AnomalyHybrid: A Domain-agnostic Generative Framework for General Anomaly Detection

Ying Zhao
Ricoh Software Research Center (Beijing) Co., Ltd., China
zy_deepwhite_zy@hotmail.com

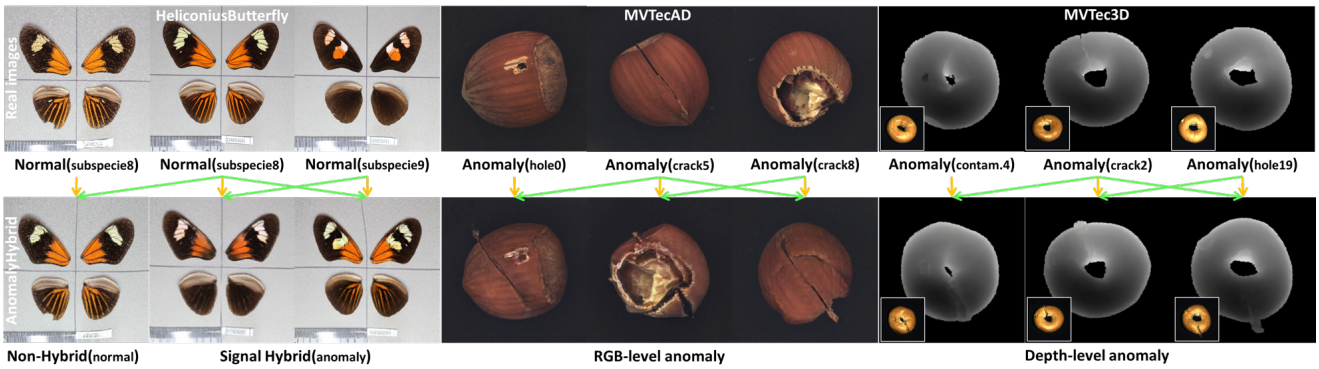

Figure 1. **AnomalyHybrid is a domain-agnostic generative framework.** Unlike prior industrial anomaly specialists, it generates general anomalies simply by combining the reference(green arrows) and target(yellow arrows) images.

Abstract

Anomaly generation is an effective way to mitigate data scarcity for anomaly detection task. Most existing works shine at industrial anomaly generation with multiple specialists or large generative models, rarely generalizing to anomalies in other applications. In this paper, we present AnomalyHybrid, a domain-agnostic framework designed to generate authentic and diverse anomalies simply by combining the reference and target images. AnomalyHybrid is a Generative Adversarial Network(GAN)-based framework having two decoders that integrate the appearance of reference image into the depth and edge structures of target image respectively. With the help of depth decoders, AnomalyHybrid achieves authentic generation especially for the anomalies with depth values changing, such a s protrusion and dent. More, it relaxes the fine granularity structural control of the edge decoder and brings more diversity. Without using annotations, AnomalyHybrid is easily trained with sets of color, depth and edge of same images having different augmentations. Extensive experiments carried on HeliconiusButterfly, MVTecAD and MVTec3D datasets demonstrate that AnomalyHybrid surpasses the GAN-based state-of-the-art on anomaly generation and its downstream anomaly classification, detection and segmentation tasks. On MVTecAD dataset, AnomalyHybrid achieves 2.06/0.32 IS/LPIPS for anomaly generation, 52.6 Acc for anomaly classification with ResNet34, 97.3/72.9 AP for image/pixel-level anomaly detection with a simple UNet.

1. Introduction

Visual anomaly detection benefits the work and economic efficiency for manufacturing industries. As anomaly is infrequent, it is barely possible to gather all kinds of real anomaly samples for training anomaly detectors. The performance of anomaly detectors are greatly constrained by the scarcity of real anomalies. Besides that, the normal appearance of a same product can also varies from sample to sample. With limited training data, it is challenging to construct a robust anomaly detector to handle unseen cases. The emergence of model-free anomaly synthesis [5, 14, 21, 28, 30, 40] and model-based anomaly generation methods [9, 11, 33, 38, 39] has catalyzed significant strides in anomaly detection.

The model-free anomaly synthesis methods [5, 14, 28, 30] are basically based on image processing paradigm of fusing selected anomaly regions to the normal image. They mainly differ in strategies of region selection, anomaly sourcing and image fusion. While evolving with different fusion strategies [36, 37], the synthetic anomalies produced by model-free methods are far from realistic. The model-based anomaly generation methods output more re-

AnomalyHybrid：一种领域无关的通用异常检测生成框架

应赵 理光软件研究开发中心（
北京）有限公司，中国 zy_deepwhite_zy@hotmail.com

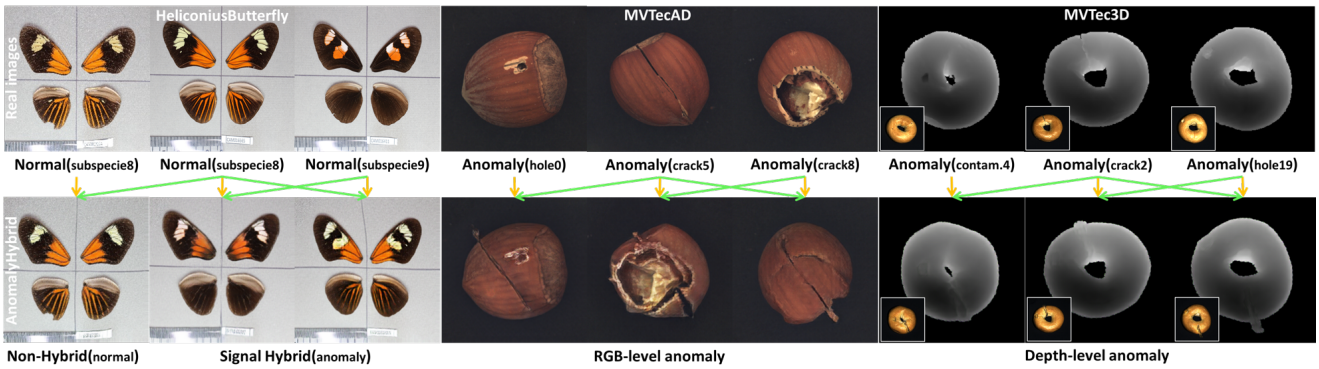

图1. AnomalyHybrid是一个领域无关的生成框架。与以往的工业异常检测专家不同，它能够生成通用 al
仅通过结合参考（绿色箭头）和目标（黄色箭头）图像即可检测异常。

摘要

Anomaly generation is an effective way to mitigate data scarcity for anomaly detection task. Most existing works shine at industrial anomaly generation with multiple specialists or large generative models, rarely generalizing to anomalies in other applications. In this paper, we present AnomalyHybrid, a domain-agnostic framework designed to generate authentic and diverse anomalies simply by combining the reference and target images. AnomalyHybrid is a Generative Adversarial Network(GAN)-based framework having two decoders that integrate the appearance of reference image into the depth and edge structures of target image respectively. With the help of depth decoders, AnomalyHybrid achieves authentic generation especially for the anomalies with depth values changing, such a s protrusion and dent. More, it relaxes the fine granularity structural control of the edge decoder and brings more diversity. Without using annotations, AnomalyHybrid is easily trained with sets of color, depth and edge of same images having different augmentations. Extensive experiments carried on HeliconiusButterfly, MVTecAD and MVTec3D datasets demonstrate that AnomalyHybrid surpasses the GAN-based state-of-the-art on anomaly generation and its downstream anomaly classification, detection and segmentation tasks. On MVTecAD dataset, AnomalyHybrid achieves 2.06/0.32 IS/LPIPS for anomaly generation, 52.6 Acc for anomaly classification with ResNet34, 97.3/72.9 AP for image/pixel-level anomaly detection with a simple UNet.

1. 引言

视觉异常检测有助于提升制造业的工作效率与经济效益。由于异常情况较为罕见，几乎不可能收集到所有类型的真实异常样本来训练异常检测器。异常检测器的性能在很大程度上受限于真实异常样本的稀缺性。此外，同一产品的正常外观也可能因样本而异。在训练数据有限的情况下，构建一个能够处理未知情况的鲁棒异常检测器具有挑战性。无模型异常合成方法[5, 14, 21, 28, 30, 40]与基于模型的异常生成方法[9, 11, 33, 38, 39]的出现，极大地推动了异常检测领域的发展。

无模型异常合成方法[5, 14, 28, 30]基本遵循将选定异常区域融合至正常图像的图像处理范式，其主要差异在于区域选择策略、异常源获取及图像融合方式。尽管通过不同融合策略[36, 37]持续演进，但无模型方法生成的合成异常仍与真实情况相去甚远。基于模型的异常生成方法则能输出更真

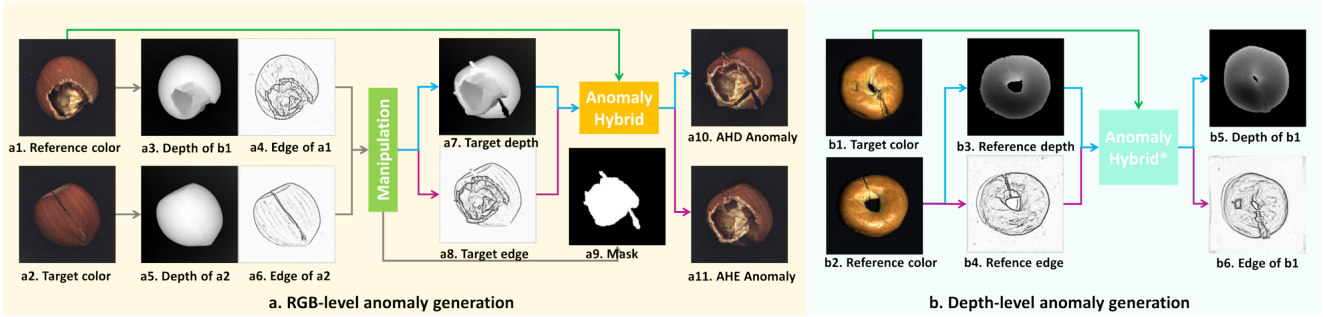

Figure 2. **Inference workflow of AnomalyHybrid.** AnomalyHybrid combines the appearance of reference image with the depth and edge structural target image. It generates global and local anomalies without and with applying manipulations on target depth and edge maps.

alistic samples by employing generative models, like Generative Adversarial Networks (GANs) and Diffusion models, in supervised [9, 11] and unsupervised [33, 38, 39] ways. Though effective, the supervised methods barely can generate unseen anomaly types. On the other hand, the unsupervised generative methods focus only on image-level anomaly generation but ignore the depth-level. They are still struggling to generate realistic anomalies along with depth values changing, such as protrusion and dent. Besides generation quality, efficiency and generalization capability of anomaly generation methods are also worthy of attention. With increasing amounts of objects from versatile datasets, it is infeasible to learn multiple large and dedicated generative models per category or dataset.

To solve aforementioned problems, we propose AnomalyHybrid that is a simple framework designed to generate diversity and authentic anomalies across application domains with color, depth and edge conditional controls. It is a GAN-based framework having two decoders that integrate the appearance of reference image into the depth and edge structures of target image respectively. To demonstrate the generation quality and generalization ability of AnomalyHybrid, Fig.1 visualizes the results produced by models trained on 3 datasets across application domains. For HeliconiusButterfly dataset, the anomaly is the hybrid butterfly of two non-hybrid subspecies. The hybrid butterfly contains appearance information of its parents. On the contrary, for the industrial anomaly datasets, such as MVtecAD and MVtec3D, the anomaly is local regions having different in depth or color values, such as crack and hole anomalies. Without network structure modification, the proposed AnomalyHybrid can easily transfer to generate anomalies for these different application domains.

In summary, we make following main contributions:

- We propose a domain-agnostic framework, AnomalyHybrid, that not only works for industrial anomaly scenarios but also can be easily transfer to generate anomalies for broad applications. Experiments carried on industrial and biological datasets valid its generalization ability.
- AnomalyHybrid consists of depth and edge decoders that substantiate each other to generates diverse and authentic both normal and anomaly images. We achieve 2.06/0.32 and 1.85/0.24 IS/LPIPS for anomaly generation on MVTecAD and MVTec3D, that is better than recent GAN-based and diffusion-based SOTA.
- We conduct extensive experiments to demonstrate that our generated images bring benefits to downstream anomaly detection tasks. On MVTecAD dataset, we achieve 52.6 Acc for anomaly classification with ResNet34, 97.3/72.9 AP for image/pixel-level anomaly detection with a simple UNet, that surpasses the GAN-based SOTA.

2. Related works

Image-level Anomaly Synthesis. With the merits of simple and efficient, anomaly synthesis is widely used in unsupervised anomaly detection methods [14, 21, 28, 36, 37, 40]. CutPaste [14] syntheses anomalies by creating local discontinuous regions with the cut-paste processing. It cuts local rectangular regions from normal images and directly paste them back at random positions. SPD [40] and NSA [21] improve it by adding different strategies to smooth the pasting boundary. To increase diversity of synthetic anomaly, Draem [28] extracts anomaly source from an extra DTD [6] dataset in irregular regions obtained by using binarized Perlin noise. To make the synthesis more naturally, JNLD [36] simulates different levels of anomalies based on the just noticeable distortion [25]. OmniAL [37] extends JNLD [36] by controlling the portion of synthetic anomalies with a panel-guided strategy.

