STRUCTRDM: PROTOTYPE-ANCHORED RESIDUAL DIFFUSION FOR TITANIUM ALLOY MICROSTRUCTURE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Titanium alloy microstructures are jointly influenced by alloy composition, processing parameters and mechanical properties. In limited-data regimes, however, these low-dimensional conditions underdetermine the spatial arrangement of phase regions and the geometry of phase boundaries in high-dimensional microstructure images, making end-to-end conditional generation prone to fragmented phase regions and unstable boundaries. To address these issues, we propose StructRDM, a reliability-aware, prototype-anchored residual diffusion framework. StructRDM follows a "structure first, texture second" strategy. It first predicts structural statistics from multi-source material conditions and estimates their reliability through conformal calibration. It then retrieves type-consistent structural prototypes through reliability-weighted matching and generates the target phase structure using latent residual diffusion. Finally, the generated phase-structure map provides spatial guidance for microstructural texture synthesis. Compared with the direct condition-to-image diffusion baseline, StructRDM improves F1 by 26.9% while reducing Gradient W1 and Autocorrelation MAE by 21.0% and 35.4%, respectively, indicating broader distributional coverage and improved spatial-statistical consistency.

1 INTRODUCTION

Titanium alloys are widely used in aerospace, marine engineering, and biomedicine because of their high specific strength, excellent corrosion resistance, thermal stability, and biocompatibility (Pushp et al., 2022; Marin & Lanzutti, 2023). Their composition and processing conditions jointly determine phase distributions, grain sizes, and boundary morphologies, which further influence performance. Microstructure therefore serves as a crucial link between composition, processing, and properties (Lütjering & Williams, 2007; Kang & Yang, 2019). Generating microstructure images from material conditions could provide a low-cost auxiliary tool for materials design, process screening, and data augmentation (Wodo et al., 2016; Robertson et al., 2023).

In recent years, deep generative models have shown considerable promise for microstructure reconstruction and conditional generation (White et al., 2024; Nimmal Haribabu et al., 2023; Azqadan et al., 2023). Nevertheless, material-condition-driven generation of titanium alloy microstructures still faces two key challenges. First, material conditions such as alloy composition, processing parameters, and mechanical properties primarily characterize global sample attributes and thus provide insufficient direct constraints on phase-region arrangements and boundary morphologies. Second, with limited and heterogeneous data, the relationship between material conditions and microstructural organization is highly uncertain, further complicating stable structure generation. The central question addressed in this work is therefore how to extract an effective structural prior from multi-source material conditions and use it to constrain subsequent image generation.

To address these challenges, we propose StructRDM, a reliability-aware, prototype-anchored residual diffusion method that adopts a "structure first, texture second" strategy. In the first stage, StructRDM predicts structural statistics from multi-source material conditions and calibrates the predictive reliability of individual structural attributes. It then uses the structural predictions and their reliability estimates to retrieve a real, type-consistent structure from the training set as a prototype.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

A latent residual diffusion model subsequently corrects this prototype to produce a phase-structure map corresponding to the target material conditions. In the second stage, the generated phase-structure map serves as a spatial condition, with multi-scale conditioning guiding the generation of grayscale texture and local morphology. This decomposition converts direct condition-to-image generation into prototype-anchored phase-structure generation followed by structure-conditioned texture generation, thereby reducing the learning difficulty under limited-data conditions.

Our main contributions are as follows:

1. **A structure-first microstructure generation framework for limited multi-source materials data.** It uses an explicit phase structure as intermediate representation and employs multi-scale conditioning to guide the generation of grayscale texture and local morphology.
2. **A prototype-anchored latent residual diffusion method.** It learns conditional residuals around real structural prototypes rather than directly generating complete phase structures.
3. **A reliability-aware mechanism for structural prototype retrieval and generation.** It estimates structural-prediction reliability through out-of-sample residual learning and split-conformal calibration, and accordingly modulates the role of each structural attribute in prototype retrieval and residual generation.

2 Related Work

2.1 Titanium Alloy Microstructure Data and Structural Representation

Titanium alloy microstructures contain structural information such as phase composition, grain size, and boundary distribution. Traditional methods typically combine microscopic characterization, threshold-based segmentation, morphological processing, and statistical analysis to quantify features such as the α/β phase fraction and grain size (Tiley et al., 2004; Searles et al., 2005). Subsequent computational characterization methods further employ morphological descriptors and spatial statistics to construct computable representations of microstructures (Bostanabad et al., 2018).

With the development of materials informatics, titanium alloy datasets have gradually evolved into multi-source data systems integrating alloy compositions, processing conditions, microstructural features, and mechanical properties (Wodo et al., 2016; Salvador et al., 2022; Zhao et al., 2023). To address limited sample sizes, heterogeneous data sources, and missing fields, previous studies have employed strategies such as data imputation, transfer learning, and generative data augmentation (Luo et al., 2026). For example, Goh et al. (2023) used nearest-neighbor and multiple-imputation methods to complete processing and property data for Ti-6Al-4V, while Zhou et al. (2024) applied cross-material transfer learning to few-shot property prediction. In contrast, we address missing spatial annotations by extracting structural pseudo-labels and statistics from micrographs. We treat these pseudo-labels as potentially noisy intermediate representations rather than metallurgical ground truth. These representations provide spatial cues for subsequent structure modeling.

2.2 Microstructure Reconstruction and Conditional Generation

Early microstructure generation primarily relied on phase-field models, Monte Carlo methods, and statistical reconstruction. Phase-field and Monte Carlo methods simulate phase evolution and grain growth (Cogswell & Carter, 2011; Anderson et al., 1984), whereas spatial-correlation-based methods reconstruct structures whose prescribed statistics match those of reference samples (Yeong & Torquato, 1998; Fullwood et al., 2008). These methods have clear physical or statistical foundations but typically rely on specific modeling assumptions or incur high computational costs.

In recent years, variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models have increasingly been applied to data-driven microstructure generation (Cang et al., 2017; Yang et al., 2018). VAEs have been used to learn and reconstruct Ti–6Al–4V microstructures, but their outputs tend to be overly smooth (White et al., 2024). StyleGAN2-ADA can generate multiple classes of Ti–6Al–4V microstructures from limited data (Nimmal Haribabu et al., 2023), although GANs remain susceptible to training instability and mode collapse (Thanh-Tung & Tran, 2020; Durall et al., 2020). In contrast, diffusion models offer stable training and flexible conditioning mechanisms (Ho et al., 2020) and have been applied to complex microstructure reconstruction

and conditional generation (Azqadan et al., 2023; Düreth et al., 2023; Fernandez-Zelaia et al., 2024; Lee & Yun, 2024), as well as to modeling the relationships among heat-treatment conditions, microstructures, and mechanical properties in the T55511 titanium alloy (Zhu et al., 2025).

Existing approaches primarily address microstructure reconstruction, data augmentation, or generation from a single type of condition (Iyer et al., 2019; Nimmal Haribabu et al., 2023; Azqadan et al., 2023). In tasks jointly driven by composition, processing conditions, mechanical properties, and microstructure type, low-dimensional global conditions cannot directly constrain phase-region layouts, boundary locations, or local morphology. StructRDM therefore introduces a phase-structure map as an intermediate representation and separates structural modeling from texture generation to improve structural plausibility and controllability under multi-source conditions.

3 METHOD

The overall framework of StructRDM is illustrated in Figure 1.

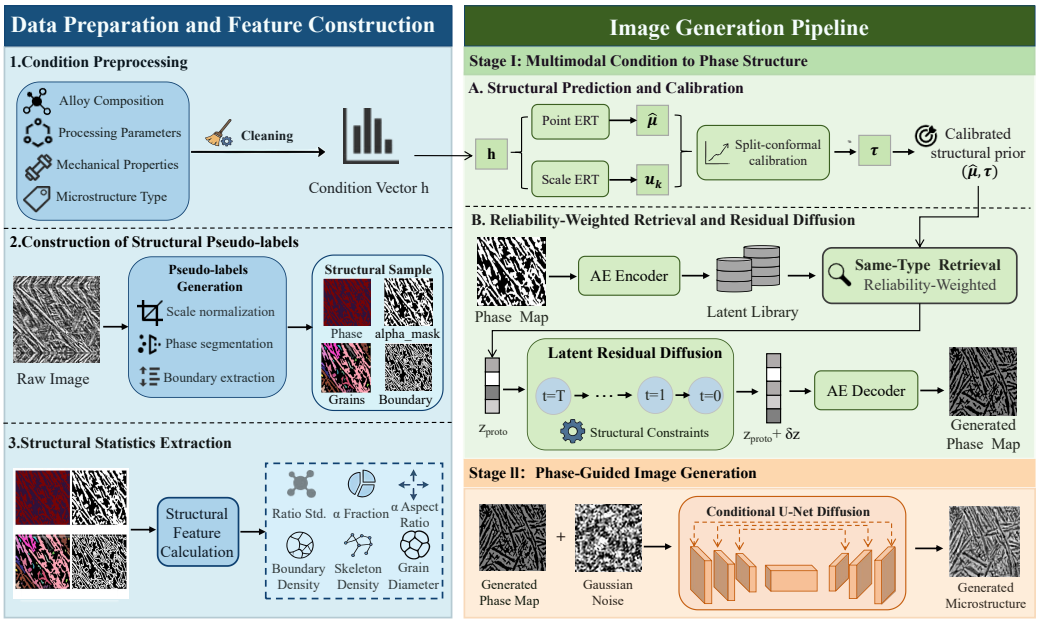

Figure 1: Overall framework of StructRDM.

