

DeCo: Zero-Shot Industrial Anomaly Generation through Decoupling and Recoupling

Shilei Zeng, Xurui Li, Yaohan Tang, and Yu Zhou✉

School of Electronic Information and Communications,
Huazhong University of Science and Technology
{shlzensg,xrli_plus,yhtang_,yuzhou}@hust.edu.cn

Abstract. Industrial anomaly inspection is severely hindered by the scarcity of real anomalous data. Zero-shot industrial anomaly generation addresses this by generating anomalies on specific products without requiring any of their real anomalous images. However, existing methods suffer from two critical limitations, i.e., inaccurate anomaly information acquisition and uncontrolled anomaly-product fusion. To overcome these challenges, we propose DeCo, which decouples the anomaly structure from its source product, and explicitly recouples it with the normal textures of the target product. During anomaly information acquisition, Dual-Routing Flow (DR-Flow) binds the texture-invariant anomaly structure to an abnormal token, while a parallel constraint, Product-Invariant Flow (PI-Flow), prevents the abnormal token from binding the source product. During anomaly-product fusion, we propose a hybrid injection to recouple the acquired anomaly structure with the target product, and Product Compatibility Correction (PCC) to compensate for the incompatibility between the acquired anomaly structure and the product. Extensive experiments demonstrate that DeCo establishes a new state-of-the-art. Training downstream detection models on our generated data yields massive pixel AP improvements of 5.1% on MVTec AD and 8.2% on VisA. Code is available at <https://github.com/HUST-SLOW/DeCo>.

Keywords: Zero-shot industrial anomaly generation · Decoupling and recoupling

1 Introduction

Industrial anomaly inspection is hindered by the scarcity of real-world defects. To mitigate data scarcity, anomaly synthesis methods [5, 20, 29, 37, 38] rely on data augmentation to generate pseudo-anomalies, but their authenticity is often limited. Few-shot anomaly generation methods [12, 16, 19, 39] offer a promising alternative by learning from a few anomaly images. However, collecting even a small set of target anomalies remains challenging in cold-start scenarios.

To solve this, zero-shot industrial anomaly generation aims to generate anomalies on a specific product without requiring any real anomaly images of that product. Current methods primarily fall into two categories, namely, text-based [31, 32] and reference-based [13, 18] generation. Although promising anomalies have

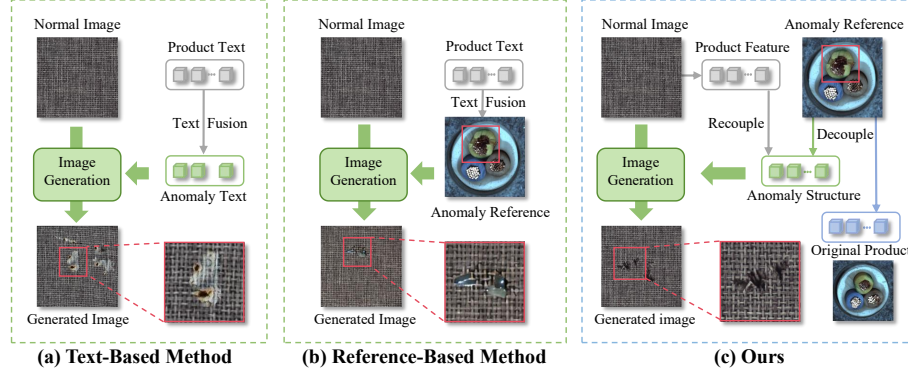


Fig. 1: (a) Text-based methods face a semantic gap when using fixed text descriptions, yielding unrealistic patterns for complex anomalies. **(b) Reference-based methods** struggle to disentangle the anomaly from the original product, causing unwanted textures to leak into the target generation. **(c) Our method (Ours)** explicitly decouples the anomaly structure from its original product, and seamlessly recouples it with the specific product, generating high-fidelity anomalies.

been generated, they still face two main challenges, **(1) Inaccurate Anomaly Information Acquisition**. Text-based methods acquire anomaly information from fixed text prompts, and use it to guide the generation. However, due to the semantic gap between the fixed text descriptions and the complex anomalies, these methods often generate unrealistic patterns, as shown in Fig. 1(a). In reference-based methods, visual features are first extracted from anomalies in other products, and then used to guide generation. Nevertheless, they struggle to disentangle anomalies from the original product, causing unwanted textures to leak into the generated image, as shown in Fig. 1(b). **(2) Uncontrolled Anomaly-Product Fusion**. Existing methods lack an explicit mechanism to fuse anomaly information to a specific product. They typically rely on fused text prompts (e.g., “a cut on carpet”), which often results in poor anomaly-product fusion.

To address these challenges, we propose DeCo, a novel zero-shot industrial anomaly generation framework. As illustrated in Fig. 1(c), our core idea is to first decouple the anomaly structure from its original product context, and then explicitly recouple it with the target product. For anomaly structure acquisition, we first introduce Dual-Routing Flow (DR-Flow), which comprises AP-Routing and QK-Routing. AP-Routing isolates the abnormal text token, while QK-Routing extracts the anomaly structure. Thus DR-Flow guarantees the anomaly structure is explicitly bound to the abnormal text token. In addition, Product-Invariant Flow (PI-Flow) is further proposed, which prevents the abnormal token from directly inheriting product features from the source reference. During anomaly-product fusion in the generation process, we perform hybrid injection to explicitly recouple the acquired anomaly structures with the specific product. To

address potential poor incompatibility, we introduce Product Compatibility Correction (PCC), which corrects the mismatch between the anomaly structures and the product. Extensive experiments demonstrate the superiority of DeCo in zero-shot anomaly generation. By training downstream detection models on our generated data, we establish a new state-of-the-art, outperforming existing zero-shot anomaly generation methods with a pixel AP increase of 5.1% on the MVTec AD dataset and 8.2% on the VisA dataset.