Depth-level Anomaly Synthesis. Recent methods flourish the Perlin noise based anomaly synthesis paradigm of Draem [28] from various aspects, such as EasyNet [5], DBRN [3], 3DSR [30] and 3Draem [29] extend it to depth-level anomaly synthesis. EasyNet [5] takes the Perlin noise as the anomaly depth values and injects them to the selected regions of depth image to produce depth-level anomaly. Meanwhile, it regards random texture as the anomaly image values and uses the same Perlin noise to guide the image-level anomaly synthesis. Slight differently, DBRN [3] simulates the anomaly depth values by normalizing the

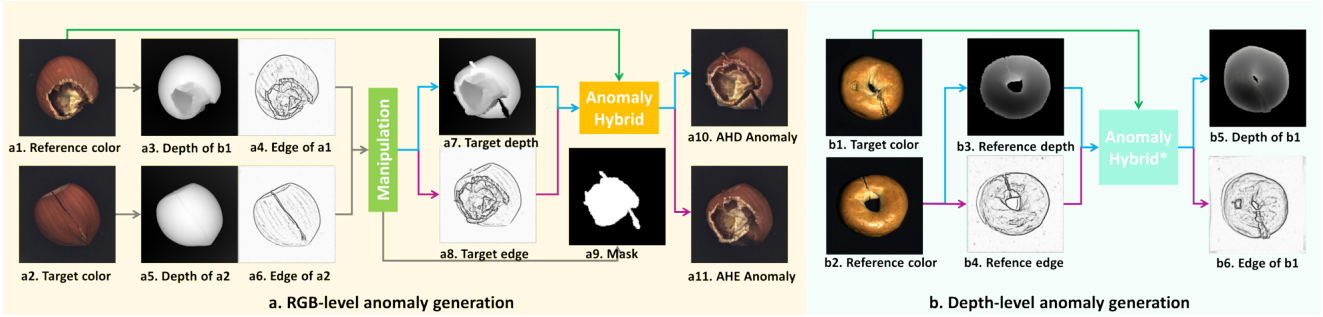

图2. AnomalyHybrid的推理工作流程。AnomalyHybrid将参考图像的外观与深度和边缘信息相结合。 e结构目标图像。它在不对目标深度和边缘图应用操作的情况下生成全局和局部异常，并在应用操作时生成全局和局部异常。

通过采用生成模型，如生成对抗网络（GAN）和扩散模型，以有监督[9, 11]和无监督[33, 38, 39]的方式生成逼真的异常样本。尽管有效，但有监督方法几乎无法生成未见过的异常类型。另一方面，无监督生成方法仅关注图像级别的异常生成，却忽略了深度级别。它们仍难以生成随深度值变化（如凸起和凹陷）的真实异常。除了生成质量外，异常生成方法的效率和泛化能力也值得关注。随着多样化数据集中物体数量的增加，为每个类别或数据集学习多个大型专用生成模型是不切实际的。

为解决上述问题，我们提出了AnomalyHybrid——这是一个简洁的框架，旨在通过色彩、深度和边缘条件控制，跨应用领域生成多样且真实的异常数据。该框架基于生成对抗网络，配备双解码器结构，可分别将参考图像的外观特征融合至目标图像的深度与边缘结构中。为展示AnomalyHybrid的生成质量与泛化能力，图1可视化呈现了跨三个应用领域数据集训练的模型输出结果。在HeliconiusButterfly数据集中，异常表现为两个非杂交亚种蝴蝶的混合体，这种杂交蝴蝶承载了亲代的外观信息。相反，在工业异常数据集（如MVtec AD与MVtec3D）中，异常则表现为局部区域的深度或色彩数值差异，例如裂纹与孔洞缺陷。无需调整网络结构，所提出的AnomalyHybrid框架即可轻松迁移至不同应用领域生成相应的异常样本。

总而言之，我们做出了以下主要贡献：

- 我们提出了一个领域无关的框架——AnomalyHybrid，它不仅适用于工业异常检测场景，还能轻松迁移至广泛的应用中以生成异常数据。在工业和生物数据集上进行的实验验证了其泛化能力。
- AnomalyHybrid由深度和边缘解码器组成，它们相互补充以生成多样且真实的我们实现了在MVTecAD和MVTec3D上异常生成的2.06/0.32和1.85/0.24 IS/LPIPS指标，这优于近期基于GAN和扩散模型的SOTA方法。
- 我们进行了大量实验，以证明我们生成的图像对下游异常检测任务具有显著增益。在MVTecAD数据集上，使用ResNet34实现了52.6%的异常分类准确率，通过简单的UNet在图像/像素级异常检测中分别达到97.3%和72.9%的平均精度，这些结果超越了基于GAN的当前最优方法。

2. 相关工作

图像级异常合成。凭借简单高效的优势，异常合成技术被广泛应用于无监督异常检测方法中[14, 21, 28, 36, 37, 40]。CutPaste[14]通过剪切-粘贴处理创建局部不连续区域来合成异常，该方法从正常图像中切割局部矩形区域，并直接随机粘贴至其他位置。SPD[40]和NSA[21]通过采用不同策略平滑粘贴边界来改进该方法。为增加合成异常的多样性，Draem[28]从额外DTD[6]数据集中提取异常源，并利用二值化Perlin噪声生成不规则区域进行合成。为使合成效果更自然，JNLD[36]基于恰可察觉失真模型[25]模拟了不同等级的异常。OmniAL[37]通过面板引导策略控制合成异常的比例，进一步扩展了JNLD[36]的方法。

深度级异常合成。近期方法从多个方面丰富了基于Perlin噪声的Draem[28]异常合成范式，例如EasyNet[5]、DBRN[3]、3DSR[30]和3Draem[29]将其扩展至深度级异常合成。EasyNet[5]将Perlin噪声作为异常深度值注入深度图像的选定区域以生成深度级异常，同时将随机纹理视为异常图像值，并使用相同的Perlin噪声指导图像级异常合成。略有不同的是，DBRN[3]通过归一化操作模拟异常深度值。

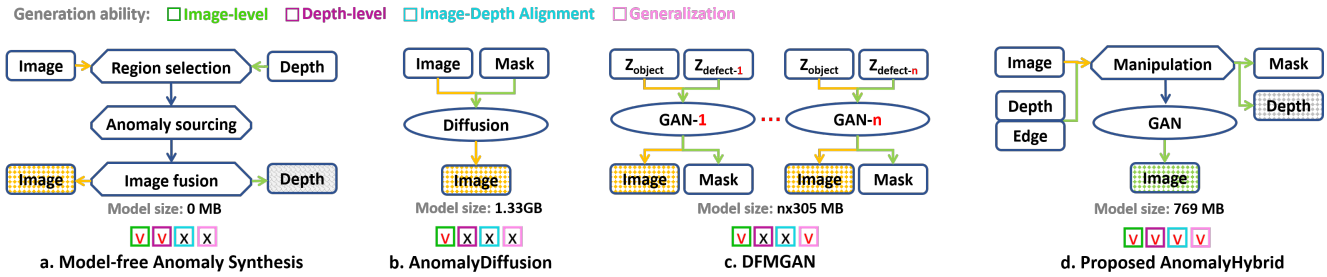

Figure 3. **Comparison of related frameworks.** (a) summarizes the three key components in model-free anomaly synthesis methods, such as Draem [28] and 3DSR [30]. (b) relies on large diffusion model. (c) achieves authentic anomaly generation by learning multiple defect-aware specialists. Comparing to previous workflows, (d) our proposed AnomalyHybrid has more comprehensive generation ability.

same texture image used for image-level anomaly synthesis. Based on the handcrafted principles of anomaly depth values, changing gradually, capturing local changes and variable average object depth, 3DSR [30] forges the depth-level anomaly by adapting the Perlin noise image with a randomized affine transform. Similarly, 3Draem [29] uses the Perlin noise generator to create anomaly regions and smooths the simulated depth values to ensure more consistent local depth changes. Though the synthetic image-level and depth-level anomalies are in same location, there is no guarantee that the texture and depth have the same changing tendency. Without considering alignment of depth-level and image-level information, these methods usually generate unrealistic anomalies.

Image-level Unsupervised Anomaly Generation. RealNet [33] proposes a diffusion process-based synthesis strategy that generates anomaly samples by blending the normal image with the diffusion generated anomalous texture. To mimic real anomalies distribution, it carefully selects the parameter that controls the strength of anomaly generation. GRAD [7] proposes a diffusion model to generate both structural and logical anomaly patterns. It generates contrastive patterns by preserving the local structures while disregarding the global structures present in normal images. Moreover, it uses a self-supervised re-weighting mechanism to handle the challenge of long-tailed and unlabeled synthetic contrastive patterns. LogicAL [39] proposes to generate logical and structural anomalies with a GAN-based framework by manipulating edges in semantic and arbitrary regions. AnomalyFactory [38] designs a GAN-based network architecture that combines structure of a target map and appearance of a reference color image with the guidance of a learned heatmap. It has strong scalability in generating various types of samples with anomaly heatmaps for training an unified anomaly predictor for multiple categories of different datasets. Due to lack of depth information, these methods barely generate anomalies having realistic depth changing, such as protrusion and dent.

Image-level Supervised Anomaly Generation. To obtain realistic anomalies, more and more methods [9, 11, 18, 31] are equipped with the powerful generative models, like GANs and Diffusion models. SDGAN [18] generates surface defects with GANs trained by cycle consistency loss on a small number of real defect images. DefectGAN [31] generates realistic defect samples with GANs by simulating the defacement and restoration processes with a layer-wise composition strategy. DFMGAN [9] attaches defect-aware residual blocks to the pre-trained StyleGAN2 [13] backbone and manipulates the features within the learnt defect masks. AnoDiffusion [11] proposes a diffusion-based few-shot anomaly generation model that separately learns the anomaly appearance and location information, then generates the anomaly on the masked normal samples. These supervised anomaly generation methods, though effective, rely on real anomalous images and cannot generate unseen anomaly types.

Depth-to-image generation. According to recent survey [17] on multimodal unsupervised anomaly detection, there is no method to generate depth-level anomaly with generative models. We further investigate generation methods of depth-to-image. To enhance the controllability of pre-trained text-to-image diffusion models, many efforts [15, 20, 32, 35] focus on incorporating it with image-based conditional controls, such as depth map. UniControl [20] introduces a mixture of expert (MOE)-style adapter and a task-aware HyperNet to modulate the diffusion models, enabling the adaptation to different condition-to-image tasks simultaneously. Uni-ControlNet [35] leverages two lightweight adapters to enable local and global controls over pre-trained text-to-image diffusion models. With shared local and global condition encoder adapters, it injects multi-scale local condition and concatenates global visual conditional tokens with text tokens respectively. ControlNet [32] proposes a neural network architecture to add spatial conditioning controls to large, pretrained text-to-image diffusion models. The neural architecture is connected with zero-initialized convolution layers that progressively grow the parameters from zero and ensure that no harmful noise could affect the fine-tuning. ControlNet++ [15] improves controllable generation of ControlNet [32] by explicitly optimizing pixel-level cycle consistency between generated images and conditional controls.

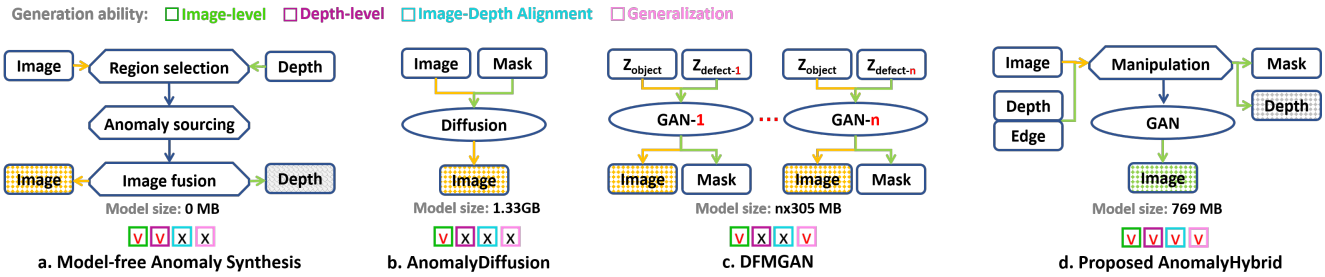

图3. 相关框架对比。(a) 总结了无模型异常合成方法中的三个关键组件，例如Draem [28]和3DSR [30]。(b) 依赖于大型扩散模型。(c) 通过学习多个缺陷感知专家实现逼真的异常生成。与先前工作流程相比，(d) 我们提出的AnomalyHybrid具有更全面的生成能力。

同一纹理图像被用于图像级异常合成。基于手工设计的异常深度值原则——逐渐变化、捕捉局部变化以及可变平均物体深度，3DSR [30] 通过采用随机仿射变换调整Perlin噪声图像来伪造深度级异常。类似地，3Draem [29] 使用Perlin噪声生成器创建异常区域，并平滑模拟的深度值以确保更一致的局部深度变化。尽管合成的图像级和深度级异常位于同一位置，但无法保证纹理和深度具有相同的变化趋势。由于未考虑深度级与图像级信息的对齐，这些方法通常生成不真实的异常。