3.1 PROBLEM FORMULATION AND OVERALL FRAMEWORK

We study titanium alloy microstructure generation conditioned on multi-source material information. Given a dataset

$$\mathcal{D} = \{(\mathbf{c}_i, \mathbf{p}_i, \mathbf{m}_i, o_i, \mathbf{x}_i)\}_{i=1}^{N}, \tag{1}$$

where $\mathbf{c}_i$, $\mathbf{p}_i$, $\mathbf{m}_i$, and o_i denote the alloy composition, processing parameters, mechanical properties, and microstructure type, respectively, and $\mathbf{x}_i \in \mathbb{R}^{H\times W}$ is the corresponding grayscale microstructure image. We denote the multi-source material condition as

$$\mathbf{y}_i = (\mathbf{c}_i, \mathbf{p}_i, \mathbf{m}_i, o_i). \tag{2}$$

Our objective is to learn the conditional distribution $p(\mathbf{x} \mid \mathbf{y})$. Because low-dimensional global conditions cannot directly constrain phase-region layouts, connectivity, and boundary morphology, StructRDM introduces a phase-structure map $\mathbf{s}$ as an intermediate spatial representation and factorizes the generation process as

$$p(\mathbf{x} \mid \mathbf{y}) = \sum_{\mathbf{s}} p_\psi(\mathbf{x} \mid \mathbf{s})\, p_\phi(\mathbf{s} \mid \mathbf{y}), \tag{3}$$

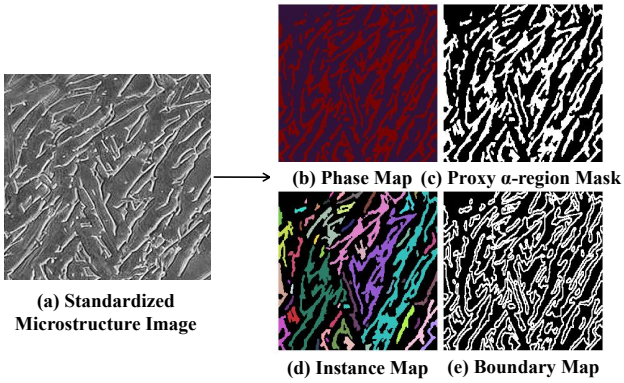

Figure 2: Automatically constructed structural pseudo-labels and auxiliary representations.

where $p_\phi(\mathbf{s} \mid \mathbf{y})$ represents the first stage, which generates a phase-structure map from multi-source material conditions, and $p_\psi(\mathbf{x} \mid \mathbf{s})$ represents the second stage, which generates grayscale textures and local morphology using the phase-structure map as a spatial condition.

3.2 Structural Pseudo-Labels and Statistical Representation

Because pixel-level phase annotations are unavailable for titanium alloy microstructure images, we construct structural pseudo-labels using an automated image-processing pipeline. Images collected from different sources are first normalized to a common reference physical scale and resampled to 256×256 pixels, followed by mild local contrast enhancement and Gaussian smoothing. Discrete phase-structure maps $\mathbf{s}_i$ are then generated by combining intensity- and gradient-based features, unsupervised clustering, and morphological refinement. Based on these maps, additional structural representations, including proxy α-region masks, boundaries, skeletons, and connected regions, are extracted, as illustrated in Figure 2. These automatically constructed representations serve only as structural pseudo-labels rather than expert-validated metallurgical phase annotations.

From these structural pseudo-labels, we further extract a six-dimensional structural-statistics vector:

$$\mathbf{a}_i = \left[a_i^{\mathrm{prs}}, a_i^{\alpha}, a_i^{\mathrm{ar}}, a_i^{\mathrm{gd}}, a_i^{\mathrm{bd}}, a_i^{\mathrm{bsl}}\right]^\top. \tag{4}$$

Its components respectively denote the standard deviation of phase proportions, proxy α-region area fraction, mean aspect ratio of proxy α regions, mean equivalent diameter of proxy α regions, boundary density, and boundary-skeleton length density. This vector summarizes phase composition, regional morphology, structural scale, and boundary complexity, and serves as a low-dimensional descriptor linking material-condition prediction to spatial-structure generation.

3.3 Multi-Source Condition Representation and Uncertainty-Aware Structural-Statistics Prediction

We use alloy composition, processing parameters, mechanical properties, and microstructure type as conditions for structural prediction. The numerical features comprise 13 elemental-composition variables, eight heat-treatment parameters, and five mechanical-property variables. Missing processing and mechanical-property fields are zero-filled and accompanied by binary validity masks. This yields a 39-dimensional numerical representation:

$$\mathbf{u}_i = [\mathbf{c}_i, \widetilde{\mathbf{p}}_i, \mathbf{q}_i^p, \widetilde{\mathbf{m}}_i, \mathbf{q}_i^m] \in \mathbb{R}^{39}, \tag{5}$$

where $\mathbf{q}_i^p$ and $\mathbf{q}_i^m$ are the corresponding validity masks. The microstructure type is represented using one-hot encoding. The final multi-source condition representation is defined as

$$\mathbf{h}_i = [\mathbf{u}_i, \mathrm{OneHot}(o_i)]\,. \tag{6}$$

Based on this representation, we employ two Extremely Randomized Trees (ERT) regressors (Geurts et al., 2006) to predict the structural-statistics point estimates and conditional error scales, respec-

tively. The point predictor R_θ outputs the conditional mean of the six-dimensional structural-statistics vector:

$$\widehat{\boldsymbol{\mu}}_i = R_\theta(\mathbf{h}_i). \tag{7}$$

To prevent in-sample residuals from underestimating prediction errors, we perform five-fold cross-fitting within the proper-training set to obtain out-of-fold residuals. The error-scale predictor U_ω learns the absolute out-of-fold residuals after normalization by the corresponding training-set standard deviations and log transformation. Its outputs are then converted back to the original units to obtain the dimension-wise conditional error estimates $u_k(\mathbf{y})$.

We then calibrate the predicted error scales on an independent calibration set using adaptive split-conformal prediction (Lei et al., 2018). For each calibration sample, the nonconformity score is defined as

$$s_{i,k} = \frac{|a_{i,k} - \widehat{\mu}_{i,k}|}{\max\left(u_{i,k}, \epsilon_k\right)}, \tag{8}$$

which yields the following prediction interval at the target coverage level $1-\alpha$:

$$C_k(\mathbf{y}) = \left[\widehat{\mu}_k(\mathbf{y}) - q_{1-\alpha,k} u_k(\mathbf{y}),\ \widehat{\mu}_k(\mathbf{y}) + q_{1-\alpha,k} u_k(\mathbf{y})\right], \tag{9}$$

where $q_{1-\alpha,k}$ is the corresponding finite-sample quantile of the calibration scores, and the interval endpoints are clipped to the feasible domain when necessary. Under the exchangeability assumption, each interval provides marginal coverage for the corresponding structural statistic. Subsequent generation uses the point prediction $\widehat{\boldsymbol{\mu}}$ to characterize the target structure and the half-width of the 90% prediction interval, $\tau_k = q_{0.9,k} u_k$, to represent dimension-wise reliability. A larger τ_k indicates lower confidence in the corresponding attribute during prototype retrieval and residual generation.

3.4 Uncertainty-Weighted Prototype Retrieval and Latent Residual Diffusion

Given the predicted structural statistics $\widehat{\boldsymbol{\mu}}$ and the calibrated half-widths $\boldsymbol{\tau}$ of their 90% prediction intervals, StructRDM first retrieves real structural prototypes that match the target condition from the training structure library. It then employs latent residual diffusion to generate conditional corrections from the retrieved prototypes toward the target structure. Specifically, a pretrained and frozen structure autoencoder (E_s, D_s) encodes phase-structure maps into a continuous latent space, from which the prototype library is constructed as

$$\mathcal{B} = \{(\mathbf{z}_i, o_i, \mathbf{a}_i)\}_{i=1}^{N_{\text{train}}}, \qquad \mathbf{z}_i = E_s(\mathbf{s}_i), \tag{10}$$

where o_i and $\mathbf{a}_i$ denote the microstructure class and structural statistics, respectively. Retrieval is restricted to candidate samples of the same microstructure class, and the query sample itself is excluded during training.

After standardizing the structural statistics and interval half-widths using training-set statistics, the reliability weight of structural dimension k and the distance to candidate prototype i are defined as

$$\rho_k(\mathbf{y}) = \text{clip}\left(\frac{1}{\widetilde{\tau}_k(\mathbf{y})^2 + \epsilon}, \rho_{\min}, \rho_{\max}\right), \qquad d_i(\mathbf{y}) = \frac{\sum_k w_k \rho_k(\mathbf{y}) \left(\widetilde{\mu}_k(\mathbf{y}) - \widetilde{a}_{i,k}\right)^2}{\sum_k w_k \rho_k(\mathbf{y}) + \epsilon}. \tag{11}$$

A wider prediction interval results in a lower weight for the corresponding structural attribute during prototype matching. The model retrieves the top-K prototypes according to d_i. During training, a single prototype is sampled according to the retrieval probabilities, whereas during inference, each of the top-K prototypes is used separately as an anchor to generate a candidate structure.