In summary, our contributions are threefold:

- We propose DeCo, a novel zero-shot industrial anomaly generation framework. During anomaly information acquisition, it decouples the anomaly structure from its source product and explicitly binds it to an abnormal token. During anomaly-product fusion, it performs hybrid injection to explicitly recouple the acquired abnormal structure with the target product.
- The strength of DeCo stems from its multiple novel designs. DR-Flow employs AP-Routing to isolate the abnormal text token and QK-Routing to extract the anomaly structure, ensuring explicit binding between them. PI-Flow is introduced to prevent the token from inheriting source product features. At inference, Product Compatibility Correction (PCC) compensates for incompatibilities between the acquired anomaly structures and the target product.
- Extensive experiments demonstrate that DeCo establishes a new state-of-the-art for downstream zero-shot anomaly detection. Training detection models on our generated data yields significant improvements, outperforming existing methods with a pixel AP increase of 5.1% on the MVTec AD dataset and 8.2% on the VisA dataset.

2 Related Work

2.1 Industrial Anomaly Generation

Industrial anomaly inspection is limited by the scarcity of real-world anomalies. Anomaly synthesis methods [5, 20, 29, 37, 38] rely on data augmentation to generate pseudo-anomalies. Few-shot anomaly generation methods [7, 12, 16, 19, 39] address this by generating anomalies from limited samples. While zero-shot anomaly generation aims to generate anomalies on a specific product without requiring any real anomaly images of that product. Existing methods face two limitations. Specifically, they struggle with anomaly information acquisition and anomaly-product fusion. Text-driven approaches [31, 32] struggle with inaccurate anomaly acquisition. Due to the semantic gap between the text descriptions and the anomalies, they often yield unrealistic patterns. Reference-based methods [13, 18] use visual features extracted from reference anomalies as guidance during generation. They also suffer from inaccurate acquisition by failing to disentangle the anomaly from its original product. Furthermore, both text and reference-based methods exhibit uncontrolled anomaly-product fusion, lacking an explicit mechanism to seamlessly fuse the acquired anomaly information

with the specific product. In contrast, our method explicitly addresses these dual challenges. For anomaly information acquisition, we extract a product-invariant structure through DR-Flow and PI-Flow, leaving unwanted backgrounds behind. For anomaly-product fusion, we recouple the acquired anomaly structure and product via a hybrid LoRA injection, utilizing Product Compatibility Correction to address potential poor incompatibility.

2.2 LoRA Composition in Image Generation

LoRA composition primarily explores two directions: object integration [6, 11, 17, 24, 36] and content-style fusion. The latter relies on decoupling and coupling to separate content and styles and apply them to new targets. Early works achieve this by merging or selectively activating pre-trained LoRAs [26, 30, 34]. More recently, attention has shifted to training-time decoupling. B-LoRA [10] identifies distinct roles for attention modules and achieves object-style separation. UnzipLoRA [22] and QR-LoRA [35] jointly optimize content and style LoRA, the former through parallel branch training and the latter via orthogonal constraints. Previous LoRA-based decoupling methods rely on strong pre-trained semantics (e.g., “dog” or “watercolor”) to establish an initial binding for decoupling. In contrast, industrial anomalies are highly abstract and lack such semantic priors. To overcome this gap, we shift from semantic-level to structural-level binding. Instead of relying on the text encoder’s inherent knowledge, DR-Flow directly extracts the anomaly structure from the reference image and binds it to the isolated abnormal text token. This allows us to decouple the anomaly structure from its original product context.

3 Preliminaries

Rectified Flow and SD3. Stable Diffusion 3 (SD3) [9] is built on Rectified Flow [23], a specific form of Flow Matching [21]. Given a real image x_0 , a VAE encoder first encodes it to a latent z_0 . SD3 learns a linear path $z_t = (1 - t)z_0 + tz_1$ to connect the clean latent z_0 and the Gaussian noise $z_1 \sim \mathcal{N}(0, I)$ across timesteps $t \in [0, 1]$. Accordingly, the target velocity field is constant: $v_t = \frac{dz_t}{dt} = z_1 - z_0$. SD3 utilizes a Multimodal Diffusion Transformer (MM-DiT) to predict this velocity. Given a text prompt $\mathcal{P}$, the predicted velocity $\hat{v}_\theta$ is trained using the flow matching objective:

$$\mathcal{L} = \mathbb{E}_{z_0, z_1, t} [\|\hat{v}_\theta(z_t, t, \mathcal{P}) - (z_1 - z_0)\|_2^2]. \quad (1)$$

During inference, the generated latent z_0 is reconstructed by solving the ODE $\frac{dz_t}{dt} = \hat{v}_\theta(z_t, t, \mathcal{P})$ starting from noise z_1 .

LoRA Fine-tuning in SD3. Low-Rank Adaptation (LoRA) [15] is a lightweight fine-tuning method that introduces trainable low-rank matrices into large pre-trained models. Given a frozen pretrained weight $W^0 \in \mathbb{R}^{m \times n}$, the update is

parameterized by two low-rank matrices $B \in \mathbb{R}^{m \times r}$ and $A \in \mathbb{R}^{r \times n}$ with rank $r \ll \min(m, n)$:

$$W = W^0 + \Delta W = W^0 + BA. \quad (2)$$

In the MMDiT architecture of SD3, LoRA modules are typically applied to the linear projections within the attention blocks $(W_Q^0, W_K^0, W_V^0, W_{Out}^0)$, then the output of each projection is augmented as:

$$Y_i = (W_i^0 + \Delta W_i)h, \quad i \in \{Q, K, V, Out\}, \quad (3)$$

where h denotes the input hidden representation. This formulation enables efficient parameter adaptation while maintaining the knowledge of the pretrained model. For brevity, we omit the output projection *Out* in subsequent discussions.

Self-Attention Mechanism in Diffusion Models. The denoising model of SD3 is an MM-DiT [9], which processes image and text tokens through a sequence of transformer blocks. Each block contains modulation mechanisms and self-attention layers. The attention mechanism is formulated as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V, \quad (4)$$

where Q, K, V are the query, key, and value projected from the tokens. Functionally, the Value (V) carries the features to be aggregated, while the Key (K) serves as the indexing feature for retrieval. The Query (Q) determines the attention map by computing matching scores against K , deciding which features from V are retrieved and how they are assembled. Recent works [3, 4, 33] have identified distinct semantic roles for query, key, and value features. The Q governs the spatial layout and structure, determining the arrangement of features, while the K and V provide the appearance content to be aggregated.

