MORPHCONTROL: NUCLEAR FUEL ROD DEFECT IMAGE GENERATION VIA MORPHOLOGY PRIOR ADAPTATION AND QUALITY-FEEDBACK PRIOR REGULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Surface defect datasets for nuclear fuel rods are often scarce and class-imbalanced. Moreover, defects within the same category exhibit substantial variations in shape, scale, and local texture. These characteristics make it difficult for generative models to simultaneously preserve class consistency, intra-class diversity, and the structural fidelity of industrial backgrounds. Existing methods often rely on instance-level conditioning without explicitly modeling morphological patterns shared across samples of the same class. Furthermore, fixed conditioning strengths cannot readily accommodate different defect categories, backgrounds, and stochastic sampling states, potentially resulting in weak defect expression, excessive enhancement, or local artifacts. To address these limitations, we propose MorphControl, an industrial defect generation framework that integrates defect-morphology-prior-driven multi-condition adaptation with quality-feedback prior regulation. During training, MorphControl attenuates the original defects in reference images to construct defect-suppressed background conditions while keeping the pretrained diffusion backbone frozen. It also learns class-level defect morphology priors jointly from samples of the same category. These priors are combined with instance-dependent industrial backgrounds, geometric layouts, and textual semantics, allowing class-shared morphological knowledge to be recombined with diverse generation conditions. A region-weighted diffusion objective and a background-morphology decorrelation constraint further improve fine-grained defect modeling while reducing information coupling between conditions. During inference, MorphControl adaptively regulates the injection strength of the defect morphology prior according to regional feature consistency and artifact risk. It then selects the best candidate satisfying the quality constraints. Experiments show that MorphControl outperforms existing methods in generation fidelity and intra-class diversity, achieving a lower KID and a higher IC-LPIPS. Augmenting the training set with the generated samples further improves downstream defect detection, increasing both mAP_{50} and $\mathrm{mAP}_{50:95}$, with larger gains for under-represented defect categories. These results demonstrate that class-level defect morphology priors and inference-time quality-feedback regulation improve both synthesis quality and downstream utility under limited and class-imbalanced industrial data.

Keywords: Industrial Defect Generation; Nuclear Fuel Rods; Diffusion Models; Morphology Priors; Quality Feedback.

1 INTRODUCTION

Nuclear fuel-rod cladding forms a primary barrier against the release of fission products, making its manufacturing quality important to fuel integrity. Surface inspection must therefore identify dents, scratches, and adhered foreign material before nonconforming fuel rods enter service. Developing reliable inspection models, however, requires sufficiently diverse defect data. Such data are difficult to collect in manufacturing environments because defects are rare and unevenly distributed across categories. In our dataset, dents account for 62% of the samples, whereas foreign matter and

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.

scratches account for only 20% and 18%, respectively. This imbalance limits the observed variations of minority classes and motivates generative data augmentation.

Existing synthesis methods provide only partial solutions to this problem. Image-manipulation approaches construct pseudo-defects using predefined transformations, but their diversity and realism are constrained by hand-crafted operations (Li et al., 2021; Zavrtanik et al., 2021). GAN-based methods learn defect distributions from data, yet adversarial training can be unstable when anomalous samples are limited (Duan et al., 2023; Wei et al., 2023). Pretrained diffusion models improve generation quality and controllability by leveraging large-scale generative priors (Hu et al., 2024; Zhang et al., 2024; Jin et al., 2025; Li et al., 2026; Dai et al., 2025). These methods commonly condition generation on text, spatial masks, or individual reference images (Zhang et al., 2023; Mou et al., 2024; Ye et al., 2023; He et al., 2024), but three coupled limitations remain in the fuel-rod inspection setting. First, globally encoding a defect-containing reference image can entangle localized defect cues with background appearance. Second, instance-level conditions do not explicitly aggregate morphological patterns shared by samples of the same category and may overemphasize the appearance of a particular reference instance. Third, fixed conditioning strengths cannot adapt to variations in defect category, background, and sampling state, potentially producing weak defects, excessive enhancement, or local artifacts.

These limitations suggest that defect synthesis should separate category-level morphology from instance-specific generation conditions. A class-level morphology representation should aggregate patterns shared across same-category samples without copying a particular defect instance. It should then be recombined with diverse industrial backgrounds and spatial layouts, while its contribution is adjusted according to the quality of the generated result. This formulation preserves category consistency through shared morphology statistics and instance-level diversity through backgrounds, layouts, and stochastic sampling.

Based on this insight, we propose MorphControl, a diffusion-based framework that combines Defect Morphology Prior-driven Multi-condition Adaptation (DMPA) with Quality-Feedback Prior Regulation (QFPR). During training, DMPA suppresses the original defect in each reference image to obtain a cleaner background-domain condition and learns a category-specific morphology prior jointly from same-category samples. Background, morphology, geometry, and text conditions are separately encoded and injected into a frozen SDXL backbone. A region-weighted diffusion loss strengthens the learning signal from small defect regions, while a background–morphology decorrelation loss reduces similarity between the two condition representations. During inference, QFPR keeps the background, category, prompt, target region, and random noise fixed and adjusts only the morphology-prior strength. It evaluates candidate results using target-region feature consistency and artifact risk, and selects the best candidate satisfying the quality constraint.

