

SeaS：基于分离与共享微调的少样本工业异常图像生成

戴哲伟^{1,*}、曾世雷^{1,*}、刘昊天¹、李旭瑞¹、薛峰⁴、周宇^{1,2,3,†} ¹华中科技大学电子信息与通信学院 ²华中科技大学智能互联网技术湖北省重点实验室 ³武汉精测电子集团股份有限公司人工智能研究院 ⁴特伦托大学信息工程与计算机科学系

{zw dai, shl zeng, ht liu master, xr li plus, yu zhou}@hust.edu.cn, xuefengbupt@gmail.com

摘要

We introduce SeaS, a unified industrial generative model for automatically creating diverse anomalies, authentic normal products, and precise anomaly masks. While extensive research exists, most efforts either focus on specific tasks, i.e., anomalies or normal products only, or require separate models for each anomaly type. Consequently, prior methods either offer limited generative capability or depend on a vast array of anomaly-specific models. We demonstrate that U-Net’s differentiated learning ability captures the distinct visual traits of slightly-varied normal products and diverse anomalies, enabling us to construct a unified model for all tasks. Specifically, we first introduce an Unbalanced Abnormal (UA) Text Prompt, comprising one normal token and multiple anomaly tokens. More importantly, our Decoupled Anomaly Alignment (DA) loss decouples anomaly attributes and binds them to distinct anomaly tokens of UA, enabling SeaS to create unseen anomalies by recombining these attributes. Furthermore, our Normal-image Alignment (NA) loss aligns the normal token to normal patterns, making generated normal products globally consistent and locally varied. Finally, SeaS produces accurate anomaly masks by fusing discriminative U-Net features with high-resolution VAE features. SeaS sets a new benchmark for industrial generation, significantly enhancing downstream applications, with average improvements of +8.66% pixel-level AP for synthesis-based AD approaches, +1.10% image-level AP for unsupervised AD methods, and +12.79% IoU for supervised segmentation models. Code is available at <https://github.com/HUST-SLOW/SeaS>.

1. 引言

在工业场景中，生成模型被用于合成各种视觉元素，以满足需求——

*Contributed Equally, † Corresponding Authors.

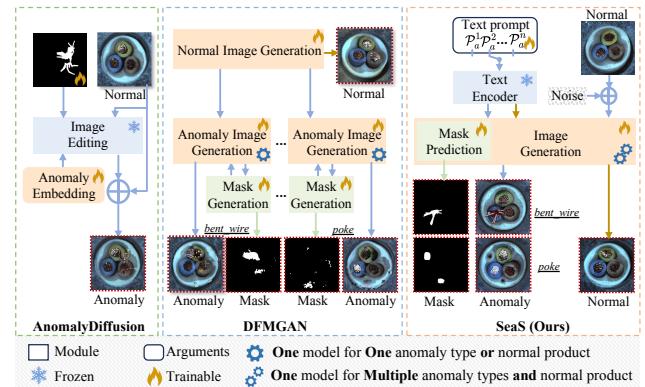


图1. (a) AnomalyDiffusion仅生成异常区域，并依据预定义的掩码将其编辑到输入的正常图像上。(b) DFMGAN为每种异常类型或正常产品训练多个专用生成器，且无法生成精确的掩码。(c) SeaS训练了一个统一模型，能够为多种异常类型生成异常图像和掩码，同时也能够生成正常图像。

不同异常检测（AD）方法和监督分割模型的评估如下。
◦

- 为基于合成的异常检测方法生成伪异常[9, 41]。
- 为无监督异常检测方法生成伪正常图像[15, 26, 32]。
- 生成完整的异常图像及对应的精确掩码，用于训练有监督的分割模型。

上述要求已被先前算法所覆盖。CutPaste [21] 通过将裁剪的正常区域粘贴到正常产品图像（正常图像的简称）上来创建异常。AnomalyDiffusion [17] 使用扩散模型生成异常，并根据预定义掩码将异常编辑到正常图像上（图1(a)），但它无法创建伪正常图像，且可能因掩码未对齐而受到影响。DFMGAN [11] 为每种正常产品和异常类型分别训练模型。

(图1(b)) 但无法生成精确的掩码，限制了其在训练有监督分割模型时的有效性。总而言之，现有方法要么仅关注正常产品或异常样本的生成，要么需要多个独立模型分别完成所有任务。它们无法灵活运用统一模型处理全部需求，即同时实现多样化异常、逼真的正常产品及像素级精确掩码的生成。本文研究统一的工业异常生成方案，以满足各类下游任务的需求。

本工作的新颖性源于对单一工业生产线的关键观察：*normal products exhibit a globally consistent surface with minor local variations, whereas anomalies exhibit high variability*

。这些特征能够被U-Net有效捕捉，得益于其在扩散过程中具备的差异化学习能力。基于这一洞见，我们提出了一种分离与共享微调方法(SeaS)，使用共享的U-Net对不同的变化进行建模。首先，为显式建模正常产品和异常的变化，我们提出了非平衡异常(UA)文本提示。其非平衡设计包含一个正常标记和多个异常标记，从而解耦了轻微变化的正常产品表面与多样异常语义的表征。其次，为学习高度多样化的异常，我们提出解耦异常对齐(DA)损失，将异常属性绑定到UA的不同异常标记上。重组这些解耦的属性可能产生训练数据集中从未出现过的异常，从而增加了生成异常的多样性。第三，针对轻微变化的正常产品，我们提出了正常图像对齐(NA)损失。它使网络能够从正常图像中学习正常产品的关键特征，从而使UA的正常标记能够表达产品的全局一致性。上述两个训练过程相互分离，但在共享的U-Net上进行。SeaS使U-Net能够同时建模正常产品和异常中的不同变化，表征用于掩码预测的判别性特征。然而，U-Net的低分辨率特征在直接预测时会导致粗糙的掩码。因此，我们提出了一个精细化掩码预测(RMP)分支。它结合U-Net特征与高分辨率VAE特征，逐步生成精确且清晰的掩码。生成的异常图像在IS得分上达到1.88(MVTec AD)、1.27(VisA)和1.95(MVTec 3D AD)，IC-LPIPS分别为0.34、0.26和0.30。在多个数据集上，SeaS生成的图像将基于合成的异常检测方法的像素级平均精度提升了+8.66%，将无监督异常检测方法的图像级平均精度提高了+1.10%，并将有监督分割模型的平均交并比提升了+12.79%。

总而言之，我们方法的关键贡献在于：

- 我们提出了一种适用于工业的统一生成模型

视觉元素。它仅使用一个模型就实现了多样化的异常、全局一致的正品和像素级精确的掩码，为该领域树立了新标准。

- 新设计的分离与共享微调模型能够区分正常产品和异常产品的不同变体，从而实现对它们生成的精确控制，并获取用于掩码预测的判别性特征。
- SeaS极大地提升了多种基于合成和无监督异常检测方法的性能，并赋能监督式分割模型，使其展现出相当不错的性能。

2. 相关工作

异常图像生成。早期的非生成方法[10, 21, 41]依赖数据增强来合成伪异常，但由于异常模式不一致而存在保真度低的问题。一些生成方法[13, 17, 34]仅生成异常并将其融合到真实正常图像中。NSA[34]使用泊松图像编辑[30]来融合裁剪的正常区域。然而，这些方法无法创建伪正常图像，且需要异常掩码作为输入，不合理的掩码位置会损害保真度与一致性。基于GAN的方法[11, 28, 42]生成完整的异常图像。DFMGAN[11]为每种异常类型训练多个独立模型来生成正常图像和异常图像，其产生的掩码常与异常区域未能精确对齐，限制了其在训练监督分割模型中的实用性。不同于现有方法，我们提出一种基于稳定扩散的统一生成模型，可生成多样化异常、全局一致的正规产品及像素级精确的掩码。

微调扩散模型。微调是一种增强预训练扩散模型特定能力的有效策略[6, 12, 43]。个性化方法[8, 12, 33]利用少量图像集微调扩散模型，从而生成相同对象的图像。多种多概念图像微调方法[1, 14, 19, 20, 37]使用交叉注意力图将嵌入与图像中的独立概念对齐。然而，这些方法未考虑不同图像区域中的 $\{v^*\}$ 差异变化，这对工业异常图像生成至关重要。因此，我们提出一种分离共享式微调策略，以建模异常品与正常产品不同程度的变化，该策略在共享U-Net上独立学习产品与异常特征。

基于生成方法的掩码预测。先前针对生成图像的掩码预测方法主要基于GANs的特征[22, 44]。然而，这些方法无法保证在极小数据集上生成精确的掩码。基于Stable Diffusion [31]，近期一些方法（例如Diffu-

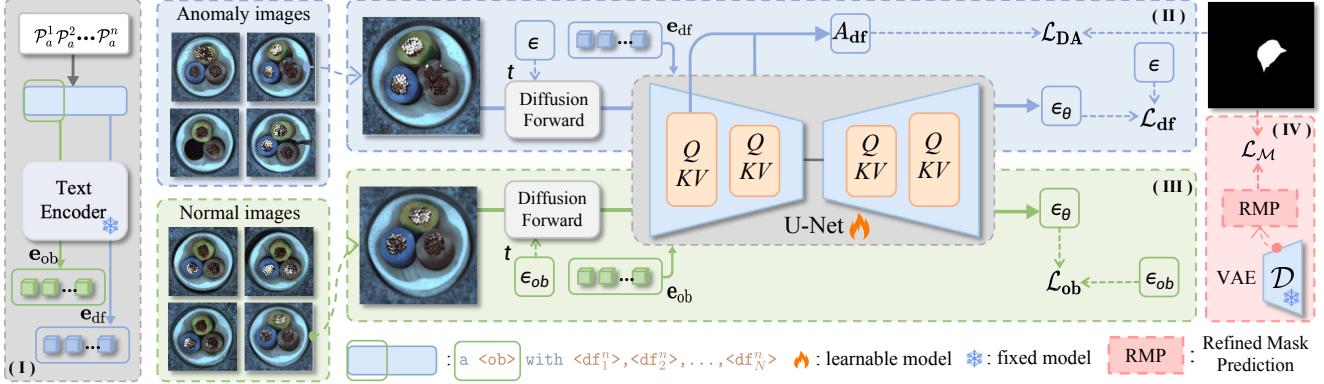


图2. SeaS的整体框架。它包含四个部分：(I) 不平衡异常文本提示，(II) 解耦异常对齐，用于将异常标记 $\text{<df}_n\text{>}$ 对齐到异常图像的异常区域，(III) 正常图像对齐，通过正常图像保持真实性，以及(IV) 精细化掩码预测分支，用于生成精确的掩码。

Mask [36]、DatasetDM [35]和DatasetDiffusion [27]通过利用交叉注意力图的潜力来生成掩码。然而，由于交叉注意力图的分辨率较低，它们被直接插值到更高分辨率以匹配图像尺寸，且未借助任何辅助信息，这导致了显著的边界不确定性。我们引入VAE解码器的高分辨率特征作为辅助信息进行分辨率恢复，将其与U-Net解码器特征融合——这些特征因对正常产品和异常的不同变化进行建模而具有强区分性——从而生成精确的高分辨率掩码。

3. 方法

所提出的分离与共享（SeaS）微调策略的训练阶段如图2所示。在第3.1节中，我们介绍了方法的预备知识。在第3.2节中，我们首先设计了一种非平衡异常文本提示，其中包含一组分别表征正常产品和异常特征的标记。随后，我们提出解耦异常对齐（DA）损失，将异常图像区域与异常标记绑定，并利用正常图像对齐（NA）损失使正常标记能够表达全局一致的正常产品表面。这两个训练过程分别在异常图像和正常图像上实施，但基于共享的U-Net架构。接着，基于训练良好的U-Net，我们在第3.3节中设计了一个精细化掩码预测分支，用于生成与生成的异常图像对应的精确掩码。最后，我们在第3.4节中详细说明了异常图像-掩码对及正常图像的生成过程。

3.1. 预备知识

稳定扩散。给定输入图像 x_0 ，稳定扩散[31]首先将 x_0 转换为潜在空间中的 $z = \varepsilon(x_0)$ ，随后加入随机采样的噪声 $\epsilon \sim N(0, I)$ 得到 $\hat{z}_t = \alpha_t z + \beta_t \epsilon$ ，其中 t 是随机采样的时间步长。接着，使用U-Net来预测噪声。

设 $c_\theta(\mathcal{P})$ 为 CLIP 文本编码器，其将条件文本提示 \mathcal{P} 映射为条件向量 \mathbf{e} 。Stable Diffusion 的训练损失可表述如下：

$$\mathcal{L}_{SD} = \mathbb{E}_{z=\varepsilon(x_0), \mathcal{P}, \epsilon \sim N(0, I), t} \left[\|\mathbf{e} - \epsilon_\theta(\hat{z}_t, t, \mathbf{e})\|_2^2 \right] \quad (1)$$

其中 ϵ_θ 是预测的噪声。

U-Net中的交叉注意力图。为了控制生成过程，条件机制通过计算条件向量 $\{\mathbf{v}^*\}$ 与 U-Net 内部层图像特征 $\{\mathbf{v}\}$ 之间的交叉注意力来实现[7, 16, 39]。交叉注意力图 $\{\mathbf{v}^*\}$ 的计算方式为：

$$A^{m,l} = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right), Q = \phi_q(\mathbf{v}), K = \phi_k(\mathbf{e}) \quad (2)$$