For training sample i, the prototype-anchored latent residual is defined as $\mathbf{r}_i = \mathbf{z}_i - \mathbf{z}_{\text{proto},i}$ and is standardized channel-wise using statistics computed from the training residuals, yielding $\bar{\mathbf{r}}_i$. Forward diffusion is then performed in the standardized residual space:

$$\bar{\mathbf{r}}_t = \sqrt{\bar{\alpha}_t}\, \bar{\mathbf{r}}_i + \sqrt{1-\bar{\alpha}_t}\, \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{12}$$

The residual diffusion model is jointly conditioned on the prototype latent representation and a reliability-aware structural condition:

$$\mathbf{c}(\mathbf{y}) = \left[\widetilde{\boldsymbol{\mu}}(\mathbf{y}), \log\left(1 + \widetilde{\boldsymbol{\tau}}(\mathbf{y})\right)\right], \qquad \widehat{\boldsymbol{\epsilon}} = \epsilon_\phi\left(\bar{\mathbf{r}}_t, t, \mathbf{z}_{\text{proto}}, \mathbf{c}(\mathbf{y})\right). \tag{13}$$

The calibrated interval half-widths thus serve two roles in structure generation: they downweight unreliable structural attributes during prototype retrieval and communicate the dimension-wise predictive reliability to the residual diffusion model.

The overall training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{noise}} + \lambda_{\text{clean}}\mathcal{L}_{\text{clean}} + \lambda_{\text{proto}}\mathcal{L}_{\text{proto}} + \gamma(e)\mathcal{L}_{\text{struct}}. \tag{14}$$

Here, $\mathcal{L}_{\text{clean}}$ constrains the recovered residual, $\mathcal{L}_{\text{proto}}$ prevents the generated structure from deviating excessively from the retrieved prototype, and $\mathcal{L}_{\text{struct}}$ enforces consistency in structural statistics and morphology. Their detailed definitions and weights are provided in the appendix. The structural-regularization weight $\gamma(e)$ gradually increases during training to avoid interfering with residual learning at early stages.

During inference, the model generates a residual anchored to each of the top-K retrieved prototypes. After inverse standardization and magnitude constraint, each residual is added back to its corresponding prototype latent representation. The frozen structure decoder then produces the candidate phase-structure maps:

$$\widehat{\mathbf{s}}^{(j)} = D_s\left(\mathbf{z}^{(j)}\right), \qquad j = 1, \ldots, K. \tag{15}$$

3.5 Phase-Structure-Conditioned Microstructure Generation

The second stage models $p_\psi(\mathbf{x} \mid \mathbf{s})$. Apart from the diffusion timestep, it uses the phase-structure map as its spatial condition and learns to synthesize grayscale textures corresponding to the given phase layout. For a target image $\mathbf{x}_0$ and its corresponding structure map $\mathbf{s}$, the forward process and noise prediction are

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \qquad \widehat{\boldsymbol{\epsilon}} = \epsilon_\psi(\mathbf{x}_t, t, \mathbf{s}). \tag{16}$$

The structure map is normalized to $[-1, 1]$ and concatenated with the noisy image along the channel dimension. It is then projected to three spatial resolutions and injected into the corresponding U-Net encoder layers through multiplicative gating and residual addition. At feature level l,

$$\widetilde{\mathbf{F}}_l = \mathbf{F}_l \odot \text{sigmoid}(G_l(\mathbf{s}_l)) + G_l(\mathbf{s}_l), \tag{17}$$

where $\mathbf{s}_l$ is the structure map at the resolution of feature $\mathbf{F}_l$, G_l is the condition-projection module, and sigmoid is the gating function. The U-Net residual blocks are otherwise modulated only by the diffusion-timestep embedding.

Training uses structures automatically constructed from the original micrographs, whereas inference uses phase structures generated by the first stage. To improve robustness to this distribution shift, noise, blur, morphological transformations, translation, and local dropout are introduced progressively during training. The objective combines min-SNR-weighted noise prediction (Hang et al., 2023), image reconstruction, and structural and textural auxiliary constraints:

$$\mathcal{L}_{\text{img}} = \mathcal{L}_{\text{MinSNR}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + w(t)\lambda_{\text{tex}}\mathcal{L}_{\text{tex}}. \tag{18}$$

$\mathcal{L}_{\text{rec}}$ constrains the grayscale and low-frequency distributions of the reconstructed image, while $\mathcal{L}_{\text{tex}}$ aggregates edge, gradient, high-frequency-energy, local-contrast, and orientation-texture constraints. The timestep weight $w(t)$ increases at low noise levels so that auxiliary constraints act primarily when texture is more recoverable. Detailed terms and weights are provided in the appendix. At inference, DDIM sampling (Song et al., 2021) generates a grayscale microstructure image for each first-stage candidate:

$$\widehat{\mathbf{x}}^{(j)} \sim p_\psi(\mathbf{x} \mid \widehat{\mathbf{s}}^{(j)}), \qquad j = 1, \ldots, K. \tag{19}$$

4 Experiments

We evaluate StructRDM from three aspects: structural-statistics prediction and uncertainty calibration, phase-structure generation, and final microstructure generation.

4.1 Experimental Setup

4.1.1 Dataset and Data Splits

We collected and curated 3,187 titanium-alloy microstructure images from publicly available literature, covering four representative microstructure classes: Widmanstätten, bimodal, basketweave, and equiaxed microstructures. After preliminary screening and multi-source information matching, we retained 1,109 paired condition–microstructure samples with sufficiently complete information for condition–structure modeling, including alloy composition, processing parameters, mechanical properties, and microstructure type. To reduce potential information leakage caused by visually related images, the data were split by source publication, such that samples originating from the same paper were assigned to the same subset. Of these samples, 170 were reserved as the test set, while the remaining 939 samples were used for model development. The test set was excluded from training, validation, conformal calibration, and prototype-library construction.

4.1.2 Baselines and Ablations

We compare StructRDM with two conditional diffusion baselines. **Ikeda-cDDPM** (Ikeda et al., 2025) directly generates microstructures conditioned on alloy composition, processing parameters, mechanical properties, and microstructure type, without using structural statistics, phase structures, or prototypes. **Stats-DDPM** instead uses the predicted six-dimensional structural statistics as global conditions. Our ablations include **w/o Reliability Weighting**, which assigns equal reliability to all structural dimensions and removes prediction-interval conditioning from residual diffusion, and **w/o Prototype Anchoring**, which removes prototype retrieval and the prototype-residual representation and directly generates the complete latent phase structure under the same conditions. **StructRDM** denotes the full model.

4.1.3 Implementation Details

The structural-statistics point predictor and error-scale predictor use ERTs with 800 and 600 trees, respectively. We construct 90% and 95% split-conformal prediction intervals. Phase-structure generation uses a 200-step diffusion schedule and Top-3 prototype retrieval. During training, one prototype is sampled according to the retrieval probabilities; during inference, the three retrieved prototypes are used separately to generate candidate structures, without latent fusion. Final image generation uses a 200-step diffusion schedule and 80-step DDIM sampling to produce 256×256 grayscale images. Additional implementation details are provided in the Appendix.

4.1.4 Evaluation Metrics

We evaluate structural descriptor prediction using MAE, RMSE, R^2, and Pearson correlation, and uncertainty calibration using error correlation, grouped MAE ratio, PICP, and NMPIW. Phase-structure generation is assessed using Macro Dice, SCE, Small Component Error, S_2 and L_2 errors, polynomial-kernel MMD^2, and scientific precision and recall. Final microstructure generation is evaluated using MAE, Gradient W1, Autocorrelation MAE, polynomial-kernel MMD^2, and precision and recall. Detailed definitions are provided in the Appendix.

4.2 Structural-Statistics Prediction and Uncertainty Calibration

The point-prediction results are presented in Table 1. The mean R^2 and Pearson correlation coefficient are 0.1848 and 0.4351, respectively, confirming the limited predictability of sample-specific structures from global conditions. Among them, boundary-skeleton length density and boundary density show relatively higher predictability, whereas the mean aspect ratio of proxy α regions is less predictable. This variation further motivates dimension-wise reliability modeling.

The predictive uncertainty and interval-calibration results are presented in Table 2. Predictive uncertainty is positively correlated with the absolute prediction error, and samples with high uncertainty exhibit substantially larger mean prediction errors than those with low uncertainty. Meanwhile, the empirical coverage rates of the 90% and 95% prediction intervals are 0.9059 and 0.9500, respectively, closely matching their nominal coverage levels. These results indicate that split-conformal calibration provides reasonable marginal coverage.

Table 1: Prediction performance for the six structural descriptors.

Structural descriptor	MAE ↓	RMSE ↓	R^2 ↑	Corr ↑
Std. of phase proportions	0.0597	0.0767	0.2026	0.4561
Proxy α-region area fraction (%)	15.7827	19.6796	0.1243	0.3924
Mean aspect ratio of proxy α regions	0.8540	1.2390	0.0652	0.2963
Mean equivalent diameter of proxy α regions (μm)	3.9732	9.1407	0.1329	0.3785
Boundary density	0.0707	0.0931	0.2564	0.5122
Boundary-skeleton length density	0.0226	0.0299	0.3272	0.5750
Overall	—	—	0.1848	0.4351

Table 2: Structural-prediction uncertainty and interval-calibration results.