Our main contributions are summarized as follows:

1. We introduce MorphControl, which recombines a class-level defect morphology prior with instance-specific backgrounds and spatial layouts, and regulates the contribution of this prior according to generation quality.
2. We develop DMPA to learn category-specific morphology priors from same-category samples while reducing defect leakage through the background branch. Region-weighted learning and background–morphology decorrelation further support localized defect modeling and condition separation.
3. We develop QFPR to adapt the morphology-prior strength at inference time using regional consistency and artifact feedback. Experiments show that MorphControl improves generation fidelity and intra-class diversity over the compared baselines, while the generated samples also improve downstream defect-detection performance.

2 Related Work

2.1 Industrial Defect Synthesis

Early industrial anomaly-synthesis methods construct pseudo-defects through patch manipulation, texture transfer, or procedural perturbations. CutPaste pastes cropped patches onto normal images, while DRAEM combines external textures with irregular masks (Li et al., 2021; Zavrtanik et al.,

2021). These methods require no real anomalous samples but remain constrained by predefined transformations, which may produce limited morphology or inconsistent boundaries. GAN-based methods instead learn defect appearances from data. DFMGAN adapts a data-efficient StyleGAN2 backbone using defect-aware residual blocks, whereas DCDGANc controls defect categories and styles through class labels and random codes (Duan et al., 2023; Wei et al., 2023). Their effectiveness can nevertheless be limited by unstable training and insufficient mode coverage when few anomalous samples are available.

3 METHOD

3.1 PRELIMINARIES AND PROBLEM FORMULATION

Latent Diffusion and SDXL. Stable Diffusion XL (SDXL) is a text-to-image generative model built on the latent diffusion model (LDM) framework (Rombach et al., 2022; **?**). Its main components include a variational autoencoder (VAE), a conditional denoising U-Net, and text encoders. SDXL first uses the encoder $\mathcal{E}$ of a pretrained VAE to compress an image x into a low-dimensional latent representation $z_0 = \mathcal{E}(x)$, where $z_0 \in \mathbb{R}^{h\times w\times c}$. Following the forward diffusion process (Ho et al., 2020), Gaussian noise is then added to z_0 to obtain z_t.

SDXL encodes a prompt T using two pretrained text encoders and combines the token-level and global text representations into $C^{\text{text}} = \tau_\varphi(T)$, where $C^{\text{text}} \in \mathbb{R}^{m\times d_\tau}$. Given the noisy latent z_t, diffusion timestep t, and text condition C^{text}, the conditional denoising U-Net ϵ_θ predicts the injected noise. Its basic training objective is $\mathcal{L}_{\text{SDXL}} = \mathbb{E}_{x,T,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(z_t, t, C^{\text{text}})\|_2^2\right]$. Cross-attention in the U-Net allows the denoising process to respond to the scene and defect semantics described by the prompt. During inference, the model starts from a Gaussian latent z_t and iteratively denoises it to obtain $\hat{z}_0$, which is mapped back to pixel space by the VAE decoder as $\hat{x} = \mathcal{D}(\hat{z}_0)$.

Although SDXL provides a strong general-purpose image-generation prior, text conditioning alone cannot precisely represent the fine-grained morphology, industrial imaging style, and local spatial layout of nuclear fuel-rod defects. Inspired by the industrial-domain adaptation strategy of PDIG (Li et al., 2025) and the decoupled image-conditioning mechanism of IP-Adapter (Ye et al., 2023), we augment SDXL with a class-level defect morphology prior, a defect-suppressed background condition, and a geometric layout condition. During training, the VAE, text encoders, image encoder, SDXL U-Net, and a pretrained fuel-rod-domain LoRA remain frozen. We optimize only the background projection, morphology prior bank, geometric encoder, and lightweight residual parameters in the multi-branch attention module. This design preserves the pretrained generative prior while reducing the cost of few-shot industrial-domain adaptation.

Problem Formulation. Given a training set with defect-category and bounding-box annotations,

$$\mathcal{D} = \{(I_i, y_i, B_i, T_i)\}_{i=1}^{N}, \tag{1}$$

where I_i is a real defect image, $y_i \in \{1, \ldots, K\}$ denotes its defect category, $B_i = (c_x, c_y, w, h)$ is a normalized bounding box, and T_i is the corresponding text description. Our objective is to learn the conditional defect distribution

$$p_\theta\Big(\hat{I} \mid I^{\text{bg}}, y, B, T\Big), \tag{2}$$

where I^{bg} is a defect-suppressed industrial background, y specifies the target defect category, B provides the region-level spatial layout, and T describes the industrial object and defect semantics.

Rather than using a single reference defect as a local appearance condition, MorphControl learns a class-level defect morphology prior jointly from training samples of the same category and combines it with the background, layout, and text conditions. During data generation, diverse defects within the same category can be produced by varying the random noise, background, and region layout. Within a single QFPR feedback loop, these conditions remain fixed and only the morphology-prior strength is adjusted. The input bounding box provides region-level layout guidance and is not assumed to be the exact pixel boundary of the generated defect.