其中 $Q \in \mathbb{R}^{r \times r \times C}$ 表示通过线性层 ϕ_q 从 \mathbf{v} 投影得到的查询， r 是 U-Net 中特征图的分辨率， l 是 U-Net 内部层的索引。 $K \in \mathbb{R}^{Z \times C}$ 表示通过另一个线性层 ϕ_k 从 \mathbf{e} 得到的键， Z 是填充后的文本嵌入数量。

3.2. 分离与共享微调

不平衡异常文本提示。通过实验观察，我们发现典型的文本提示（如“有缺陷的瓶子照片”[18]或“损坏的瓶子”[46]）对于工业异常生成并非最优。针对正常产品和异常情况的平衡语义词汇可能无法捕捉它们之间的差异变化程度。因此，我们为每种产品的每种异常类型设计了不平衡异常（UA）文本提示，即：

\mathcal{P} = 一个具有 $\text{<df}_1\text{>}、\text{<df}_2\text{>} \cdots \text{<df}_N\text{>}$ 的 <ob> ，其中 <ob> 和 $\text{<df}_n\text{>} (n \in \{1, 2, \dots, N\})$ 分别是工业正常产品（简称正常令牌）和异常产品（简称异常令牌）的标记。我们为每种异常类型使用一组 N 个异常令牌，不同集合对应不同的

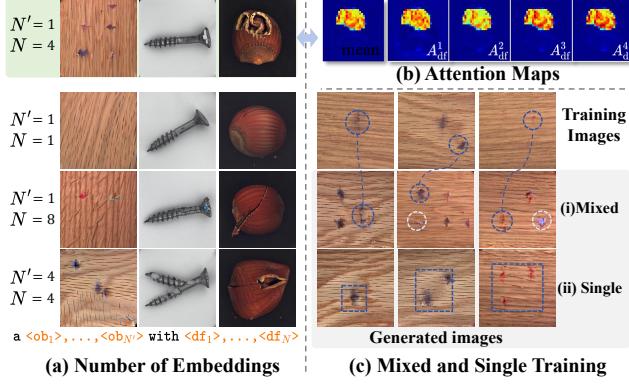


图3. (a) 使用不同数量标记生成的图像。 (b) 交叉注意力图。 (c) 多样化生成图像示例。

异常类型。如图3所示，在SeaS中，我们分别使用正常图像训练对应 $\langle ob \rangle$ 的嵌入表示，并使用异常图像训练对应 $\langle df_n \rangle$ 的嵌入表示。实验观察表明，一个 $\langle ob \rangle$ 足以表达正常产品，而需要多个 $\langle df_n \rangle$ 来控制异常的生成。如图3(a)所示，当我们使用UA提示 \mathcal{P} （(a图中的绿色虚线框)）时，(b)中的交叉注意力图显示不同标记在异常区域具有不同的响应，这表明它们关注异常的不同属性，对交叉注意力图进行平均操作会产生未见过的异常。当我们仅使用一个 $\langle df \rangle$ 时，很难将其与属于同一类别的多个不同异常对齐。因此，在推理过程中，如果去噪后的异常特征与 $\langle df \rangle$ 距离较大，U-Net会为其分配较小的响应，导致“异常缺失”现象，例如在($N' = 1, N = 1$)情况下生成的图像。此外，如果我们使用大量 $\langle df_n \rangle$ ，会发现每个 $\langle df_n \rangle$ 可能只关注异常的某些局部特性，这种情况增加了多样性但可能降低异常的真实性，如 $N' = 1, N = 8$ 案例所示。类似地，如果我们使用多个可学习的 $\langle ob \rangle$ （例如 $N' = 4, N = 4$ ），每个 $\langle ob \rangle$ 会关注正常产品的局部特征，这可能降低正常产品的全局一致性。

解耦异常对齐。给定少量异常图像 x_{df} 及其对应的掩码，我们的目标是通过调整U-Net和与 $\langle df_n \rangle$ 对应的可学习嵌入，将异常标记 $\langle df_n \rangle$ 对齐到 x_{df} 的异常区域。因此，我们提出了解耦异常对齐(DA)损失，即：

$$\mathcal{L}_{\text{DA}} = \sum_{l=1}^L \left(\left\| \frac{1}{N} \sum_{n=1}^N A_{df}^{n,l} - M^l \right\|^2 + \| A_{ob}^l \odot M^l \|^2 \right) \quad (3)$$

其中 $A_{df}^{n,l} \in \mathbb{R}^{r \times r \times 1}$ 是与第 n 个异常标记 $\langle df_n \rangle$ 对应的交叉注意力图， N 是数

\mathcal{P} 中的异常标记数量。 L 是用于对齐的U-Net总层数。 M^l 是分辨率 $r \times r$ 的二进制掩码，其中异常区域为1，背景为0。 $A_{ob}^l \in \mathbb{R}^{r \times r \times 1}$ 是对应于正常标记 $\langle ob \rangle$ 的交叉注意力图， \odot 是逐元素乘积。DA损失执行异常与正常产品的强制解耦。DA损失的第一项是根据掩码 M^l 将异常区域对齐到 $\langle df_n \rangle$ 。DA损失的第二项降低 A_{ob}^l 在异常区域的响应值，这防止 $\langle ob \rangle$ 对齐到 x_{df} 的异常区域。关于DA损失如何确保异常多样性的进一步分析见附录A.2。因此，异常图像 x_{df} 的总损失为：

$$\mathcal{L}_{df} = \mathcal{L}_{\text{DA}} + \|\epsilon_{df} - \epsilon_{\theta}(\hat{z}_{df}, t_{df}, e_{df})\|_2^2 \quad (4)$$

在公式4的第二项中，我们使用随机噪声 ϵ_{df} 和时间步 t_{df} 对异常图像 x_{df} 执行前向扩散，从而获得含噪潜变量 \hat{z}_{df} 。条件向量 $e_{df} \in \mathbb{R}^{Z \times C_1}$ 用于引导U-Net预测噪声，随后与噪声 ϵ_{df} 计算损失。

法线图像对齐。正如我们所讨论的，增加法线标记 $\langle ob \rangle$ 的数量会带来更高的多样性，但可能会降低生成法线图像的真实性和破坏全局一致性。然而，仅将一个 $\langle ob \rangle$ 与少量训练图像对齐可能会面临过拟合的问题。因此，我们添加了法线图像对齐(NA)损失来克服这一困境，其表述如下，

$$\mathcal{L}_{ob} = \|\epsilon_{ob} - \epsilon_{\theta}(\hat{z}_{ob}, t_{ob}, e_{ob})\|_2^2 \quad (5)$$

在计算NA损失时，我们并未将 x_{df} 的正常区域与 $\langle ob \rangle$ 对齐，而是使用随机噪声 ϵ_{ob} 和时间步 t_{ob} 对正常图像 x_{ob} 进行前向扩散。随后，将含噪潜变量 \hat{z}_{ob} 与对应 \mathcal{P} 正常标记（即“a $\langle ob \rangle$ ”）的嵌入向量 e_{ob} 输入U-Net进行噪声预测，进而通过 ϵ_{ob} 计算NA损失。

混合训练。基于异常图像的分离DA损失和正常图像的NA损失，分离与共享微调的目标被构建为：

$$\mathcal{L} = \mathcal{L}_{df} + \mathcal{L}_{ob} \quad (6)$$

在训练过程中，我们并非为每种异常类型单独训练一个U-Net模型，而是为每个产品训练一个统一的U-Net模型。具体而言，给定一个产品图像集，其中包含 G 个异常类别的掩码异常图像及部分正常图像。我们将该产品的所有异常图像统一归入集合

$X_{df} = \{x_{df}^1, x_{df}^2, \dots, x_{df}^H\}$ 。针对每种异常类型，我们使用 \mathcal{P} 并配合不同的异常标记集合。此外，我们采样固定数量的正常图像构成正常训练集

$X_{ob} = \{x_{ob}^1, x_{ob}^2, \dots, x_{ob}^P\}$ 。在每次

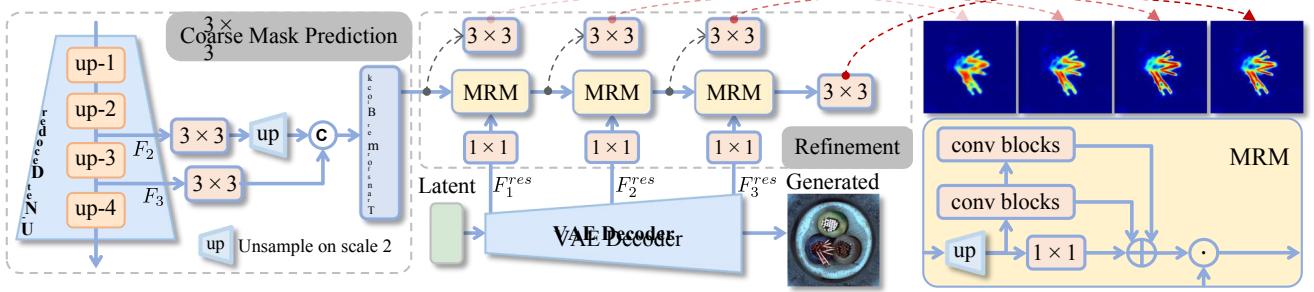


图4. 推理过程中的精细化掩码预测（RMP）分支。粗粒度特征提取利用U-Net解码器的up-2和up-3层特征提取粗粒度特征。级联掩码优化模块（MRM）在VAE解码器高分辨率特征的辅助下，进一步获取与异常区域精确对齐的掩码。

在我们微调过程的步骤中，我们从 X_{df} 和 X_{ob} 中采样相同数量的图像，并将它们混合到一个批次中。我们发现，这种混合训练策略不仅缓解了因每种异常类型数量有限而导致的过拟合问题，还增加了异常图像的多样性，同时仍保持了合理的真实性。如图3(c)所示，(i)表明混合训练的模型可能生成新的异常，例如白色虚线框内的异常。相比之下，(ii)中的异常则过度拟合了训练图像。关于混合训练策略的更多消融研究见附录A.5中的表12。

3.3. 精细化掩码预测

分离与共享方法的设计使U-Net能够同时建模正常产品和异常情况中不同程度的变化，为掩码预测提供判别性特征。为进一步获得像素级精确的掩码，我们设计了级联式精细化掩码预测（RMP）分支，该分支被嫁接至SeaS框架内训练的U-Net上（如第3.2节所述）。如图4所示，RMP包含两个步骤：首先从U-Net中捕获判别性特征，随后将其与VAE解码器的高分辨率特征相结合，生成与异常区域匹配的掩码。

粗粒度特征提取。第一步旨在从U-Net解码器中提取异常检测所需的粗粒度但高区分性特征。具体而言，令 $F_1 \in \mathbb{R}^{32 \times 32 \times 1280}$ 和 $F_2 \in \mathbb{R}^{64 \times 64 \times 640}$ 分别表示U-Net解码器中“up-2”和“up-3”层的输出特征。我们首先利用 1×1 卷积块将 F_1 和 F_2 的通道数分别压缩至 $F_1 \in \mathbb{R}^{32 \times 32 \times 128}$ 和 $F_2 \in \mathbb{R}^{64 \times 64 \times 64}$ 。随后，将 F_1 上采样至 64×64 分辨率并与 F_2 进行拼接。最后，采用四层Transformer融合拼接后的特征，得到统一的粗粒度特征 $\hat{F} \in \mathbb{R}^{64 \times 64 \times 192}$ 。

掩码细化模块。直接将粗糙特征 \hat{F} 上采样至高分辨率会导致异常细节的丢失。因此，我们设计了掩码细化模块（MRM），以在

渐进式方式。如图4所示，每个多分辨率模块（MRM）接收两个特征，即来自VAE的高分辨率特征和待优化的判别性特征。首先，将判别性特征上采样以与VAE的高分辨率特征对齐。为保持判别能力，上采样后的特征通过两个级联的卷积块处理以捕获多尺度异常特征，并通过一个 1×1 卷积捕获局部特征。这些特征随后相加，并与VAE特征逐元素相乘以增强异常边界。最后，MRM采用一个 3×3 卷积融合叠加后的特征，并输出优化后的特征。

为了优化 \hat{F} ，我们采用了三个按顺序排列的MRM。每个MRM都将前一个MRM的输出作为待优化的判别特征，而第一个MRM则以 \hat{F} 作为判别输入。对于每个MRM的另一个输入，我们分别使用VAE解码器的第一、第二和第三个“上采样块”的输出。通过这种方式，最后一个MRM所获得的特征兼具高分辨率和高判别性的优势。最后，我们使用一个 3×3 卷积层和softmax函数，基于最后一个MRM的输出生成优化后的异常掩码 $\hat{M}'_{df} \in \mathbb{R}^{512 \times 512 \times 2}$ 。