Metric	Result
Uncertainty–absolute-error Pearson	0.2406
Uncertainty–absolute-error Spearman	0.2493
High/low-uncertainty MAE ratio	1.5413
90% PICP	0.9059
90% NMPIW	2.494
95% PICP	0.9500
95% NMPIW	3.190

4.3 Effects of Reliability Weighting and Prototype-Anchored Residual Representation

All methods are evaluated using the generation corresponding to the Rank-1 candidate. Quantitative results are reported in Table 3, and representative visual comparisons are shown in Figure 3.

Table 3: Effects of reliability weighting and prototype anchoring on phase-structure generation.

Method	Macro Dice ↑	SCE ↓	Small Comp. Err. ↓	S_2 MAE ↓	L_2 MAE ↓	MMD^2 ↓	Sci. Prec. ↑	Sci. Rec. ↑
w/o Reliability Weighting	0.3191	**0.6822**	0.1143	0.0826	0.0534	0.4490	0.8529	**0.7471**
w/o Prototype Anchoring	0.2899	0.8069	0.4206	0.1184	0.0689	3.2286	0.0000	0.0000
StructRDM	**0.3212**	0.6853	**0.1058**	**0.0796**	**0.0516**	**0.4151**	**0.9176**	0.7000

Table 3 shows that, compared with w/o Prototype Anchoring, StructRDM reduces Small Component Error from 0.4206 to 0.1058, and reduces the S_2 error, L_2 error, and polynomial-kernel MMD^2 from 0.1184, 0.0689, and 3.2286 to 0.0796, 0.0516, and 0.4151, respectively. These results demonstrate that the prototype-anchored residual representation effectively suppresses structural fragmentation and improves spatial continuity. Reliability weighting lowers MMD^2, S_2 error, L_2 error, and Small Component Error, while increasing Scientific Precision from 0.8529 to 0.9176, indicating complementary gains in local integrity and distributional precision.

As shown in Figure 3, w/o Prototype Anchoring produces more fragmented phase regions, with pronounced local discontinuities along region boundaries, whereas StructRDM preserves more continuous phase layouts and clearer dominant morphology. The overall structures generated by w/o Reliability Weighting and the full StructRDM are similar, with their differences mainly concentrated in local connectivity, boundary discontinuities, and small-region distributions.

4.4 Effect of Explicit Phase-Structure Conditioning

Table 4 and Figure 4 present the quantitative and qualitative results, respectively. StructRDM achieves the highest Recall and F1 and the lowest Gradient W1 and Autocorrelation MAE, although its Precision, paired MAE, and MMD^2 are less favorable than those of Ikeda-cDDPM. This trade-off indicates broader distributional coverage and improved spatial-statistical consistency, while reflecting the difficulty of pixel-wise comparison in one-to-many conditional generation. Figure 4 further shows that StructRDM generates microstructures with more continuous lamellar structures, clearer region contours, and better-organized boundary morphologies.

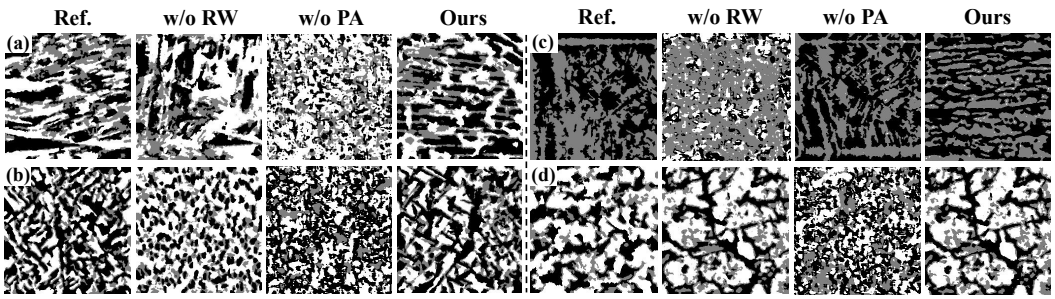

Figure 3: Qualitative comparison of phase-structure generation under different ablation settings. From left to right: reference pseudo-label (Ref.), StructRDM without reliability weighting (w/o RW), StructRDM without prototype anchoring (w/o PA), and the full StructRDM (Ours).

Table 4: Paired error, distributional quality, and spatial morphology of the generated microstructures under different conditioning representations.

Method	MAE ↓	MMD^2 ↓	Precision ↑	Recall ↑	F1 ↑	Gradient W1 ↓	Autocorrelation MAE ↓
Ikeda-cDDPM	0.2291	**5.3387**	**0.8529**	0.2706	0.4108	0.1269	0.0356
Stats-DDPM	**0.2225**	12.8275	0.7706	0.0882	0.1583	0.1495	0.0418
StructRDM	0.2829	7.0181	0.5588	**0.4882**	**0.5212**	**0.1002**	**0.0230**

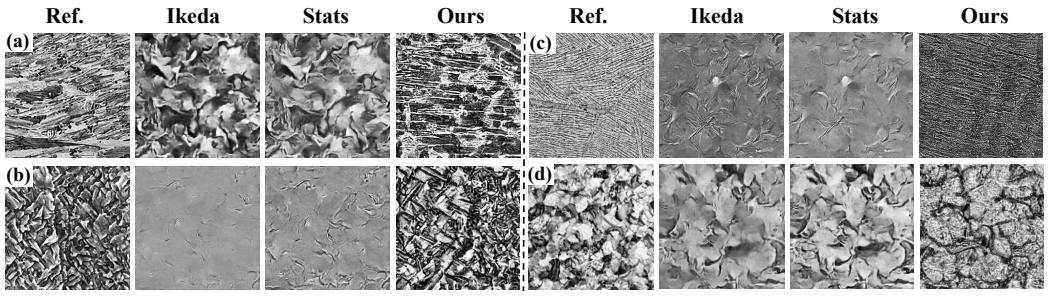

Figure 4: Qualitative comparison of final microstructure generation under different condition representations. From left to right: reference microstructure (Ref.), Ikeda-cDDPM (Ikeda), Stats-DDPM (Stats), and StructRDM (Ours).

Overall, global structural statistics describe aggregate structural attributes, whereas an explicit phase-structure map additionally provides spatial-layout information, thereby improving the structural and morphological consistency of generated microstructures.

5 Conclusion

We propose StructRDM, a structure-first framework for titanium alloy microstructure generation from limited multi-source materials data. StructRDM generates phase structures through reliability-aware prototype retrieval and latent residual diffusion, followed by structure-conditioned texture generation. Experimental results show that prototype-anchored residual modeling effectively reduces structural deviations and phase-region fragmentation, while calibrated uncertainty provides complementary reliability information for prototype retrieval and residual generation. Explicit phase-structure conditioning further helps preserve the boundary distribution and overall morphological consistency of the generated microstructures. The current method remains limited by noise in structural pseudo-labels, limited dataset size, and class imbalance. Future work will explore expert-validated structural annotations and more efficient prototype-conditioned generation methods.

AI USE STATEMENT

Generative AI tools were used during manuscript preparation to assist with Chinese–English translation, language editing, and improving the clarity and organization of the presentation. The authors critically reviewed and revised all AI-assisted text and independently verified its consistency with the methodology, experimental results, and underlying research. AI assistance was limited to manuscript preparation and did not extend to the execution of the research or the production of experimental results. The authors take full responsibility for the final content of this paper.

REFERENCES

M. P. Anderson, D. J. Srolovitz, G. S. Grest, and P. S. Sahni. Computer simulation of grain growth—i. kinetics. *Acta Metallurgica*, 32(5):783–791, 1984. doi: 10.1016/0001-6160(84)90151-2.

Erfan Azqadan, Hamid Jahed, and Arash Arami. Predictive microstructure image generation using denoising diffusion probabilistic models. *Acta Materialia*, 261:119406, 2023. doi: 10.1016/j.actamat.2023.119406.

Ramin Bostanabad, Yichi Zhang, Xiaolin Li, Tucker Kearney, L. Catherine Brinson, Daniel W. Apley, Wing Kam Liu, and Wei Chen. Computational microstructure characterization and reconstruction: Review of the state-of-the-art techniques. *Progress in Materials Science*, 95:1–41, 2018. doi: 10.1016/j.pmatsci.2018.01.005.

Ruijin Cang, Yaopengxiao Xu, Shaohua Chen, Yongming Liu, Yang Jiao, and Max Yi Ren. Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design. *Journal of Mechanical Design*, 139(7):071404, 2017. doi: 10.1115/1.4036649.

Daniel A. Cogswell and W. Craig Carter. Thermodynamic phase-field model for microstructure with multiple components and phases: The possibility of metastable phases. *Physical Review E*, 83: 061602, 2011. doi: 10.1103/PhysRevE.83.061602.

Ricard Durall, Agris Chatzimichailidis, Peter Labus, and Janis Keuper. Combating mode collapse in GAN training: An empirical analysis using hessian eigenvalues. *arXiv preprint arXiv:2012.09673*, 2020.

Christian Düreth, Paul Seibert, Dennis Rücker, Stephanie Handford, Markus Kästner, and Maik Gude. Conditional diffusion-based microstructure reconstruction. *Materials Today Communications*, 35:105608, 2023. doi: 10.1016/j.mtcomm.2023.105608.

Patxi Fernandez-Zelaia, Jiahao Cheng, Jason Mayeur, Amir Koushyar Ziabari, and Michael M. Kirka. Digital polycrystalline microstructure generation using diffusion probabilistic models. *Materialia*, 33:101976, 2024. doi: 10.1016/j.mtla.2023.101976.