4 Method

To achieve zero-shot industrial anomaly generation, our core idea is to first decouple the reference anomaly structure from its original product context, and then explicitly recouple it with the target product. For accurate anomaly information acquisition, Dual-Routing Flow (DR-Flow) is proposed to bind the anomaly structure to abnormal tokens. Further, we introduce Product-Invariant Flow (PI-Flow) as a parallel constraint. It prevents the abnormal token from inheriting product features from the source reference. During anomaly-product fusion, to recouple the acquired anomaly structure with the specific product, we perform a hybrid LoRA injection. We also propose Product Compatibility Correction (PCC) to address the incompatibility between the acquired anomaly and the specific product.

4.1 Problem Statement

We focus on zero-shot industrial anomaly generation, which aims to generate anomalies on a specific product $\mathcal{T}$ without requiring any real anomaly images

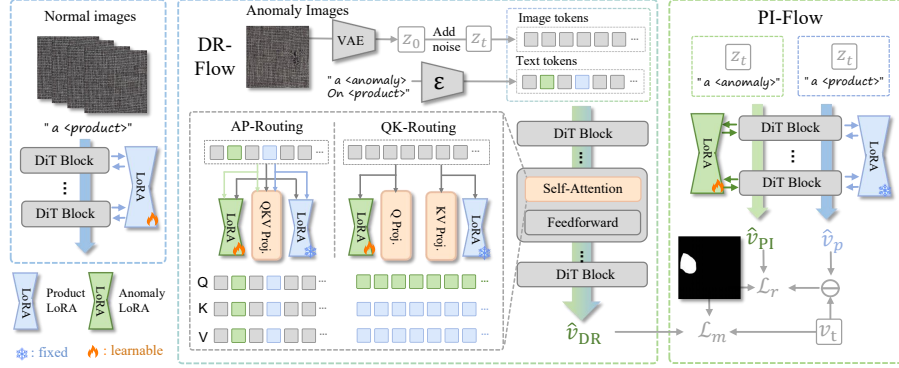


Fig. 2: Training pipeline of DeCo. It employs a Dual-LoRA design with a frozen Product LoRA (blue, pre-trained on the **left**) and a trainable Anomaly LoRA (green). **DR-Flow** (**Middle**) binds the anomaly structure to the abnormal token via AP-Routing and QK-Routing. **PI-Flow** (**Right**) acts as a constraint to prevent the abnormal token from directly inheriting product features from the source reference.

of $\mathcal{T}$. We acquire anomaly information from a reference product $\mathcal{R}$. Formally, we have access to the reference $\mathcal{R}$, which includes a set of anomalous images $X_{\mathcal{R}}^a = \{x_a^1, \dots, x_a^H\}$ with corresponding masks $M_{\mathcal{R}}$, along with normal images $X_{\mathcal{R}}^n = \{x_n^1, \dots, x_n^W\}$. For the specific product $\mathcal{T}$, only normal images $X_{\mathcal{T}}^n = \{I_n^1, \dots, I_n^L\}$ are available. No real anomaly images of $\mathcal{T}$ are used during the entire process.

To acquire the pure anomaly structure from the reference images and explicitly recouple it with the specific product, we define AP-LoRA as follows, **A-LoRA**, stands for Anomaly LoRA. It contains a text anomaly LoRA $\Delta W_{\text{T}}^{\text{A}}$, and an image anomaly LoRA $\Delta W_{\text{I}}^{\text{A}}$, which are formally stated as follows,

$$\Delta W_{\text{T}}^{\text{A}} = \{\Delta W_{\text{TQ}}^{\text{A}}, \Delta W_{\text{TK}}^{\text{A}}, \Delta W_{\text{TV}}^{\text{A}}\} \quad (5)$$

$$\Delta W_{\text{I}}^{\text{A}} = \{\Delta W_{\text{IQ}}^{\text{A}}, \Delta W_{\text{IK}}^{\text{A}}, \Delta W_{\text{IV}}^{\text{A}}\} \quad (6)$$

A-LoRA is incorporated into each block of MM-DiT following Eq. (2), and is used to align abnormal tokens with anomaly structure.

P-LoRA, refers to the Product LoRA, comprises a text LoRA $\Delta W_{\text{T}}^{\text{P}}$ and an image LoRA $\Delta W_{\text{I}}^{\text{P}}$, formulated as:

$$\Delta W_{\text{T}}^{\text{P}} = \{\Delta W_{\text{TQ}}^{\text{P}}, \Delta W_{\text{TK}}^{\text{P}}, \Delta W_{\text{TV}}^{\text{P}}\} \quad (7)$$

$$\Delta W_{\text{I}}^{\text{P}} = \{\Delta W_{\text{IQ}}^{\text{P}}, \Delta W_{\text{IK}}^{\text{P}}, \Delta W_{\text{IV}}^{\text{P}}\} \quad (8)$$

Similarly, we integrate P-LoRA into the MM-DiT blocks, and establish the alignment between normal products and their text descriptions.

4.2 Anomaly Structure Binding via Dual-Routing Flow

The key to accurate anomaly acquisition lies in decoupling the anomaly structure from the reference images and binding it to the text tokens. To this end, we introduce two mechanisms to train A-LoRA: an AP-Routing that separates text tokens to isolate the abnormal token, and a QK-Routing that restricts A-LoRA to mapping only the structure features. Thus DR-Flow effectively binds the extracted anomaly structure to the isolated abnormal token.

Product Information Binding. As shown in Fig.2, given the normal product text prompt (e.g., “a <product>”) and the normal image set $X_{\mathcal{R}}^n$, we train P-LoRA (ΔW_I^P and ΔW_T^P) following Eq. (1) to bind $X_{\mathcal{R}}^n$ to normal product text prompt. This step explicitly binds the normal product features to product text via P-LoRA, and we freeze P-LoRA in all subsequent processing.