损失函数。在训练过程中，我们使用 x_{df} 和 x_{ob} 作为输入。对于 x_{df} ，我们从粗粒度特征提取模块获得粗掩码 $\hat{M}_{df} \in \mathbb{R}^{64 \times 64 \times 2}$ ，以及经过多分辨率模块后的 \hat{M}'_{df} 。类似地，对于 x_{ob} ，我们从粗粒度特征提取模块获得 $\hat{M}_{ob} \in \mathbb{R}^{64 \times 64 \times 2}$ ，并直接上采样至原始分辨率，记为 $\hat{M}'_{ob} \in \mathbb{R}^{512 \times 512 \times 2}$ 。随后，我们对低分辨率和高分辨率预测结果进行监督，具体如下：

$$\mathcal{L}_M = \mathcal{F}(\hat{M}_{df}, M_{df}) + \mathcal{F}(\hat{M}_{ob}, M_{ob}) + \mathcal{F}(\hat{M}'_{df}, M'_{df}) + \mathcal{F}(\hat{M}'_{ob}, M'_{ob}) \quad (7)$$

其中 \mathcal{F} 表示Focal Loss[24]。 $M_{ob} \in \mathbb{R}^{64 \times 64 \times 1}$ 和 $M'_{ob} \in \mathbb{R}^{512 \times 512 \times 1}$ 用于抑制正常图像中的噪声，每个像素值均设为0。 $M_{df} \in \mathbb{R}^{64 \times 64 \times 1}$ 和 $M'_{df} \in \mathbb{R}^{512 \times 512 \times 1}$ 是异常图像的真实掩码。更多关于

训练RMP分支时正常图像的效果见附录A.5中的表16和图11。

3.4. 推理

在生成过程中，为进一步确保正常产品的全局一致性，我们从 X_{ob} 中随机选取一张正常图像 x_{ob} 作为输入，并向 x_{ob} 添加随机噪声，从而得到初始的噪声潜变量 ζ_0 。接着，为生成异常图像，将 ζ_0 输入U-Net进行噪声预测，该过程由公式4)中提到的条件向量 e_{df} (引导，该向量对应完整的UA文本提示 P)。为生成正常图像以进一步增强无监督异常检测方法，我们使用与 P 的正常标记对应的条件向量 e_{ob} 。关于异常对应的掩码，在最后三个去噪步骤中，RMP分支(第3.3节)利用U-Net解码器和VAE解码器的特征生成最终的异常掩码。具体而言，我们平均这些步骤中精炼的异常掩码以获得精炼掩码 $\hat{M}'_{df} \in \mathbb{R}^{512 \times 512 \times 2}$ 。随后对 \hat{M}'_{df} 的第二通道取阈值 τ 以分割出最终异常掩码 $M_{df} \in \mathbb{R}^{512 \times 512 \times 1}$ 。 τ 对下游监督分割模型的影响见附录A.5中的表18。在最后一个去噪步骤中，生成模型的输出被用作生成的异常图像。

4. 实验

4.1. 实验设置

实现细节。我们通过微调预训练的Stable Diffusion v1-4 [31]来训练SeaS。在异常图像生成实验中，每种异常类型在训练中使用60张正常图像和5张带掩码的异常图像。我们为每个产品训练一个生成模型，覆盖所有异常类型。在推理过程中，我们为单一异常类型生成1,000个异常图像-掩码对。更多细节见附录A.3。

数据集。我们在MVTec AD数据集[3]、VisA数据集[47]和MVTec 3D AD数据集(仅RGB图像)[4]上进行实验。MVTec AD数据集包含15个产品类别，每个类别最多有8种不同的异常类型。VisA数据集涵盖3个领域中的12个对象。MVTec 3D AD数据集包含10个产品类别，每个类别最多有4种不同的异常类型。该数据集包含更多挑战，例如光照条件变化和产品姿态变化。

评估指标。对于图像生成，与现有方法[11, 17]仅评估完整异常图像不同，我们的评估包含三个层面：异常图像、正常图像及异常区域，采用4项指标：(1) 初始分数 (IS) 和簇内成对LPIPS距离 (IC-LPIPS) [29]，用于衡量异常图像的真实性与多样性；(2) KID [5] 用于评估正常图像的真实性；(3) 仅在异常区域计算的IC-LPIPS (简称为

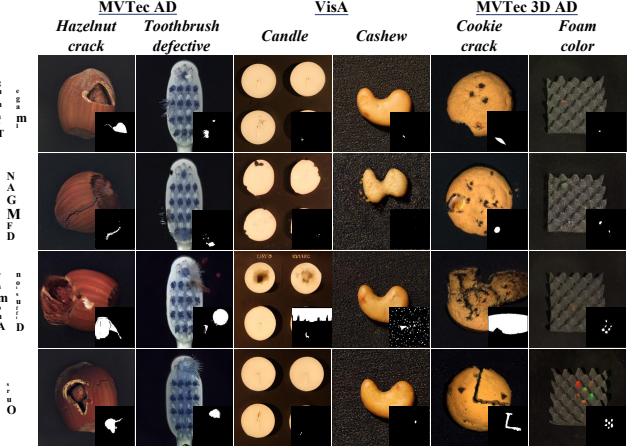


图5. MVTec AD、VisA和MVTec 3D AD上的生成结果可视化。右下角的子图像为生成的掩码，none表示该方法无法生成掩码。

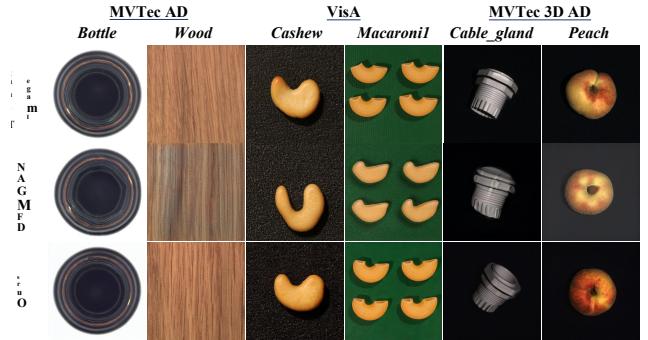


图6. 在MVTec AD、VisA和MVTec 3D AD上生成正常图像的可视化结果。

IC-LPIPS(a)) 用于衡量异常多样性。对于像素级异常分割和图像级异常检测，我们使用三个指标：接收者操作特征曲线下面积 (AUROC)、平均精度 (AP) 以及最优阈值下的 F_1 -分数 (F_1 -max)。我们还报告了分割的交并比 (IoU)。

4.2. 异常图像生成中的比较

比较方法。在图像生成方面，我们将SeaS与当前异常图像生成方法进行比较，例如Crop&Paste [23]、SDGAN [28]、Defect-GAN [42] (均未开源)、DFMGAN [11]和AnomalyDiffusion [17]，评估指标包括保真度和多样性。针对多样化的生成异常，我们将SeaS生成的异常与基于合成的异常检测方法如DRAEM [41]和GLASS [9]结合使用。对于真实的生成正常图像，我们利用SeaS生成的正常图像来增强无监督异常检测方法的训练集，例如HVQ-Trans [26]、PatchCore [32]和MambaAD [15]。针对异常图像-掩码对，我们使用DFMGAN、AnomalyDiffusion和SeaS分别生成数据，以训练分割模型如BiSeNet V2 [40]、UPerNet [38]和LFD [45]。与需要为每个类别训练一个分割模型的AnomalyDiffusion不同，

表1. 在MVTec AD、VisA和MVTec AD 3D数据集上IS和IC-LPIPS的对比结果。粗体表示最佳性能。

Methods	MVTec AD				VisA				MVTec 3D AD			
	IS↑	IC-LPIPS↑	KID↓	IC-LPIPS(a)↑	IS↑	IC-LPIPS↑	KID↓	IC-LPIPS(a)↑	IS↑	IC-LPIPS↑	KID↓	IC-LPIPS(a)↑
Crop&Paste[23]	1.51	0.14	-	-	-	-	-	-	-	-	-	-
SDGAN[28]	1.71	0.13	-	-	-	-	-	-	-	-	-	-
Defect-GAN[42]	1.69	0.15	-	-	-	-	-	-	-	-	-	-
DFMGAN[11]	1.72	0.20	0.12	0.14	1.25	0.25	0.24	0.05	1.80	0.29	0.19	0.08
AnomalyDiffusion[17]	1.80	0.32	-	0.12	1.26	0.25	-	0.04	1.61	0.22	-	0.07
SeaS	1.88	0.34	0.04	0.18	1.27	0.26	0.02	0.06	1.95	0.30	0.06	0.09

表2. 梳状结构对比

结合生成异常与基于合成的异常检测方法

多种数据集上的方法。

Segmentation Models	MVTec AD						VisA						MVTec 3D AD								
	Image-level			Pixel-level			Image-level			Pixel-level			Image-level			Pixel-level					
	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU		
DRAEM [41]	98.00	98.45	96.34	97.90	67.89	66.04	60.30	86.28	85.30	81.66	92.92	17.15	22.95	13.57	79.16	90.90	89.78	86.73	14.02	17.00	12.42
DRAEM + SeaS	98.64	99.40	97.89	98.11	76.55	72.70	58.87	88.12	87.04	83.04	98.45	49.05	48.62	35.00	85.45	93.58	90.85	95.43	20.09	26.10	17.07
GLASS [9]	99.92	99.98	99.60	99.27	74.09	70.42	57.14	97.68	96.89	93.03	98.47	45.58	48.39	39.92	92.34	96.85	93.37	98.46	48.46	49.13	45.03
GLASS + SeaS	99.97	99.99	99.81	99.29	76.82	72.38	57.45	97.88	97.39	93.21	98.43	48.06	49.32	40.00	92.95	97.38	93.21	98.73	48.55	49.28	46.02
Average [9, 41]	98.96	99.22	97.97	98.59	70.99	68.23	58.72	91.98	91.10	87.35	95.70	31.37	35.67	26.75	85.75	93.88	91.58	92.60	31.24	33.07	28.73
Average(+ SeaS)	99.31	99.70	98.85	98.70	76.69	72.54	58.16	93.00	92.22	88.13	98.44	48.56	48.97	37.50	89.20	95.48	92.03	97.08	34.32	37.69	31.55

表3. 在多个数据集上结合生成正常图像与无监督异常检测方法的比较。

Segmentation Models	MVTec AD						VisA						MVTec 3D AD								
	Image-level			Pixel-level			Image-level			Pixel-level			Image-level			Pixel-level					
	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU		
HVQ-Trans [26]	96.38	98.09	95.30	97.60	47.95	53.32	45.03	90.11	88.18	84.08	98.10	28.67	35.05	24.03	68.15	84.38	85.20	96.40	17.23	24.59	20.51
HVQ-Trans + SeaS	97.25	98.48	95.78	97.58	48.53	53.84	44.61	92.12	90.35	86.23	98.15	29.52	36.00	23.60	71.26	90.35	89.23	96.56	19.34	26.40	20.47
PatchCore [32]	98.63	99.47	98.18	98.37	56.13	58.83	49.45	94.84	95.98	91.69	98.38	48.58	49.69	42.44	83.44	94.89	92.24	98.55	34.52	39.09	39.29
PatchCore + SeaS	98.64	99.48	98.22	98.37	63.98	64.07	55.43	94.97	96.06	91.81	98.41	48.60	49.72	42.46	83.88	94.97	92.32	98.56	34.65	39.41	39.43
MambaAD [15]	98.54	99.52	97.77	97.67	56.23	59.34	51.31	94.19	94.44	89.55	98.49	39.27	44.18	37.68	85.92	95.69	92.51	98.57	37.30	41.08	39.44
MambaAD + SeaS	98.80	99.64	98.40	97.66	56.86	59.70	51.51	94.23	94.65	89.93	98.70	39.33	43.99	36.62	88.67	96.60	93.41	98.74	35.46	39.59	39.51
Average [15, 26, 32]	97.85	99.03	97.08	97.88	53.44	57.16	48.60	93.05	92.87	88.44	98.32	38.84	42.97	34.72	79.17	91.65	89.98	97.84	29.68	34.92	33.08
Average(+ SeaS)	98.23	99.20	97.47	97.87	56.46	59.20	50.52	93.77	93.69	89.32	98.42	39.15	43.24	34.23	81.27	93.97	91.65	97.95	29.82	35.13	33.14

表4. 在多个数据集上对经过训练的有监督分割模型进行异常检测与分割的性能比较。

Segmentation Models	Generative Models	MVTec AD						VisA						MVTec 3D AD								
		Image-level			Pixel-level			Image-level			Pixel-level			Image-level			Pixel-level					
		AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU		
DFMGAN	90.90	94.43	90.33	94.57	60.42	60.54	45.83	63.07	62.63	66.48	75.91	9.17	15.00	9.66	61.88	81.80	84.44	75.89	15.02	21.73	15.68	
BiSeNet V2	AnomalyDiffusion	90.08	94.84	91.84	96.27	64.50	62.27	42.89	76.11	77.74	73.13	89.29	34.16	37.93	15.93	61.49	81.35	85.36	92.39	15.15	20.09	14.70
UPerNet	AnomalyDiffusion	96.62	98.61	96.21	96.87	69.92	66.95	50.80	83.18	84.08	78.88	95.00	39.92	45.37	20.53	76.56	90.42	87.35	88.48	28.95	35.81	25.04
LFD	AnomalyDiffusion	91.08	95.40	90.58	94.91	67.06	65.09	45.49	65.38	62.25	66.59	81.21	15.14	18.70	6.44	62.23	82.17	85.38	72.15	9.54	14.29	14.81
[45]	SeaS	95.15	97.78	94.66	96.30	69.77	66.99	45.77	81.97	82.36	77.35	88.00	30.86	38.56	16.61	77.06	89.44	87.20	92.68	24.29	32.74	19.90
DFMGAN	Average	90.91	94.75	90.43	93.94	61.50	60.85	45.99	66.71	65.51	67.92	77.40	12.24	17.41	10.52	63.89	83.83	84.94	74.39	14.70	20.69	16.42
AnomalyDiffusion	SeaS	93.95	97.08	94.24	94.68	68.06	65.40	46.49	80.42	81.39	76.45	90.76	34.98	40.62	17.69	71.70	87.07	86.64	91.18	22.80	29.55	19.88
SeaS	96.72	98.41	95.97	97.72	73.59	69.86	57.66	86.34	86.75	80.69	95.32	47.38	49.29	29.40	78.38	90.52	87.27	91.32	34.93	39.87	35.37	