图像级无监督异常生成。Re-alNet [33]提出了一种基于扩散过程的合成策略，通过将正常图像与扩散生成的异常纹理融合来生成异常样本。为了模拟真实异常分布，该方法精心选择了控制异常生成强度的参数。GRAD [7]提出了一种扩散模型，用于生成结构和逻辑异常模式。它通过保留局部结构而忽略正常图像中的全局结构来生成对比模式。此外，该方法采用自监督重加权机制来处理长尾分布且未标记的合成对比模式带来的挑战。LogicAL [39]提出通过基于GAN的框架，在语义区域和任意区域中操纵边缘来生成逻辑和结构异常。AnomalyFactory [38]设计了一种基于GAN的网络架构，该架构结合了目标图的结构和参考彩色图像的外观，并在学习得到的热力图指导下进行生成。该方法在生成具有异常热力图的各种类型样本方面具有很强的可扩展性，可用于训练针对多个不同数据集类别的统一异常预测器。由于缺乏深度信息，这些方法几乎无法生成具有真实深度变化的异常，例如凸起和凹陷。

图像级监督异常生成。为了获得逼真的异常，越来越多的方法[9, 11, 18, 31]配备了强大的生成模型，例如GAN与扩散模型。SDGAN [18] 通过少量真实缺陷图像上的循环一致性损失训练GAN生成表面缺陷。DefectGAN [31] 采用分层组合策略模拟缺陷形成与修复过程，利用GAN生成逼真的缺陷样本。DFMGAN [9] 在预训练的StyleGAN2 [13] 骨干网络上添加缺陷感知残差块，并在学习到的缺陷掩码内操纵特征。AnoDiffusion [11] 提出基于扩散的少样本异常生成模型，分别学习异常外观与位置信息，随后在掩码后的正常样本上生成异常。这些有监督的异常生成方法虽有效，但依赖真实异常图像，无法生成未见过的异常类型。

深度到图像的生成。根据近期关于多模态无监督异常检测的综述[17]，目前尚无利用生成模型生成深度级异常的方法。我们进一步研究了深度到图像的生成方法。为增强预训练文本到图像扩散模型的可控性，许多研究[15, 20, 32, 35]致力于将其与基于图像的条件控制（如深度图）相结合。UniControl[20]引入了专家混合（MOE）风格的适配器和任务感知超网络来调制扩散模型，使其能同时适应不同的条件到图像任务。Uni-ControlNet[35]利用两个轻量级适配器实现对预训练文本到图像扩散模型的局部与全局控制。通过共享的局部与全局条件编码器适配器，它注入多尺度局部条件，并将全局视觉条件标记与文本标记分别拼接。ControlNet[32]提出了一种神经网络架构，为大型预训练文本到图像扩散模型添加空间条件控制。该架构通过零初始化卷积层连接，使参数从零逐步增长，确保有害噪声不会影响微调过程。ControlNet++[15]通过显式优化生成图像与条件控制之间的像素级循环一致性，改进了ControlNet[32]的可控生成能力。

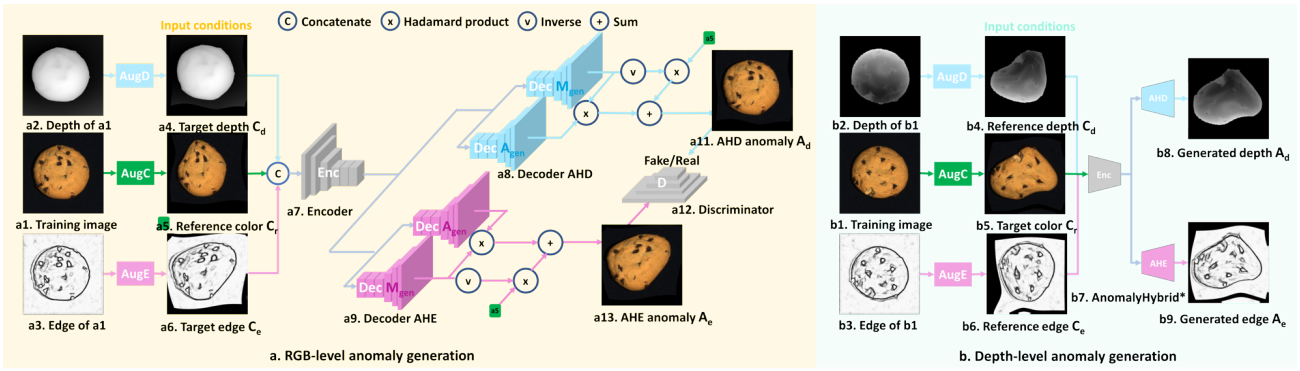

Figure 4. **Training workflow of AnomalyHybrid.** AnomalyHybrid is trained with sets of depth, color, edge of same images but having different augmentations. It consists of an encoder, two decoders and a discriminator. All decoders consist of anomaly texture and mask branches. The two-branch architecture forces the network to inject the appearance of reference to the structural of target depth and edge.

As summarized in Fig.3, different with the aforementioned methods, such as [28, 30, 33, 38], we propose a GAN-based framework that generates and aligns different levels of anomalies across versatile application domains.

3. Methods

3.1. Overview

AnomalyHybrid has a GAN-based network architecture. Fig.4 demonstrates the training workflow of AnomalyHybrid for both RGB-level and depth-level anomaly generation. Without using annotations, AnomalyHybrid is trained in an unsupervised way using sets of RGB, depth and edge of the same images with different augmentations. Since most datasets contain only RGB images, the edge and depth maps are extracted by pre-trained PiDiNet [22] and DepthAnythingV2 [26] respectively. As shown in Fig.4a, during training phase, AnomalyHybrid learns to generate RGB images for the target depth and edge maps conditioning on the reference RGB images. Thanks to the heavy augmentations, AnomalyHybrid learns to convert any depth and edge maps into RGB images that share appearance of the reference RGB images. It brings benefits that AnomalyHybrid can generate local anomalies by simply manipulating the edge and depth maps during inference phase, as shown in Fig.2a. As illustrated in Fig.4b, the depth-level anomaly generator, AnomalyHybrid*, is trained in a similar way but with different learning targets. It learns to extract depth and edge maps for the target RGB images referring to the input depth and edge maps. With AnomalyHybrid and AnomalyHybrid*, we can get aligned RGB-level and depth-level anomalies for 3D datasets, such as MVTec3D [2] demonstrated in Fig.2b.

3.2. Network architecture

AnomalyHybrid's network architecture is motivated by the observations of task representation and data flow. Anomaly generation A_{out}, in different levels, can be generally taken as a fusion of generated anomaly source A_{gen} and input reference content C_{in} under the guidance of a generated fusion map M_{gen}. It can be defined as following formulation.

$$A_{out} = C_{in} \cdot (1 - M_{gen}) + A_{gen} \cdot M_{gen} \quad (1)$$

To generate diverse and authentic anomaly source A_{gen}, we consider to simultaneously use depth and edge conditions to control local structure and image condition to control global appearance. Therefore, we design a GAN-based network architecture shown in Fig.4.

Generally, AnomalyHybrid follows the encoder-decoder architecture that is broadly used in conditional generative adversarial network (cGAN) models, such as pix2pixHD [24]. It mainly evolves four scales features encoding and decoding with basic convolution blocks and ResnetBlocks. The encoder extracts multi-scale features of concatenated conditions of image, depth and edge. Two groups of dedicated decoders, AHD and AHE, target to translate the encoded features into generations that are controlled by the depth and edge conditions respectively. Each group of decoders has two branches to generate the anomaly source A_{gen} and fusion map M_{gen} corresponding to Eq.1. The fusion results, AHD and AHE anomalies, are fed into a discriminator to distinguish the generation quality comparing to the real inputs.

3.3. Training data preparation

Edge. We extract edge maps with pre-trained PiDiNet [22] that is one of the appealing edge detectors that achieve a better trade-off between accuracy and efficiency. It integrates the advantages of traditional edge detectors and deep CNNs by using the well-designed pixel difference convolution. By learning from different annotations, it can produce four granularities of edge maps in which the first one contains the most details. As shown in the bottom of Fig.4a3 and Fig.4b3, the first granularity edge maps of PiDiNet [22] mainly contain high-level semantic edges, such as contours of deer, riverside and forest. Therefore, we take only the first granularity edge maps as our edge condition controls.

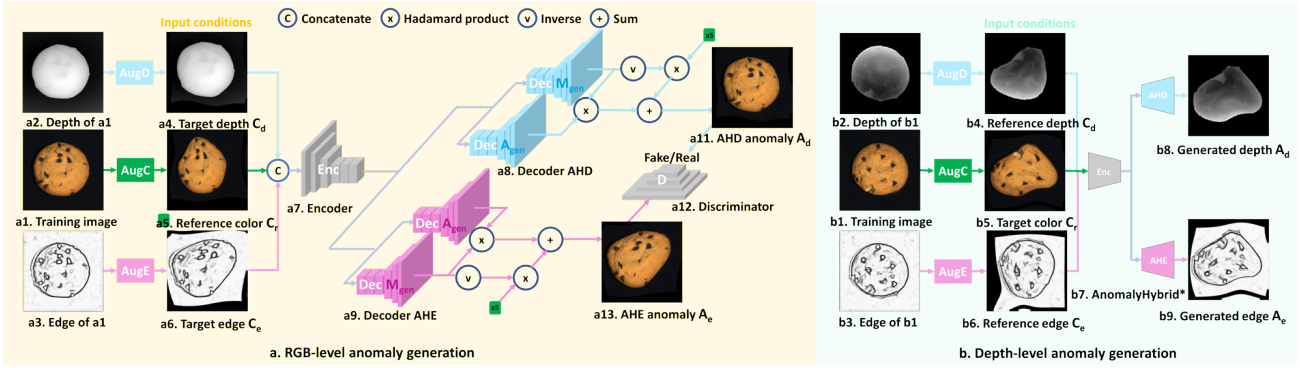

图4. AnomalyHybrid的训练流程。AnomalyHybrid使用同一图像经过不同增强处理的深度、颜色和边缘集合进行训练。它包含一个编码器、两个解码器和一个判别器。所有解码器均由异常纹理分支和掩码分支构成。这种双分支架构迫使网络将参考图像的外观注入目标深度和边缘的结构中。

如图3总结所示，与前述方法（如[28, 30, 33, 38]）不同，我们提出了一种基于GAN的框架，该框架能够生成并跨多种应用领域对齐不同级别的异常。

3. 方法

3.1. 概述

AnomalyHybrid采用基于GAN的网络架构。图4展示了AnomalyHybrid在RGB层级和深度层级异常生成中的训练流程。该方法无需标注，通过使用同一图像经过不同增强处理的RGB、深度及边缘数据集进行无监督训练。由于多数数据集仅包含RGB图像，边缘与深度图分别通过预训练的PiDiNet[22]和DepthAnythingV2[26]提取。如图4a所示，在训练阶段，AnomalyHybrid学习以参考RGB图像为条件，为目标深度图和边缘图生成对应的RGB图像。借助强数据增强，模型能够将任意深度图和边缘图转换为具有参考RGB图像外观特征的RGB图像。这使得AnomalyHybrid在推理阶段（如图2a所示）仅需操纵边缘图和深度图即可生成局部异常。如图4b所示，深度层级异常生成器AnomalyHybrid*采用类似训练方式但具有不同学习目标：它学习根据输入的深度图和边缘图，为目标RGB图像提取对应的深度与边缘信息。通过AnomalyHybrid与AnomalyHybrid*，我们可为三维数据集（如图2b所示的MVTec3D[2]）生成对齐的RGB层级与深度层级异常。

3.2. 网络架构

AnomalyHybrid的网络架构设计灵感来源于对任务表示与数据流的观察。在不同层面上，异常生成A_{out}通常可被视为作为生成的异常源 A_{gen} 与输入参考内容 C_{in} 在生成的融合图 M_{gen} 引导下的融合。其可定义为以下公式。

$$A_{out} = C_{in} \cdot (1 - M_{gen}) + A_{gen} \cdot M_{gen} \quad (1)$$

为了生成多样且真实的异常源A_{gen}，我们考虑同时利用深度和边缘条件来控制局部结构，以及图像条件来调控全局外观。因此，我们设计了如图4所示的基于GAN的网络架构。

通常，AnomalyHybrid遵循在条件生成对抗网络（cGAN）模型中广泛使用的编码器-解码器架构，例如pix2pixHD [24]。它主要通过基础卷积块和ResnetBlock实现四个尺度特征的编码与解码演进。编码器提取图像、深度和边缘拼接条件的多尺度特征。两组专用解码器——AHD与AHE——旨在将编码特征分别转换为受深度和边缘条件控制的生成结果。每组解码器包含两个分支，用于生成对应公式1的异常源A_{gen}与融合图M_{gen}。融合结果（即AHD与AHE异常）被送入判别器，通过与真实输入对比来区分生成质量。

3.3. 训练数据准备

边缘。我们使用预训练的PiDiNet [22]提取边缘图，这是一种在精度和效率之间实现更好平衡的吸引人的边缘检测器。它通过精心设计的像素差分卷积，融合了传统边缘检测器和深度CNN的优势。通过学习不同的标注，它可以生成四种粒度的边缘图，其中第一种包含最多的细节。如图4a3和图4b3底部所示，PiDiNet [22]的第一种粒度边缘图主要包含高级语义边缘，例如鹿的轮廓、河岸和森林。因此，我们仅将第一种粒度边缘图作为我们的边缘条件控制。

Depth. We estimate depth maps also with the pre-trained model, DepthAnythingV2 [26]. It is a powerful foundation model for monocular depth estimation. It produces robust predictions for complex scenes with fine details. Comparing to previous methods, its most critical modification is replacing all labelled real images with precise synthetic images. It overcomes the drawbacks of using real labelled images that contain noise and overlooks certain details in depth maps. Following this guidance, we use pseudo depth maps extracted by DepthAnythingV2 [26] for all datasets in RGB-level anomaly generation. However, the pseudo depth maps is much less accurate than the real ones, as demonstrated in Fig.4a2 and Fig.4b2. Therefore, we use the real depth maps in depth-level anomaly generation for MVTec3D [2] dataset.

