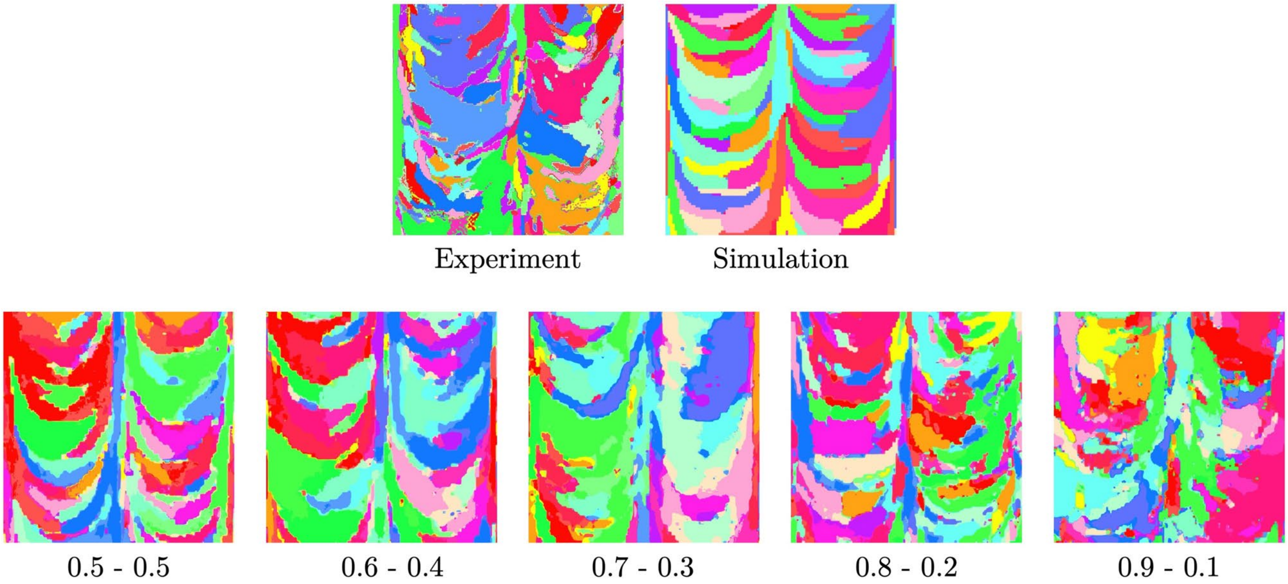

Experiment
Simulation
0.5 - 0.5
0.6 - 0.4
0.7 - 0.3
0.8 - 0.2
0.9 - 0.1

Fig. 23 Cross-section generated results under test condition (150 W, 0.5 m/s) with different settings of hyperparameters in distillation. The captions under the second row indicate the values of λ_1 and λ_2

5 Conclusion

In this study, we propose sim-to-real diffusion models that are initially pre-trained using extensive simulation data and subsequently fine-tuned or distilled with a small amount of experiment data, in order to achieve sim-to-real microstructure prediction in metal AM. To demonstrate the effectiveness of the proposed method, we consider two prediction scenarios: cross-section and longitudinal-section microstructure generation. The pre-training data are obtained from numerical simulations, where CFD and PF models are developed to generate synthetic microstructural images. The experiment data used for model refinement are collected from EBSD measurements. Specifically, the cross-section data are obtained from our in-house experiments, while the longitudinal-section data is from an independent study [57]. The quality of the generated microstructures is evaluated using both machine learning and physics-based metrics. The FID score is employed to quantify the similarity between the generated and experimental microstructures in the feature space. The physics-based metrics, which are used to assess microstructural fidelity at the grain level, include grain inclination angle, length, width, area, and aspect ratio. For longitudinal predictions, melt pool depth is also included as an additional metric. Comparative analyses are conducted among microstructures obtained from experiments, simulations, and three diffusion models trained with different strategies: baseline, fine-tuning and distillation. The results demonstrate that both the fine-tuned and distilled models effectively overcome the limitations of numerical simulation and the baseline model trained solely on experiment data, and are capable of generating microstructures that are realistic and physically consistent with experimental observations under both seen and unseen conditions.

The proposed sim-to-real training strategies developed in this study effectively harness the complementary strengths of both simulation and experiment data while addressing their respective limitations. By integrating the diversity and scalability of simulation data with the physical realism of experimental observations, our approach enables the high-fidelity generation of realistic microstructures, including those under previously unseen processing conditions. This capability is particularly beneficial for the metal AM industry, where acquiring comprehensive experiment datasets is often costly and time-consuming. Furthermore, the proposed sim-to-real diffusion models exhibit remarkable adaptability to new datasets within the metal AM community, significantly reducing the need for manual re-calibration of numerical simulations or retraining machine learning models when facing variations in materials, processing parameters, or experimental setups. This adaptability not only improves efficiency but also enhances the practical value of our models in real-world manufacturing scenarios. Overall, the proposed sim-to-real framework offers a novel and promising perspective on the application of diffusion models to engineering problems, bridging the gap between data-driven generative modeling and physics-based design requirements.