针对每个产品单独建模，我们为所有产品训练了一个统一的监督式分割模型，这更具挑战性。

异常图像生成质量。在表1中，我们将SeaS与一些最先进的异常图像生成方法在保真度 (IS和KID) 和多样性

(IC-LPIPS和IC-LPIPS(a)) 上进行了比较。SeaS在IS和IC-LPIPS方面优于其他方法，显示出卓越的保真度和多样性。它在生成真实的正常图像和多样化的异常方面也表现出色。与无法生成正常图像的AnomalyDiffusion相比，SeaS在IC-LPIPS(a)方面领先。SeaS在KID和IC-LPIPS(a)方面也超越了DFMGAN。我们在图5中展示了生成的异常图像，SeaS生成的异常图像具有更高的多样性。

保真度（例如，*hazelnut_crack*）。与其他方法相比，SeaS能够生成具有不同类型、颜色和形状的异常图像，而不是过度拟合训练图像（例如，*foam_color*）。SeaS生成的掩码也与异常区域精确对齐（例如，*tooth-brush_defective*）。我们还在图6中展示了真实生成的正常图像。更多定性和定量的异常图像生成结果见附录A.6。

将生成的异常与基于合成的异常检测方法相结合。我们将DRAEM [41]和GLASS [9]中合成的伪异常替换为SeaS生成的异常。如表2所示，SeaS生成的异常具有足够的多样性，能够持续改

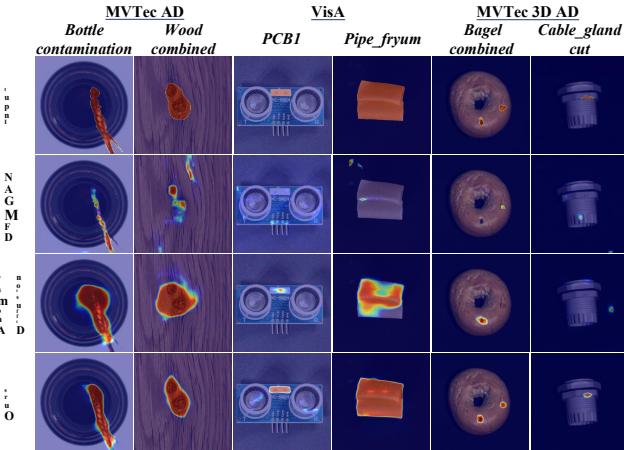


图7. 在MVTec AD数据集上使用BiSeNet V2的定性监督异常分割结果。通过抑制假阴性验证了基于合成的异常检测方法，从而在多个数据集上实现了更优性能。更多训练细节见附录A.3。

结合生成的正样本图像与异常检测方法。我们利用SeaS生成的正样本图像来补充现有最先进的无监督异常检测方法的训练集，结果如表3所示。通过使用具有局部细微变化和全局一致性的SeaS生成正样本图像，无监督异常检测方法减少了误报，并在多项指标上表现优异，提升了多种数据集上的工业异常检测性能。更多训练细节见附录A.3。

训练用于异常分割与检测的监督式分割模型。我们为每种异常类型生成1000张图像-掩码对，并将其与原始训练集中的所有正常图像一同用于训练统一的监督式分割模型。模型在训练集未包含的剩余图像上进行测试。所有方法均使用相同数量的图像及训练设置进行训练，具体细节见附录A.4。如表4所示，分割结果一致表明我们的方法在所有分割模型中均优于其他方法，平均IoU分别提升11.17%（MVTec AD）、11.71%（VisA）和15.49%（MVTec 3D AD）。分割异常图如图7所示。使用我们生成的图像-掩码对训练BiSeNet V2时，*wood_combined*中的误报更少，*bottle_contamination*中的漏报更少。我们还将分割异常图的最大值作为图像级异常分数用于异常检测，在图像-AUROC指标上分别获得2.77%（MVTec AD）、5.92%（VisA）和6.68%（MVTec 3D AD）的提升。更多定性对比结果见附录A.7与附录A.8。

4.3. 消融研究

我们训练了额外的模型来评估每个组件的影响：(a) 使用预定义典型文本提示并固定通用语义词的模型（在表格中简称为with TP）。

表5. 生成模型的消融实验。

Method	Metrics					
	IS	IC-L	AUROC	AP	F_1 -max	IoU
(a) with TP	1.72	0.33	94.72	57.16	55.67	50.46
(b) w/o Mixed	1.79	0.32	95.82	66.07	64.50	53.11
(c) w/o NA	1.67	0.31	96.20	66.03	64.09	53.97
(d) w/o ST	1.86	0.33	96.44	67.73	65.23	54.99
(e) All (Ours)	1.88	0.34	97.21	69.21	66.37	55.28

5); (b) 同一产品中不混合不同类型异常图像的模型; (c) 不使用NA损失的模型; (d) 不使用公式3中DA损失第二项的模型（即表5中的ST缩写）; (e) 我们的完整模型。我们使用这些模型为每种异常类型生成1000张异常图像-掩码对，并训练BiSeNet V2进行有监督的异常分割。表5结果显示，省略任何组件都会导致生成图像的保真度和多样性下降，分割结果也会变差。这验证了我们提出的各组件有效性。关于SeaS的更多消融实验见附录A.5。

精炼掩码预测分支。为验证RMP分支中各组件的有效性，我们对多分辨率融合模块（MRM）、渐进式精炼粗糙特征方法（表6中简称为PM）以及粗糙掩码监督

（表6中简称为CMS）进行了消融实验：1) 仅使用CMS的模型，即不通过MRM融合高分辨率特征，而是直接对粗糙特征 $\hat{P} \in \mathbb{R}^{64 \times 64 \times 192}$ 进行卷积和双线性插值上采样以获取掩码；2) 使用MRM的模型；3) 以渐进方式采用三个MRM精炼粗糙特征的模型；4) 我们的完整模型。表6展示了BiSeNet V2的实验结果，证明RMP中每个组件对于下游监督式异常分割任务均不可或缺。更多关于RMP的消融实验详见附录A.5。

表6. RMP分支的消融实验。

Method	Metrics						
	MRM	PM	CMS	AUROC	AP	F_1 -max	IoU
			✓	97.00	65.28	62.56	53.93
✓				94.54	60.52	59.06	49.42
✓	✓			94.04	62.04	59.82	50.44
✓	✓	✓		97.21	69.21	66.37	55.28

5. 结论

本文提出了一种名为SeaS的统一生成方法。我们探究了一个隐含特性：异常表现出高变异性，而正常产品保持全局一致性。我们设计了一种分离与共享微调策略，以建模正常产品和异常的不同变化，使精细化掩码预测分支能够利用判别性特征预测精确掩码。该方法显著提升了基于合成与有监督的异常检测方法，并增强了有监督分割模型的性能。

6. 致谢

本研究得到国家自然科学基金（项目编号：62176098）的资助。计算工作在华中科技大学高性能计算平台上完成。

参考文献

- [1] Omri Avrahami, Kfir Aberman, Ohad Fried, Daniel Cohen-Or, 与 Dani Lischinski。Break-a-scene：从单张图像中提取多个概念。收录于 *SIGGRAPH Asia 2023 Conference Papers*, 第1–12页, 2023年。2[2] Shane Barratt 与 Rishi Sharma。关于初始分数的说明。arXiv preprint arXiv:1801.01973, 2018年。12[3] Paul Bergmann, Michael Fauser, David Sattlegger, 与 Carsten Steger。Mvtec ad——一个用于无监督异常检测的综合真实世界数据集。收录于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第9592–9600页, 2019年。6[4] Paul Bergmann., Xin Jin., David Sattlegger., 与 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*, 第202–213页, 2022年。6[5] Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, 与 Arthur Gretton。揭秘MMD GANs。收录于 *International Conference on Learning Representations*, 2018年。6, 12[6] Tim Brooks, Aleksander Holynski, 与 Alexei A Efros。InstructPix2Pix：学习遵循图像编辑指令。收录于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第18392–18402页, 2023年。2[7] Hila Chefer, Yuval Alaluf, Yael Vinker, Lior Wolf, 与 Daniel Cohen-Or。Attend-and-Excite：基于注意力的文本到图像扩散模型语义引导。ACM Transactions on Graphics, 42(4):1–10, 2023年。3[8] Hong Chen, Yipeng Zhang, Xin Wang, Xuguang Duan, Yuwei Zhou, 与 Wenwu Zhu。DisenBooth：用于主体驱动文本到图像生成的解耦参数高效调优。收录于 *International Conference on Learning Representations*, 2024年。2[9] Qiyu Chen, Huiyuan Luo, Chengkan Lv, 与 Zhengtao Zhang。一种用于工业异常检测与定位的、基于梯度上升的统一异常合成策略。收录于 *European Conference on Computer Vision*, 第37–54页。Springer, 2024年。1, 6, 7, 12
- [10] Terrance DeVries 与 Graham W Taylor。使用 Cutout 改进卷积神经网络的正则化。arXiv preprint arXiv:1708.04552, 2017年。2
- [11] 段宇轩, 洪岩, 牛力, 张立清. 通过缺陷感知特征操作的少样本缺陷图像生成. 于 *Proceedings of the AAAI Conference on Artificial Intelligence*, 571–578页, 2023. 1, 2, 6, 7, 12, 13, 17, 26
- [12] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit Haim Bermano, Gal Chechik 和 Daniel Cohen-or。一种

一图胜千言：基于文本反演的个性化文本到图像生成。于 *International Conference on Learning Representations*, 2022年。2, 35 [13] 桂冠, 高彬彬, 刘军, 王成杰, 吴云生。面向异常分类与分割的少样本异常驱动生成方法。于 *European Conference on Computer Vision*, 第210–226页, 2024年。2 [14] 韩立功, 李寅晓, 张晗, Peyman Milanfar, Dimitris Metaxas, 杨峰。Svdiff：用于扩散模型微调的紧凑参数空间。于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第7323–7334页, 2023年。2 [15] 何浩洋, 白宇虎, 张江宁, 何庆东, 陈鸿旭, 甘振业, 王成杰, 李祥泰, 田冠中, 谢磊。Mambaaad：探索状态空间模型在多类别无监督异常检测中的应用。Advances in Neural Information Processing Systems, 37:71162–71187, 2025年。1, 6, 7, 12 [16] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, Daniel Cohen-or。基于交叉注意力控制的提示到提示图像编辑。于 *International Conference on Learning Representations*, 2022年。3 [17] 胡腾, 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024年。1, 2, 6, 7, 12, 13, 17, 26, 35 [18] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, Onkar Dabeer。Winclip：零样本/少样本异常分类与分割。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第19606–19616页, 2023年。3 [19] 金晨, Ryutaro Tanno, Amrutha Saseendran, Tom Diethe, Philip Teare。一图值多词：使用多概念提示学习学习对象级概念。于 *International Conference on Machine Learning*, 2024年。2 [20] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, Jun-Yan Zhu。文本到图像扩散的多概念定制。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第1931–1941页, 2023年。2 [21] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第9664–9674页, 2021年。1, 2 [22] 戴庆李, 凌欢, Seung Wook Kim, Karsten Kreis, Sanja Fidler, Antonio Torralba。Bigdatasetgan：合成带有像素级标注的ImageNet。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第21330–21340页, 2022年。2 [23] 林东云, 曹彦鹏, 朱文斌, 李逸群。利用丰富正常样本通过正常背景正则化与裁剪粘贴操作的少样本缺陷分割。于 *2021 IEEE International Conference on Multimedia and Expo*, 第1–6页, 2021年。6, 7, 17