3.4. Unsupervised training

Without using annotations, we construct sets of image, depth and edge conditions having different augmentations for training. Generally, the input conditions are applied different augmentations and the generate contents share the same augmentations with the target conditions. The consistency of augmentations are indicated with same color shown in Fig.4. As illustrated in Fig.4a, the generated AHD and AHE anomalies are applied same augmentations with the target depth and edge conditions accordingly.

Augmentation. Following [23, 38, 39], our augmentations mainly consist of local thin-plate-spline (TPS) [8] warps, resize-translation-padding and top-bottom/left-right flip. The local TPS randomly shifts 3x3 control points from a local region in the horizontal and vertical directions. Compare to selecting control points globally, the local TPS brings smaller spatial range of warps. The resize-translation-padding augmentation is mainly design to counter the drastically edge manipulation on texture categories, such as editing edges of the most parts of Hazelnut [1] category. The flip augmentation brings the benefits of direction-agnostic authentic generation. By using these three augmentations, we get a generator that is robust to the drastic manipulations, as shown in Fig.5.

Losses. Following [23, 38, 39], we use the VGG perceptual loss [12] L_{vgg} to measure the fidelity of generation $G(C_x, A_y)$ and the conditional GAN loss L_D to measure the differentiate between the generated and true images. The loss of anomaly generation L_A is defined as follow.

$$L_G(C_x, A_y; G) = L_{vgg}(G(C_x), A_y) \tag{2}$$

$$\begin{aligned} L_D(C_x, A_y; D, G) = log(D(C_x, A_y)) \\ +log(1 - D(C_x, G(C_x))) \end{aligned} \tag{3}$$

$$L_A = L_G(C_x, A_y; G) + L_D(C_x, A_y; D, G) \tag{4}$$

Where, $C_{x=\{d,r,e\}}$ indicate the input depth C_d, color C_r and edge C_e conditions, $A_{y=\{d,e\}}$ denote the target images A_d and A_e for AHD and AHE anomaly generations, G is the generator and D is the discriminator.

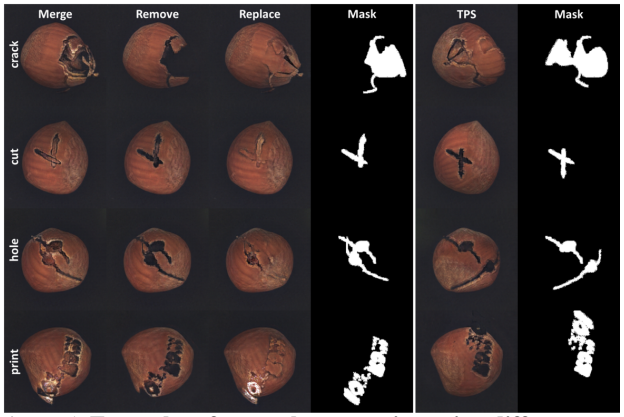

Figure 5. **Examples of anomaly generation using different manipulations on Hazelnut of MVTecAD [1].**

3.5. Anomaly generation

As shown in Fig.1, AnomalyHybrid can generate global and local anomalies for different applications. For global anomaly generation, it directly combines appearance and structure features of the reference and target images, such as the butterfly hybrid. As demonstrated in Fig.7, the AHD and AHE decoders bring diverse generations by focusing on depth and edge controlled butterfly hybrid(anomaly) respectively. In terms of local anomaly generation, AnomalyHybrid applies manipulations on depth and edge maps to make the anomalies more diverse and authentic. Fig.2a summarizes the workflow of generating anomalies by applying local manipulations on the depth and edge maps before feeding them to the generator.

Following [38, 39], the basic manipulation consists of simply removing, replacing, merging conditions and applying TPS on conditions in local regions.

- **Mask.** The local anomaly regions are the combination of augmented anomaly regions of reference and target images. The augmentation basically consists of resize, flip and crop.
- **Merge.** The depth and edge values in the anomaly regions of reference and target images are merged together to forge new anomaly textures.
- **Remove.** To forge the typical defects, such as crack, cut and hole, the depth and edge values in the anomaly regions are set as background values.
- **Replace.** To reduce the intensity level of manipulation, the depth and edge values in the anomaly regions of target image are replaced by the values in the reference image.
- **TPS.** To increase anomaly diversity, we apply TPS on anomaly regions and get various anomaly textures.

Fig.5 and Table 7 illustrate the anomaly generations with different manipulations for four types of Hazelnut defects.

深度。我们同样使用预训练模型DepthAnythingV2 [26] 来估计深度图。这是一个用于单目深度估计的强大基础模型，能够为包含精细细节的复杂场景生成鲁棒的预测。与先前方法相比，其最关键改进在于将所有标注的真实图像替换为精确的合成图像，从而克服了使用真实标注图像时存在的噪声干扰以及忽略深度图中某些细节的问题。遵循这一思路，我们在RGB层级异常生成中，对所有数据集均采用由DepthAnythingV2 [26] 提取的伪深度图。然而，如图4a2和图4b2所示，伪深度图的准确性远低于真实深度图。因此，在针对MVTec3D [2] 数据集的深度层级异常生成中，我们使用了真实深度图。

3.4. 无监督训练

在不使用标注的情况下，我们构建了具有不同增强效果的图像、深度和边缘条件集用于训练。通常，输入条件会应用不同的增强处理，而生成的内容则与目标条件共享相同的增强方式。增强的一致性通过图4中相同的颜色标示。如图4a所示，生成的AHD和AHE异常会分别与目标深度及边缘条件采用相同的增强处理。

增强。遵循[23, 38, 39]的方法，我们的增强主要包括局部薄板样条（TPS）[8]扭曲、缩放-平移-填充以及上下/左右翻转。局部TPS会随机在水平和垂直方向上移动局部区域内的3x3控制点。与全局选择控制点相比，局部TPS带来的扭曲空间范围更小。缩放-平移-填充增强主要是为了应对纹理类别中剧烈的边缘操作，例如对Hazelnut [1]类别大部分边缘的编辑。翻转增强则带来了方向无关的真实生成优势。通过使用这三种增强方法，我们得到了一个对剧烈操作具有鲁棒性的生成器，如图5所示。

损失函数。遵循[23, 38, 39]的方法，我们使用VGG感知损失[12] L_{vgg}来衡量生成图像$G(C_x, A_y)$的保真度，并使用条件GAN损失L_D来衡量生成图像与真实图像之间的区分度。异常生成损失L_A的定义如下。

$$L_G(C_x, A_y; G) = L_{vgg}(G(C_x), A_y) \tag{2}$$

$$L_D(C_x, A_y; D, G) = log(D(C_x, A_y)) + log(1 - D(C_x, G(C_x))) \tag{3}$$

$$L_A = L_G(C_x, A_y; G) + L_D(C_x, A_y; D, G) \tag{4}$$

其中，$C_{x=\{d,r,e\}}$表示输入深度C_d、颜色C_r和边缘C_e条件，$A_{y=\{d,e\}}$代表AHD和AHE异常生成的目标图像A_d和A_e，G为生成器，D为判别器。

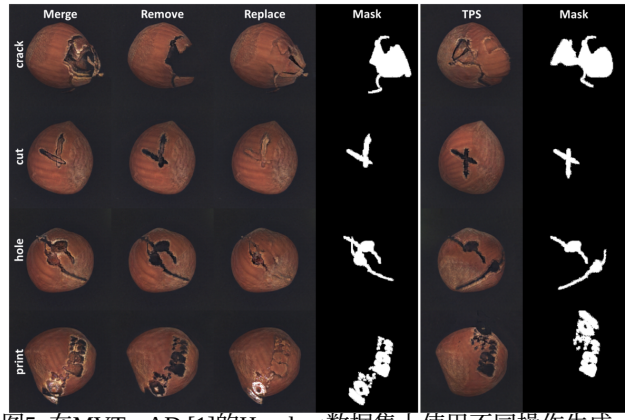

图5. 在MVTecAD [1]的Hazelnut数据集上使用不同操作生成异常样本的示例。

3.5. 异常生成

如图1所示，AnomalyHybrid可为不同应用生成全局与局部异常。在全局异常生成方面，该方法直接融合参考图像与目标图像的外观及结构特征，例如蝴蝶混合异常。图7表明，AHD与AHE解码器分别通过聚焦深度控制和边缘控制的蝴蝶混合异常，实现了多样化的生成效果。对于局部异常生成，AnomalyHybrid通过对深度图和边缘图进行操控，使异常形态更具多样性与真实性。图2a总结了在深度图与边缘图输入生成器前施加局部操控以生成异常的工作流程。

根据[38, 39]的研究，基本操作包括直接删除、替换、合并条件，并在局部区域的条件上应用TPS。

- 掩码。局部异常区域是参考图像和目标图像增强异常区域的组合。增强主要包括调整大小、翻转和裁剪操作。
- 合并。参考图像和目标图像异常区域中的深度和边缘值被合并在一起，以伪造新的异常纹理。
- 移除。为了伪造典型缺陷，如裂纹、切口和孔洞，异常区域中的深度和边缘值被设置为背景值。
- 替换。为了降低操作的强度级别，目标图像异常区域的深度和边缘值被参考图像中的值所替换。
- TPS。为了增加异常多样性，我们在异常区域应用TPS，获得多种异常纹理。

图5和表7展示了针对四种榛子缺陷类型进行不同操作时的异常生成情况。

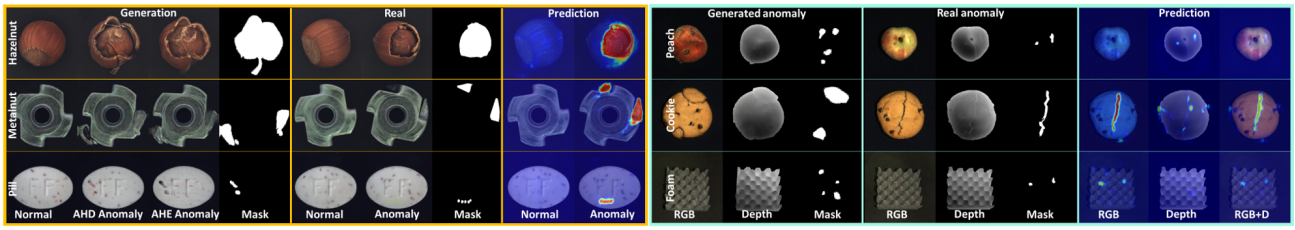

Figure 6. **Examples of anomaly generation and detection by AnomalyHybrid on (Left)MVTecAD [1] and (Right)MVTec3D [2].**

Table 1. Comparison of anomaly generation and classification performance using same ResNet34 on MVTecAD [1]. (AnoDiffusion is excluded for ranking, denoted as gray, since it trains classifiers with selected generation images.)

Category (NO.defects)	DiffAug[34]	CropPaste[16]	SDGAN[27]	DGAN[31]	DFMGAN[9]	**AnomalyHybrid**	AnoDiffusion[11]
	GAN-based						Diffusion-based
	IS↑ (Inception Score) / **LPIPS**↑ (Learned Perceptual Image Patch Similarity) / Classification **Accuracy**↑						
bottle(3)	1.59/0.03/48.8	1.43/0.04/52.7	1.57/0.06/48.8	1.39/0.07/53.5	1.62/0.12/56.6	**2.01/0.23/62.5**	1.58/0.19/90.7
cable(8)	1.72/0.07/21.4	1.74/0.25/32.8	1.89/0.19/21.9	1.70/0.22/21.4	1.96/0.25/**45.3**	**2.75/0.42**/41.1	2.13/0.41/67.2
capsule(5)	1.34/0.03/34.7	1.23/0.05/32.9	1.49/0.03/30.2	1.59/0.04/32.0	1.59/0.11/37.2	**2.33/0.29/40.0**	1.59/0.21/66.7
carpet(5)	1.19/0.06/35.5	1.17/0.11/28.0	1.18/0.11/21.5	1.24/0.12/29.0	1.23/0.13/**47.3**	**1.43/0.29**/33.3	1.16/0.24/58.1
grid(5)	1.96/0.06/28.3	2.00/0.12/28.3	1.95/0.10/30.8	**2.01**/0.12/27.5	1.97/0.13/**40.8**	1.92/**0.28**/40.0	2.04/0.44/42.5
hazelnut(4)	1.67/0.05/65.3	1.74/0.21/59.0	1.85/0.16/43.8	1.87/0.19/61.1	1.93/0.24/**81.9**	**2.16/0.33**/77.6	2.13/0.31/85.4
leather(5)	2.07/0.06/40.7	1.47/0.14/34.4	2.04/0.12/38.1	2.12/0.14/42.3	2.06/0.17/49.7	**2.56/0.38/62.1**	1.94/0.41/61.9
metal nut(4)	1.58/0.29/58.9	1.56/0.15/60.0	1.45/0.28/44.3	1.47/0.30/56.8	1.49/**0.32**/64.6	**1.88**/0.27/**68.3**	1.96/0.30/59.4
pill(7)	1.53/0.05/29.9	1.49/0.11/26.7	1.61/0.07/20.5	1.61/0.10/28.5	1.63/0.16/29.5	**2.06/0.26/43.8**	1.61/0.26/59.4
screw(5)	1.10/0.10/25.1	1.12/0.16/28.8	1.17/0.10/26.8	**1.19**/0.12/28.8	1.12/0.14/**37.5**	1.14/**0.24**/34.2	1.28/0.30/48.2
tile(5)	1.93/0.09/59.7	1.83/0.20/68.4	2.53/0.21/42.7	2.35/0.22/26.9	2.39/0.22/74.9	**2.89/0.49/88.5**	2.54/0.55/84.2
toothbrush(1)	1.33/0.06/-	1.30/0.08/-	1.78/0.03/-	1.85/0.03/-	1.82/0.18/-	**1.95/0.25**/-	1.68/0.21/-
transistor(4)	1.34/0.05/38.1	1.39/0.15/41.7	1.76/0.13/32.1	1.47/0.13/35.7	1.64/0.25/**52.4**	**2.13/0.41**/45.8	1.57/0.34/60.7
wood(5)	2.05/0.30/41.3	1.95/0.23/47.6	2.12/0.25/31.0	**2.19**/0.29/24.6	2.12/0.35/49.2	2.09/**0.38/64.9**	2.33/0.37/71.4
zipper(7)	1.30/0.05/22.8	1.23/0.11/26.4	1.25/0.10/21.5	1.25/0.10/18.7	1.29/**0.27**/27.6	**1.65/0.25/34.7**	1.39/0.25/69.5
Mean	1.58/0.09/39.3	1.51/0.14/40.6	1.71/0.13/32.4	1.69/0.15/34.8	1.72/0.20/49.6	**2.06/0.32/52.6**	1.80/0.32/66.1