D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi. Microstructure reconstructions from 2-point statistics using phase-recovery algorithms. *Acta Materialia*, 56(5):942–948, 2008. doi: 10.1016/j.actamat.2007.10.044.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine Learning*, 63:3–42, 2006. doi: 10.1007/s10994-006-6226-1.

G. D. Goh, X. Huang, S. Huang, J. L. J. Thong, J. J. Seah, and W. Y. Yeong. Data imputation strategies for process optimization of laser powder bed fusion of Ti6Al4V using machine learning. *Materials Science in Additive Manufacturing*, 2(1):50, 2023.

Tiankai Hang, Shuyang Gu, Chen Li, Jianmin Bao, Dong Chen, Han Hu, Xin Geng, and Baining Guo. Efficient diffusion training via min-SNR weighting strategy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7441–7451, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

A. Ikeda, R. Higuchi, T. Yokozeki, K. Endo, Y. Kojima, M. Suzuki, and M. Muramatsu. Conditional diffusion model for inverse prediction of process parameters and dendritic microstructures from mechanical properties. *Scientific Reports*, 15(1):37147, 2025.

Akshay Iyer, Bapi Dey, Abhishek Dasgupta, Wei Chen, and Anirban Chakraborty. A conditional generative model for predicting material microstructures from processing methods. *arXiv preprint arXiv:1910.02133*, 2019.

LiMei Kang and Chao Yang. A review on high-strength titanium alloys: Microstructure, strengthening, and properties. *Advanced Engineering Materials*, 21(8):1801359, 2019. doi: 10.1002/adem.201801359.

Kang Hyun Lee and Gun Jin Yun. Microstructure reconstruction using diffusion-based generative models. *Mechanics of Advanced Materials and Structures*, 31(18):4443–4461, 2024. doi: 10.1080/15376494.2023.2198528.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116.

J.-J. Luo, B. Qin, and B. Wang. Generative adversarial network-based data augmentation for improving creep life prediction of high-temperature titanium alloys. *Materials Science and Engineering: A*, 950:149513, 2026.

Gerd Lütjering and James C. Williams. *Titanium*. Springer Berlin Heidelberg, 2 edition, 2007. doi: 10.1007/978-3-540-73036-1.

Elia Marin and Alex Lanzutti. Biomedical applications of titanium alloys: A comprehensive review. *Materials*, 17(1):114, 2023. doi: 10.3390/ma17010114.

Gowtham Nimmal Haribabu, Jeyapriya Thimukonda Jegadeesan, Chiranjib Bhattacharya, and Bikramjit Basu. A deep adversarial approach for the generation of synthetic titanium alloy microstructures with limited training data. *Computational Materials Science*, 230:112512, 2023. doi: 10.1016/j.commatsci.2023.112512.

Pallavi Pushp, S. M. Dasharath, and C. Arati. Classification and applications of titanium and its alloys. *Materials Today: Proceedings*, 54:537–542, 2022. doi: 10.1016/j.matpr.2022.01.008.

Andreas E. Robertson, Conlain Kelly, Michael Buzzy, and Surya R. Kalidindi. Local–global decompositions for conditional microstructure generation. *Acta Materialia*, 253:118966, 2023. doi: 10.1016/j.actamat.2023.118966.

Camilo A. F. Salvador, Eloa L. Maia, Fernando H. Costa, Julian D. Escobar, and João P. Oliveira. A compilation of experimental data on the mechanical properties and microstructural features of Ti-alloys. *Scientific Data*, 9:188, 2022. doi: 10.1038/s41597-022-01283-9.

T. Searles, J. Tiley, A. Tanner, R. Williams, B. Rollins, E. Lee, S. Kar, R. Banerjee, and H. L. Fraser. Rapid characterization of titanium microstructural features for specific modelling of mechanical properties. *Measurement Science and Technology*, 16(1):60–69, 2005.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Hoang Thanh-Tung and Truyen Tran. Catastrophic forgetting and mode collapse in GANs. In *2020 International Joint Conference on Neural Networks*, pp. 1–10. IEEE, 2020.

J. Tiley, T. Searles, E. Lee, S. Kar, R. Banerjee, J. C. Russ, and H. L. Fraser. Quantification of microstructural features in α/β titanium alloys. *Materials Science and Engineering: A*, 372(1–2):191–198, 2004. doi: 10.1016/j.msea.2003.12.008.

Michael D. White, Gowtham Nimmal Haribabu, Jeyapriya Thimukonda Jegadeesan, Bikramjit Basu, Philip J. Withers, and Chris P. Race. Exploring descriptors for titanium microstructure via digital fingerprints from variational autoencoders. *Computational Materials Science*, 240: 112992, 2024. doi: 10.1016/j.commatsci.2024.112992.

Olga Wodo, Scott Broderick, and Krishna Rajan. Microstructural informatics for accelerating the discovery of processing–microstructure–property relationships. *MRS Bulletin*, 41(8):603–609, 2016. doi: 10.1557/mrs.2016.161.

Zijiang Yang, Xiaolin Li, L. Catherine Brinson, Alok N. Choudhary, Wei Chen, and Ankit Agrawal. Microstructural materials design via deep adversarial learning methodology. *Journal of Mechanical Design*, 140(11):111416, 2018. doi: 10.1115/1.4041371.

C. L. Y. Yeong and Salvatore Torquato. Reconstructing random media. *Physical Review E*, 57: 495–506, 1998. doi: 10.1103/PhysRevE.57.495.

P. Zhao, Y. Wang, B. Jiang, H. Zhang, X. Cheng, and Q. Fan. Neural network modeling of titanium alloy composition–microstructure–property relationships based on multimodal data. *Materials Science and Engineering: A*, 879:145202, 2023. doi: 10.1016/j.msea.2023.145202.

Changlu Zhou, Ruihao Yuan, Baolong Su, Jiangkun Fan, Bin Tang, Pingxiang Zhang, and Jinshan Li. Creep rupture life prediction of high-temperature titanium alloy using cross-material transfer learning. *Journal of Materials Science & Technology*, 178:39–47, 2024.

Yongcheng Zhu, Zhaoxin Du, Tianhao Gong, Zhiyong Yue, Shuzhi Zhang, Jun Cheng, Xudong Kang, Qiang Zeng, and Jingshun Liu. Accurate reconstruction and prediction of T55511 titanium alloy microstructure using DDPM model and quantitative evaluation. *Progress in Natural Science: Materials International*, 35(4):712–723, 2025. doi: 10.1016/j.pnsc.2025.04.002.

A Dataset and Structural Pseudo-Label Construction

A.1 Dataset Overview

We collected and curated 3,187 titanium-alloy microstructure images from publicly available research literature and simultaneously extracted the material conditions associated with each image. Each sample contains alloy composition, processing and heat-treatment parameters, mechanical properties, microstructure type, and the corresponding microstructure image. The material conditions used by the model are summarized in Table A.1.

After image-quality inspection, multi-source material-information matching, and duplicate removal, 1,109 condition–microstructure paired samples were retained. Their distribution across microstructure types is reported in Table A.2. The class imbalance mainly reflects differences in the availability of different microstructure types in the published literature.

Missing processing, heat-treatment, and mechanical-property values in the literature are zero-filled and distinguished from actual zero values using corresponding validity masks. The vector representation and encoding of the material conditions are described in the main paper. Image preprocessing and structural pseudo-label construction are detailed in Appendix A.2.

A.2 Structural Pseudo-Label Construction

Image preprocessing and pixel descriptors. First, according to the scale information associated with each image, the microstructure images are normalized to a reference physical resolution of $0.25\ \mu\mathrm{m/pixel}$ and resampled to 256×256 pixels. Color images are converted to grayscale. Contrast-limited adaptive histogram equalization (CLAHE) is then applied to enhance local contrast, with a clip limit of 1.8 and an 8×8 tile grid, followed by Gaussian smoothing using a 5×5 kernel.

For each pixel p, we construct a four-dimensional feature descriptor consisting of the smoothed grayscale intensity, the mean intensity within a 15×15 neighborhood, the magnitude of the 3×3 Sobel gradient, and the absolute response of the 3×3 Laplacian:

$$\mathbf{x}_p = \left[I_p,\ \bar{I}_p^{15\times15},\ G_p^{\mathrm{Sobel}},\ \left| L_p^{\mathrm{Laplacian}} \right| \right]^{\mathrm{T}}. \tag{A.1}$$

To reduce image-to-image variations in brightness and contrast, each feature channel is standardized independently within each image:

$$\widetilde{x}_{p,k} = \frac{x_{p,k} - \mu_k}{\sigma_k + 10^{-6}}, \tag{A.2}$$

where μ_k and σ_k denote the mean and standard deviation of the k-th feature channel in the current image, respectively.

Unsupervised pixel clustering and label refinement. K-means is applied to the standardized pixel descriptors. The number of clusters is set to 3 for bimodal microstructures and 2 for the other three microstructure types. The number of initializations is 10, and the random seed is fixed to 42. After clustering, the cluster labels are reordered according to the mean grayscale intensity of each cluster, from low to high, to maintain a consistent label ordering across images.