3.2 MORPHCONTROL OVERVIEW

As shown in Figure 1, MorphControl consists of Defect Morphology Prior-Driven Multi-Condition Adaptation (DMPA) and Quality-Feedback Prior Regulation (QFPR). During training, DMPA learns

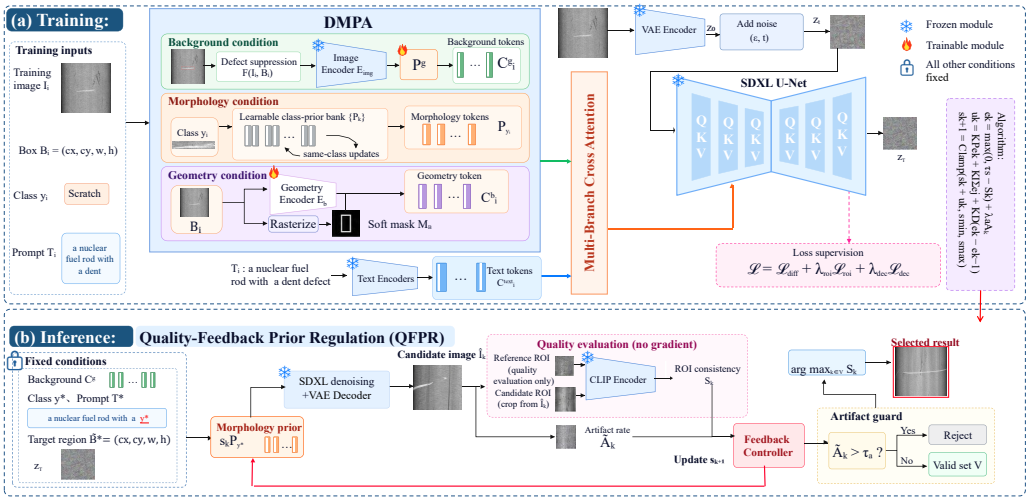


Figure 1: Overall framework of MorphControl. During training, DMPA learns and injects composable background-domain, class-level morphology, geometric, and textual conditions into a frozen SDXL backbone. During inference, QFPR adjusts the morphology-prior strength using region-level quality feedback.

composable representations of the industrial background, class-level defect morphology, geometric layout, and text semantics. During inference, QFPR adjusts the morphology-prior strength according to target-region feature consistency and artifact risk. The two components address the tendency of instance-level conditioning to reproduce a specific reference appearance and the inability of a fixed conditioning strength to consistently avoid under-generation and over-generation.

During training, the annotated defect is first suppressed to obtain an image from which the background-domain condition is extracted. Samples from the same category jointly update their corresponding morphology prior, while the bounding box and prompt provide region layout and high-level semantics, respectively. A multi-branch adapter injects these conditions into the frozen SDXL backbone. Training jointly optimizes a Min-SNR-weighted diffusion objective, a defect-region-weighted objective, and a background–morphology decorrelation constraint. During inference, a target region is sampled from the class-conditional empirical geometry distribution. QFPR then holds the background, category, prompt, target region, and random noise fixed while iteratively adjusting the morphology-prior strength and selecting a candidate that satisfies the quality constraints.

3.3 Defect Morphology Prior-Driven Multi-Condition Adaptation (DMPA)

MorphControl first constructs a defect-suppressed background condition. Directly encoding a defect-containing reference image as a global condition allows the original defect appearance to leak through the background branch, which interferes with the class-level morphology prior and geometric condition. To reduce this leakage, we create a feathered mask from an expanded annotation box and apply Gaussian blurring within the defect region while preserving the overall fuel-rod structure. The defect-suppressed background is defined as

$$I^{\mathrm{bg}} = F(I, B), \tag{3}$$

where I is the reference image, B denotes its defect bounding box, and $F(\cdot)$ is the defect-suppression operator. A frozen image encoder E_{img} and a trainable projection module P^g then extract the background-domain condition:

$$C^g = P^g\big(E_{\mathrm{img}}\big(I^{\mathrm{bg}}\big)\big)\,. \tag{4}$$

This branch preserves the material texture, structural arrangement, and imaging style of the fuel rod without explicitly representing the morphology of a particular defect. Background purification therefore retains the industrial context while reducing contamination of the global domain representation by the original local defect. A single reference defect contains both category-level regularities

and instance-specific appearance. Using it directly as a local condition can cause the model to reproduce the shape or texture of a particular instance. MorphControl instead maintains a learnable class-level defect morphology prior bank:

$$\mathcal{P} = \{P_1, \ldots, P_K\}, \qquad P_k \in \mathbb{R}^{M \times d}, \tag{5}$$

where K is the number of defect categories, M is the number of prior tokens assigned to each category, and d is the condition dimension. For a training sample i with category label y_i, the corresponding morphology prior is retrieved as

$$C_i^p = P_{y_i}. \tag{6}$$

All categories share the same diffusion backbone but maintain category-specific morphology prior parameters. Gradients from samples of the same category jointly shape the corresponding prior, allowing it to capture stable morphological patterns shared within that category. Instance-level variations in defect shape, scale, and texture arise from the random noise, background condition, geometric layout, and the resulting denoising trajectory. The learned prior therefore captures shared statistics of the class-level morphology distribution rather than encoding a compressed copy of any individual reference image. The class label y selects the corresponding category-specific morphology prior P_y, which encodes the visual characteristics of the target defect category but is not explicitly parameterized by its spatial location or extent. We represent the target region using a normalized bounding box

$$B = (c_x, c_y, w, h), \tag{7}$$

where c_x and c_y denote the normalized center coordinates, and w and h denote the normalized width and height. The bounding box is encoded as a geometry token and rasterized into a soft spatial mask:

$$C^b = E_b(B), \qquad M_B = \mathrm{Rasterize}(B). \tag{8}$$

The geometry token C^b represents the center, scale, and aspect ratio of the target region. At each U-Net block, M_B is resized to the corresponding feature resolution and used to spatially route the morphology-branch response toward the target region. The frozen SDXL text encoders map the prompt T to C^{text}, which represents the fuel-rod scene and the target defect category. Although both the class label and the text prompt identify the target category, they serve complementary roles: the class label retrieves the image-derived morphology prior P_y, whereas the text condition provides scene-level semantics and a textual description of the defect category.