Acknowledgements Ziyuan Xie and Tianju Xue acknowledge the support from the Young Scientist Fund (YSF) by the National Natural Science Foundation of China (NSFC) (Project No. 12402143). Zichuan Fu, Jingchi Zhang, and Kaihao Zhang acknowledge the support from Guangdong Basic and Applied Basic Research Funding Scheme-Provincial and Municipal Joint Fund (No. 2022A1515110735), Guangzhou-HKUST(GZ) Joint Funding Program (No. 2023A03J0003). The experimental work used HKUST(GZ) Materials Characterization and Preparation Facility (MCPF) and Materials Design and Manufacturing Facility (MDMF).

Author Contributions Ziyuan Xie - Writing - original draft, Methodology, Experiment, Software, Validation. Zichuan Fu - Writing - review & editing, Experiment. Jingchi Zhang - Writing - review & editing, Experiment. Kaihao Zhang - Writing – review & editing, Methodology, Supervision. Tianchen Zeng - Writing – review & editing, Software. Yu Wu - Writing – review & editing, Methodology, Supervision. Shuheng Liao - Writing - review & editing, Software. Xiang Li - Writing - review & editing. Tianju Xue - Writing - review & editing, Conceptualization, Methodology, Supervision.

Funding Open access funding provided by Hong Kong University of Science and Technology

Data Availability Data is provided within the manuscript.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

1. Buchanan C, Gardner L (2019) Metal 3d printing in construction: A review of methods, research, applications, opportunities and challenges. Eng Struct 180:332–348
2. Blakey-Milner B, Gradl P, Snedden G, Brooks M, Pitot J, Lopez E, Leary M, Berto F, Plessis AD (2021) Metal additive manufacturing in aerospace: A review. Materials & Design 209:110008