To remove isolated pixels and small noisy regions, a 3×3 median filter is first applied to the cluster-label map, followed by 8-connected component analysis for each label. Connected regions with an area smaller than 35 pixels are treated as local fragments and reassigned to the most frequent label in their 3×3 dilated neighborhood. The resulting discrete phase-structure map is denoted by

$$s(p) \in \{0, \ldots, C-1\}, \tag{A.3}$$

where C is the number of clusters associated with the current microstructure type.

Proxy α region and phase-boundary construction. After morphological refinement, the cluster with the highest mean grayscale intensity is operationally defined as the proxy α region. Its binary mask is written as

$$M_\alpha(p) = \mathbb{I}[s(p) = l_\alpha], \tag{A.4}$$

Table A.1: Material conditions used by the model.

Category	Variables	Unit
Alloy composition	Ti, Al, V, Mo, Nb, Zr, Sn, Cr, Fe, Si, O, C, N	wt.%
Processing and heat treatment	Solution temperature, solution time, primary-aging temperature, primary-aging time, secondary-aging temperature, secondary-aging time, annealing temperature, annealing time	°C / h
Mechanical properties	Yield strength (YS), ultimate tensile strength (UTS), elongation (EL), hardness, elastic modulus (E)	MPa, MPa, %, HV, GPa
Microstructure type	Bimodal, basketweave, equiaxed, Widmanstätten	–

Table A.2: Sample distribution of the condition–microstructure paired dataset.

Microstructure type	Samples	Proportion
Bimodal	761	68.62%
Basketweave	137	12.35%
Equiaxed	113	10.19%
Widmanstätten	98	8.84%
Total	**1,109**	**100.00%**

where l_α denotes the cluster label with the highest mean grayscale intensity. An 8-connected component analysis is applied to M_α, and only connected regions with an area of at least 10 pixels are retained to reduce the influence of isolated noise on region-level statistics. The proxy α region is used solely to construct a consistent structural intermediate representation and should not be interpreted as a true α phase confirmed by compositional analysis, diffraction characterization, or manual annotation.

The phase-boundary map is determined by label transitions in the discrete phase-structure map. A pixel location is marked as a boundary if horizontally or vertically adjacent pixels have different labels. The initial boundary is skeletonized to obtain a single-pixel-wide boundary skeleton and is then dilated once using a 3×3 structuring element to produce the binary phase-boundary map used for structural statistics and spatial conditioning.

Each microstructure image therefore yields a discrete phase-structure map s, a proxy α mask M_α, and a binary phase-boundary map M_b. These representations characterize the spatial distribution of major structural regions, the geometry of the proxy α regions, and the interfaces between different structural regions, respectively.

B Six-Dimensional Structural Statistical Representation

Based on the phase-structure map, proxy α mask, and boundary representation constructed in Appendix A, we further extract a six-dimensional structural statistical vector:

$$\mathbf{a}_i = \left[a_i^{\mathrm{prs}}, a_i^{\alpha}, a_i^{\mathrm{ar}}, a_i^{\mathrm{gd}}, a_i^{\mathrm{bd}}, a_i^{\mathrm{bsl}}\right]^{\mathrm{T}}. \tag{B.1}$$

The six components correspond to the standard deviation of phase ratios, proxy α-region area fraction, mean aspect ratio of proxy α regions, mean equivalent diameter of proxy α regions, boundary density, and boundary-skeleton length density, respectively. Together, they characterize regional composition, morphology, characteristic scale, and interfacial complexity.

B.1 Regional Composition Statistics

Let the phase-structure map s have spatial size $H \times W$ and contain C structural classes. The area fraction of class c is defined as

$$r_c = \frac{1}{HW} \sum_{u=1}^{H} \sum_{v=1}^{W} \mathbb{I}[s(u,v) = c], \qquad c = 0, \ldots, C-1. \tag{B.2}$$

Table A.3: Main parameters used for structural pseudo-label construction.

Item	Setting
Reference physical resolution	$0.25\ \mu\mathrm{m/pixel}$
Image resolution	256×256
CLAHE	clip limit $= 1.8$, 8×8 tiles
Gaussian smoothing	5×5
Local-mean window	15×15
Gradient feature	Sobel 3×3
Laplacian feature	3×3
K-means initializations	10
K-means random seed	42
Clusters for bimodal microstructures	3
Clusters for other microstructures	2
Label smoothing	3×3 median filtering
Small-region correction threshold	35 pixels
Connected-component definition	8-connectivity
Valid proxy α region threshold	10 pixels
Boundary extraction	Horizontal/vertical label transitions
Boundary processing	Skeletonization + one 3×3 dilation

Because the class area fractions sum to one, their mean is $1/C$. The standard deviation of phase ratios is defined as

$$a^{\mathrm{prs}} = \sqrt{\frac{1}{C}\sum_{c=0}^{C-1}\left(r_c - \frac{1}{C}\right)^2}. \tag{B.3}$$

This quantity measures the dispersion of area fractions across structural classes. A larger value indicates that the image area is more concentrated in one or a few structural classes, whereas a smaller value indicates a more balanced distribution.

Let M_α denote the binary mask of the proxy α region. Its area fraction is defined as

$$a^{\alpha} = 100 \times \frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W} M_\alpha(u, v). \tag{B.4}$$

The quantity a^α is expressed as a percentage and describes the fraction of the image occupied by proxy α regions. The proxy α mask is obtained using the existing segmentation pipeline, and its area fraction is then computed by the feature-extraction pipeline from the resulting binary mask. Because it is computed from structural pseudo-labels, it represents an image-level proxy-region fraction rather than the true α-phase volume fraction determined by independent phase characterization.

B.2 Regional Morphology and Scale Statistics

We perform 8-connected component analysis on the proxy α mask and retain valid regions with an area of at least 10 pixels. Let w_j and h_j denote the width and height of the axis-aligned bounding box of the j-th valid region. Its aspect ratio is defined as

$$q_j = \frac{\max(w_j, h_j)}{\max(\min(w_j, h_j), 1)}. \tag{B.5}$$

The mean aspect ratio of proxy α regions in an image is

$$a^{\mathrm{ar}} = \frac{1}{J}\sum_{j=1}^{J} q_j, \tag{B.6}$$

where J is the number of valid regions. The connected regions and their bounding boxes are obtained from the proxy α mask by the existing segmentation and feature-extraction pipeline. This quantity describes the elongation of proxy α regions: values close to 1 indicate relatively equiaxed regions, whereas larger values indicate increasingly elongated or lamellar morphology.

The mean equivalent diameter describes the two-dimensional characteristic scale of proxy α regions. We extract the external contours of the proxy α mask and let A_j denote the pixel area enclosed by the j-th valid contour. The physical pixel size is

$$\delta = 0.25\ \mu\mathrm{m/pixel}. \tag{B.7}$$

The equivalent circular diameter of region j is defined as

$$d_j = 2\delta\sqrt{\frac{A_j}{\pi}}. \tag{B.8}$$

Averaging over all valid regions gives

$$a^{\mathrm{gd}} = \frac{1}{J}\sum_{j=1}^{J} d_j. \tag{B.9}$$

The unit of a^{gd} is μm. This statistic is the equivalent circular diameter computed from two-dimensional proxy α regions and is used to compare relative structural scales across images. It should not be interpreted as a true three-dimensional grain size measured using standardized metallographic procedures.

B.3 Boundary Statistics

Let M_b denote the binary phase-boundary map obtained in Appendix A. The boundary density is defined as

$$a^{\mathrm{bd}} = \frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W} M_b(u,v). \tag{B.10}$$

This metric measures the fraction of the image occupied by the boundary region after one 3×3 dilation and is therefore influenced by both the amount of interface and its local spatial distribution.

Let M_{skel} denote the corresponding single-pixel-wide boundary skeleton. The boundary-skeleton length density is defined as

$$a^{\mathrm{bsl}} = \frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W} M_{\mathrm{skel}}(u,v). \tag{B.11}$$

This metric uses the number of skeleton pixels per unit image area as an approximation of boundary-network length density. Horizontal, vertical, and diagonal connections are not assigned different physical length weights during computation.

Overall, the six-dimensional statistical vector provides a compact description of structural pseudo-labels from four perspectives: regional composition, regional morphology, characteristic scale, and boundary complexity. It is subsequently used for structural-statistics prediction, reliability modeling, prototype retrieval, and conditional generation.

C Model Implementation Details

This section provides additional implementation details for the phase-structure autoencoder and the two diffusion models. The overall generation pipeline, condition construction, and spatial-condition injection strategy are described in the main paper.

C.1 Phase-Structure Autoencoder

The phase-structure autoencoder is a deterministic convolutional autoencoder rather than a variational or vector-quantized model. Each 256×256 phase map is converted to a three-channel one-hot representation. The encoder uses a base width of 32 channels and three stride-2 convolutional downsampling layers. Its feature widths are 32, 64, 128, and 128, followed by a 1×1 projection to an eight-channel latent tensor of spatial size 32×32. Each convolutional block contains two 3×3 convolutions, Group Normalization, and GELU activation. The decoder mirrors the three resolution

changes using nearest-neighbor upsampling and produces three-channel phase logits. Consequently, the latent representation used by the residual diffusion model has shape $8 \times 32 \times 32$ for each sample.