- [24] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 用于密集目标检测的Focal Loss。载于 *Proceedings of the IEEE International Conference on Computer Vision*, 第2980–2988页, 2017年。5[25] Ilya Loshchilov and Frank Hutter. 解耦权重衰减正则化。载于 *International Conference on Learning Representations*, 2018年。12[26] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. 用于多类无监督异常检测的分层向量量化Transformer。
Advances in Neural Information Processing Systems, 第36卷, 第8487–8500页, 2023年。1, 6, 7, 12[27] Quang Nguyen, Truong Vu, Anh Tran, and Khoi Nguyen. 数据集扩散：基于扩散的像素级语义分割合成数据生成。*Advances in Neural Information Processing Systems*, 第36卷, 2024年。3[28] Shuanglong Niu, Bin Li, Xinggang Wang, and Hui Lin. 使用GAN生成缺陷图像样本以改进缺陷识别。
IEEE Transactions on Automation Science and Engineering, 第17卷第3期, 第1611–1622页, 2020年。2, 6, 7, 17[29] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A. Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. 通过跨域对应关系进行少样本图像生成。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第10743–10752页, 2021年。6, 12[30] Patrick Pérez, Michel Gangnet, and Andrew Blake. 泊松图像编辑。载于 *ACM SIGGRAPH 2003*, 第313–318页。2003年。2[31] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 基于潜在扩散模型的高分辨率图像合成。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第10684–10695页, 2022年。2, 3, 6[32] Karsten Roth, Latha Pamula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. 迈向工业异常检测的完全召回。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第14318–14328页, 2022年。1, 6, 7, 12[33] Natael Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. DreamBooth：针对主体驱动生成的文本到图像扩散模型微调。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第22500–22510页, 2023年。2[34] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. 用于自监督异常检测与定位的自然合成异常。载于 *European Conference on Computer Vision*, 第474–489页。Springer, 2022年。2[35] Weijia Wu, Yuzhong Zhao, Hao Chen, Yuchao Gu, Rui Zhao, Yefei He, Hong Zhou, Mike Zheng Shou, and Chunhua Shen. DatasetDM：使用扩散模型合成带有感知注释的数据。*Advances in Neural Information Processing Systems*, 第36卷, 第54683–54695页, 2023年。3[36] Weijia Wu, Yuzhong Zhao, Mike Zheng Shou, Hong Zhou, and Chunhua Shen. DiffuMask：使用扩散模型合成带有像素级注释的图像以进行语义分割。载于 *Proceedings of the IEEE/CVF Interna-*

tional Conference on Computer Vision, 第1206–1217页, 2023年。3 [37] 肖光轩、尹天威、William T Freeman、Frédéric Durand 和韩松。Fastcomposer：基于局部注意力的免调用多主体图像生成。*arXiv preprint arXiv:2305.10431*, 2023年。2 [38] 肖特特、刘应成、周博磊、蒋宇宁和孙健。用于场景理解的统一感知解析。载于 *Proceedings of the European Conference on Computer Vision*, 第418–434页, 2018年。6, 7, 12, 26 [39] 谢金衡、李越翔、黄雅雯、刘浩哲、张文天、郑晔峰和Mike Zheng Shou。Boxdiff：基于免训练框约束扩散的文本到图像合成。载于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第7452–7461页, 2023年。3 [40] 于长谦、高长新、王敬波、于刚、沈春华和桑农。Bisenet v2：用于实时语义分割的引导聚合双边网络。
International Journal of Computer Vision, 第129卷: 3051–3068页, 2021年。6, 7, 12, 26 [41] Vitjan Zavrtanik、Matej Kristan 和 Danijel Skočaj。Draem- 一种用于表面异常检测的判别性训练重建嵌入。载于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第8330–8339页, 2021年。1, 2, 6, 7, 12, 35 [42] 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。载于 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 第2524–2534页, 2021年。2, 6, 7, 17 [43] 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。载于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第3836–3847页, 2023年。2 [44] 张宇轩、凌欢、高俊、尹康学、Jean-Francois Lafleche、Adela Barriuso、Antonio Torralba 和Sanja Fidler。Datasetgan：以最小人力实现高效标注数据工厂。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第10145–10155页, 2021年。2 [45] 周欢、薛峰、李雨聪、龚实、李逸群和周宇。利用低层表示进行超快速道路分割。
IEEE Transactions on Intelligent Transportation Systems, 2024年。6, 7, 12, 26 [46] 周启航、庞冠松、田宇、何世波和陈继明。Anomalyclip：面向零样本异常检测的对象无关提示学习。载于 *International Conference on Learning Representations*, 2024年。3 [47] 邹阳、Jongheon Jeong、Latha Pamula、张东庆和Onkar Dabeer。用于异常检测与分割的找差异自监督预训练。载于 *European Conference on Computer Vision*, 第392–408页。Springer, 2022年。6

A. 附录

A.1. 概述

本补充材料包括：

- 关于解耦异常对齐损失与多令牌的分析（附录A.2节）。
- 更多实现细节（见附录A.3节）。
- 下游监督分割模型实现与使用的更多细节（附录A.4节）。
- 更多消融研究（见附录A.5节），包括对非平衡异常文本提示设计、分离与共享微调损失、训练图像最小尺寸要求、SeaS训练策略、解耦异常对齐的交叉注意力图、粗粒度特征提取所用特征、用于精细化掩码预测的VAE特征、精细化掩码预测的正样本图像监督、掩码优化模块以及掩码二值化阈值的研究。
- 更多异常图像生成的定性和定量结果（见附录A.6）。
- 基于不同异常生成方法（第A.7节）生成的图像-掩码对训练的监督分割模型的定性比较结果。
- 在由SeaS（见附录A.8）生成的图像-掩码对数据上训练的不同监督分割模型的定性比较结果。
- 与文本反转（附录A.9）的比较。
- 更多关于光照条件的实验（见附录A.10）。
- 关于小缺陷生成的更多结果（见附录A.11）。
- 关于未见异常类型生成的更多分析（附录A.12）。
- 更多与DRAEM比较的实验（见附录A.13节）

A.2. 解耦异常对齐损失与多令牌分析

这里我们对DA损失的学习过程进行更详细的分析。根据公式3，直观上，DA损失可能会使异常标记彼此相似。然而，Stable Diffusion中的U-Net使用了多头注意力机制，这确保了不同的异常标记能覆盖异常的不同属性。在公式3中，交叉注意力图是U-Net特征图与异常标记的乘积。在多头注意力的实现中，异常标记的可学习嵌入和U-Net特征都沿着通道维度被分解为若干组。例如，对应异常标记的条件向量 $e_a \in \mathbb{R}^{1 \times C_1}$ 被划分为 $\{e_{a,i} \in \mathbb{R}^{1 \times \frac{C_1}{q}} | i \in [1, q]\}$ ，图像特征 $v \in \mathbb{R}^{r \times r \times C_2}$ 被划分为 $\{v_i \in \mathbb{R}^{1 \times \frac{C_2}{q}} | i \in [1, q]\}$ ，其中 q 是多头注意力中的头数。然后对应组相乘，最终输—

所有头部的输出被平均。 e_a 的注意力图 A 的计算方式为：
：

$$A = \frac{1}{q} \sum_{i=1}^q \text{softmax}\left(\frac{Q_i K_{a,i}^\top}{\sqrt{d}}\right), Q_i = \phi_q(v_i), K_{a,i} = \phi_k(e_{a,i}). \quad (8)$$

因此，在缺陷区域，DA损失仅确保每个头部的平均值趋近于1，但并未要求异常标记彼此相同。此外，每个 e_a 各不相同，并通过 $e_{a,i}$ 组合。每个 $e_{a,i}$ 的更新方向与 v_i 相关，并覆盖了缺陷的部分特征，它从多个角度囊括了异常属性，从而提供了多样化的信息。

我们在图8中提供了更多示例，其中生成的新异常在颜色和形状上与训练样本存在显著差异。例如，我们展示了具有新颖形状的*bottle_contamination*、*hazelnut_print*和*tile_gray_stroke*，具有新颖颜色的*wood_color*和*metal_nut_scratch*，以及具有新形状的*pill_crack*——其呈现多处裂纹，而训练样本仅显示单一裂纹。这些示例证明了模型能够通过重组解耦属性来创建未见过的异常。

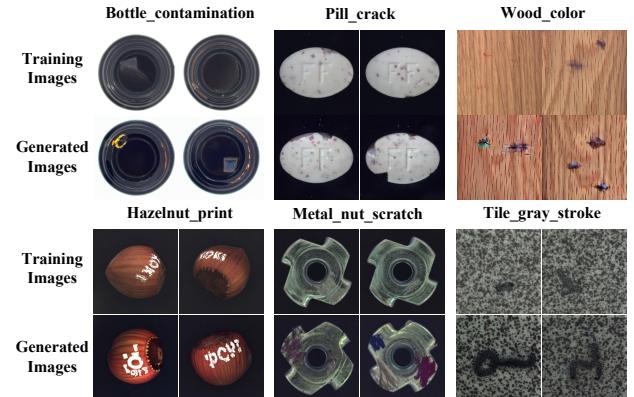


图8. MVTec AD上未见异常生成结果的可视化。

A.3. 更多实现细节

更多训练细节。对于不平衡异常文本提示，我们将多重 $\langle df_n \rangle$ 的数量 N 设为4，并将 $\langle ob \rangle$ 的数量 N' 设为1，这些参数在所有产品类别中均保持不变。针对特定类型的异常，我们使用带有不同异常标记集的不平衡异常（UA）文本提示 \mathcal{P}_n 作为条件，以生成指定类型的异常。

$$\mathcal{P}_n = a \langle ob \rangle \text{ with } \langle df_{4 \times n-3} \rangle, \langle df_{4 \times n-2} \rangle, \langle df_{4 \times n-1} \rangle, \langle df_{4 \times n} \rangle$$

其中 n 代表产品中异常类型的索引。为生成正常图像，我们使用与 P 的正常标记相对应的嵌入 e_{ob} ，即“a <ob>”，来引导U-Net预测噪声。例如，对于正常标记<ob>，给定查找表 $\mathcal{U} \in \mathbb{R}^{b \times 768}$ （其中 b 是预训练文本编码器存储的文本嵌入数量），我们使用占位符字符串“ob1”作为输入。首先，“ob1”在分词器中被转换为标记ID $s_{ob1} \in \mathbb{R}^{1 \times 1}$ ；其次， $s_{ob1} \in \mathbb{R}^{1 \times 1}$ 被转换为独热向量 $\mathcal{S}_{ob1} \in \mathbb{R}^{1 \times (b+1)}$ ；第三，将与 s_{ob1} 对应的一个可学习新嵌入 $g \in \mathbb{R}^{1 \times 768}$ 插入查找表 \mathcal{U} ，得到 $\mathcal{U}' \in \mathbb{R}^{(b+1) \times 768}$ 。此处 $g \in \mathbb{R}^{1 \times 768}$ 是<ob>的可学习嵌入。这些嵌入和U-Net在微调过程中均为可学习的。

训练图像生成模型。对于每个产品，我们执行 $800 \times G$ 步进行微调，其中 G 代表产品的异常类别数量。训练图像生成模型的批次大小设置为4。在微调过程的每一步中，我们从异常训练集 X_{df} 中采样2张图像，并从正常训练集 X_{ob} 中采样2张图像。我们使用AdamW [25]优化器，U-Net的学习率为 4×10^{-6} 。文本嵌入的学习率为 4×10^{-5} 。

训练精炼掩码预测分支。我们设计了一个级联的精炼掩码预测（RMP）分支，该分支被嫁接在根据SeaS训练的U-Net上。对于每个产品，我们为RMP模型执行 $80 \times G$ 步训练，其中 G 代表该产品的异常类型数量。训练RMP分支的批量大小设置为4。在微调过程的每一步中，我们从异常训练集 X_{df} 中采样2张图像及其对应的掩码，并从正常训练集 X_{ob} 中采样2张图像。用于抑制正常图像噪声的掩码的每个像素值均设置为0。RMP模型的学习率为 5×10^{-4} 。

更多推理细节。对于所有实验，我们使用 $t = 1500$ 对正常图像进行扩散前向处理以获取初始噪声。采样时我们采用 $T = 25$ 步。

指标。对于异常图像生成，我们报告了4项指标：使用初始分数（IS）和簇内成对LPIPS距离（IC-LPIPS）来评估异常图像，采用KID[5]来评估正常图像的真实性，以及仅在异常区域计算的IC-LPIPS（简称IC-LPIPS(a)）来评估多样性。初始分数（IS）由[2]提出，通过测量输入样本与其预测类别之间的互信息，作为评估生成图像保真度和多样性的独立指标。IC-LPIPS[29]用于评估生成图像的多样性，它量化了同一簇中图像块之间的感知相似性。对于像素级异常分割和图像级异常检测，我们报告了3项指标：面积

表7. 资源需求与时间消耗对比。

Methods	Training	Inference
	Overall Time	Time (per image)
DFMGAN[11]	414 hours	48 ms
AnomalyDiffusion[17]	249 hours	3830 ms
SeaS	73 hours	720 ms

在接收者操作特征曲线下面积（AUROC）、平均精度（AP）以及最优阈值下的 F_1 分数（ F_1 -max）等指标上。所有这些指标均使用`scikit-learn`库进行计算。此外，我们还计算了交并比（IoU）以更精确地评估异常分割结果。

更多关于异常检测方法的训练细节。在本节中，我们提供了本文表2和表3中对比异常检测方法的更多训练细节。对于DRAEM [41]、GLASS [9]和HVQ-Trans [26]，我们使用其在MVTec AD数据集上的官方检查点，而其他方法由于缺乏官方检查点均为自行训练。对于GLASS，由于VisA和MVTec 3D AD数据集的官方前景掩码不可用，因此未使用此操作。对于PatchCore [32]，我们采用256的图像尺寸且不进行中心裁剪，因为部分异常出现在边缘区域。对于MambaAD [15]，我们使用其提供的官方检查点。

资源需求与时间消耗。我们在NVIDIA Tesla A100 40G GPU上按产品类别顺序进行训练，每类训练约占用20G内存。时间消耗的对比如表7所示。在MVTec AD数据集上，我们的训练耗时73小时，短于AnomalyDiffusion所需的249小时和DFMGAN所需的414小时。在推理时间方面，SeaS每张图像耗时720毫秒，短于基于扩散模型的方法AnomalyDiffusion所需的每张图像3830毫秒。基于GAN的方法DFMGAN的推理时间为每张图像48毫秒。