3. Martin JH, Yahata BD, Hundley JM, Mayer JA, Schaedler TA, Pollock TM (2017) 3d printing of high-strength aluminium alloys. Nature 549(7672):365–369
4. Shamsaei N, Yadollahi A, Bian L, Thompson SM (2015) An overview of direct laser deposition for additive manufacturing; part ii: Mechanical behavior, process parameter optimization and control. Addit Manuf 8:12–35
5. Herzog D, Seyda V, Wycisk E, Emmelmann C (2016) Additive manufacturing of metals. Acta Mater 117:371–392
6. Gong J, Wei K, Liu M, Song W, Li X, Zeng X (2022) Microstructure and mechanical properties of alsi10mg alloy built by laser powder bed fusion/direct energy deposition hybrid laser additive manufacturing. Addit Manuf 59:103160
7. Trosch T, Strößner J, Völkl R, Glatzel U (2016) Microstructure and mechanical properties of selective laser melted inconel 718 compared to forging and casting. Mater Lett 164:428–431
8. Andreau O, Koutiri I, Peyre P, Penot J-D, Saintier N, Pessard E, De Terris T, Dupuy C, Baudin T (2019) Texture control of 316l parts by modulation of the melt pool morphology in selective laser melting. J Mater Process Technol 264:21–31
9. Heiden MJ, Jensen SC, Koepke JR, Saiz DJ, Dickens SM, Jared BH (2022) Process and feedstock driven microstructure for laser powder bed fusion of 316l stainless steel. Materialia 21:101356
10. Tan JHK, Sing SL, Yeong WY (2020) Microstructure modelling for metallic additive manufacturing: a review. Virtual and Physical Prototyping 15(1):87–105
11. Akram J, Chalavadi P, Pal D, Stucker B (2018) Understanding grain evolution in additive manufacturing through modeling. Addit Manuf 21:255–268
12. Lian Y, Lin S, Yan W, Liu WK, Wagner GJ (2018) A parallelized three-dimensional cellular automaton model for grain growth during additive manufacturing. Comput Mech 61:543–558
13. Rodgers TM, Madison JD, Tikare V (2017) Simulation of metal additive manufacturing microstructures using kinetic monte carlo. Comput Mater Sci 135:78–89
14. Liu P, Wang Z, Xiao Y, Horstemeyer MF, Cui X, Chen L (2019) Insight into the mechanisms of columnar to equiaxed grain transition during metallic additive manufacturing. Addit Manuf 26:22–29
15. Yang M, Wang L, Yan W (2021) Phase-field modeling of grain evolutions in additive manufacturing from nucleation, growth, to coarsening. NPJ Comput Mater 7(1):56
16. Xue T, Gan Z, Liao S, Cao J (2022) Physics-embedded graph network for accelerating phase-field simulation of microstructure evolution in additive manufacturing. npj Computational Materials, 8(1):201,
17. Choi JY, Xue T, Liao S, Cao J (2024) Accelerating phase-field simulation of three-dimensional microstructure evolution in laser powder bed fusion with composable machine learning predictions. Additive Manufacturing, 79:103938
18. Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114
19. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Advances in neural information processing systems, 27
20. Rezende D, Mohamed S (2015) Variational inference with normalizing flows. In International conference on machine learning, pages 1530–1538. PMLR
21. Yang L, Zhang Z, Song Y, Hong S, Runsheng X, Zhao Y, Zhang W, Cui B, Yang M-H (2023) Diffusion models: A comprehensive survey of methods and applications. ACM Comput Surv 56(4):1–39
22. Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. Adv Neural Inf Process Syst 33:6840–6851
23. Song J, Meng C, Ermon S (2020) Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502
24. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B (2020) Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456
25. Azqadan E, Jahed H, Arami A (2023) Predictive microstructure image generation using denoising diffusion probabilistic models. Acta Mater 261:119406
26. Lyu X, Ren X (2024) Microstructure reconstruction of 2d/3d random materials via diffusion-based deep generative models. Sci Rep 14(1):5041
27. Hoffman N, Diniz C, Liu D, Rodgers T, Tran A, Fuge M (2025) Grainpaint: A multi-scale diffusion-based generative model for microstructure reconstruction of large-scale objects. Acta Mater 288:120784
28. Plimpton S, Battaile C, Chandross M, Holm L, Thompson A, Tikare V, Wagner G, Webb E, Zhou X, Cardona CG, et al. (2009) Crossing the mesoscale no-man's land via parallel kinetic monte carlo. Sandia Report SAND2009-6226, 1
29. Shu D, Li Z, Farimani AB (2023) A physics-informed diffusion model for high-fidelity flow field reconstruction. Journal of Computational Physics, 478:111972
30. Qu Y, Nathaniel J, Li S, Gentine P (2024) Deep generative data assimilation in multimodal setting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 449–459,
31. Gao J, Zheng C, Jeni LA, Erickson Z (2025) Disrt-in-bed: Diffusion-based sim-to-real transfer framework for in-bed human mesh recovery. In Proceedings of the Computer Vision and Pattern Recognition Conference, pages 1829–1838
32. Peng XB, Andrychowicz M, Zaremba W, Abbeel P (2018) Sim-to-real transfer of robotic control with dynamics randomization. In 2018 IEEE international conference on robotics and automation (ICRA), pages 3803–3810. IEEE
33. Zhao W, Queralta JP, Westerlund T (2020) Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In 2020 IEEE symposium series on computational intelligence (SSCI), pages 737–744. IEEE
34. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298
35. Hinton G, Vinyals O, Dean J (2015) Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531
36. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S (2017) Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30
37. Wang L, Zhang Y, Chia HY, Yan W (2022) Mechanism of keyhole pore formation in metal additive manufacturing. NPJ Computational Materials, 8(1):22
38. Gan Z, Jones KK, Lu Y, Liu WK (2021) Benchmark study of melted track geometries in laser powder bed fusion of inconel 625. Integrating Materials and Manufacturing Innovation 10(2):177–195
39. Liao S, Xue T, Jeong J, Webster S, Ehmann K, Cao J (2023) Hybrid thermal modeling of additive manufacturing processes using physics-informed neural networks for temperature prediction and parameter identification. Comput Mech 72(3):499–512
40. Modeling of powder bed fusion (2018) T Mukherjee, HL Wei, A De, and Tarasankar DebRoy. Heat and fluid flow in additive manufacturing-part i. Comput Mater Sci 150:304–313
41. Gan Z, Lian Y, Lin SE, Jones KK, Liu WK, Wagner GJ (2019) Benchmark study of thermal behavior, surface topography, and dendritic microstructure in selective laser melting of inconel 625. Integrating Materials and Manufacturing Innovation, 8:178–193