The autoencoder is trained through three reconstruction paths. The main path decodes the latent tensor without encoder skip connections, matching its use during latent diffusion. An auxiliary path uses encoder skip features to facilitate the learning of local boundaries, and a robustness path decodes a perturbed latent tensor. For each path, the reconstruction objective is

$$\mathcal{L}_{\text{branch}} = \mathcal{L}_{\text{CE}} + 0.5\mathcal{L}_{\text{Dice}} + 0.5\mathcal{L}_{\text{edge}}, \tag{C.1}$$

where $\mathcal{L}_{\text{CE}}$ is class-weighted cross-entropy, $\mathcal{L}_{\text{Dice}}$ is the soft multiclass Dice loss, and $\mathcal{L}_{\text{edge}}$ is the sum of horizontal and vertical first-order probability-gradient errors. Class weights are inversely proportional to the square root of the training-set pixel frequencies, normalized to unit mean, and clipped to $[0.25, 5]$.

The complete autoencoder objective is

$$\mathcal{L}_{\text{AE}} = \mathcal{L}_{\text{noskip}} + 0.2\mathcal{L}_{\text{skip}} + 0.5\mathcal{L}_{\text{noisy}} + 10^{-5}\|\mathbf{z}\|_2^2. \tag{C.2}$$

Gaussian noise with standard deviations 0.01 and 0.03 is applied to the one-hot input and latent tensor, respectively. Training uses AdamW for up to 150 epochs with batch size 8, an initial learning rate of 2×10^{-4}, weight decay 10^{-4}, gradient clipping at 1.0, and ReduceLROnPlateau scheduling. The checkpoint with the lowest no-skip validation loss is retained. After training, the encoder and decoder are frozen; latent diffusion uses only the no-skip decoder path $D_s(\mathbf{z})$.

C.2 Prototype Residual Diffusion Network

The prototype residual diffusion model employs a lightweight U-Net. The noisy latent residual and prototype latent are concatenated along the channel dimension and fed into the network. The encoder contains two residual blocks, with feature channels increasing from 64 to 128. The middle block maintains 128 channels, and the decoder restores the original latent spatial resolution through one upsampling operation together with a skip connection from the encoder. The output layer maps the features back to the original latent-channel dimension.

Each residual block consists of Group Normalization, SiLU activation, and 3×3 convolutions, with dropout set to 0.05. Diffusion timesteps are represented using sinusoidal positional embeddings. The reliability condition is mapped through a two-layer multilayer perceptron and added to the timestep embedding, followed by linear projection into each residual block.

During latent-space composition, the generated residual is element-wise clipped to $[-1.2, 1.2]$, multiplied by a residual scaling factor of 0.4, and then applied to the prototype latent. A prototype anchoring coefficient of 0.5 is used during composition to limit the deviation of the generated structure from the retrieved prototype.

C.3 Phase-Structure-Conditioned Image Generation Network

The final image-generation model adopts a two-level U-Net with a base channel width of 96. The encoder uses feature-channel widths of 96, 192, and 384, while the decoder follows a symmetric channel configuration and combines encoder features at matching resolutions through skip connections.

Residual blocks use Group Normalization and SiLU activation. Diffusion timesteps are first mapped through sinusoidal positional encoding and a two-layer multilayer perceptron, and then modulate each residual block using FiLM. The output head consists of a 3×3 convolution, SiLU activation, and a 1×1 convolution, producing a single-channel noise prediction.

Perturbation of the conditioning phase map starts from epoch 10 and increases linearly from zero to full strength over the subsequent 40 epochs. This curriculum first allows the model to learn the basic correspondence between phase structure and grayscale appearance and then gradually improves its robustness to first-stage generation errors.

C.4 Training Configuration

Both diffusion models are trained using AdamW, automatic mixed precision, gradient clipping, and exponential moving average (EMA). The main training settings are summarized in Table C.1.

Table C.1: Main training configurations of the two diffusion models.

Configuration	Prototype residual diffusion	Phase-structure-conditioned image generation
Epochs	150	150
Batch size	16	4
Initial learning rate	2×10^{-4}	3×10^{-5}
Weight decay	1×10^{-4}	1×10^{-4}
Learning-rate schedule	ReduceLROnPlateau	Cosine annealing
Minimum learning rate	–	1×10^{-6}
Gradient clipping	1.0	1.0
EMA decay	0.999	0.999
Inference sampling steps	50	80
Validation split	15%	10%
Early-stopping patience	25 validations	30 validations
Random seed	42	42

For prototype residual diffusion, the learning rate is halved when the validation loss does not improve for 7 consecutive epochs, and the best EMA parameters are selected according to the total validation loss. The phase-structure-conditioned image-generation model uses cosine learning-rate decay and likewise saves the best EMA parameters according to the total validation loss.

D Implementation of the Loss Functions

The main paper presents the overall objectives for the two training stages. This section provides the implementation details, weights, and schedules of the auxiliary losses that are not expanded in the main text.

D.1 Auxiliary Losses for Phase-Structure Generation

Let $\bar{\mathbf{r}}$ denote the standardized ground-truth latent residual and $\widehat{\bar{\mathbf{r}}}_0$ the latent residual recovered by the model from the noisy input. The latent residual reconstruction loss is

$$\mathcal{L}_{\text{clean}} = \left\| \widehat{\bar{\mathbf{r}}}_0 - \bar{\mathbf{r}} \right\|_1, \tag{D.1}$$

with a weight of 0.5.

The predicted latent, target latent, and prototype latent are converted into phase-probability maps by a frozen structural decoder, denoted by $\widehat{\mathbf{P}}$, $\mathbf{P}_{\text{tar}}$, and $\mathbf{P}_{\text{pro}}$, respectively. The prototype-space constraint consists of phase-structure reconstruction, prototype preservation, relative improvement, and residual-magnitude regularization:

$$\mathcal{L}_{\text{proto}} = \mathcal{L}_{\text{phase}} + 0.15\mathcal{L}_{\text{keep}} + 0.25\mathcal{L}_{\text{better}} + 0.0008\mathcal{L}_{\Delta}. \tag{D.2}$$

The phase-structure reconstruction term is a confidence-weighted probability-map discrepancy:

$$\mathcal{L}_{\text{phase}} = \frac{\sum_p \omega_p \left\| \widehat{\mathbf{P}}(p) - \mathbf{P}_{\text{tar}}(p) \right\|_1}{\sum_p \omega_p}, \tag{D.3}$$

where

$$\omega_p = \text{clip}\left(1 - \frac{H(\mathbf{P}_{\text{tar}}(p))}{\log C},\, 0.05,\, 1\right). \tag{D.4}$$

Pixels with lower entropy therefore receive larger weights.

The prototype-preservation and relative-improvement terms are

$$\mathcal{L}_{\text{keep}} = \left\| \widehat{\mathbf{P}} - \mathbf{P}_{\text{pro}} \right\|_1, \tag{D.5}$$

Table D.1: Weights of auxiliary losses for phase-structure generation.

Loss term	**Weight**
Latent residual reconstruction $\mathcal{L}_{\mathrm{clean}}$	0.5
Phase-structure reconstruction $\mathcal{L}_{\mathrm{phase}}$	1.0
Prototype preservation $\mathcal{L}_{\mathrm{keep}}$	0.15
Relative improvement $\mathcal{L}_{\mathrm{better}}$	0.25
Residual-magnitude regularization $\mathcal{L}_{\Delta}$	0.0008
Phase-ratio standard deviation	2.0
Proxy α area fraction	2.0
Boundary density	4.0
Probability entropy	0.020
Boundary curvature	0.002
First-order smoothness	0.010

and

$$\mathcal{L}_{\mathrm{better}} = \max\left[0,\, \mathrm{MAE}(\widehat{\mathbf{P}}, \mathbf{P}_{\mathrm{tar}}) - \mathrm{MAE}(\mathbf{P}_{\mathrm{pro}}, \mathbf{P}_{\mathrm{tar}})\right]. \tag{D.6}$$

The term $\mathcal{L}_{\mathrm{keep}}$ prevents the generated structure from deviating excessively from the retrieved prototype, while $\mathcal{L}_{\mathrm{better}}$ encourages the corrected result not to be worse than the prototype. The term $\mathcal{L}_{\Delta}$ is the mean absolute magnitude of the de-standardized latent residual and is used to constrain the correction amplitude.

Structural constraints are computed on decoded soft phase-probability maps and include three differentiable statistics: phase-ratio standard deviation, proxy α area fraction, and boundary density. In addition, probability entropy, boundary curvature, and first-order smoothness are used to suppress ambiguous labels and local boundary oscillations. The corresponding weights are listed in Table D.1.

The structural constraints are activated from epoch 5 and increased linearly over the following 25 epochs. During training, the structural term is adaptively scaled according to the magnitudes of the base diffusion loss and the prototype-space loss, targeting approximately 35% of their sum while being capped at 70% of the sum of the remaining losses.

The mean aspect ratio, mean equivalent diameter, and boundary-skeleton length density depend on discrete operations such as connected-component analysis or skeletonization and are therefore not directly used for gradient-based optimization. The complete six-dimensional structural statistics are nevertheless used for structural prediction, prototype retrieval, and experimental evaluation.

D.2 Auxiliary Losses for Microstructure Image Generation

The final image-generation model uses the Min-SNR noise-prediction loss with a clipping parameter of 5. The reconstruction component described in the main paper consists of image reconstruction and low-frequency reconstruction, while the texture component includes soft edges, multi-scale gradients, Laplacian response, high-frequency statistics, local contrast, gradient orientation, and Gabor response. The loss weights used in training are listed in Table D.2.