A.4. 监督式分割模型的更多细节

如实验部分所述，我们选择了三种有监督分割模型（BiSeNet V2 [40]、UPerNet [38]、LFD [45]），以验证生成的图像-掩码对在下游有监督异常分割及检测任务中的有效性。对于BiSeNet V2和UPerNet，我们基本遵循MMsegmentation提供的实现方案；对于LFD，我们也采用了官方实现。

具体而言，对于BiSeNet V2，我们选择了一个骨干网络

一个细节分支的结构，包含三个阶段，通道数分别为6、64和128；以及一个语义分支，包含四个阶段，通道数分别为16、32、64和128，并配备一个解码头部和四个辅助头部（对应语义分支的阶段数）。至于UPerNet，我们选择ResNet-50作为主干网络，配备一个解码头部和一个辅助头部。

在训练用于下游任务的监督分割模型时，我们采用了为所有产品类别训练统一监督分割模型的策略，而非为每个类别单独训练监督分割模型。实验结果如表8所示，表明统一监督分割模型的性能超越了多个独立监督分割模型的表现。

表8. 监督式分割模型训练策略的消融实验。

Models	Multiple Models				Unified Model			
	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU
BiSeNet V2	96.00	67.68	65.87	54.11	97.21	69.21	66.37	55.28
UPerNet	96.77	73.88	70.49	60.37	97.87	74.42	70.70	61.24
LFD	93.02	72.97	71.56	55.88	98.09	77.15	72.52	56.47
Average	95.26	71.51	69.31	56.79	97.72	73.59	69.86	57.66

A.5. 更多消融实验

关于不平衡异常文本提示设计的消融研究

在工业异常图像生成的提示设计中，我们通过实验验证了针对每种产品各异常类型的不平衡异常（UA）文本提示的有效性。我们将可学习的 $\langle df_n \rangle$ 数量设置为 N ，并将可学习的 $\langle ob_j \rangle$ 数量设置为 N' 。如表9所示，通过使用UA文本提示，即

\mathcal{P} = 一个具有 $\langle ob \rangle$ 和 $\langle df_1 \rangle$ 、 $\langle df_2 \rangle$ 、 $\langle df_3 \rangle$ 、 $\langle df_4 \rangle$ 的模型，我们能够为下游有监督异常分割任务提供高保真且多样化的图像，从而在分割指标上取得最佳性能。

分离与共享微调损失的消融研究

在分离与共享微调的DA损失和NA损失设计中，我们进行了两组实验：(a) 移除DA损失中的第二项（表10中简写为w/o ST）；(b) 将DA损失中的第二项替换为NA损失中的另一项（表10中简写为with AT），该项根据掩码将背景区域与标记 $\langle ob \rangle$ 对齐：

$$\mathcal{L}_{ob} = \sum_{l=1}^L (\|A_{ob}^l - (1 - M^l)\|^2) + \|\epsilon_{ob} - \epsilon_\theta(\hat{z}_{ob}, t_{ob}, \mathbf{e}_{ob})\|^2 \quad (9)$$

其中 $A_{ob}^l \in \mathbb{R}^{r \times r \times 1}$ 是与正常标记 $\langle ob \rangle$ 对应的交叉注意力图。如表10所示，实验结果表明，我们采用的损失设计在下游监督分割任务中取得了最佳性能。

表9. 不平衡异常文本提示设计的消融实验。

Prompt	AUROC	AP	F_1 -max	IoU
$N' = 1, N = 1$	96.48	63.69	62.50	52.02
$N' = 1, N = 4$ (Ours)	97.21	69.21	66.37	55.28
$N' = 4, N = 4$	96.55	66.28	63.95	54.07

表10. 分离与共享微调损失的消融实验。

Loss	AUROC	AP	F_1 -max	IoU
w/o ST	96.44	67.73	65.23	54.99
with AT	96.42	63.99	62.43	53.36
Ours	97.21	69.21	66.37	55.28

关于训练图像最小尺寸要求的消融实验

在少样本设置中，为公平比较，我们遵循DFMGAN [11]和AnomalyDiffusion [17]中的通用设置，即在训练中对每种异常类型使用三分之一的异常图像-掩码对。在此设置下，异常训练图像的最小数量为2。若采用3样本设置，则需重新组织测试集。为确保测试集不被重组以实现公平比较，我们在训练中对所有异常类型采用1样本和2样本设置，即 $H = 1$ 和 $H = 2$ ，其中 H 为图像数量。结果如表11和图9所示。显然，通过1样本和2样本设置训练的模型仍能生成具有良好多样性和真实性的异常图像。

表11. 训练图像最小尺寸要求的消融实验。

Size	IS	IC-L
$H = 1$	1.790	0.311
$H = 2$	1.794	0.314
$H = \frac{1}{3} \times H_0$	1.876	0.339

关于SeaS训练策略的消融实验

在微调过程的每一步中，我们从异常训练集 X_{df} 和正常训练集 X_{ob} 中采样相同数量的图像。为了探究该策略的有效性，我们进行了三组不同的实

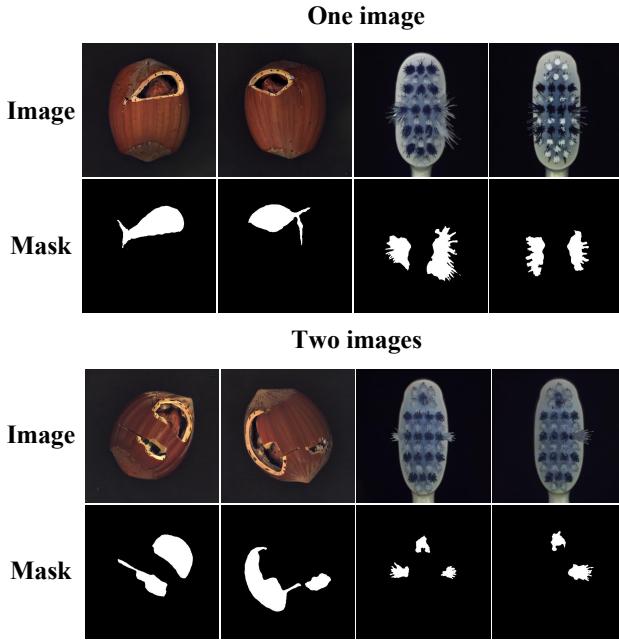


图9. 关于训练图像最小尺寸要求的消融研究可视化。图中第一行为生成图像，第二行为生成掩码。

实验设置如下：(a) 优先使用异常图像训练，随后使用正常图像训练（表12中简称为异常-正常）；(b) 优先使用正常图像训练，随后使用异常图像训练（表12中简称为正常-异常）；(c) 在每个批次中混合使用正常与异常图像进行训练（表12中简称为异常&正常）。如表12所示，SeaS在异常图像生成方面表现出优越性能，其生成图像兼具高保真度与多样性。

表12. SeaS训练策略的消融实验。

Strategy	IS	IC-L
Abnormal-Normal	1.53	0.28
Normal-Abnormal	1.70	0.32
Abnormal&Normal (Ours)	1.88	0.34

针对解耦异常对齐的交叉注意力图消融研究

在解耦异常对齐 (DA) 损失中，我们利用了U-Net编码器各层的交叉注意力图。具体而言，我们研究了整合不同交叉注意力图（记为 $A^1 \in \mathbb{R}^{64 \times 64}$ 、 $A^2 \in \mathbb{R}^{32 \times 32}$ 、 $A^3 \in \mathbb{R}^{16 \times 16}$ 和 $A^4 \in \mathbb{R}^{8 \times 8}$ ）的影响。这些分别对应“down-1”、“down-2”、“down-3”和“down-4”层的交叉注意力图。

编码器在U-Net中的层数。如表13所示，实验结果表明，采用 $\{A^2, A^3\}$ 组合作为DA损失，在下游监督分割任务中取得了最佳性能。

表13. 解耦异常对齐中交叉注意力图的消融研究。

A^l	AUROC	AP	F_1 -max	IoU
$l = 1, 2, 3$	96.42	68.92	66.24	54.52
$l = 2, 3, 4$	95.71	64.51	62.33	52.46
$l = 2, 3$ (Ours)	97.21	69.21	66.37	55.28

关于粗粒度特征提取特征的消融研究

在粗糙特征提取过程中，我们从U-Net解码器中提取用于异常检测的粗糙但高区分度的特征。具体而言，我们研究了整合不同特征（记为 $F_1 \in \mathbb{R}^{16 \times 16 \times 1280}$ 、 $F_2 \in \mathbb{R}^{32 \times 32 \times 1280}$ 、 $F_3 \in \mathbb{R}^{64 \times 64 \times 640}$ 和 $F_4 \in \mathbb{R}^{64 \times 64 \times 320}$ ）的影响。这些特征分别对应U-Net编码器中“up-1”、“up-2”、“up-3”和“up-4”层的输出特征。

如图10所示，我们利用U-Net解码器中“up-2”和“up-3”层的输出特征，通过卷积块和拼接操作，得到统一的粗粒度特征 $\hat{F} \in \mathbb{R}^{64 \times 64 \times 192}$ ，该特征可用于预测异常图像对应的掩码。如表14所示，实验结果表明，采用 $\{F_2, F_3\}$ 组合进行粗粒度特征提取，在下游监督分割任务中取得了最佳性能。

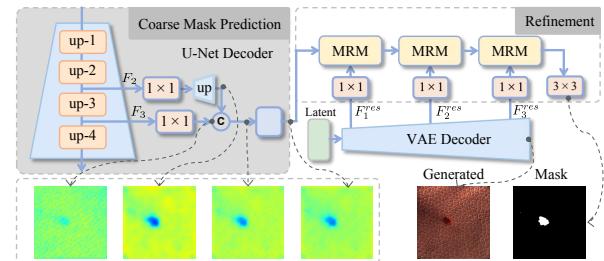


图10. U-Net解码器特征在掩码预测过程中的可视化。

表14. 用于粗粒度特征提取的特征消融实验。

F_y	AUROC	AP	F_1 -max	IoU
$y = 1, 2, 3$	94.35	63.58	60.54	52.36
$y = 2, 3, 4$	96.93	67.42	64.26	55.31
$y = 2, 3$ (Ours)	97.21	69.21	66.37	55.28

针对精细化掩码预测的VAE特征消融研究

在精细化掩码预测中，我们将VAE解码器的高分辨率特征与U-Net的判别性特征相结合，以生成精确对齐的异常图像-掩码对。此外，我们也可以使用VAE编码器特征作为高分辨率特征。如表15所示，实验结果表明，使用VAE解码器特征在下游监督分割任务中能获得更优的性能。

表15. 用于精细化掩码预测的VAE特征消融实验。

F^{res}	AUROC	AP	F_1 -max	IoU
VAE encoder	96.14	66.26	63.48	54.99
VAE decoder	97.21	69.21	66.37	55.28

关于精炼掩码预测中正常图像监督的消融研究

在精细化掩码预测分支中，我们为正常图像预测掩码，作为掩码预测的监督信号。我们进行了两组实验：(a) 移除RMP损失函数中的第二项和第四项，即正常图像监督（表16中简称为NIA）；(b) 使用RMP分支损失函数的完整形式，即采用正常图像进行监督，如公式(10)所示：

$$\begin{aligned} \mathcal{L}_{\mathcal{M}} = & \mathcal{F}(\hat{M}_{df}, \mathbf{M}_{df}) + \mathcal{F}(\hat{M}_{ob}, \mathbf{M}_{ob}) \\ & + \mathcal{F}(\hat{M}'_{df}, \mathbf{M}'_{df}) + \mathcal{F}(\hat{M}'_{ob}, \mathbf{M}'_{ob}) \end{aligned} \quad (10)$$

如表16所示，实验结果表明，使用法线图像进行监督在下游监督分割任务中取得了更好的性能。我们还进一步提供了法线图像监督（图11中简称为NIA）对MVTec AD影响的定性结果。

表16. 关于精修掩码预测中正常图像监督的消融实验。

F^{res}	AUROC	AP	F_1 -max	IoU
w/o NIA	96.20	66.03	64.09	53.97
with NIA (Ours)	97.21	69.21	66.37	55.28

掩码细化模块的消融研究

在精细化掩码预测分支中，我们采用掩码优化模块(MRM)来生成精细化掩码。如图中所示，我们为MRM设计了不同的结构。

12种情况，包括案例a)：无卷积块的情况，案例b)：含一个卷积块的情况，以及案例c)：含链式卷积块的情况。如图13所示，我们发现使用案例b)中的卷积块——该结构由两个 1×1 卷积和一个 3×3 卷积组成——能帮助模型更准确地学习缺陷区域的特征，而非像案例a)中仅使用单一卷积时那样过度关注背景区域。基于此观察，我们进一步设计了案例c)中的链式卷积块结构，所提取的特征能更好地反映缺陷区域。这种逐级递进的残差学习有助于模型对缺陷区域特征实现更优的残差校正效果。如附录表17所示，与案例b)相比，案例c)在AUR OC指标上提升0.28%，在AP指标上提升2.29%，在F-max指标上提升2.29%，在IoU指标上提升0.32%。我们通过下游监督分割实验的结果，证实了所设计的MRM结构的优越性。