To reduce interference among heterogeneous conditions, MorphControl injects the text, background, morphology, and geometry conditions through four separate cross-attention branches. For an intermediate U-Net feature H, the branches share a query derived from H, while their keys and values are computed independently from C^{text}, C^g, C^p, and C^b. We denote the resulting branch outputs by A_{text}, A_g, A_p, and A_b, respectively. These outputs are fused as

$$\widetilde{H} = H + \alpha_t A_{\mathrm{text}} + \alpha_g A_g + \alpha_b A_b + s\alpha_p \left(M_B \odot A_p\right), \tag{9}$$

where α_t, α_g, α_b, and α_p denote the weights of the text, background, geometry, and morphology branches, respectively. The morphology-prior strength s is fixed during training and adjusted by QFPR during inference. The resized soft mask M_B spatially gates the morphology branch, providing a soft regional constraint rather than an exact boundary of the generated defect.

During training, the frozen VAE encodes a real image I into a latent representation z_0. Gaussian noise ϵ is added at a randomly sampled timestep t to obtainz_t. The conditional diffusion network predicts the added noise under the complete condition set

$$\mathcal{C} = \left\{C^{\mathrm{text}}, C^g, C^p, C^b, M_B\right\}. \tag{10}$$

Because normal background regions occupy most of the image, their gradients can dominate the training objective and weaken the learning signal from small defect regions. We therefore increase the optimization weight within the annotated region and constrain the overlap between the background and morphology representations. The complete DMPA training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{diff}} + \lambda_{\mathrm{roi}}\mathcal{L}_{\mathrm{roi}} + \lambda_{\mathrm{dec}}\mathcal{L}_{\mathrm{dec}}, \tag{11}$$

where λ_{roi} and λ_{dec} control the contributions of region-weighted learning and condition decorrelation, respectively.

The diffusion term is a Min-SNR-weighted noise-prediction objective:

$$\mathcal{L}_{\text{diff}} = w_t \left\| \epsilon - \hat{\epsilon}_\theta \left(z_t, t; \mathcal{C} \right) \right\|_2^2, \tag{12}$$

where w_t is the Min-SNR weighting coefficient at timestep t, and $\hat{\epsilon}_\theta$ denotes the noise predicted by the conditional diffusion network.

For the region-weighted term, the soft bounding-box mask is resized to the latent-space resolution and used to linearly weight the noise-prediction error:

$$\mathcal{L}_{\text{roi}} = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{h,w} M^{\text{lat}}_{B_i,h,w} \left\| \epsilon_{i,:,h,w} - \hat{\epsilon}_{\theta,i,:,h,w} \right\|_2^2}{\sum_{h,w} M^{\text{lat}}_{B_i,h,w} + \delta}, \tag{13}$$

where $M^{\text{lat}}_{B_i}$ denotes the latent-space soft mask for sample i, N is the batch size, and δ is a small constant for numerical stability. This term prevents the large normal background from overwhelming the learning signal associated with fine-grained local defects.

The decorrelation term reduces redundant information shared by the background condition and the morphology prior. Let $\bar{C}^g_i$ and $\bar{C}^p_i$ denote the mean background token and mean morphology token of sample i, respectively. Their absolute cosine similarity is minimized as

$$\mathcal{L}_{\text{dec}} = \frac{1}{N} \sum_{i=1}^{N} \left| \cos \left(\bar{C}^g_i, \bar{C}^p_i \right) \right|, \tag{14}$$

where N is the batch size. This constraint encourages the background branch to retain domain and structural information while allowing the morphology prior to focus on category-specific defect appearance.

Class-balanced sampling is employed so that morphology priors associated with low-frequency categories receive sufficient updates. The SDXL backbone, VAE, text encoders, and image encoder remain frozen throughout training. Only the background projection module, morphology prior bank, geometric encoder, branch gates, and lightweight residual parameters of the multi-branch attention module are optimized.

3.4 Quality-Feedback Prior Regulation(QFPR)

After training, QFPR adaptively determines the injection strength of the class-level defect morphology prior without updating the network parameters. Given a target category y^*, the model first samples an initial region B^* from a class-conditional geometry distribution estimated solely from the training annotations. A constrained random perturbation then produces the target region $\widetilde{B}^*$:

$$B^* \sim \hat{p}(B \mid y^*), \qquad \widetilde{B}^* = \Pi_\Omega(B^* + \Delta B). \tag{15}$$

The distribution $\hat{p}(B \mid y^*)$ preserves category-dependent statistics of defect position, scale, and aspect ratio, while Π_Ω projects the perturbed region onto the valid image domain. During inference, $\widetilde{B}^*$ is rasterized into a soft mask that routes the morphology response toward the target region. The text and background branches continue to operate over the entire image. This region is internal guidance and is not assumed to be an exact detection boundary for the generated defect.