42. Guermond J-L, Minev P, Shen J (2006) An overview of projection methods for incompressible flows. Comput Methods Appl Mech Eng 195(44–47):6011–6045
43. Cherry JA, Davies HM, Mehmood S, Lavery NP, Brown SGR, Sienz JJTIJoAMT (2015) Investigation into the effect of process parameters on microstructural and physical properties of 316l stainless steel parts by selective laser melting. The International Journal of Advanced Manufacturing Technology, 76:869–879
44. Moelans N, Blanpain B, Wollants P (2008) Quantitative analysis of grain boundary properties in a generalized phase field model for grain growth in anisotropic systems. Physical Review B-Condensed Matter and Materials Physics 78(2):024113
45. Quey R, Dawson PR, Barbe F (2011) Large-scale 3d random polycrystals for the finite element method: Generation, meshing and remeshing. Computer Methods in Applied Mechanics and Engineering, 200(17–20), 1729–1745
46. Turnbull D (1950) Formation of crystal nuclei in liquid metals. J Appl Phys 21(10):1022–1028
47. Kalikmanov VI (2012) Classical nucleation theory. In Nucleation theory, pages 17–41. Springer
48. Simmons JP, Shen C, Wang Y (2000) Phase field modeling of simultaneous nucleation and growth by explicitly incorporating nucleation events. Scripta Mater 43(10):935–942
49. Roberts IA, Wang CJ, Esterlein R, Stanford M, Mynors DJ (2009) A three-dimensional finite element analysis of the temperature field during laser melting of metal powders in additive layer manufacturing. International Journal of Machine Tools and Manufacture, 49(12–13), 916–923
50. Johnstone DN, Martineau BH, Crout P, Midgley PA, Eggeman AS (2020) Density-based clustering of crystal (mis) orientations and the orix python library. Applied Crystallography, 53(5), 1293–1298,
51. Bradbury J, Frostig R, Hawkins P, Johnson MJ, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S, Zhang Q (2018) JAX: composable transformations of Python+NumPy programs
52. Bertoli US, Wolfer AJ, Matthews MJ, Delplanque J-P R, Schoenung JM (2017) On the limitations of volumetric energy density as a design parameter for selective laser melting. Materials & Design, 113:331–340
53. Ahmed N, Barsoum I, Haidemenopoulos G, Al-Rub RKA (2022) Process parameter selection and optimization of laser powder bed fusion for 316l stainless steel: A review. J Manuf Process 75:415–434
54. Nayak SK, Mishra SK, Paul CP, Jinoop AN, Bindra KS (2020) Effect of energy density on laser powder bed fusion built single tracks and thin wall structures with 100 μm preplaced powder layer thickness. Optics & Laser Technology 125:106016
55. Niederreiter H (1988) Low-discrepancy and low-dispersion sequences. Journal of number theory 30(1):51–70
56. Tang C, Tan JL, Wong CH (2018) A numerical investigation on the physical mechanisms of single track defects in selective laser melting. Int J Heat Mass Transf 126:957–968
57. Wang Y, Chenfan Yu, Xing L, Li K, Chen J, Liu W, Ma J, Shen Z (2020) Grain structure and texture of the slm single track. J Mater Process Technol 281:116591
58. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer
59. Platen Pv, Patil S, Lozhkov A, Cuenca P, Lambert N, Rasul K, Davaadorj M, Nair D, Paul S, Berman W, Xu Y, Liu S, Wolf T (2022) Diffusers: State-of-the-art diffusion models. https://github.com/huggingface/diffusers
60. Tancik M, Srinivasan P, Mildenhall B, Fridovich-Keil S, Raghavan N, Singhal U, Ramamoorthi R, Barron J, Ng R (2020) Fourier features let networks learn high frequency functions in low dimensional domains. Adv Neural Inf Process Syst 33:7537–7547
61. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587
62. Iglovikov V, Shvets A (2018) Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. arXiv preprint arXiv:1801.05746
63. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826
64. Seitzer M (2020) pytorch-fid: FID Score for PyTorch. https://github.com/mseitzer/pytorch-fid, August Version 0.3.0
65. Levine L, Lane B, Becker C, Belak J, Carson R, Deisenroth D, Glaessgen E, Gnaupel-Herold T, Gorelik M, Greene G et al (2024) Outcomes and conclusions from the 2022 am bench measurements, challenge problems, modeling submissions, and conference. Integrating Materials and Manufacturing Innovation 13(3):598–621
66. Abdi H, Williams LJ (2010) Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4):433–459
67. Yi M, Sun J, Li Z (2023) On the generalization of diffusion model. arXiv preprint arXiv:2305.14712
68. Li P, Li Z, Zhang H, Bian J (2023) On the generalization properties of diffusion models. Adv Neural Inf Process Syst 36:2097–2127
69. Kadkhodaie Z, Guth F, Simoncelli EP, Mallat S (2023) Generalization in diffusion models arises from geometry-adaptive harmonic representations. arXiv preprint arXiv:2310.02557
70. Zhang H, Zhou J, Lu Y, Guo M, Shen L, Qu Q (2024) The emergence of reproducibility and consistency in diffusion models. International Conference on Machine Learning
71. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L(2009) Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee
72. Michopoulos JG, Iliopoulos AP, Steuben JC, Birnbaum AJ, Lambrakos SG (2018) On the multiphysics modeling challenges for metal additive manufacturing processes. Addit Manuf 22:784–799