4. Experiments

4.1. Datasets

We conduct extensive experiments on 2 industrial datasets, MVTecAD [1] and MVTec3D [2], and 1 biological dataset, HeliconiusButterfly [4]. **MVTecAD** [1] contains 3,629 high-resolution color images from 15 different categories of industrial objects and textures in trainset. Its testset contains 70 types of structural anomalies in different categories, including broken, crack, contamination and misplacement. **MVTec3D** [2] contains over 4,147 high-resolution scans of 10 categories acquired by an industrial 3D sensor that acquires RGB data. There are 894 anomalous containing various defects that are visible in either RGB or 3D data. **HeliconiusButterfly** [4] contains high-resolution (5184x3456) images of non-hybrid (normal) and hybrid (anomaly) subspecies of Heliconius butterfly. The trainset contains 2,084 images of 14 non-hybrid and 1 hybrid subspecies. The testset has 2,350 images of 16 non-hybrid and 7 hybrid subspecies. According to the number of hybrid subspecies, it split them into the Signal Hybrid and Non-Signal Hybrid. The unseen hybrid in testset is called as Mimic Hybrid. The visual appearances (e.g., color patterns on the wings) of these subspecies can be drastically different. More details are shown in the Supplementary.

4.2. Metrics

Anomaly generation Following [11, 38], we utilize Inception Score(**IS**) and cluster-based Learned Perceptual Image Patch Similarity(**LPIPS**) to evaluate the realistic and diversity of our generation. IS measures the realistic and diversity of generated images but is independent of the given real anomaly data. A higher IS indicates better realistic and greater diversity. LPIPS computes the similarity between the features of two image patches extracted from a pre-trained network. The higher LPIPS the greater variety generated images are.

Anomaly detection For butterfly hybrid detection, we use the harmonic mean of the Signal Hybrid Recall, Non-Signal Hybrid Recall, and Mimic Hybrid Recall as the final score. The true positive rate (TPR) at the true negative rate (TNR) is set as 0.95. That is recall of hybrid cases, with a score threshold set to recognizing non-hybrid cases with 0.95 accuracy. For industrial anomaly inspection, we utilize AUROC, Average Precision (AP), and the F1-max score to evaluate the accuracy of image-level anomaly detection and pixel-level anomaly localization.

4.3. Main results

Anomaly generation. Following previous works[9, 11], we randomly choose 1/3 of the dataset images from each defect

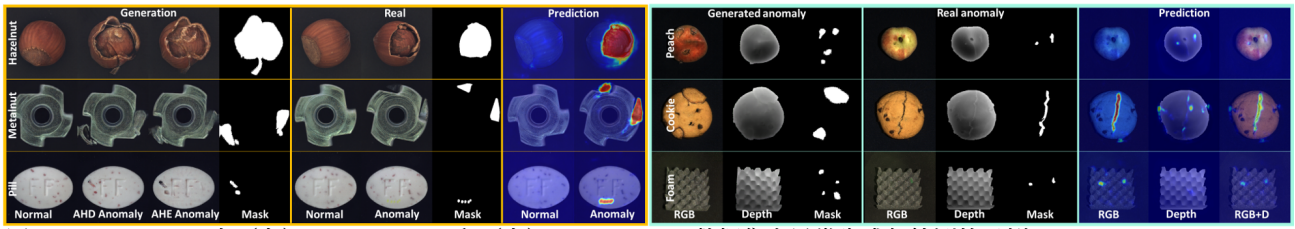

图6. AnomalyHybrid在（左）MVTecAD [1] 和（右）MVTec3D [2] 数据集上异常生成与检测的示例。

表1. 在MVTecAD [1]上使用相同ResNet34的异常生成与分类性能对比。（AnoDiffusion因使用筛选生成图像训练分类器，在排名中予以排除，以灰色标示。）

Category (NO.defects)	DiffAug[34]	CropPaste[16]	SDGAN[27]	DGAN[31]	DFMGAN[9]	**AnomalyHybrid**	AnoDiffusion[11]
	GAN-based						Diffusion-based
	IS↑ (Inception Score) / **LPIPS**↑ (Learned Perceptual Image Patch Similarity) / Classification **Accuracy**↑						
bottle(3)	1.59/0.03/48.8	1.43/0.04/52.7	1.57/0.06/48.8	1.39/0.07/53.5	1.62/0.12/56.6	**2.01/0.23/62.5**	1.58/0.19/90.7
cable(8)	1.72/0.07/21.4	1.74/0.25/32.8	1.89/0.19/21.9	1.70/0.22/21.4	1.96/0.25/**45.3**	**2.75/0.42**/41.1	2.13/0.41/67.2
capsule(5)	1.34/0.03/34.7	1.23/0.05/32.9	1.49/0.03/30.2	1.59/0.04/32.0	1.59/0.11/37.2	**2.33/0.29/40.0**	1.59/0.21/66.7
carpet(5)	1.19/0.06/35.5	1.17/0.11/28.0	1.18/0.11/21.5	1.24/0.12/29.0	1.23/0.13/**47.3**	**1.43/0.29**/33.3	1.16/0.24/58.1
grid(5)	1.96/0.06/28.3	2.00/0.12/28.3	1.95/0.10/30.8	**2.01**/0.12/27.5	1.97/0.13/**40.8**	1.92/**0.28**/40.0	2.04/0.44/42.5
hazelnut(4)	1.67/0.05/65.3	1.74/0.21/59.0	1.85/0.16/43.8	1.87/0.19/61.1	1.93/0.24/**81.9**	**2.16/0.33**/77.6	2.13/0.31/85.4
leather(5)	2.07/0.06/40.7	1.47/0.14/34.4	2.04/0.12/38.1	2.12/0.14/42.3	2.06/0.17/49.7	**2.56/0.38/62.1**	1.94/0.41/61.9
metal nut(4)	1.58/0.29/58.9	1.56/0.15/60.0	1.45/0.28/44.3	1.47/0.30/56.8	1.49/**0.32**/64.6	**1.88**/0.27/**68.3**	1.96/0.30/59.4
pill(7)	1.53/0.05/29.9	1.49/0.11/26.7	1.61/0.07/20.5	1.61/0.10/28.5	1.63/0.16/29.5	**2.06/0.26/43.8**	1.61/0.26/59.4
screw(5)	1.10/0.10/25.1	1.12/0.16/28.8	1.17/0.10/26.8	**1.19**/0.12/28.8	1.12/0.14/**37.5**	1.14/**0.24**/34.2	1.28/0.30/48.2
tile(5)	1.93/0.09/59.7	1.83/0.20/68.4	2.53/0.21/42.7	2.35/0.22/26.9	2.39/0.22/74.9	**2.89/0.49/88.5**	2.54/0.55/84.2
toothbrush(1)	1.33/0.06/-	1.30/0.08/-	1.78/0.03/-	1.85/0.03/-	1.82/0.18/-	**1.95/0.25**/-	1.68/0.21/-
transistor(4)	1.34/0.05/38.1	1.39/0.15/41.7	1.76/0.13/32.1	1.47/0.13/35.7	1.64/0.25/**52.4**	**2.13/0.41**/45.8	1.57/0.34/60.7
wood(5)	2.05/0.30/41.3	1.95/0.23/47.6	2.12/0.25/31.0	**2.19**/0.29/24.6	2.12/0.35/49.2	2.09/**0.38/64.9**	2.33/0.37/71.4
zipper(7)	1.30/0.05/22.8	1.23/0.11/26.4	1.25/0.10/21.5	1.25/0.10/18.7	1.29/**0.27**/27.6	**1.65/0.25/34.7**	1.39/0.25/69.5
Mean	1.58/0.09/39.3	1.51/0.14/40.6	1.71/0.13/32.4	1.69/0.15/34.8	1.72/0.20/49.6	**2.06/0.32/52.6**	1.80/0.32/66.1

4. 实验

4.1. 数据集

我们在2个工业数据集MVTecAD [1]和MVTec3D [2]以及1个生物数据集HeliconiusButterfly [4]上进行了大量实验。MVTecAD [1]的训练集包含来自15个不同类别工业物体和纹理的3,629张高分辨率彩色图像。其测试集涵盖不同类别中70种结构异常，包括破损、裂纹、污染和错位。MVTec3D [2]通过工业3D传感器采集RGB数据，包含10个类别超过4,147次高分辨率扫描。其中894个异常样本包含在RGB或3D数据中可见的各类缺陷。HeliconiusButterfly [4]包含海利科尼乌斯蝴蝶非杂交（正常）与杂交（异常）亚种的高分辨率（5184x3456）图像。训练集包含14个非杂交亚种和1个杂交亚种的2,084张图像。测试集包含16个非杂交亚种和7个杂交亚种的2,350张图像。根据杂交亚种数量，将其分为信号杂交与非信号杂交。测试集中未见的杂交亚种被称为拟态杂交。这些亚种的视觉特征（如翅膀上的颜色图案）可能存在显著差异。更多细节见补充材料。

4.2. 评估指标

异常生成 遵循[11, 38]的方法，我们采用初始分数（IS）和基于聚类的学习感知图像块相似度（LPIPS）来评估生成结果的真实性与多样性。IS衡量生成图像的真实性和多样性，但独立于给定的真实异常数据。IS值越高，表明生成图像越真实且多样性越丰富。LPIPS通过预训练网络提取两个图像块的特征并计算其相似度。LPIPS值越高，表示生成图像的多样性越强。

对于蝴蝶杂交检测，我们使用信号杂交召回率、非信号杂交召回率和模拟杂交召回率的调和平均数作为最终得分。真负率（TNR）设定为0.95时的真正率（TPR）即为杂交案例的召回率，其得分阈值设定为以0.95的准确率识别非杂交案例。对于工业异常检测，我们利用AUROC、平均精度（AP）和F1-max分数来评估图像级异常检测和像素级异常定位的准确性。

4.3. 主要结果

异常生成。遵循先前的工作[9, 11]，我们从每个缺陷类别的数据集中随机选择1/3的图像。

Table 2. Comparison of anomaly localization and detection performance using same UNet on MVTecAD [1]. AnoHybrid trains UNet with images generated by both depth and edge decoders. AnoHybrid+ indicates additionally using generated normal images for training.