AP-Routing for Text Token Split. Given a text prompt $\mathcal{P}$ such as “a <anomaly> on <product>”, SD3 first tokenizes $\mathcal{P}$ to form the text tokens $T \in \mathbb{R}^{N \times C}$, where N is the number of text tokens and C is the channel number. Then, we decompose T into three groups, namely, $T_A \in \mathbb{R}^{1 \times C}$, $T_P \in \mathbb{R}^{1 \times C}$ and $T_O \in \mathbb{R}^{(N-2) \times C}$, which indicate the abnormal text “<anomaly>”, product text “<product>” and other texts like “a”, “on” and padding tokens. Accordingly, the corresponding Q_T , K_T , and V_T are computed using the weights $W_T = \{W_T^A, W_T^P, W_T^O\}$ as follows:

$$\begin{aligned} (Q_T^A, K_T^A, V_T^A) &= W_T^A T_A = (W_T^0 + \Delta W_T^A) T_A, \\ (Q_T^P, K_T^P, V_T^P) &= W_T^P T_P = (W_T^0 + \Delta W_T^P) T_P, \\ (Q_T^O, K_T^O, V_T^O) &= W_T^O T_O = (W_T^0 + \Delta W_T^A + \Delta W_T^P) T_O. \end{aligned} \quad (9)$$

where W_T^0 is the pre-trained weight, and text P-LoRA ΔW_T^P is learned on product $\mathcal{R}$. Both of them are fixed in Eq. (9), and the text A-LoRA ΔW_T^A is learnable. We then concatenate $Q_T^A \in \mathbb{R}^{1 \times C}$, $Q_T^P \in \mathbb{R}^{1 \times C}$, and $Q_T^O \in \mathbb{R}^{(N-2) \times C}$, resulting in $Q_T \in \mathbb{R}^{N \times C}$. We obtain $K_T \in \mathbb{R}^{N \times C}$ and $V_T \in \mathbb{R}^{N \times C}$ in the same way. AP-Routing maps the anomaly and product tokens separately, ensuring that each of them is processed exclusively by its own LoRA. We explicitly separate them because industrial anomaly structure carries limited semantic information, unlike the objects in natural scenes (e.g., “dog”). By isolating the text routing, we ensure that the abstract anomaly structure is bound solely to the abnormal text token, independent of the product context.

QK-Routing for Structure Extraction. Given the image token $T_I \in \mathbb{R}^{M \times C}$, we expand the pre-trained weight W_I^0 as $\{W_{IQ}^0, W_{IK}^0, W_{IV}^0\}$, and propose a QK-Routing mechanism as follows,

$$\begin{aligned} Q_I &= (W_{IQ}^0 + \Delta W_{IQ}^A) T_I, \\ K_I &= (W_{IK}^0 + \Delta W_{IK}^P) T_I, \\ V_I &= (W_{IV}^0 + \Delta W_{IV}^P) T_I \end{aligned} \quad (10)$$

In Eq. (10), we map T_I to K and V using the frozen P-LoRA parameters. As discussed in Sec. 3, the query (Q) controls the spatial structure, while the key

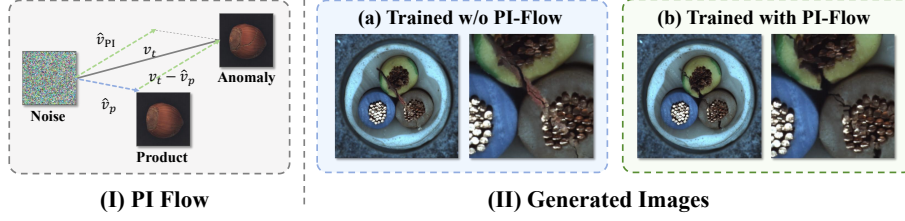


Fig. 3: (I) PI-Flow. (II) Without PI-Flow, the generated anomalies may carry over the source product’s appearance.

(K) and value (V) provide the visual texture. By computing K_I and V_I using only the frozen P-LoRA, we limit the available textures to the normal product. At the same time, we compute Q_I using the trainable A-LoRA. Through this design, the trainable A-LoRA extracts the reference anomaly’s structure into Q_I .

Within the standard MM-DiT block, the queries (Q_T and Q_I), keys (K_T and K_I), and values (V_T and V_I) are concatenated to construct a unified set $\{Q_{T,I}, K_{T,I}, V_{T,I}\}$ for joint self-attention. During this process, the isolated abnormal text token acts as the sole text condition for the A-LoRA. Driven by the generation loss, this attention interaction inherently forces A-LoRA to bind the anomaly structure to the abnormal text token. After the joint self-attention, these features are processed by the remaining MM-DiT blocks to predict the velocity $\hat{v}_{DR}$. The output $\hat{v}_{DR}$ is supervised using a velocity-based flow-matching loss restricted to the anomaly region:

$$\mathcal{L}_{DR} = \mathbb{E}_{z_0, z_1, t, m} \left[\|m \odot (\hat{v}_{DR}(z_t, t, \mathcal{P}) - v_t)\|_2^2 \right], \quad (11)$$

where $\hat{v}_{DR}$ is the predicted velocity and $v_t = z_1 - z_0$ is the constant target velocity. The variable z_t represents the noise latent at timestep t , m is the binary anomaly mask, and $\mathcal{P}$ denotes the condition prompt.

4.3 Product-Invariant Flow

In DR-Flow, we train A-LoRA to bind the anomaly structure to the abnormal token. However, to reduce the training loss, A-LoRA might simply align the abnormal token to the original reference anomaly that contains the product’s texture. As a result, the generated anomaly carries over the source product’s appearance, as shown in Fig. 3(II)(a).

To further ensure accurate and pure anomaly structure acquisition, we introduce Product-Invariant Flow (PI-Flow) as a parallel constraint. This constraint prevents the abnormal token from binding the source product. Specifically, given the prompt “a <product>”, and the noise anomaly latent z_t as input, we activate the fixed P-LoRA alone to predict a product flow $\hat{v}_p$ in the anomaly region, thus PI-Flow is stated as follows:

$$v_{PI} = v_t - \hat{v}_p \quad (12)$$

As shown in Fig. 3(I), since P-LoRA is fixed and trained on normal images, $\hat{v}_p$ is a constant velocity that solely reconstructs the normal product. By explicitly subtracting this product-specific velocity $\hat{v}_p$ from the total target velocity v_t , v_{PI} acts as a residual flow. This residual flow represents the pure anomaly, completely separated from the normal product.