The low-frequency reconstruction term compares images after 11×11 average pooling. Local contrast is computed using the standard deviation within a 9×9 neighborhood, and directional texture is measured using multi-orientation 15×15 Gabor filters. The final weight of the total-variation loss is set to zero to avoid additional smoothing of high-frequency microstructural morphology.

Except for global grayscale contrast, texture-related losses are weighted according to the diffusion timestep:

$$w(t) = \mathrm{clip}\left(1 - \frac{t}{T-1},\, 0.25,\, 1\right). \tag{D.7}$$

Consequently, high-noise timesteps focus primarily on diffusion denoising, whereas edge and texture constraints contribute more strongly at low-noise timesteps where fine details can be recovered.

Table D.2: Weights of auxiliary losses for microstructure image generation.

Loss term	**Weight**
Image reconstruction	0.20
Low-frequency reconstruction	0.08
Soft edge	0.15
Multi-scale gradient	0.08
Laplacian response	0.01
High-frequency response	0.25
Local contrast	0.10
Gradient orientation	0.08
Gabor response	0.12
Global grayscale contrast	0.05
Total variation	0

E Evaluation Metrics and Computation Settings

All methods are evaluated on the same fixed set of 170 test samples. For phase-structure generation, the main results use the candidate associated with the Rank-1 retrieved prototype; the ground-truth test image is never used to select among top-K candidates. Paired metrics compare corresponding generated and reference samples, whereas set-level metrics compare the generated and real test distributions.

E.1 Structural-Statistics Prediction and Uncertainty

Point-prediction performance is evaluated using MAE, RMSE, R^2, and Pearson correlation. Each metric is computed separately for the six structural dimensions; the overall R^2 and correlation reported in the main paper are arithmetic means across dimensions.

For sample i and structural dimension k, the absolute prediction error is

$$e_{i,k} = |a_{i,k} - \widehat{a}_{i,k}| \tag{E.1}$$

where $u_{i,k}$ denotes the corresponding predicted error scale. Pearson and Spearman correlations between $u_{i,k}$ and $e_{i,k}$ are computed per dimension and then averaged over the six dimensions.

Using the median predicted error scale, test samples are also divided into high- and low-uncertainty groups, $\mathcal{H}_k$ and $\mathcal{L}_k$. Their MAE ratio is

$$R_k^{\mathrm{HL}} = \frac{\mathrm{MAE}(\mathcal{H}_k)}{\mathrm{MAE}(\mathcal{L}_k)}, \tag{E.2}$$

A value greater than 1 indicates that the high-uncertainty group has a larger empirical prediction error.

For calibrated intervals $[L_{i,k}, U_{i,k}]$, prediction interval coverage probability (PICP) and normalized mean prediction interval width (NMPIW) are

$$\mathrm{PICP}_k = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left[a_{i,k} \in [L_{i,k}, U_{i,k}]\right], \tag{E.3}$$

and

$$\mathrm{NMPIW}_k = \frac{1}{N s_k}\sum_{i=1}^{N} \left(U_{i,k} - L_{i,k}\right). \tag{E.4}$$

where s_k is the training-set standard deviation of statistic k. Results are averaged over the six dimensions. PICP should approach the nominal 90% or 95% coverage; at comparable coverage, lower NMPIW indicates tighter intervals.

E.2 Phase-Structure Generation

Phase-structure generation is evaluated in terms of paired consistency, local integrity, spatial correlation, and set-level distribution.

Paired consistency. Macro Dice is computed over valid phase classes and then averaged across classes and test samples. Higher values indicate greater pixel-wise overlap. Structural Condition Error (SCE) measures agreement with the predicted six-dimensional condition:

$$\mathrm{SCE}_i = \frac{1}{6}\sum_{k=1}^{6} \frac{\left|a_{i,k}^{\mathrm{gen}} - \widehat{\mu}_{i,k}\right|}{\max(\sigma_k, \varepsilon)}, \tag{E.5}$$

where $a_{i,k}^{\mathrm{gen}}$ is extracted from the generated map, $\widehat{\mu}_{i,k}$ is the predicted target, and σ_k is the training-set standard deviation. Lower SCE indicates better condition satisfaction and, unlike Macro Dice, does not require pixel-wise alignment.

Local integrity. For non-background classes, we use 8-connectivity, discard components smaller than 10 pixels, and define valid components of at most 32 pixels as small. Small Component Error is the test-set mean absolute difference between the generated and reference proportions of small components. Lower values indicate better agreement in local fragmentation.

Spatial correlation. For each phase indicator, S_2 is its periodic two-point autocorrelation computed by FFT. We radially average it into 24 equal-width bins over $[0, \min(H, W)/2)$ and report the mean absolute difference between generated and reference curves. The lineal-path function $L_2(\ell)$ is the probability that a segment of length ℓ lies entirely within one phase. It is evaluated horizontally and vertically, without periodic wrapping, at 16 rounded lengths spanning 1–48 pixels. Both errors are averaged over three fixed phase-label slots and all test samples; lower values indicate better spatial-scale and continuity agreement.

Set-level distribution. Each phase map is represented by a 175-dimensional vector containing, for each of three fixed phase-label slots, area fraction, Euler density, a 24-dimensional radial S_2 curve, and a 32-dimensional L_2 curve, plus global boundary density:

$$3 \times (1 + 1 + 24 + 32) + 1 = 175. \tag{E.6}$$

For two-label structures, the absent third slot is represented by an empty mask and hence by zero-valued features. Features are standardized dimension-wise using the real test-set mean and standard deviation. For real features $\{\mathbf{x}_i\}_{i=1}^n$ and generated features $\{\mathbf{y}_j\}_{j=1}^m$, the unbiased squared Maximum Mean Discrepancy is

$$\widehat{\mathrm{MMD}}_u^2 = \frac{\sum_{i\neq i'} k(\mathbf{x}_i, \mathbf{x}_{i'})}{n(n-1)} + \frac{\sum_{j\neq j'} k(\mathbf{y}_j, \mathbf{y}_{j'})}{m(m-1)} - \frac{2\sum_{i=1}^n \sum_{j=1}^m k(\mathbf{x}_i, \mathbf{y}_j)}{nm}, \tag{E.7}$$

using the third-order polynomial kernel

$$k(\mathbf{x}, \mathbf{y}) = \left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{y}}{d} + 1\right)^3, \tag{E.8}$$

where $d = 175$. Lower MMD^2 indicates closer agreement between the two distributions.

Scientific Precision and Recall use Euclidean distance in the same standardized feature space. After excluding self-distances, each sample's neighborhood radius is its third-nearest-neighbor distance. Precision is the fraction of generated samples covered by at least one real neighborhood; Recall is the fraction of real samples covered by at least one generated neighborhood. Higher values are better.

E.3 Final Microstructure Images

All grayscale images are normalized to $[0, 1]$. Pixel MAE is reported only as an auxiliary paired measure because conditional microstructure generation is inherently one-to-many.

Gradient distribution. Gradient W1 uses 3×3 Sobel magnitudes and normalized 32-bin histograms. Each image uses its own 99.5th gradient-magnitude percentile as the histogram upper bound. The one-dimensional Wasserstein distance is computed over bin indices, divided by 31, and averaged over the test set. It therefore measures relative histogram-shape differences rather than absolute gradient-scale differences; lower values are better.

Table E.1: Feature representation for set-level evaluation of final microstructure images.

Feature	Dim.	Primary characterization
Grayscale histogram	32	Grayscale distribution
Gradient-magnitude histogram	32	Local grayscale variation
Gradient-orientation histogram	18	Orientation distribution
Uniform LBP histogram	10	Local texture
Radial autocorrelation curve	24	Spatial-correlation scale
Radial log-power spectrum	24	Frequency distribution
Structure-tensor statistics	3	Orientation coherence and edge density
Gray-level co-occurrence matrix statistics	16	Contrast and homogeneity
Total	**159**	–

Spatial correlation. Autocorrelation MAE subtracts the image mean, computes periodic two-dimensional autocorrelation by FFT, normalizes by the zero-shift value, and radially averages the result into 24 bins. It is the mean absolute difference between paired radial curves; lower values indicate better agreement in texture periodicity and spatial-correlation scale.

Set-level distribution. Final images are compared in the 159-dimensional feature space summarized in Table E.1.

The grayscale and gradient-magnitude histograms use 32 bins; the magnitude-weighted orientation histogram uses 18 bins over $[0, \pi)$; and uniform LBP uses eight neighbors at radius 1 and 10 bins. Radial autocorrelation and radial log-power spectra each use 24 bins. Structure-tensor features use Gaussian smoothing with $\sigma = 2$ and comprise the mean and standard deviation of coherence and the fraction of pixels above the image-specific 75th gradient percentile. After quantization to 32 gray levels, symmetric normalized gray-level co-occurrence matrices are computed at distances 1 and 3 and orientations 0, $\pi/4$, $\pi/2$, and $3\pi/4$; contrast and homogeneity are retained.

The 159-dimensional features are standardized using the real test-set statistics. MMD^2 uses Equation E.7 with $d = 159$, while Precision and Recall use the same Euclidean third-nearest-neighbor construction as above. Their harmonic mean is

$$F_1 = \frac{2\ \mathrm{Precision}\,\mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{E.9}$$

Lower MMD^2 and higher Precision, Recall, and F_1 indicate better set-level agreement.