表17. 掩码优化模块的消融实验。

Model	AUROC	AP	F_1 -max	IoU
with MRM (a)	96.75	68.18	64.96	55.51
with MRM (b)	96.93	66.92	64.08	54.96
with MRM (c)	97.21	69.21	66.37	55.28

掩膜二值化阈值的消融研究

在精细化掩码预测分支中，我们采用阈值 τ 对精细化异常掩码 \hat{M}'_{df} 的第二通道进行分割，以生成最终异常掩码。我们使用 τ 设置在0.1至0.5范围内的异常掩码训练有监督分割模型。如表18所示，结果表明将 τ 设为0.2时模型性能达到最优。

A.6. 更多定性与定量异常图像生成结果

更详细的定量结果 在本节中，我们报告MVTec AD数据集、VisA数据集和MVTec 3D AD数据集中每个类别的详细生成结果，这些结果分别展示在表19、表20和表21中。

更多定性生成结果

我们在MVTec AD数据集上提供了从图15到图16每个类别的进一步定性结果。我们展示了SeaS针对不同类型异常生成的异常图像结果。第一列为生成的异常图像，第二列为对应的生成掩码，第三列为未使用掩码优化模块生成的掩码。

我们提供每个类别的进一步定性结果

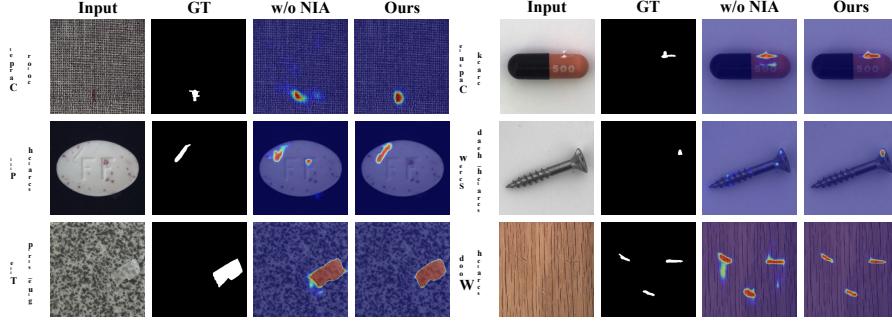


图11. 正常图像监督对MVTec AD影响的定性结果。

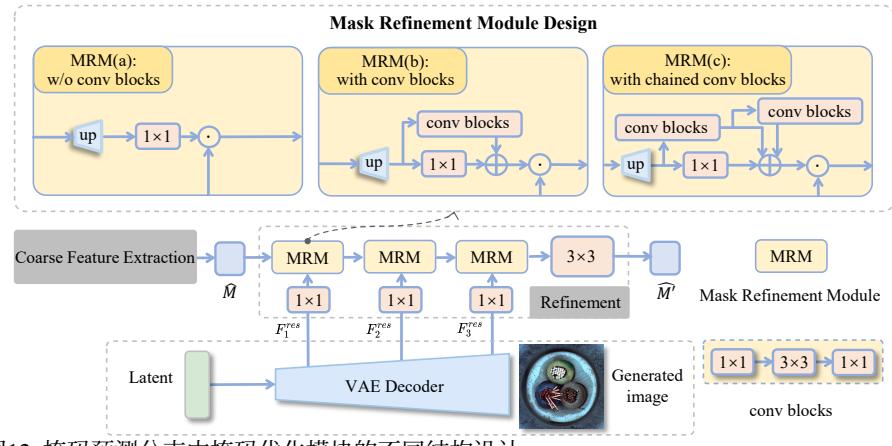


图12. 掩码预测分支中掩码优化模块的不同结构设计。

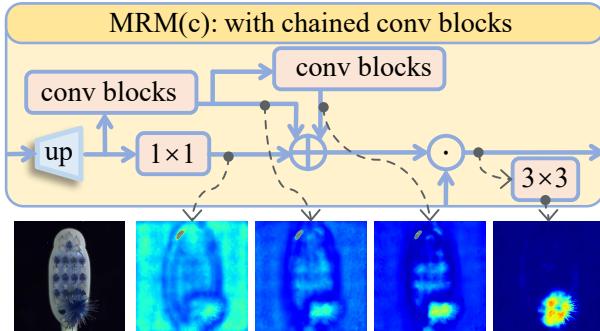


图13. MRM模块中间结果的可视化。顶部为MRM结构示意图，底部依次为输入图像、MRM中间过程的特征图及预测掩码。

在图17的MVTec 3D AD数据集上，我们展示了SeaS针对不同类型异常生成的异常图像结果。第一列为生成的异常图像，第二列为对应的生成掩码。

表18. 关于掩码二值化阈值的消融实验。

threshold	AUROC	AP	F_1 -max	IoU
$\tau = 0.1$	97.56	65.33	63.38	52.40
$\tau = 0.2$ (Ours)	97.21	69.21	66.37	55.28
$\tau = 0.3$	97.20	66.92	64.35	54.68
$\tau = 0.4$	95.31	63.55	61.97	53.03
$\tau = 0.5$	94.11	60.85	59.92	50.87

A.7. 基于不同异常生成方法生成的图像-掩码对训练的有监督分割模型的更多定性与定量比较结果

我们在MVTec AD数据集上提供了采用不同异常生成方法的进一步定性结果。我们展示了SeaS在各类别中针对不同类型异常的生成结果，相关结果见图18至图21。

我们在生成数据训练的下游监督分割任务上提供了进一步的定性比较。

表19. MVTec AD数据集上IS和IC-LPIPS的对比结果。**粗体**表示最佳性能，下划线表示次优性能。
结果。

Category	Crop& Paste [23]		SDGAN [28]		Defect-GAN [42]		DFMGAN [11]		Anomaly Diffusion [17]		Ours	
	IS ↑	IC-L ↑	IS ↑	IC-L ↑	IS ↑	IC-L ↑	IS ↑	IC-L ↑	IS ↑	IC-L ↑	IS ↑	IC-L ↑
bottle	1.43	0.04	1.57	0.06	1.39	0.07	<u>1.62</u>	0.12	1.58	<u>0.19</u>	1.78	0.21
cable	1.74	0.25	1.89	0.19	1.70	0.22	1.96	0.25	2.13	<u>0.41</u>	<u>2.09</u>	0.42
capsule	1.23	0.05	1.49	0.03	1.59	0.04	1.59	0.11	1.59	<u>0.21</u>	<u>1.56</u>	0.26
carpet	1.17	0.11	1.18	0.11	1.24	0.12	<u>1.23</u>	0.13	1.16	<u>0.24</u>	1.13	0.25
grid	2.00	0.12	1.95	0.10	2.01	0.12	1.97	<u>0.13</u>	<u>2.04</u>	0.44	2.43	0.44
hazelnut	1.74	0.21	1.85	0.16	1.87	0.19	<u>1.93</u>	<u>0.24</u>	2.13	0.31	1.87	0.31
leather	1.47	0.14	2.04	0.12	2.12	0.14	<u>2.06</u>	0.17	1.94	0.41	2.03	<u>0.40</u>
metal_nut	1.56	0.15	1.45	0.28	1.47	0.30	1.49	0.32	1.96	0.30	<u>1.64</u>	<u>0.31</u>
pill	1.49	0.11	1.61	0.07	1.61	0.10	1.63	0.16	1.61	<u>0.26</u>	<u>1.62</u>	0.33
screw	1.12	0.16	1.17	0.10	1.19	0.12	1.12	0.14	<u>1.28</u>	<u>0.30</u>	1.52	0.31
tile	1.83	0.20	2.53	0.21	2.35	0.22	2.39	0.22	<u>2.54</u>	0.55	2.60	<u>0.50</u>
toothbrush	1.30	0.08	1.78	0.03	<u>1.85</u>	0.03	1.82	0.18	1.68	<u>0.21</u>	1.96	0.25
transistor	1.39	0.15	1.76	0.13	1.47	0.13	<u>1.64</u>	<u>0.25</u>	1.57	0.34	1.51	<u>0.34</u>
wood	1.95	0.23	2.12	0.25	2.19	0.29	2.12	0.35	<u>2.33</u>	<u>0.37</u>	2.77	0.46
zipper	1.23	0.11	1.25	0.10	1.25	0.10	1.29	<u>0.27</u>	<u>1.39</u>	0.25	1.63	0.30
Average	1.51	0.14	1.71	0.13	1.69	0.15	1.72	0.20	<u>1.80</u>	<u>0.32</u>	1.88	0.34

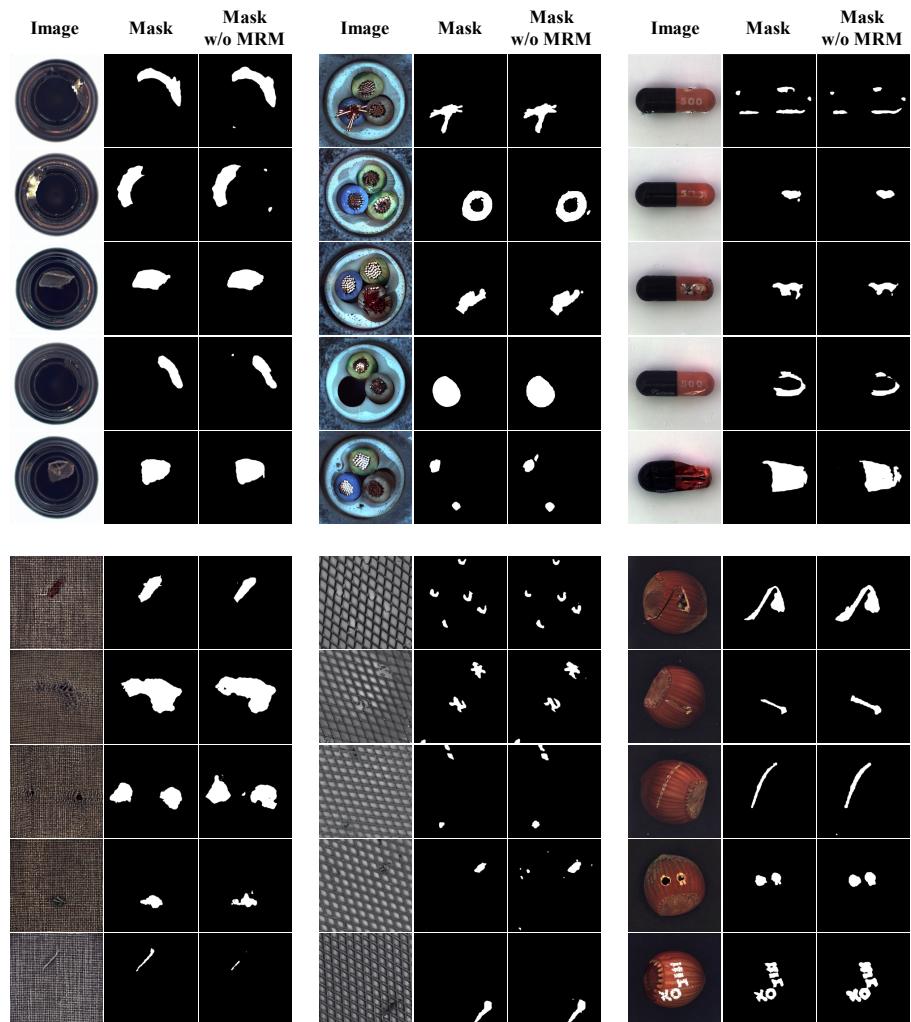
表20. VisA数据集上IS与IC-LPIPS指标对比结果。**粗体**表示最佳性能。

图像。分割异常图如图22所示。当BiSeNet V2使用我们方法生成的图像-掩码对进行训练时，误报（例如 *potato_combined*）和漏报（例如 *bagel_contamination*）均有所减少。

Category	DFMGAN [11]		AnomalyDiffusion [17]		Ours	
	IS ↑	IC-L ↑	IS ↑	IC-L ↑	IS ↑	IC-L ↑
candle	1.19	0.23	1.28	0.17	1.20	0.12
capsules	1.25	0.22	1.39	0.50	1.58	0.60
cashew	1.25	0.24	1.27	0.26	1.21	0.28
chewinggum	1.33	0.24	1.15	0.19	1.29	0.27
fryum	1.28	0.20	1.20	0.14	1.14	0.21
macaroni1	1.14	0.24	1.15	0.14	1.15	0.18
macaroni2	1.47	0.38	1.56	0.38	1.57	0.39
pcb1	1.12	0.16	1.18	0.35	1.18	0.26
pcb2	1.12	0.26	1.26	0.21	1.25	0.27
pcb3	1.19	0.18	1.21	0.24	1.22	0.21
pcb4	1.21	0.28	1.14	0.25	1.15	0.22
pipe_fryum	1.43	0.32	1.29	0.17	1.31	0.16
Average	1.25	0.25	1.26	0.25	1.27	0.26

表21. MVTec 3D AD数据集上IS与IC-LPIPS指标对比。**粗体**表示最佳性能。

Category	DFMGAN [11]		AnomalyDiffusion [17]		Ours	
	IS ↑	IC-L ↑	IS ↑	IC-L ↑	IS ↑	IC-L ↑
bagel	1.07	0.26	1.02	0.22	1.28	0.29
cable_gland	1.59	0.25	1.79	0.19	2.21	0.19
carrot	1.94	0.29	1.66	0.17	2.07	0.22
cookie	1.80	0.31	1.77	0.29	2.07	0.38
dowel	1.96	0.37	1.60	0.20	1.95	0.26
foam	1.50	0.17	1.77	0.30	2.20	0.39
peach	2.11	0.34	1.91	0.23	2.40	0.28
potato	3.05	0.35	1.92	0.17	1.98	0.22
rope	1.46	0.29	1.28	0.25	1.53	0.41
tire	1.53	0.25	1.35	0.20	1.81	0.31
Average	1.80	0.29	1.61	0.22	1.95	0.30