In addition, given the prompt $\mathcal{P}_a$ (“*a <anomaly>*”) and the noise anomaly latent z_t as input, we activate the trainable A-LoRA ΔW^A alone to predict the product-invariant flow $\hat{v}_{PI}$. The objective is defined as follows:

$$\mathcal{L}_{PI} = \mathbb{E}_{z_0, z_1, t, m} \left[\|m \odot (\hat{v}_{PI}(z_t, t, \mathcal{P}_a) - v_{PI})\|_2^2 \right]. \quad (13)$$

Consequently, ΔW^A is explicitly constrained to bind these product-invariant anomaly features to the abnormal token. This prevents the token from inheriting product features, as illustrated in Fig. 3(II)(b).

The total training objective is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{DR} + \lambda_{PI} \cdot \mathcal{L}_{PI}. \quad (14)$$

Here, $\mathcal{L}_{DR}$ drives the text-structure binding within the DR-Flow. Furthermore, $\mathcal{L}_{PI}$ prevents the abnormal tokens from binding the source product. By adjusting the weighting factor λ_{PI} , we control the strength of this constraint. Guided by this joint objective, the A-LoRA is forced to bind product-invariant anomaly structure to the abnormal token.

4.4 Inference with Compatibility Correction

During the generation stage, given the pretrained A-LoRA $\Delta W_{\mathcal{R}}^A$ on reference product $\mathcal{R}$, we aim to recouple the acquired anomaly structure from A-LoRA with the specific product $\mathcal{T}$. To this end, we adopt the following hybrid injection strategy.

SD3 with hybrid LoRA injection. We first build the inference model upon the frozen SD3 backbone weight W_0 . To recouple the acquired anomaly structure with the specific product $\mathcal{T}$, we inject both the P-LoRA $\Delta W_{\mathcal{T}}^P$ of the specific product $\mathcal{T}$ and the A-LoRA $\Delta W_{\mathcal{R}}^A$ from the source product $\mathcal{R}$ and form the model weight W_{inf} as follow:

$$W_{inf} = W_0 + \Delta W_{\mathcal{T}}^P + \Delta W_{\mathcal{R}}^A. \quad (15)$$

We apply Eq.15 to each block of MM-DiT, resulting in the final inference model.

Hybrid Text Prompt. To guide the generation, we use a composite prompt based on the anomaly type from $\mathcal{R}$ and the product category of $\mathcal{T}$. Given the text prompt template “*a <anomaly> on <product>*”, we combine the name of acquired anomaly, e.g., “*cut*”, and the product name of $\mathcal{T}$, e.g., “*cable*”, resulting in a new composite prompt $\mathcal{P}_{mix}$, e.g., “*a cut on cable*”.

Product Compatibility Correction. While the hybrid injection enables zero-shot anomaly generation, recoupling the acquired anomaly structure with different specific products $\mathcal{T}$ often reveals varying degrees of compatibility. This

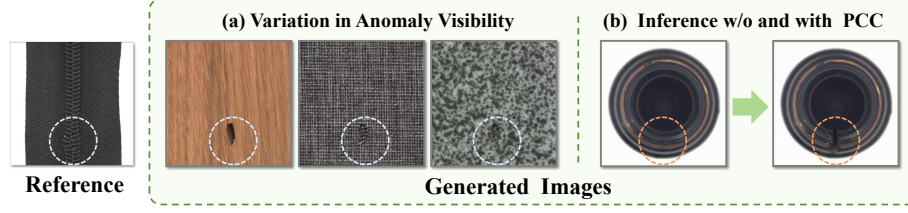


Fig. 4: Zero-shot anomaly generation using the same reference anomaly. (a) Across different products, the anomaly varies in visibility. (b) With PCC to address the incompatibility, the anomaly can be generated well.

variation arises because the two LoRAs are trained independently on different products. As a result, the same acquired anomaly can exhibit different levels of effectiveness when recoupled with different products. As shown in Fig. 4(a), in some cases, the anomaly is injected with varying visibility. For certain products, however, the injected anomaly structure exhibits incompatibility with the product, making the anomaly barely visible. This creates a misleading impression that the anomaly cannot be generated.

To address this incompatibility, we propose Product Compatibility Correction (PCC). This strategy refines the velocity prediction. Specifically, we compute a product-only velocity $\hat{v}_p$ by activating $\Delta W_{\mathcal{T}}^P$ alone. We then push the inference model prediction $\hat{v}_{DR}$ away from the product-only prediction to obtain the final inference velocity:

$$\hat{v}_{inf} = \hat{v}_p + \omega \cdot (\hat{v}_{DR} - \hat{v}_p). \quad (16)$$

Here, ω is a scalar hyperparameter that controls the degree of correction. This extrapolation resembles Classifier-Free Guidance (CFG) [14], which amplifies the difference between conditional and unconditional predictions to improve prompt fidelity. In our case, $\hat{v}_p$ represents the normal product. By pushing the final prediction away from this product-only prediction, we effectively compensate for the incompatibility between the acquired anomaly structure and the product, enabling the successful generation of anomalies on the specific product $\mathcal{T}$, as shown in Fig.4(b).

Therefore, given a normal image I_n of $\mathcal{T}$, an anomaly mask m , and the composite prompt $\mathcal{P}_{mix}$, we first encode I_n into latent z_n through VAE encoder. Then we use the composite prompt $\mathcal{P}_{mix}$ as guidance and inject anomalies via a blended diffusion [1]:

$$z_t = \hat{z}_t \odot m + z_t^{\text{ref}} \odot (1 - m), \quad (17)$$

where $\hat{z}_t$ is the denoised latent obtained using $\hat{v}_{inf}$, and z_t^{ref} is the corresponding noisy latent obtained by forward noising z_n to timestep t . This blending ensures that the acquired anomaly is injected only within the masked region, while preserving the original product elsewhere.