At feedback iteration k, the background condition C^g, target category y^*, prompt T, target region $\widetilde{B}^*$, and random seed remain fixed. Only the runtime injection strength s_k of the morphology prior is updated. The current candidate image is

$$\hat{I}_k = G_\theta \Big(z, T, C^g, s_k P_{y^*}, E_b(\widetilde{B}^*) \Big). \tag{16}$$

To determine whether the target region sufficiently expresses the defect morphology, we crop a reference region R_{ref} from a real defect image of the same category and a generated region R_k from the current candidate. A frozen CLIP image encoder E_f (Radford et al., 2021) measures their region-level feature consistency:

$$S_k = \cos(E_f(R_{\text{ref}}), E_f(R_k)). \tag{17}$$

The reference defect is used only for post-generation quality evaluation and is not supplied to the diffusion network as an instance-level appearance condition.

Maximizing region-level consistency alone can result in excessive prior injection. We therefore also compute a normalized artifact rate $\widetilde{A}_k$ and impose an artifact guard relative to the first candidate. The current implementation measures conspicuous color artifacts using the proportion of pixels with abnormal differences across RGB channels. A candidate is rejected by the safety gate when this proportion exceeds that of the first candidate by a predefined margin.

When the artifact guard is not triggered, QFPR constructs the feedback error from the shortfall between the target consistency τ_s and the current consistency S_k, with the normalized artifact rate acting as a soft quality penalty. A feedback controller then updates the morphology-prior strength (Åström & Murray, 2008):

$$e_k = \max(0, \tau_s - S_k) + \lambda_a \widetilde{A}_k, \tag{18}$$

$$u_k = K_P e_k + K_I \sum_{j=0}^{k} e_j + K_D(e_k - e_{k-1}), \tag{19}$$

$$s_{k+1} = \mathrm{Clamp}(s_k + u_k, s_{\min}, s_{\max}). \tag{20}$$

Here, τ_s is the target region consistency, λ_a controls the soft artifact penalty, and K_P, K_I, and K_D are the proportional, integral, and derivative coefficients, respectively. If the current candidate passes the artifact guard, the controller updates s_k and clamps it to $[s_{\min}, s_{\max}]$. If its artifact rate increases excessively relative to the first candidate, the candidate is rejected, the strength is reset to the baseline value of 1, and the accumulated integral and previous-error states are cleared. All other generation conditions remain unchanged across the predefined feedback iterations, so the resulting differences primarily reflect changes in the morphology-prior strength.

Let $\mathcal{V}$ denote the set of candidates that pass the artifact guard. The final result is the candidate with the highest region-level feature consistency:

$$k^* = \arg\max_{k \in \mathcal{V}} S_k. \tag{21}$$

MorphControl outputs the synthesized image $\hat{I}_{k^*}$ together with its target category y^*. The target region $\widetilde{B}^*$ is retained as internal control metadata for position validation and result analysis, but it is not treated as an exact detection annotation. QFPR neither invokes a downstream detector nor computes a detection loss, and it does not update the diffusion-model parameters. When the generated images are used to train a bounding-box-supervised detector, the candidate boxes should be corrected according to the regional response or manually verified instead of being used as unvalidated detection labels.

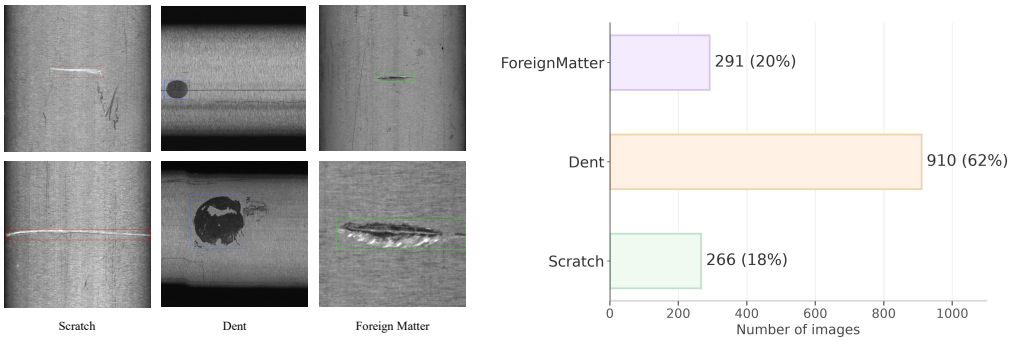
4 Experiments

4.1 Experimental Settings

Dataset. We fabricated representative physical defects and captured nuclear fuel-rod cladding surfaces using a line-scan camera to construct an in-house dataset containing dents, foreign matter, and scratches. The dataset was divided into training, validation, and test sets at a ratio of $70\% : 15\% : 15\%$; only the training split was used to optimize MorphControl. Figure 2 shows representative samples and the class distribution.

Metrics. Generation fidelity is evaluated using Kernel Inception Distance (KID), where lower values indicate closer agreement with the captured-data distribution (Bińkowski et al., 2018). Intra-class diversity is measured using Intra-cluster Pairwise LPIPS Distance (IC-LPIPS), where higher values indicate greater perceptual variation (Ojha et al., 2021; Zhang et al., 2018). Because visual artifacts can also increase perceptual distances, KID and IC-LPIPS are interpreted jointly. Downstream detection is evaluated using mAP_{50} and $\mathrm{mAP}_{50:95}$.