Category	CropPaste[16]	DFMGAN[9]	**AnoHybrid**	**AnoHybrid+**	CropPaste[16]	DFMGAN[9]	**AnoHybrid**	**AnoHybrid+**
	Pixel-level AUC/AP/F_1-max				**Image-level** AUC/AP/F_1-max			
bottle	94.5/67.4/63.5	**98.9/90.2/83.9**	98.3/77.2/72.5	98.5/78.7/74.6	85.4/95.1/90.9	99.3/99.8/97.7	99.2/99.8/98.7	**99.6/99.9/98.7**
cable	96.0/75.3/69.3	**97.2/81.0/75.4**	94.1/76.0/73.4	93.0/75.5/72.3	93.3/96.1/91.6	95.9/97.8/93.8	96.1/97.7/91.9	**97.3/98.5/94.4**
capsule	95.3/**49.2**/51.1	79.2/26.0/35.0	**98.4/51.7/54.6**	97.3/44.5/49.1	77.1/94.1/90.4	92.8/98.5/**94.5**	**94.9/98.8/95.2**	89.0/97.6/93.5
carpet	83.7/36.6/39.7	90.6/33.4/38.1	**98.6/82.8**/75.6	98.5/82.3/**76.2**	57.7/84.3/87.3	67.9/87.9/87.3	96.3/98.9/94.0	**97.6/99.3/96.4**
grid	84.7/13.1/22.4	75.2/14.3/20.5	**98.8**/58.6/59.2	98.7/**59.9/59.4**	83.0/94.1/87.6	73.0/90.4/85.4	**100/100/100**	**100/100/100**
hazelnut	88.5/38.0/42.8	**99.7/95.2/89.5**	99.6/89.4/82.6	99.5/84.8/79.3	68.8/85.0/78.0	**99.9/100/99.0**	96.7/98.3/92.6	93.8/97.1/91.8
leather	97.5/**76.0/70.8**	98.5/68.7/66.7	**99.6**/72.7/67.1	99.4/67.9/64.9	91.9/97.5/90.9	**99.9/100/99.2**	98.4/99.5/97.4	99.3/99.7/98.3
metal nut	96.3/84.2/74.0	**99.3/98.1/94.5**	98.8/93.5/87.0	98.6/93.0/87.5	92.2/98.1/93.3	99.3/99.8/**99.2**	**99.8/99.9/99.2**	**99.8/99.9/99.2**
pill	81.5/17.8/24.3	81.2/67.8/72.6	99.3/94.9/**88.4**	**99.5/94.9**/88.0	51.7/87.1/91.4	68.7/91.7/91.4	**99.1/99.8/98.9**	98.3/**99.7**/97.2
screw	**93.4/31.2/36.0**	58.8/2.2/5.3	77.0/7.8/6.4	74.3/3.9/11.1	**59.3/81.9/86.0**	22.3/64.7/85.3	44.6/72.6/84.9	50.4/75.6/85.2
tile	94.0/79.3/74.5	**99.5/97.1/91.6**	99.3/94.6/87.4	99.3/94.5/86.4	73.8/91.1/83.8	**100/100/100**	99.5/99.8/99.0	99.6/99.9/98.1
toothbrush	89.3/30.9/34.6	96.4/**75.9/72.6**	**98.7**/65.2/67.8	98.2/64.5/67.0	81.2/91.0/88.9	**100/100/100**	**100/100/100**	**100/100/100**
transistor	85.9/52.5/52.1	**96.2/81.2/77.0**	98.1/80.8/74.2	96.8/73.8/70.2	85.9/81.8/80.0	90.8/92.5/**88.9**	92.9/90.4/86.3	**95.2/93.4**/86.4
wood	84.0/45.7/48.0	95.3/70.7/65.8	**95.8**/70.7/64.8	**96.4/73.1/66.6**	49.5/81.2/86.6	**98.4/99.4/98.8**	96.6/98.7/98.7	96.6/98.8/97.3
zipper	94.8/47.6/51.4	92.9/**65.6**/64.9	**99.1/82.3/74.9**	98.9/81.7/73.7	59.4/82.8/88.9	99.7/99.9/99.4	99.9/**100**/99.3	100/**100/100**
Mean	90.4/48.4/49.4	90.0/62.7/62.1	**96.9/72.9/69.1**	96.5/71.5/68.4	74.0/89.4/87.7	87.2/94.8/94.7	94.3/96.9/95.7	**94.4/97.3/95.8**

Table 3. Comparison of anomaly generation on MVTec3D [2].

Category	DFMGAN[9]	AnoDiffusion[11]	**AnomalyHybrid**	
	RGB-level			**Depth-level**
	IS ↑/LPIPS ↑			
bagel	**1.07/0.26**	1.02/0.22	1.05/0.23	1.52/0.14
cableG	1.59/**0.25**	1.79/0.19	**2.42**/0.21	2.63/0.21
carrot	1.94 /**0.29**	1.66/0.17	**2.31**/0.21	2.02/0.11
cookie	1.80/**0.31**	1.77/0.29	**1.95**/0.28	1.45/0.16
dowel	**1.96/0.37**	1.60/0.20	1.89/0.22	1.78/0.15
foam	1.50/0.17	**1.77/0.30**	1.73/0.28	1.36/0.19
peach	**2.11/0.34**	1.91/0.23	1.97/0.25	1.71/0.13
potato	**3.05/0.35**	1.92/0.17	2.31/0.18	1.63/0.09
rope	**1.46/0.29**	1.28/0.25	1.42/**0.29**	1.61/0.12
tire	**1.53/0.25**	1.35/0.20	1.47/0.22	1.44/0.11
Mean	1.80 /**0.29**	1.61/0.22	**1.85**/0.24	1.72/0.14

Table 4. Comparison of anomaly localization and detection performance on MVTec3D [2]. RGB: only RGB images, D: only depth images, +: mean of RGB and depth predictions.

	Method	Pixel-level			Image-level		
		AUC	AP	F_1-max	AUC	AP	F_1-max
RGB	DFMGAN[9]	74.4	14.7	20.7	63.7	82.8	84.9
RGB	AnoDiffusion[11]	91.2	**22.8**	**29.6**	71.7	87.1	86.6
RGB	**AnomalyHybrid**	**96.9**	16.0	23.2	**83.7**	**94.8**	**91.4**
D	**AnomalyHybrid**	94.2	12.8	19.1	82.7	94.5	92.0
+	**AnomalyHybrid**	98.4	19.5	26.4	90.1	97.1	94.0

category as the base sets, and the other 2/3 from each category are combined as the test set. With the base sets, both AHD and AHE decoders generate 500 anomaly images for each category. These generated images are used for evaluating generation quality and training models for downstream tasks. We also construct lists of 100 sampled anomaly images from the testing dataset. To calculate LPIPS, we partition the generated 1,000 images into 100 groups by finding the lowest LPIPS. We compute the mean pairwise LPIPS within each group. The average LPIPS of all groups will be the final score. Table 1 and Table 3 show the comparisons of RGB-level anomaly generation on MVTecAD [1] and MVTec3D [2]. Table 3 also demonstrates the depth-level anomaly generation performance of AnomalyHybrid. Comparing to both GAN-based and Diffusion-based SOTA, AnomalyHybrid generates RGB-level anomalies with both the highest realistic and diversity on all evaluated datasets. On MVTec3D [2], the generated depth-level anomalies achieve 1.72 IS score that is higher than the AnoDiffusion's RGB-level anomaly generation performance. As visualized in Figure 6, AnomalyHybrid not only generates different level of anomalies but also diverse normal samples.

Anomaly detection. Table 2 and Table 4 illustrate the benefit of our generated images for downstream anomaly classification, detection and segmentation tasks on MVTecAD [1] and MVTec3D [2]. On MVTecAD [1] dataset, we also evaluate the contribution of using generated normal samples for training anomaly detectors. Overall, the classifier ResNet34 trained on images generated from both AHD and AHE decoders of AnomalyHybrid achieves the highest accuracy 52.6 comparing to the GAN-based SOTA. With the same anomaly detector UNet, our generated images bring the highest performance both in image-level and pixel-level anomaly detection. By using normal images generated by AnomalyHybrid, the detector gains 0.4 percentage improvement in pixel-level AP. On MVTec3D [2] dataset, we conduct experiments of using RGB-level anomalies, depth-level anomalies and both of them to train UNet. Under these three settings, AnomalyHybrid all achieves better performance than GAN-based SOTA. By using both RGB-level and depth-level anomalies, AnomalyHybrid gains around 4 percentages improvement in pixel-level anomaly localization and 6 percentages improvement in image-level anomaly detection.

表2. 使用相同UNet在MVTecAD [1]上的异常定位与检测性能对比。AnoHybrid采用{v*}训练UNet 第由深度和边缘解码器生成的图像。AnoHybrid+表示额外使用生成的法线图像进行训练。

Category	CropPaste[16]	DFMGAN[9]	**AnoHybrid**	**AnoHybrid+**	CropPaste[16]	DFMGAN[9]	**AnoHybrid**	**AnoHybrid+**
	Pixel-level AUC/AP/F_1-max				**Image-level** AUC/AP/F_1-max			
bottle	94.5/67.4/63.5	**98.9/90.2/83.9**	98.3/77.2/72.5	98.5/78.7/74.6	85.4/95.1/90.9	99.3/99.8/97.7	99.2/99.8/98.7	**99.6/99.9/98.7**
cable	96.0/75.3/69.3	**97.2/81.0/75.4**	94.1/76.0/73.4	93.0/75.5/72.3	93.3/96.1/91.6	95.9/97.8/93.8	96.1/97.7/91.9	**97.3/98.5/94.4**
capsule	95.3/**49.2**/51.1	79.2/26.0/35.0	**98.4/51.7/54.6**	97.3/44.5/49.1	77.1/94.1/90.4	92.8/98.5/**94.5**	**94.9/98.8/95.2**	89.0/97.6/93.5
carpet	83.7/36.6/39.7	90.6/33.4/38.1	**98.6/82.8**/75.6	98.5/82.3/**76.2**	57.7/84.3/87.3	67.9/87.9/87.3	96.3/98.9/94.0	**97.6/99.3/96.4**
grid	84.7/13.1/22.4	75.2/14.3/20.5	**98.8**/58.6/59.2	98.7/**59.9/59.4**	83.0/94.1/87.6	73.0/90.4/85.4	**100/100/100**	**100/100/100**
hazelnut	88.5/38.0/42.8	**99.7/95.2/89.5**	99.6/89.4/82.6	99.5/84.8/79.3	68.8/85.0/78.0	**99.9/100/99.0**	96.7/98.3/92.6	93.8/97.1/91.8
leather	97.5/**76.0/70.8**	98.5/68.7/66.7	**99.6**/72.7/67.1	99.4/67.9/64.9	91.9/97.5/90.9	**99.9/100/99.2**	98.4/99.5/97.4	99.3/99.7/98.3
metal nut	96.3/84.2/74.0	**99.3/98.1/94.5**	98.8/93.5/87.0	98.6/93.0/87.5	92.2/98.1/93.3	99.3/99.8/**99.2**	**99.8/99.9/99.2**	**99.8/99.9/99.2**
pill	81.5/17.8/24.3	81.2/67.8/72.6	99.3/94.9/**88.4**	**99.5/94.9**/88.0	51.7/87.1/91.4	68.7/91.7/91.4	**99.1/99.8/98.9**	98.3/**99.7**/97.2
screw	**93.4/31.2/36.0**	58.8/2.2/5.3	77.0/7.8/6.4	74.3/3.9/11.1	**59.3/81.9/86.0**	22.3/64.7/85.3	44.6/72.6/84.9	50.4/75.6/85.2
tile	94.0/79.3/74.5	**99.5/97.1/91.6**	99.3/94.6/87.4	99.3/94.5/86.4	73.8/91.1/83.8	**100/100/100**	99.5/99.8/99.0	99.6/99.9/98.1
toothbrush	89.3/30.9/34.6	96.4/**75.9/72.6**	**98.7**/65.2/67.8	98.2/64.5/67.0	81.2/91.0/88.9	**100/100/100**	**100/100/100**	**100/100/100**
transistor	85.9/52.5/52.1	**96.2/81.2/77.0**	98.1/80.8/74.2	96.8/73.8/70.2	85.9/81.8/80.0	90.8/92.5/**88.9**	92.9/90.4/86.3	**95.2/93.4**/86.4
wood	84.0/45.7/48.0	95.3/70.7/65.8	**95.8**/70.7/64.8	**96.4/73.1/66.6**	49.5/81.2/86.6	**98.4/99.4/98.8**	96.6/98.7/98.7	96.6/98.8/97.3
zipper	94.8/47.6/51.4	92.9/**65.6**/64.9	**99.1/82.3/74.9**	98.9/81.7/73.7	59.4/82.8/88.9	99.7/99.9/99.4	99.9/**100**/99.3	100/**100/100**
Mean	90.4/48.4/49.4	90.0/62.7/62.1	**96.9/72.9/69.1**	96.5/71.5/68.4	74.0/89.4/87.7	87.2/94.8/94.7	94.3/96.9/95.7	**94.4/97.3/95.8**

表3. MVTec3D [2] 异常生成方法对比。

Category	DFMGAN[9]	AnoDiffusion[11]	**AnomalyHybrid**	
	RGB-level			**Depth-level**
	IS ↑/LPIPS ↑			
bagel	**1.07/0.26**	1.02/0.22	1.05/0.23	1.52/0.14
cableG	1.59/**0.25**	1.79/0.19	**2.42**/0.21	2.63/0.21
carrot	1.94 /**0.29**	1.66/0.17	**2.31**/0.21	2.02/0.11
cookie	1.80/**0.31**	1.77/0.29	**1.95**/0.28	1.45/0.16
dowel	**1.96/0.37**	1.60/0.20	1.89/0.22	1.78/0.15
foam	1.50/0.17	**1.77/0.30**	1.73/0.28	1.36/0.19
peach	**2.11/0.34**	1.91/0.23	1.97/0.25	1.71/0.13
potato	**3.05/0.35**	1.92/0.17	2.31/0.18	1.63/0.09
rope	**1.46/0.29**	1.28/0.25	1.42/**0.29**	1.61/0.12
tire	**1.53/0.25**	1.35/0.20	1.47/0.22	1.44/0.11
Mean	1.80 /**0.29**	1.61/0.22	**1.85**/0.24	1.72/0.14

表4. MVTec3D [2] 数据集上的异常定位与检测性能对比。RGB：仅使用RGB图像，D：仅使用深度图像，{v*}：RGB与深度预测结果的均值。

	Method	**Pixel-level**			**Image-level**		
		AUC	AP	F_1-max	AUC	AP	F_1-max
RGB	DFMGAN[9]	74.4	14.7	20.7	63.7	82.8	84.9
RGB	AnoDiffusion[11]	91.2	**22.8**	**29.6**	71.7	87.1	86.6
RGB	**AnomalyHybrid**	**96.9**	16.0	23.2	**83.7**	**94.8**	**91.4**
D	**AnomalyHybrid**	94.2	12.8	19.1	82.7	94.5	92.0
+	**AnomalyHybrid**	98.4	19.5	26.4	90.1	97.1	94.0