5 Experiments

5.1 Experimental setting

Datasets. We conduct our experiments on the MVTec AD dataset [2]. The MVTec AD dataset comprises 15 product categories, including 5 texture types and 10 object types, with each category containing up to 8 different types of anomalies, making it suitable for simulating real-world industrial scenarios. For a more extensive evaluation, we also include the VisA dataset [40], which consists of 12 objects spanning 3 distinct domains.

Implementation Details. We build our framework upon the pre-trained Stable Diffusion 3 (SD3). In our zero-shot setting, for the specific product $\mathcal{T}$, no real anomaly images are available. To generate anomalies for $\mathcal{T}$, we use the other 14 products of MVTec AD as our reference set. During the anomaly information acquisition phase, we extract product-invariant anomaly structures by training a dedicated A-LoRA for each of these 14 reference products, setting the PI-Flow loss weight $\lambda_{PI} = 0.1$. During generation, we explicitly recouple the normal images of $\mathcal{T}$ with the anomaly structures acquired from this reference set. For each category, we generate 1,000 images. To strictly preserve the zero-shot paradigm, we uniformly utilize reference masks without any similarity-based filtering. The conditional anomaly masks are derived from the reference ground-truth masks and augmented through mask combinations, specifically by taking the union of two randomly sampled masks of the same anomaly type. Following [12], these combined masks are intersected with the specific product’s foreground mask to ensure valid anomaly placement. The PCC scale ω is set to 2. More details are in the Appendix A.

Baselines. We evaluate our method against a diverse set of anomaly synthesis and generation methods. For visual generation quality, we compare our method with anomaly synthesis methods, including CutPaste [20], DRAEM [37], NSA [29], RealNet [38], and zero-shot anomaly generation methods AnomalyAny [32] and AnoStyler [31]. We also compare our method with some few-shot anomaly generation methods [8, 16, 19]. For downstream anomaly detection tasks, we compare our approach with anomaly synthesis and zero-shot anomaly generation methods. Additionally, to validate our advantages, we compare our approach with recent state-of-the-art Dual-LoRA decoupling and coupling methods (UnzipLoRA [22] and QR-LoRA [35]).

Evaluation Metrics. For image generation, following [16], we use Inception Score (IS) [28] and Intra-cluster pairwise LPIPS distance (IC-LPIPS) [25] to evaluate the authenticity and diversity of the anomaly images. Furthermore, to specifically assess the local authenticity of the generated anomalies, we compute the IS exclusively within the anomaly regions, denoted as IS(a). For the downstream anomaly detection task, we use Area Under ROC Curve (AUROC), Average Precision (AP), and F1-score at both the image level (I-AUC, I-AP, I-F1) and pixel level (P-AUC, P-AP, P-F1). Additionally, we include Per-Region Overlap (PRO) to evaluate pixel-level detection performance.

Table 1: Quantitative comparison of anomaly image generation quality on the MVTec AD dataset. We evaluate authenticity via IS and IS(a), and diversity via IC-L.

Metric	Few-Shot Generation			Anomaly Synthesis				Zero-Shot Generation		
	DFMGAN	AnoDiff	AnoGen	CutPaste	DRAEM	NSA	RealNet	AnomalyAny	AnoStyler	Ours
IS $\uparrow$	1.72	1.80	1.77	1.76	1.76	1.44	1.64	2.02	2.04	2.06
IS(a) $\uparrow$	2.80	2.77	2.84	—	—	—	—	2.98	3.12	3.98
IC-L $\uparrow$	0.20	0.27	0.32	0.22	0.25	0.26	0.22	0.33	0.32	0.33

Table 2: Comparison of anomaly detection with zero-shot anomaly synthesis and generation methods and decoupling methods on MVTec AD.

Method	Image-Level			Pixel-Level			
	I-AUC	I-AP	I-F1	P-AUC	P-AP	P-F1	PRO
<i>Zero-shot Anomaly Synthesis and Generation Methods</i>							
CutPaste [20]	89.8	92.1	89.8	88.2	51.9	50.7	76.4
DRAEM [37]	94.6	97.0	94.4	92.2	54.1	53.1	83.1
NSA [29]	93.0	95.6	91.6	92.0	52.6	52.5	82.2
RealNet [38]	95.2	97.0	95.3	94.0	57.7	56.6	85.2
AnomalyAny [32]	95.2	96.9	96.3	89.0	62.7	59.9	84.7
AnoStyler [31]	98.0	99.0	97.0	94.4	62.9	60.7	88.3
<i>Dual-LoRA Decoupling Methods</i>							
UnzipLoRA [22]	96.6	98.5	96.6	94.6	61.9	59.2	85.0
QR-LoRA [35]	95.4	97.2	95.3	93.1	60.2	58.3	84.2
Ours	98.0	99.1	96.7	95.3	68.0	64.5	89.6

5.2 Comparison in Anomaly Image Generation

We evaluate the authenticity and diversity of the generated anomalies in Tab. 1. Our method achieves the highest IS (2.06) and IC-L (0.33). Additionally, for the evaluation of anomaly authenticity, our approach obtains an IS(a) of 3.98. This indicates that the anomalies generated by our method possess higher authenticity. With the decoupled structure, our method generates diverse anomalies while keeping them realistic, achieving better results than several few-shot methods. Fig. 5 shows the visual comparisons. Anomaly synthesis methods (e.g., DRAEM, NSA) produce obvious artifacts, such as unnatural stripes on the hazelnut. Few-shot methods (e.g., DFMGAN, AnoGen) lack realism and show unnatural blending. Recent zero-shot methods like AnomalyAny and AnoStyler struggle with faithful anomalies (e.g., tile and wood). In contrast, our method generates highly authentic anomalies. The generated anomaly seamlessly blends with the product textures (e.g., screw, carpet). More generated images are provided in Appendix D.