Implementation. MorphControl uses SDXL 1.0 at a resolution of 512×512 and is trained for 25,000 steps. QFPR performs three feedback iterations under fixed generation conditions. Complete training, inference, metric, and baseline configurations are provided in Appendix C.



(a) Representative samples.

(b) Class distribution.

Figure 2: Overview of the nuclear fuel-rod surface-defect dataset.

Table 1: Generation-quality comparison. Best and second-best results are shown in bold and underlined, respectively.

Defect type	Metric	BBox Copy-Paste	StyleGAN2 -ADA	SDXL +LoRA	IP-Adapter (SDXL)	Ours
Dent	KID ↓	<u>0.1140</u>	0.8366	0.1629	0.1596	**0.1064**
	IC-LPIPS ↑	0.6238	0.3836	<u>0.6369</u>	0.5713	**0.6689**
Foreign matter	KID ↓	0.1246	0.7327	0.0901	<u>0.0876</u>	**0.0267**
	IC-LPIPS ↑	0.6122	0.4080	0.5741	<u>0.6149</u>	**0.6763**
Scratch	KID ↓	0.1193	0.6427	<u>0.1037</u>	0.1060	**0.0496**
	IC-LPIPS ↑	**0.6329**	0.3893	0.5736	0.5664	<u>0.6323</u>
Average	KID ↓	0.1193	0.7373	0.1189	<u>0.1177</u>	**0.0609**
	IC-LPIPS ↑	<u>0.6230</u>	0.3936	0.5949	0.5842	**0.6592**

4.2 Comparison with Representative Baselines

We compare MorphControl with BBox Copy-Paste, (Ghiasi et al., 2021), StyleGAN2-ADA (Karras et al., 2020), SDXL-LoRA (**?**Hu et al., 2022), and IP-Adapter with an SDXL backbone (Ye et al., 2023). BBox Copy-Paste is adapted from Simple Copy-Paste (Ghiasi et al., 2021) using annotated defect regions and normal fuel-rod backgrounds. SDXL-LoRA, IP-Adapter, and MorphControl use the same SDXL 1.0 backbone. All methods use the same training split, generate the same number of samples per category, and have no access to test images or annotations during generation.

As shown in Table 1, MorphControl achieves the lowest average KID (0.0609) and the highest average IC-LPIPS (0.6592), improving over the second-best results by 48.3% and 5.8%, respectively. For scratches, BBox Copy-Paste obtains a marginally higher IC-LPIPS, but MorphControl reduces KID from 0.1193 to 0.0496, indicating a better fidelity–diversity trade-off. Qualitative comparisons in Appendix B.3 further show the differences in defect localization and background coherence.

4.3 Effectiveness of QFPR

We compare QFPR with fixed-strength prior injection using the same trained DMPA, backgrounds, prompts, target regions, initial noise, and random seeds. The fixed-strength configuration uses $s = 1.0$, whereas QFPR adjusts s according to regional feature consistency and artifact risk. A QFPR-only variant is undefined because QFPR regulates the morphology prior introduced by DMPA.

Table 2 shows that QFPR reduces KID by 11.1%, increases IC-LPIPS by 43.5%, and improves regional feature consistency from 0.8004 to 0.8117. The artifact ratio remains unchanged at 0.003, indicating that the observed improvements do not increase the measured artifact rate.

Table 2: Effect of QFPR, averaged over the three defect categories.

Model	DMPA	QFPR	KID ↓	IC-LPIPS ↑	S_{roi} ↑	R_{art} ↓
DMPA only	✓	–	0.0685	0.4595	0.8004	**0.003**
MorphControl	✓	✓	**0.0609**	**0.6592**	**0.8117**	**0.003**

Table 3: Effect of MorphControl-generated data on detection performance on the captured-image test set. The best result in each category–metric combination is shown in bold.

Training data	Metric	Dent	Foreign matter	Scratch	Average
Real only	mAP_{50}	0.9013	0.9327	0.8718	0.9019
	$\mathrm{mAP}_{50:95}$	**0.7107**	0.8455	0.6119	0.7227
Real + MorphControl	mAP_{50}	**0.9411**	**0.9509**	**0.9379**	**0.9433**
	$\mathrm{mAP}_{50:95}$	0.7050	**0.8621**	**0.6676**	**0.7449**

4.4 Downstream Defect Detection

We generate 600 manually verified images and bounding-box annotations for each defect category. YOLOv8s is trained using either the original real training set or the training set augmented with MorphControl samples. The two settings use identical initialization and optimization configurations and are evaluated on the same captured-image test set, which is excluded from generator and detector training.

Adding MorphControl samples increases the average mAP_{50} from 0.9019 to 0.9433 and the average $\mathrm{mAP}_{50:95}$ from 0.7227 to 0.7449, corresponding to gains of 4.14 and 2.22 percentage points. The largest improvement is observed for scratches. However, dent $\mathrm{mAP}_{50:95}$ decreases slightly from 0.7107 to 0.7050, showing that augmentation does not improve every category–metric combination.