以类别作为基础集，其余2/3的每个类别数据合并为测试集。基于基础集，AHD和AHE解码器为每个类别生成500张异常图像。这些生成的图像用于评估生成质量及训练下游任务模型。我们还从测试数据集中抽取100张异常图像构建样本列表。为计算LPIPS，我们将生成的1000张图像通过寻找最低LPIPS值划分为100组，计算每组内的平均成对LPIPS值，所有组的平均LPIPS即为最终得分。表1和表3展示了对比在MVTecAD [1] 和 MVTec3D [2] 上进行的RGB级异常生成对比。表3同时展示了AnomalyHybrid的深度级异常生成性能。与基于GAN和基于扩散的SOTA方法相比，AnomalyHybrid在所有评估数据集上生成的RGB级异常均实现了最高的真实性与多样性。在MVTec3D [2]上，其生成的深度级异常取得了1.72的IS分数，甚至高于AnoDiffusion的RGB级异常生成性能。如图6所示，AnomalyHybrid不仅能生成不同级别的异常样本，还能生成多样化的正常样本。

异常检测。表2和表4展示了我们生成的图像在MVTecAD [1]和MVTec3D [2]数据集上对下游异常分类、检测与分割任务的益处。在MVTecAD [1]数据集上，我们还评估了使用生成的正常样本训练异常检测器的贡献。总体而言，使用AnomalyHybrid的AHD和AHE解码器生成的图像训练的ResNet34分类器，与基于GAN的SOTA方法相比，取得了最高的52.6%准确率。使用相同的异常检测器UNet，我们生成的图像在图像级和像素级异常检测中均带来了最佳性能。通过使用AnomalyHybrid生成的正常图像，检测器在像素级AP上获得了0.4个百分点的提升。在MVTec3D [2]数据集上，我们进行了使用RGB级异常、深度级异常以及两者结合来训练UNet的实验。在这三种设置下，AnomalyHybrid均取得了优于基于GAN的SOTA方法的性能。通过同时使用RGB级和深度级异常，AnomalyHybrid在像素级异常定位上提升了约4个百分点，在图像级异常检测上提升了6个百分点。

Figure 7. **Examples of anomaly generation by AnomalyHybrid on HeliconiusButterfly [4].**

4.4. Ablation

Decoders. As shown in Table 5, on MVTecAD [1], AHD decoder achieves higher IS/LPIPS scores that indicate more diverse anomalies than AHE decoder. However, the classifier(ResNet34) trained on anomalies generated by the AHD decoder achieves 1 percentage lower accuracy than the AHE decoder's. The reason is that AHD decoder focuses on less texture details than AHE decoder does. By using anomalies generated by both AHD and AHE decoders, the classifier gains 5.2 percentages accuracy improvement.

Table 6 illustrates the comparison of AHD and AHE anomalies for classification on HeliconiusButterfly [4] dataset. We use DINOv2 [19] to extract image features and simply use SGD and a 3 linear layers head as the detectors. The trainset contains only SignalHybrid and non-hybrid images. The testset consists of 3-type hybrids, including SignalHybrid, Non-signalHybrid and MimicHybrid. Figure 7 demonstrates 2 types of non-hybrid and 2 types of non-signal hybrid. Since the MimicHybrid is similar to the SignalHybrid, the classifier directly trained on the original trainset achieves the best performance on this type. The baseline anomaly generation method AnomalyFactory [38], having only AHE branch, improves the classification performance more than 10 percentages. Different with AnomalyFactory [38], our network consists of both AHD and AHE branches that generate diverse and authentic hybrid as shown in Figure 7. With the help of anomalies generated by AHE decoder, the classifier achieves the highest harmonic mean recall 0.551 on 3-type hybrids.

Manipulation. We evaluate the effectiveness of different manipulations for anomaly generation and classification on Hazelnut of MVTecAD [1]. Table 7 shows the performance of using different manipulations. As shown in Figure 5, there are four types of defects and three out of them are close to depth values changing. The Remove manipulation always generates hole-like defect and causes ambiguity in other types, such as cut. Therefore, it achieves the lowest performance in diverse generation and defect classification. On the contrary, the Merge and Replace manipulations generate anomalies similar to the original types but with higher diversity. They both achieves the second highest classification accuracy. By randomly applying different manipulations, we gain 0.078 higher IS score for anomaly generation and 22.4 acc improvement for anomaly classification. With TPS manipluation, we increase the variety of anomaly shapes and achieve the overall highest performance.

Table 5. Ablation study of decoders for anomaly generation and classification performance on MVTecAD [1].

Decoder		Generation		Classification
AHD	AHE	IS↑	LPIPS↑	Accuracy↑
-	v	1.88	0.35	48.2
v	-	1.99	**0.36**	47.4
v	v	**2.06**	0.32	**52.6**

Table 6. Ablation study of anomaly generation and classification on HeliconiusButterfly [4]. (∗Without using manipulation.)

Recall@	Trainset	Baseline	AHD	AHE
	Classifier: Linear/SGD/Max(Linear, SGD)			
SignalHybrid	0.847/**0.923**/0.893	0.764/0.792/0.789	0.781/0.778/0.784	0.811/0.860/0.855
Non-SignalH	0.143/0.143/0.036	0.214/0.357/0.357	0.250/0.357/0.357	0.321/**0.429**/**0.429**
MimicHybrid	**0.621**/0.605/**0.621**	0.435/0.431/0.419	0.524/0.484/0.500	0.480/0.509/0.516
HMean	0.306/0.308/0.098	0.363/0.470/0.465	0.417/0.488/0.494	0.467/0.549/**0.551**

Table 7. Ablation study of manipulation for anomaly generation and classification performance on Hazelnut of MVTecAD [1].

Manipulation				Generation		Classification
Merge	Remove	Replace	TPS	IS↑	LPIPS↑	Accuracy↑
v	-	-	-	1.716	0.320	81.6
-	v	-	-	1.668	**0.327**	53.1
-	-	v	-	1.723	0.321	81.6
v	v	v	-	1.746	0.321	75.5
v	v	v	v	**2.163**	0.326	**88.5**

5. Conclusion

In this paper, we propose a domain-agnostic framework, AnomalyHybrid, that generates diverse and authentic anomalies refer to multimodal conditional controls. It significantly optimizes existing GAN-based anomaly generation paradigm of learning multiple dedicated generative models per defect types. Extensive experiments conducted on four datasets demonstrate the superiority of AnomalyHybrid in general anomaly generation and the downstream anomaly detection tasks. With well-designed unsupervised training, AnomalyHybrid is easily generalized to other applications like edge extraction, depth estimation, and Out-of-Distribution detection. We believe that it will contribute more to downstream tasks with wilder extension.

图7. AnomalyHybrid在HeliconiusButterfly [4]上生成异常值的示例。

4.4. 消融实验

解码器。如表5所示，在MVTecAD [1]数据集上，AHD解码器获得了更高的IS/LPIPS分数，这表明其生成的异常比AHE解码器更具多样性。然而，基于AHD解码器生成的异常训练的（ResNet34）分类器，其准确率比使用AHE解码器生成的异常训练的分类器低1个百分点。原因是AHD解码器对纹理细节的关注度低于AHE解码器。通过同时使用AHD和AHE解码器生成的异常进行训练，分类器的准确率提升了5.2个百分点。

表6展示了在HeliconiusButterfly [4]数据集上AHD与AHE异常分类性能的对比。我们使用DINOv2 [19]提取图像特征，并仅采用SGD优化器和3层线性头作为检测器。训练集仅包含信号杂交种与非杂交种图像，测试集则包含信号杂交种、非信号杂交种和拟态杂交种这三类杂交样本。图7展示了两类非杂交种与两类非信号杂交种的示例。由于拟态杂交种与信号杂交种形态相似，直接在原始训练集上训练的分类器对此类样本取得了最佳性能。仅包含AHE分支的基线异常生成方法AnomalyFactory [38]将分类性能提升了超过10个百分点。与AnomalyFactory [38]不同，我们的网络同时包含AHD和AHE分支，能够生成如图7所示多样且逼真的杂交样本。借助AHE解码器生成的异常样本，分类器在三类杂交种上取得了0.551的最高调和平均召回率。

操纵。我们在MVTecAD[1]的Hazelnut数据集上评估了不同操纵方法在异常生成与分类中的有效性。表7展示了使用不同操纵方法的性能表现。如图5所示，存在四种缺陷类型，其中三种与深度值变化密切相关。移除操作总是生成孔洞状缺陷，并在其他类型（如切割缺陷）中产生歧义，因此在多样性生成和缺陷分类中表现最差。相反，合并与替换操作生成的异常更接近原始缺陷类型，同时具有更高的多样性，两者均取得了第二高的分类准确率。通过随机应用不同操纵——

表5. 在MVTecAD [1]上进行的异常生成与分类性能解码器消融研究。

Decoder		Generation		Classification
AHD	AHE	IS↑	LPIPS↑	Accuracy↑
-	v	1.88	0.35	48.2
v	-	1.99	**0.36**	47.4
v	v	**2.06**	0.32	**52.6**

表6. HeliconiusButterfly [4]上异常生成与分类的消融研究。（*未使用操纵。）

Recall@	Trainset	Baseline	AHD	AHE
	Classifier: Linear/SGD/Max(Linear, SGD)			
SignalHybrid	0.847/**0.923**/0.893	0.764/0.792/0.789	0.781/0.778/0.784	0.811/0.860/0.855
Non-SignalH	0.143/0.143/0.036	0.214/0.357/0.357	0.250/0.357/0.357	0.321/**0.429**/**0.429**
MimicHybrid	**0.621**/0.605/**0.621**	0.435/0.431/0.419	0.524/0.484/0.500	0.480/0.509/0.516
HMean	0.306/0.308/0.098	0.363/0.470/0.465	0.417/0.488/0.494	0.467/0.549/**0.551**

表7. 在MVTecAD [1]的Hazelnut数据集上，针对异常生成与分类性能的操作消融研究。

Manipulation				Generation		Classification
Merge	Remove	Replace	TPS	IS↑	LPIPS↑	Accuracy↑
v	-	-	-	1.716	0.320	81.6
-	v	-	-	1.668	**0.327**	53.1
-	-	v	-	1.723	0.321	81.6
v	v	v	-	1.746	0.321	75.5
v	v	v	v	**2.163**	0.326	**88.5**

通过插值操作，我们在异常生成上获得了0.078的IS分数提升，在异常分类上实现了22.4%的准确率改进。借助TPS变换操作，我们增加了异常形状的多样性，并取得了整体最佳性能。

5. 结论

本文提出了一种领域无关的框架AnomalyHybrid，该框架能参照多模态条件控制生成多样且真实的异常{v*}。它显著优化了现有基于GAN的异常生成范式——该范式需要为每种缺陷类型学习多个专用生成模型。在四个数据集上进行的大量实验表明，AnomalyHybrid在通用异常生成及下游异常检测任务中均具有优越性。通过精心设计的无监督训练，AnomalyHybrid可轻松推广至边缘提取、深度估计和分布外检测等其他应用。我们相信该框架将通过更广泛的扩展为下游任务做出更多贡献。

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec AD - A comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019. Computer Vision Foundation / IEEE, 2019. 5, 6, 7, 8

[2] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 5: VISAPP, Online Streaming, February 6-8, 2022*, pages 202–213. SCITEPRESS, 2022. 4, 5, 6, 7

[3] Chenyang Bi, Yueyang Li, and Haichi Luo. Dual-branch reconstruction network for industrial anomaly detection with RGB-D data. *CoRR*, abs/2311.06797, 2023. 2

[4] Elizabeth G Campolongo, Yuan-Tang Chou, Ekaterina Govorkova, Wahid Bhimji, Wei-Lun Chao, Chris Harris, Shih-Chieh Hsu, Hilmar Lapp, Mark S Neubauer, Josephine Namayanja, et al. Building machine learning challenges for anomaly detection in science. *arXiv preprint arXiv:2503.02112*, 2025. 6, 8

[5] Ruitao Chen, Guoyang Xie, Jiaqi Liu, Jinbao Wang, Ziqi Luo, Jinfan Wang, and Feng Zheng. Easynet: An easy network for 3d industrial anomaly detection. In *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*, pages 7038–7046. ACM, 2023. 1, 2

[6] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *CVPR*, pages 3606–3613, 2014. IEEE Computer Society. 2

[7] Songmin Dai, Yifan Wu, Xiaoqiang Li, and Xiangyang Xue. Generating and reweighting dense contrastive patterns for unsupervised anomaly detection. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 1454–1462. AAAI Press, 2024. 3

[8] Gianluca Donato and Serge Belongie. Approximate thin plate spline mappings. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part III 7*, pages 21–31. Springer, 2002. 5

[9] Yuxuan Duan, Yan Hong, Li Niu, and Liqing Zhang. Few-shot defect image generation via defect-aware feature manipulation. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 571–578. AAAI Press, 2023. 1, 2, 3, 6, 7

[10] HDR Imageomics Institute. Heliconius-collection-cambridge-butterfly (revision 691fd81), 2025.