5.3 Comparison in Downstream Anomaly Detection Tasks

Following [16], we train a U-Net [27] segmentation model on the generated image-mask pairs to evaluate downstream performance on the MVTec AD dataset,

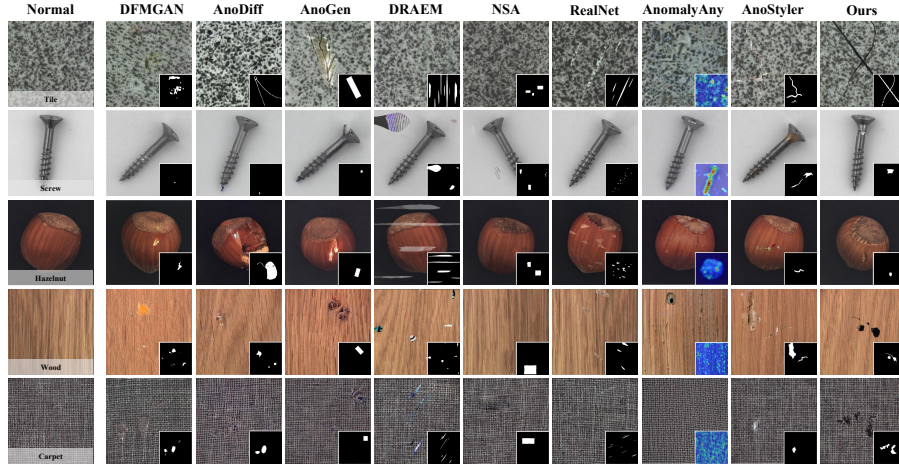


Fig. 5: Comparison of generated anomaly images.

with results summarized in Tab. 2. By establishing text-structure binding via DR-Flow, our method achieves 68.0% in P-AP and 64.5% in P-F1, outperforming the zero-shot method AnoStyler by 5.3% and 3.8%. It also obtains a PRO score of 89.6%. Furthermore, by explicitly binding the anomaly structure to the abnormal token, our method yields a 6.1% P-AP improvement over the Dual-LoRA decoupling method. For image-level metrics, our method also achieves the highest performance, with 98.0% I-AUC and 99.1% I-AP. These results indicate that the generated zero-shot anomalies are realistic and effectively contribute to downstream detection tasks.

5.4 Generalization of Acquired Anomaly Structures

To further verify the generalization of the anomaly structures acquired by our method, we introduce 12 distinct and highly complex product categories of the VisA dataset [40]. Then, we directly recouple the anomaly structures previously acquired from the MVTec AD dataset with these products. This design tests whether the structural priors acquired from the reference set can be seamlessly recoupled with the normal textures of distinct product categories. As shown in Tab. 3, our method consistently outperforms existing methods. While maintaining state-of-the-art image-level metrics (95.9% I-AUC), our structural decoupling and explicit recoupling strategy yields massive gains in pixel-level localization. Most notably, our method reaches 41.5% in P-AP and 46.4% in P-F1, eclipsing the strongest method by absolute margins of 8.2% and 5.4%, respectively. This confirms that our framework successfully extracts generalizable anomaly structures, enabling robust zero-shot generation. More results on other datasets are shown in Appendix C.

Table 3: Comparison of anomaly detection on VisA.

Method	Image-Level			Pixel-Level			
	I-AUC	I-AP	I-F1	P-AUC	P-AP	P-F1	PRO
CutPaste [20]	86.3	86.9	87.1	88.4	32.2	39.6	77.7
DRAEM [37]	91.8	92.9	88.6	91.4	29.5	37.2	81.9
NSA [29]	87.3	89.8	84.2	92.6	26.1	34.2	74.2
RealNet [38]	92.6	93.8	89.2	92.2	33.3	41.0	83.0
AnomalyAny [32]	88.9	86.2	85.9	90.4	31.2	33.0	84.6
AnoStyler [31]	93.9	95.3	90.1	93.8	31.4	36.4	84.3
Ours	95.9	97.0	92.1	96.0	41.5	46.4	88.3

Table 4: Ablation study of core components.

Row	Components				Image-Level			Pixel-Level		
	AP	QK	PI-Flow	PCC	I-AUC	I-AP	I-F1	P-AUC	P-AP	P-F1
(a)	-	-	-	-	94.0	97.3	95.4	94.2	62.7	60.5
(b)	✓	-	-	-	95.7	98.0	96.2	94.3	63.2	61.1
(c)	✓	✓	-	-	96.9	98.8	96.3	95.0	66.7	63.0
(d)	✓	✓	✓	-	97.5	99.0	96.4	95.1	67.6	63.2
(e)	✓	✓	✓	✓	98.0	99.1	96.7	95.3	68.0	64.5

5.5 Ablation Studies

Tab. 4 presents the impact of each core component on MVTec AD. Starting from a unified LoRA baseline (a), we incrementally add AP-Routing (short for AP), QK-Routing (short for QK), PI-Flow, and PCC. Introducing AP-Routing (b) significantly improves image-level detection by ensuring token split. Adding QK-Routing (c) enhances all metrics by enabling structure extraction. Incorporating PI-Flow (d) further improves image-level and pixel-level precision by preventing the abnormal token from binding the source product. Finally, the full method (e) yields the best overall performance. These results validate the effectiveness and complementarity of each proposed component. More ablation studies are provided in Appendix B.

6 Conclusion

In this paper, we propose DeCo to address the critical challenges of zero-shot industrial anomaly generation. DeCo explicitly decouples the anomaly structure from its source product and recouples it with the target product. Specifically, DR-Flow and PI-Flow work collaboratively to extract a pure anomaly structure. Then, our hybrid injection and Product Compatibility Correction seamlessly fuse this acquired anomaly structure into the target product. Extensive experiments

confirm that DeCo establishes a new state-of-the-art. Training downstream detection models on our generated data yields pixel AP improvements of 5.1% on MVTec AD and 8.2% on VisA.

7 Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant No.62176098. The computation is completed on the HPC Platform of Huazhong University of Science and Technology.