5 Conclusion

We propose MorphControl, an industrial defect generation framework that integrates defect morphology prior-driven multi-condition adaptation with quality-feedback prior regulation. During training, DMPA performs multi-branch adaptation of the defect-suppressed background, class-level defect morphology prior, geometric layout, and text semantics, enabling composable modeling of the background domain and defect morphology. The region-weighted diffusion objective and background–morphology decorrelation constraint further enhance fine-grained defect learning and suppress the leakage of reference-defect information through the background branch. During inference, QFPR treats the runtime injection strength of the class-level defect morphology prior as a feedback-controlled variable. It iteratively adjusts this strength according to target-region feature consistency and artifact risk, and selects a candidate satisfying the quality constraints while keeping the other generation conditions unchanged. This process alleviates insufficient defect expression, over-amplification, and cross-sample quality fluctuations caused by a fixed injection strength.

Experiments on our nuclear fuel-rod surface-defect dataset show that MorphControl achieves a lower average KID and a higher average IC-LPIPS than the compared baselines, providing a better balance between distributional realism and intra-class diversity. Augmenting the training data with the generated samples improves both mAP_{50} and $\mathrm{mAP}_{50:95}$ of the downstream defect detector, with more pronounced gains for defect categories containing fewer real samples. These results demonstrate that class-level defect morphology prior learning and quality-feedback regulation during inference improve the generation quality and downstream utility of nuclear fuel-rod defect images under limited-data and class-imbalanced conditions.

Reproducibility Statement

The formulation, training objectives, and inference procedure of MorphControl are described in Section 3. The data splits, training hyperparameters, frozen model components, loss weights, inference control variables, and evaluation protocols are reported in Section 4 and Appendix C. The source code, configuration files, random seeds, and bounding-box annotation and quality-control protocol will be released upon publication.

AI Use Statement

Generative AI tools were used to assist with organizing the manuscript structure, improving language clarity, and compiling the experimental checklist. They were not used to generate experimental data or reported results. All AI-assisted content was reviewed by the authors, and all research decisions, model implementations, experimental procedures, result interpretations, and references were independently verified by the authors. The authors take full responsibility for the final content of this work.

References

Karl J. Åström and Richard M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2008.

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018.

Zhewei Dai, Shilei Zeng, Haotian Liu, Xurui Li, Feng Xue, and Yu Zhou. SeaS: Few-shot industrial anomaly image generation with separation and sharing fine-tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23135–23144, 2025.

Yuxuan Duan, Yan Hong, Li Niu, and Liqing Zhang. Few-shot defect image generation via defect-aware feature manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 571–578, 2023. doi: 10.1609/aaai.v37i1.25132.

Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D. Cubuk, Quoc V. Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2918–2928, 2021.

Shidan He, Lei Liu, Xiujun Shu, Bo Wang, Yuanhao Feng, and Shen Zhao. AnomalyControl: Learning cross-modal semantic features for controllable anomaly synthesis. *arXiv preprint arXiv:2412.06510*, 2024. URL https://arxiv.org/abs/2412.06510.

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

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. AnomalyDiffusion: Few-shot anomaly image generation with diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8526–8534, 2024. doi: 10.1609/aaai.v38i8.28696.

Ying Jin, Jinlong Peng, Qingdong He, Teng Hu, Jiafu Wu, Hao Chen, Haoxuan Wang, Wenbing Zhu, Mingmin Chi, Jun Liu, and Yabiao Wang. Dual-interrelated diffusion model for few-shot anomaly image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30420–30429, 2025.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12104–12114. Curran Associates, Inc., 2020.

Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. CutPaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9664–9674, 2021. doi: 10.1109/CVPR46437.2021.00954.

D. Li, B. Su, H. Zhang, Y. Li, and H. Chen. Photovoltaic defect image generator with boundary alignment smoothing constraint for domain shift mitigation. *arXiv preprint arXiv:2505.06117*, 2025.

Dongying Li, Binyi Su, Hua Zhang, Yong Li, and Haiyong Chen. Photovoltaic defect image generator with boundary alignment smoothing constraint for domain shift mitigation. *IEEE Transactions on Multimedia*, 2026. doi: 10.1109/TMM.2026.3679929. Early Access.

Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4296–4304, 2024. doi: 10.1609/aaai.v38i5.28226.

Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A. Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot image generation via cross-domain correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10743–10752, 2021.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 8748–8763, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Jing Wei, Fei Shen, Chengkan Lv, Zhengtao Zhang, Feng Zhang, and Huabin Yang. Diversified and multi-class controllable industrial defect synthesis for data augmentation and transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 4445–4453, 2023. doi: 10.1109/CVPRW59228.2023.00467.

Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023. URL `https://arxiv.org/abs/2308.06721`.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. DRAEM: A discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8330–8339, 2021.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3836–3847, 2023.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. RealNet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16699–16708, 2024.

A Appendix

You may include other additional sections here.

B Experimental and Implementation Details

B.1 Dataset and Evaluation Protocol

The dataset contains physically fabricated defects captured from nuclear fuel-rod cladding tubes using a line-scan camera. It is divided into training, validation, and test sets at a ratio of $70\% : 15\% : 15\%$. Only the training split is used for generator optimization, baseline training, generation references, and detector training. The captured-image test set is excluded from all training and generation procedures.

Dataset details to report before submission: total image count, per-category counts, line-scan camera model, acquisition resolution, defect-fabrication procedure, annotation protocol, and whether samples from the same fuel rod or acquisition sequence are restricted to a single split.