[11] 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 *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 8526–8534. AAAI Press, 2024. 1, 2, 3, 6, 7

[12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*, pages 694–711. Springer, 2016. 5

[13] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 3

[14] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, pages 9664–9674, 2021. Computer Vision Foundation / IEEE. 1, 2

[15] Ming Li, Taojiannan Yang, Huafeng Kuang, Jie Wu, Zhaoning Wang, Xuefeng Xiao, and Chen Chen. Controlnet++: Improving conditional controls with efficient consistency feedback. In *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part VII*, pages 129–147. Springer, 2024. 3

[16] Dongyun Lin, Yanpeng Cao, Wenbin Zhu, and Yiqun Li. Few-shot defect segmentation leveraging abundant defect-free training samples through normal background regularization and crop-and-paste operation. In *2021 IEEE International Conference on Multimedia and Expo, ICME 2021, Shenzhen, China, July 5-9, 2021*, pages 1–6. IEEE, 2021. 6, 7

[17] Yuxuan Lin, Yang Chang, Xuan Tong, Jiawen Yu, Antonio Liotta, Guofan Huang, Wei Song, Deyu Zeng, Zongze Wu, Yan Wang, et al. A survey on rgb, 3d, and multimodal approaches for unsupervised industrial anomaly detection. *arXiv preprint arXiv:2410.21982*, 2024. 3

[18] Shuanlong Niu, Bin Li, Xinggang Wang, and Hui Lin. Defect image sample generation with GAN for improving defect recognition. *IEEE Trans Autom. Sci. Eng.*, 17(3):1611–1622, 2020. 3

[19] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision. *Trans. Mach. Learn. Res.*, 2024, 2024. 8

[20] Can Qin, Shu Zhang, Ning Yu, Yihao Feng, Xinyi Yang, Yingbo Zhou, Huan Wang, Juan Carlos Niebles, Caiming

参考文献

[1] Paul Bergmann, Michael Fauser, David Sattlegger, 和 Carsten Steger。Mvtec AD - 一个用于无监督异常检测的综合性真实世界数据集。收录于 *CVPR*，第9592–9600页，2019年。计算机视觉基金会 / IEEE，2019年。5, 6, 7, 8[2] Paul Bergmann, Xin Jin, David Sattlegger, 和 Carsten Steger。用于无监督3D异常检测与定位的mvtec 3d-ad数据集。收录于 *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 5: VISAPP, Online Streaming, February 6-8, 2022*，第202–213页。SCITEPRESS，2022年。4, 5, 6, 7[3] Chenyang Bi, Yueyang Li, 和 Haichi Luo。基于RGB-D数据的工业异常检测双分支重建网络。*CoRR*, abs/2311.06797, 2023年。2[4] Elizabeth G Campolongo, Yuan-Tang Chou, Ekaterina Govorkova, Wahid Bhimji, Wei-Lun Chao, Chris Harris, Shih-Chieh Hsu, Hilmar Lapp, Mark S Neubauer, Josephine Namayanja, 等。为科学异常检测构建机器学习挑战。*arXiv preprint arXiv:2503.02112*, 2025年。6, 8[5] Ruitao Chen, Guoyang Xie, Jiaqi Liu, Jinbao Wang, Ziqi Luo, Jinfan Wang, 和 Feng Zheng。Easynet：一个用于3D工业异常检测的简易网络。收录于 *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*，第7038–7046页。ACM，2023年。1, 2[6] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, 和 Andrea Vedaldi。描述自然场景中的纹理。收录于 *CVPR*，第3606–3613页，2014年。IEEE计算机学会。2[7] Songmin Dai, Yifan Wu, Xiaoqiang Li, 和 Xiangyang Xue。生成并重加权密集对比模式用于无监督异常检测。收录于 *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*，第1454–1462页。AAAI出版社，2024年。3[8] Gianluca Donato 和 Serge Belongie。近似薄板样条映射。收录于 *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part III 7*，第21–31页。Springer，2002年。5[9] Yuxuan Duan, Yan Hong, Li Niu, 和 Liqing Zhang。通过缺陷感知特征操作的少样本缺陷图像生成。收录于 *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*，第571–578页。AAAI出版社，2023年。1, 2, 3, 6, 7

[10] HDR 影像组学研究所。剑桥蝴蝶-赫利科尼乌斯收藏（修订版 691fd81），2025年。

[11] 滕虎，张江宁，易然，杜雨珍，陈旭，刘亮，王亚标，王成杰。AnomalyDiffusion：基于扩散模型的少样本异常图像生成。于 *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*，第8526–8534页。AAAI Press，2024年。1, 2, 3, 6, 7[12] Justin Johnson, Alexandre Alahi, 李飞飞。用于实时风格迁移和超分辨率的感知损失。于 *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*，第694–711页。Springer，2016年。5[13] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila。分析与改进StyleGAN的图像质量。于 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*，第8110–8119页，2020年。3[14] 李春良，Kihyuk Sohn，Jinsung Yoon，Tomas Pfister。CutPaste：用于异常检测与定位的自监督学习。于 *CVPR*，第9664–9674页，2021年。Computer Vision Foundation / IEEE。1, 2[15] 李明，杨天健，邝华峰，吴杰，王昭宁，肖雪峰，陈晨。ControlNet++：通过高效一致性反馈改进条件控制。于 *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part VII*，第129–147页。Springer，2024年。3[16] 林东云，曹彦鹏，朱文斌，李逸群。通过正常背景正则化与裁剪粘贴操作利用丰富无缺陷训练样本的少样本缺陷分割。于 *2021 IEEE International Conference on Multimedia and Expo, ICME 2021, Shenzhen, China, July 5-9, 2021*，第1–6页。IEEE，2021年。6, 7[17] 林宇轩，常阳，佟璇，于佳雯，Antonio Liotta，黄国凡，宋伟，曾德宇，吴宗泽，王岩等。无监督工业异常检测的RGB、3D及多模态方法综述。*arXiv preprint arXiv:2410.21982*，2024年。3[18] 牛栓龙，李斌，王兴刚，林辉。使用GAN生成缺陷图像样本以改进缺陷识别。*IEEE Trans Autom. Sci. Eng.*，17(3):1611–1622，2020年。3[19] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, Piotr Bojanowski。Dinov2：无需监督学习鲁棒的视觉特征。*Trans. Mach. Learn. Res.*，2024年，2024年。8[20] 覃灿，张舒，于宁，冯一豪，杨心怡，周应波，王欢，Juan Carlos Niebles，Caiming

Xiong, Silvio Savarese, Stefano Ermon, Yun Fu, and Ran Xu. Unicontrol: A unified diffusion model for controllable visual generation in the wild. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. 3

[21] Hannah M. Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXXI*, pages 474–489. Springer, 2022. 1, 2

[22] Zhuo Su, Wenzhe Liu, Zitong Yu, Dewen Hu, Qing Liao, Qi Tian, Matti Pietikäinen, and Li Liu. Pixel difference networks for efficient edge detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5117–5127, 2021. 4

[23] Yael Vinker, Eliahu Horwitz, Nir Zabari, and Yedid Hoshen. Image shape manipulation from a single augmented training sample. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13769–13778, 2021. 5

[24] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 4

[25] Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu, and Fei Qi. Just noticeable difference estimation for images with free-energy principle. *IEEE Trans. Multim.*, 15(7):1705–1710, 2013. 2

[26] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *arXiv preprint arXiv:2406.09414*, 2024. 4, 5

[27] Minghui Yang, Jing Liu, Zhiwei Yang, and Zhaoyang Wu. Slsg: Industrial image anomaly detection by learning better feature embeddings and one-class classification. *arXiv preprint arXiv:2305.00398*, 2023. 6

[28] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Dræm - A discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8310–8319, 2021. IEEE. 1, 2, 3, 4

[29] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Keep dræming: Discriminative 3d anomaly detection through anomaly simulation. *Pattern Recognit. Lett.*, 181:113–119, 2024. 2, 3

[30] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Cheating depth: Enhancing 3d surface anomaly detection via depth simulation. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV*, pages 2153–2161. IEEE, 2024. 1, 2, 3, 4

[31] Gongjie Zhang, Kaiwen Cui, Tzu-Yi Hung, and Shijian Lu. Defect-gan: High-fidelity defect synthesis for automated defect inspection. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Waikoloa, HI, USA, January 3-8, 2021*, pages 2523–2533. IEEE, 2021. 3, 6

[32] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 3813–3824. IEEE, 2023. 3

[33] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. *CoRR*, abs/2403.05897, 2024. 1, 2, 3, 4

[34] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient GAN training. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. 6

[35] Shihao Zhao, Dongdong Chen, Yen-Chun Chen, Jianmin Bao, Shaozhe Hao, Lu Yuan, and Kwan-Yee K. Wong. Uni-controlnet: All-in-one control to text-to-image diffusion models. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. 3

[36] Ying Zhao. Just noticeable learning for unsupervised anomaly localization and detection. In *ICME*, pages In Press, 2022. 1, 2

[37] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3924–3933, 2023. 1, 2

[38] Ying Zhao. Anomalyfactory: Regard anomaly generation as unsupervised anomaly localization. *arXiv preprint arXiv:2408.09533*, 2024. 1, 2, 3, 4, 5, 6, 8

[39] Ying Zhao. Logical: Towards logical anomaly synthesis for unsupervised anomaly localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4022–4031, 2024. 1, 2, 3, 5

[40] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pretraining for anomaly detection and segmentation. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings*, pages 392–408. Springer, 2022. 1, 2

熊、Silvio Savarese、Stefano Ermon、傅云和徐然。Unicontrol：一种用于野外可控视觉生成的统一扩散模型。发表于*Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*，2023年。3[21] Hannah M. Schlüter、Jeremy Tan、Benjamin Hou和Bernhard Kainz。用于自监督异常检测与定位的自然合成异常。发表于*Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXXI*，第474–489页。Springer，2022年。1, 2[22] 苏卓、刘文哲、于子桐、胡德文、廖青、田奇、Matti Pietikïnen和刘莉。用于高效边缘检测的像素差分网络。发表于*Proceedings of the IEEE/CVF international conference on computer vision*，第5117–5127页，2021年。4[23] Yael Vinker、Eliahu Horwitz、Nir Zabari和Yedid Hoshen。从单个增强训练样本进行图像形状操控。发表于*Proceedings of the IEEE/CVF International Conference on Computer Vision*，第13769–13778页，2021年。5[24] 王廷春、刘明宇、朱俊彦、Andrew Tao、Jan Kautz和Bryan Catanzaro。使用条件生成对抗网络进行高分辨率图像合成与语义操控。发表于*Proceedings of the IEEE conference on computer vision and pattern recognition*，第8798–8807页，2018年。4[25] 吴金健、石光明、林维思、刘安民和齐飞。基于自由能原理的图像恰可察觉差估计。*IEEE Trans. Multim.*，15(7):1705–1710，2013年。2[26] 杨立合、康秉一、黄子龙、赵振、徐小刚、冯佳时和赵衡爽。Depth Anything v2。*arXiv preprint arXiv:2406.09414*，2024年。4, 5[27] 杨明辉、刘静、杨志伟和吴朝阳。SLSG：通过学习更好的特征嵌入和单类分类进行工业图像异常检测。*arXiv preprint arXiv:2305.00398*，2023年。6[28] Vitjan Zavrtanik、Matej Kristan和Danijel Skocaj。Dræm - 一种用于表面异常检测的判别性训练重建嵌入。发表于*ICCV*，第8310–8319页，2021年。IEEE。1, 2, 3, 4[29] Vitjan Zavrtanik、Matej Kristan和Danijel Skocaj。持续Dræming：通过异常模拟进行判别性3D异常检测。*Pattern Recognit. Lett.*，181:113–119，2024年。2, 3[30] Vitjan Zavrtanik、Matej Kristan和Danijel Skocaj。欺骗深度：通过深度模拟增强3D表面异常检测。发表于*IEEE/CVF Winter Conference on Applications of Computer Vision, WACV*，第2153–2161页。IEEE，2024年。1, 2, 3, 4[31] 张功杰、崔凯文、Tzu-Yi Hung和卢世健。Defect-GAN：用于自动缺陷检测的高保真缺陷合成。发表于*IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Waikoloa, HI, USA, January 3-8, 2021*，第2523–2533页。IEEE，2021年。3, 6[32] 张律民、饶安一和Maneesh Agrawala。为文本到图像扩散模型添加条件控制。发表于*IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*，第3813–3824页。IEEE，2023年。3 [33] 张曦淼、徐敏、周秀壮。Realnet：一种用于异常检测的具有真实合成异常的特征选择网络。*CoRR*，abs/2403.05897，2024年。1, 2, 3, 4 [34] 赵盛宇、刘志健、林济、朱俊彦、韩松。用于数据高效GAN训练的可微分增强。收录于*Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, vir- tual*，2020年。6 [35] 赵世浩、陈东东、陈彦春、鲍建敏、郝少哲、袁路、黄冠怡。Uni-controlnet：文本到图像扩散模型的一体化控制。收录于*Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*，2023年。3 [36] 赵颖。用于无监督异常定位与检测的恰可察觉学习。收录于*ICME*，第In Press页，2022年。1, 2 [37] 赵颖。Omnial：一种用于无监督异常定位的统一CNN框架。收录于*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*，第3924–3933页，2023年。1, 2 [38] 赵颖。Anomalyfactory：将异常生成视为无监督异常定位。*arXiv preprint arXiv:2408.09533*，2024年。1, 2, 3, 4, 5, 6, 8 [39] 赵颖。Logical：面向无监督异常定位的逻辑异常合成。收录于*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*，第4022–4031页，2024年。1, 2, 3, 5 [40] 邹阳、郑钟宪、拉塔·佩穆拉、张冬青、奥恩卡尔·达比尔。用于异常检测与分割的找差异自监督预训练。收录于*Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings*，第392–408页。Springer，2022年。1, 2