References

1. Avrahami, O., Lischinski, D., Fried, O.: Blended diffusion for text-driven editing of natural images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18208–18218 (2022)
2. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9592–9600 (2019)
3. Cao, M., Wang, X., Qi, Z., Shan, Y., Qie, X., Zheng, Y.: Masactrl: Tuning-free mutual self-attention control for consistent image synthesis and editing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22560–22570 (2023)
4. Chung, J., Hyun, S., Heo, J.P.: Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8795–8805 (2024)
5. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
6. Dong, J., Liang, W., Li, H., Zhang, D., Cao, M., Ding, H., Khan, S.H., Shahbaz Khan, F.: How to continually adapt text-to-image diffusion models for flexible customization? *Advances in Neural Information Processing Systems* **37**, 130057–130083 (2024)
7. Duan, Y., Hong, Y., Niu, L., Zhang, L.: Few-shot defect image generation via defect-aware feature manipulation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 571–578 (2023)
8. Duan, Y., Hong, Y., Niu, L., Zhang, L.: Few-shot defect image generation via defect-aware feature manipulation. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 571–578 (2023)
9. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Pines, D., Sauer, A., Tja, M., et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: Forty-first International Conference on Machine Learning (2024)
10. Frenkel, Y., Vinker, Y., Shamir, A., Cohen-Or, D.: Implicit style-content separation using b-lora. In: European Conference on Computer Vision. pp. 181–198. Springer (2024)

11. Gu, Y., Wang, X., Wu, J.Z., Shi, Y., Chen, Y., Fan, Z., Xiao, W., Zhao, R., Chang, S., Wu, W., et al.: Mix-of-show: Decentralized low-rank adaptation for multi-concept customization of diffusion models. *Advances in Neural Information Processing Systems* **36**, 15890–15902 (2023)
12. Gui, G., Gao, B.B., Liu, J., Wang, C., Wu, Y.: Few-shot anomaly-driven generation for anomaly classification and segmentation. In: *European conference on computer vision*. pp. 210–226. Springer (2024)
13. He, S., Liu, L., Shu, X., Wang, B., Feng, Y., Zhao, S.: Anomalycontrol: Learning cross-modal semantic features for controllable anomaly synthesis. *arXiv preprint arXiv:2412.06510* (2024)
14. Ho, J., Salimans, T.: Classifier-free diffusion guidance. In: *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications* (2021)
15. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021)
16. Hu, T., Zhang, J., Yi, R., Du, Y., Chen, X., Liu, L., Wang, Y., Wang, C.: Anomaly-diffusion: Few-shot anomaly image generation with diffusion model. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 8526–8534 (2024)
17. Jiang, J., Zhang, Y., Feng, K., Wu, X., Li, W., Pei, R., Li, F., Zuo, W.: Mc²: Multi-concept guidance for customized multi-concept generation. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 2802–2812 (2025)
18. Jiang, Y., Luo, W., Zhang, H., Chen, Q., Yao, H., Shen, W., Cao, Y.: Anomagic: Crossmodal prompt-driven zero-shot anomaly generation. *arXiv preprint arXiv:2511.10020* (2025)
19. Jin, Y., Peng, J., He, Q., Hu, T., Wu, J., Chen, H., Wang, H., Zhu, W., Chi, M., Liu, J., et al.: 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)
20. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: 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)
21. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: *The Eleventh International Conference on Learning Representations* (2023)
22. Liu, C., Shah, V., Cui, A., Lazebnik, S.: Unziplora: Separating content and style from a single image. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 16776–16785 (2025)
23. Liu, X., Gong, C., et al.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: *The Eleventh International Conference on Learning Representations* (2023)
24. Liu, Z., Feng, R., Zhu, K., Zhang, Y., Zheng, K., Liu, Y., Zhao, D., Zhou, J., Cao, Y.: Cones: Concept neurons in diffusion models for customized generation. In: *International Conference on Machine Learning*. pp. 21548–21566 (2023)
25. Ojha, U., Li, Y., Lu, J., Efros, A.A., Lee, Y.J., Shechtman, E., Zhang, R.: 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)
26. Ouyang, Z., Li, Z., Hou, Q.: K-lora: Unlocking training-free fusion of any subject and style loras. In: *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 13041–13050 (2025)

27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention*. pp. 234–241. Springer (2015)
28. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. *Advances in Neural Information Processing Systems* **29** (2016)
29. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural synthetic anomalies for self-supervised anomaly detection and localization. In: *European Conference on Computer Vision*. pp. 474–489. Springer (2022)
30. Shah, V., Ruiz, N., Cole, F., Lu, E., Lazebnik, S., Li, Y., Jampani, V.: Ziplora: Any subject in any style by effectively merging loras. In: *European Conference on Computer Vision*. pp. 422–438. Springer (2024)
31. So, Y., Kang, S.: Anostyler: Text-driven localized anomaly generation via lightweight style transfer. In: *Proceedings of the AAAI Conference on Artificial Intelligence* (2026)
32. Sun, H., Cao, Y., Dong, H., Fink, O.: Unseen visual anomaly generation. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 25508–25517 (2025)
33. Tumanyan, N., Geyer, M., Barmatz, S., Dekel, T.: Plug-and-play diffusion features for text-driven image-to-image translation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1921–1930 (2023)
34. Wu, T., Wang, J., Zhao, Z., Wong, N.: Mixture-of-subspaces in low-rank adaptation. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. pp. 7880–7899 (2024)
35. Yang, J., Ma, Y., Di, D., Cui, J., Li, H., Chen, W., Xie, Y., Yang, X., Zuo, W.: Qr-lora: Efficient and disentangled fine-tuning via qr decomposition for customized generation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 17587–17597 (2025)
36. Yang, Y., Wang, W., Peng, L., Song, C., Chen, Y., Li, H., Yang, X., Lu, Q., Cai, D., He, X., et al.: Lora-composer: Leveraging low-rank adaptation for multi-concept customization in training-free diffusion models. *IEEE Transactions on Image Processing* **34**, 8145–8158 (2025)
37. Zavrtnik, V., Kristan, M., Skočaj, D.: 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)
38. Zhang, X., Xu, M., Zhou, X.: 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)
39. Zhewei Dai, Shilei Zeng, H.L.X.L.F.X.Y.Z.: Seas: Few-shot industrial anomaly image generation with separation and sharing fine-tuning. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2025)
40. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: *European Conference on Computer Vision*. pp. 392–408 (2022)