Metric details to report before submission: the number of real and generated images used for KID, the number and size of KID subsets, the IC-LPIPS pairing procedure, the implementation packages, and the random seeds or repeated runs used for metric estimation.

B.2 Training and Inference Configuration

All experiments are implemented using PyTorch 2.11 and Python 3.11 on a single NVIDIA GeForce RTX 5090 D GPU. MorphControl uses SDXL 1.0 with input and output resolutions of 512×512. The VAE, text encoders, U-Net, domain LoRA, and CLIP image encoder remain frozen. The background projection, morphology-prior bank, geometry encoder, branch gates, and lightweight residual parameters of the condition-injection module are optimized.

The model is trained for 25,000 steps using AdamW with a batch size of 16, an initial learning rate of 1×10^{-5}, and BF16 mixed precision. Inference uses the Euler sampler with 30 denoising steps and a classifier-free guidance scale of 7.5. QFPR performs three feedback iterations under fixed generation conditions.

The Min-SNR parameter is $\gamma = 5.0$, and the loss weights are $\lambda_{\mathrm{roi}} = 0.5$ and $\lambda_{\mathrm{dec}} = 0.1$. The controller parameters are $K_P = 0.30$, $K_I = 0.05$, and $K_D = 0.03$, with the morphology-prior strength restricted to $[0.80, 1.20]$.

Baseline details to report before submission: the training iterations, optimizers, learning rates, pretrained checkpoints, random seeds, and generation settings used for StyleGAN2-ADA, SDXL-LoRA, IP-Adapter, and BBox Copy-Paste.

B.3 Qualitative Results

Figure 3 shows that BBox Copy-Paste can introduce discontinuous boundaries or mismatches between pasted defects and their backgrounds, while StyleGAN2-ADA can distort the global surface structure. SDXL-LoRA and IP-Adapter improve background coherence but may generate weak, duplicated, or incorrectly scaled defects. MorphControl produces localized defects while preserving the surrounding fuel-rod structure and texture.

C Additional Implementation Details

SDXL uses OpenCLIP ViT-bigG/14 and CLIP ViT-L/14 as its dual text encoders. CLIP ViT-H/14 pretrained on LAION-2B is employed as the image encoder. The trainable components include the background projection module, class-level defect morphology prior, geometric encoder, and multi-branch condition-injection module.

The training objective combines the Min-SNR-weighted diffusion loss, defect-region-weighted loss, and background–morphology decorrelation regularization. We use AdamW with a weight decay of 0.01 and cosine learning-rate decay. The Min-SNR parameter is set to $\gamma = 5.0$, while the loss weights are set to $\lambda_{\mathrm{roi}} = 0.5$ and $\lambda_{\mathrm{div}} = 0.1$. Validation is performed every 500 training steps.

During inference, the target region is sampled from the empirical geometric distribution of the corresponding defect category and perturbed with a magnitude of 0.10. QFPR adjusts the morphology-

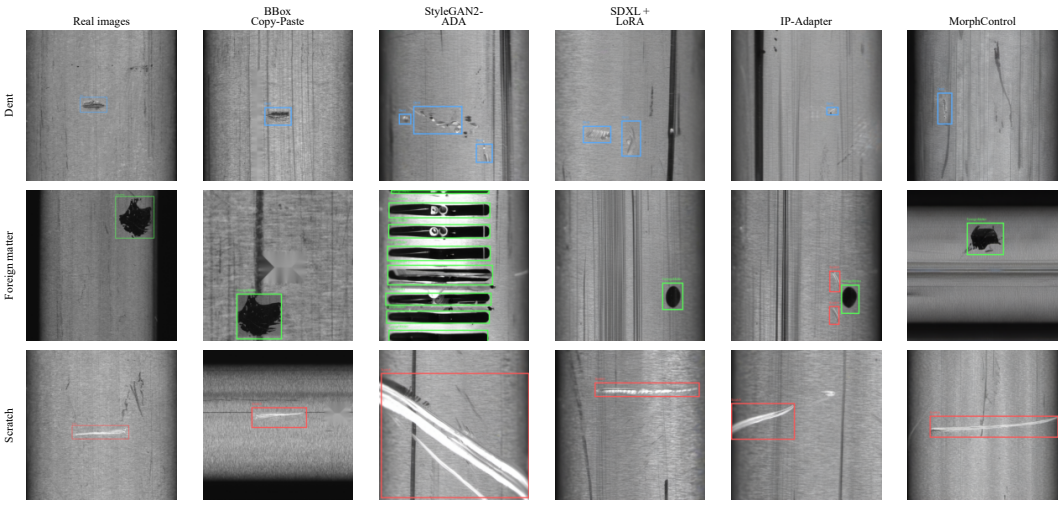


Figure 3: Qualitative comparison on the three defect categories. Blue, green, and red boxes indicate dents, foreign matter, and scratches, respectively.

prior strength according to the CLIP feature consistency within the target region and the artifact ratio. Its PID parameters are set to $K_{\mathrm{P}} = 0.30$, $K_{\mathrm{I}} = 0.05$, and $K_{\mathrm{D}} = 0.03$, with the prior strength restricted to $[0.80, 1.20]$. The candidate satisfying the artifact constraint and achieving the highest regional feature consistency is selected as the final output.