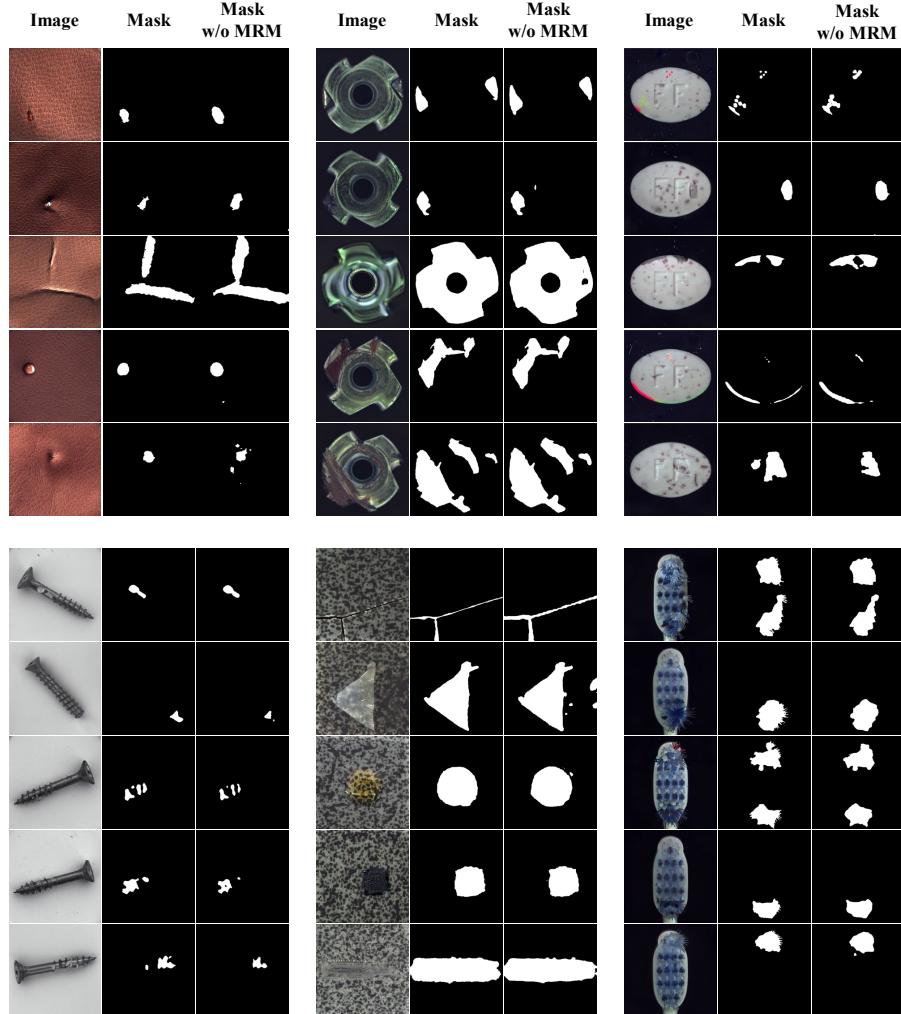


图15. 我们在MVTec AD数据集上的异常图像生成定性结果。第一行从左至右依次展示了结果f
leather、metal_nut和pill类别。第二行从左到右依次是screw、tile和toothbrush类别的结果。

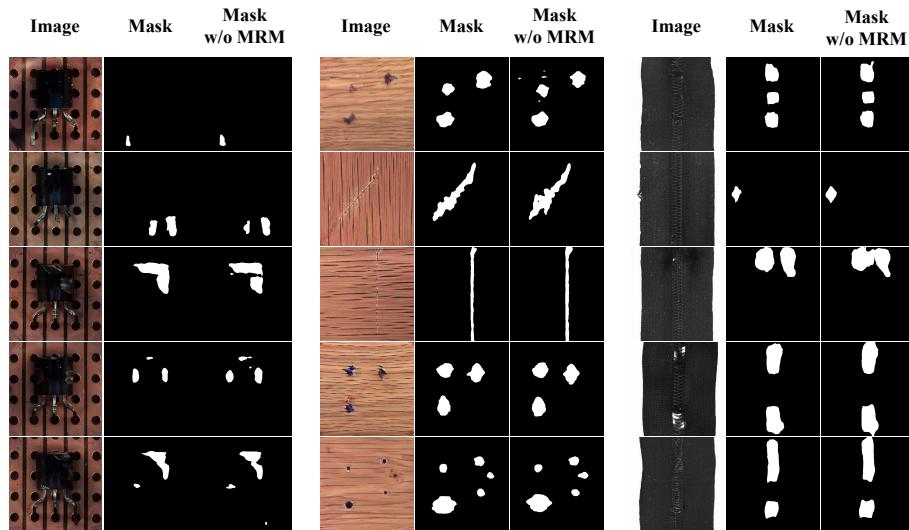


图16. 我们在MVTec AD数据集上的异常图像生成定性结果。第一行从左至右依次为transistor、wood和zipper类别的生成结果。

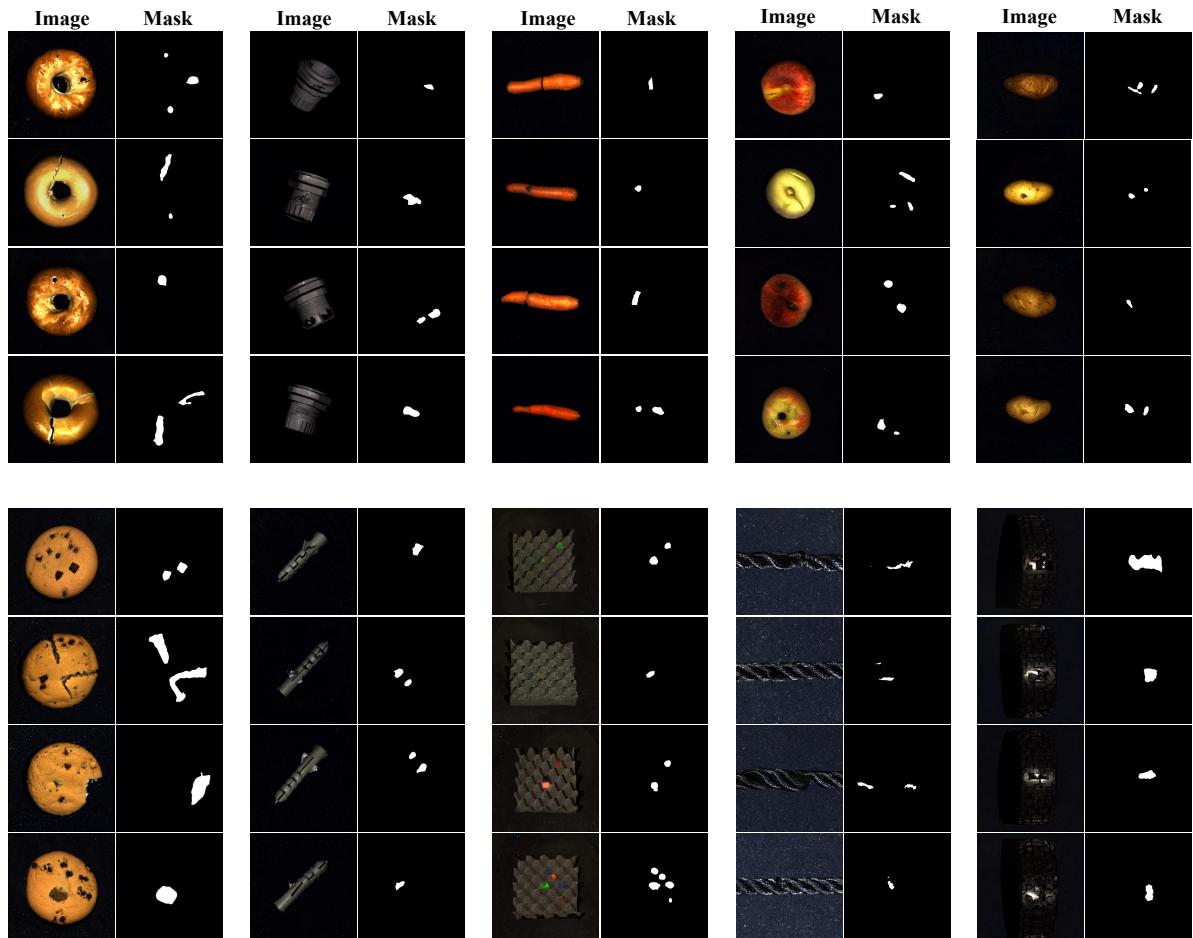


图17. 我们在MVTec 3D AD数据集上的异常图像生成定性结果。第一行从左至右依次为`bagel`、`cable_gland`、`carrot`、`peach`和`potato`类别的结果。第二行从左至右依次为`cookie`、`dowel`、`foam`、`rope`和`tire`类别的结果。

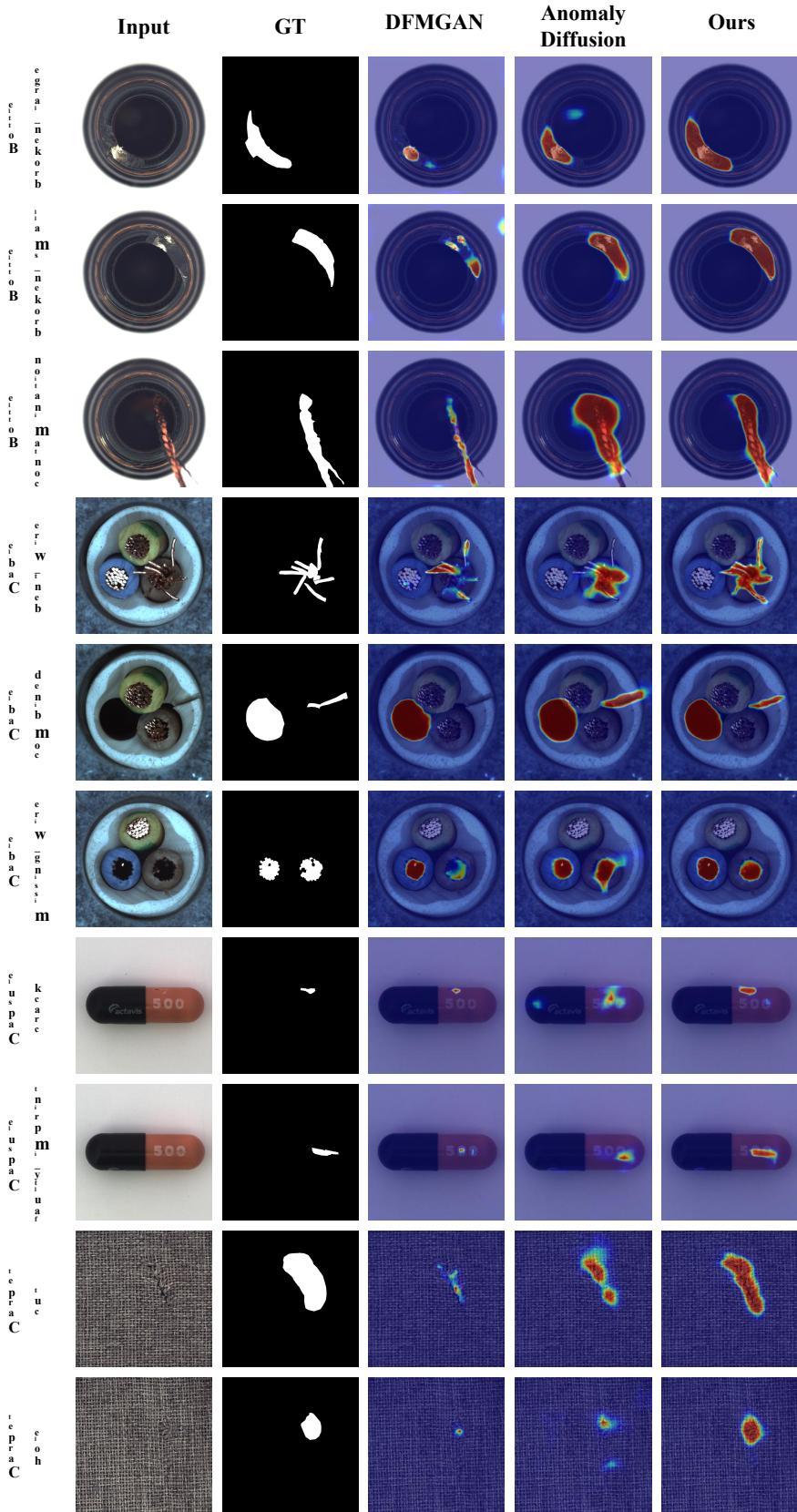


图18. 在MVTec AD上与异常监督分割模型BiSeNet V2的对比结果。图中从上到下依次为{v*}、cable、capsule和carpet类别的结果。

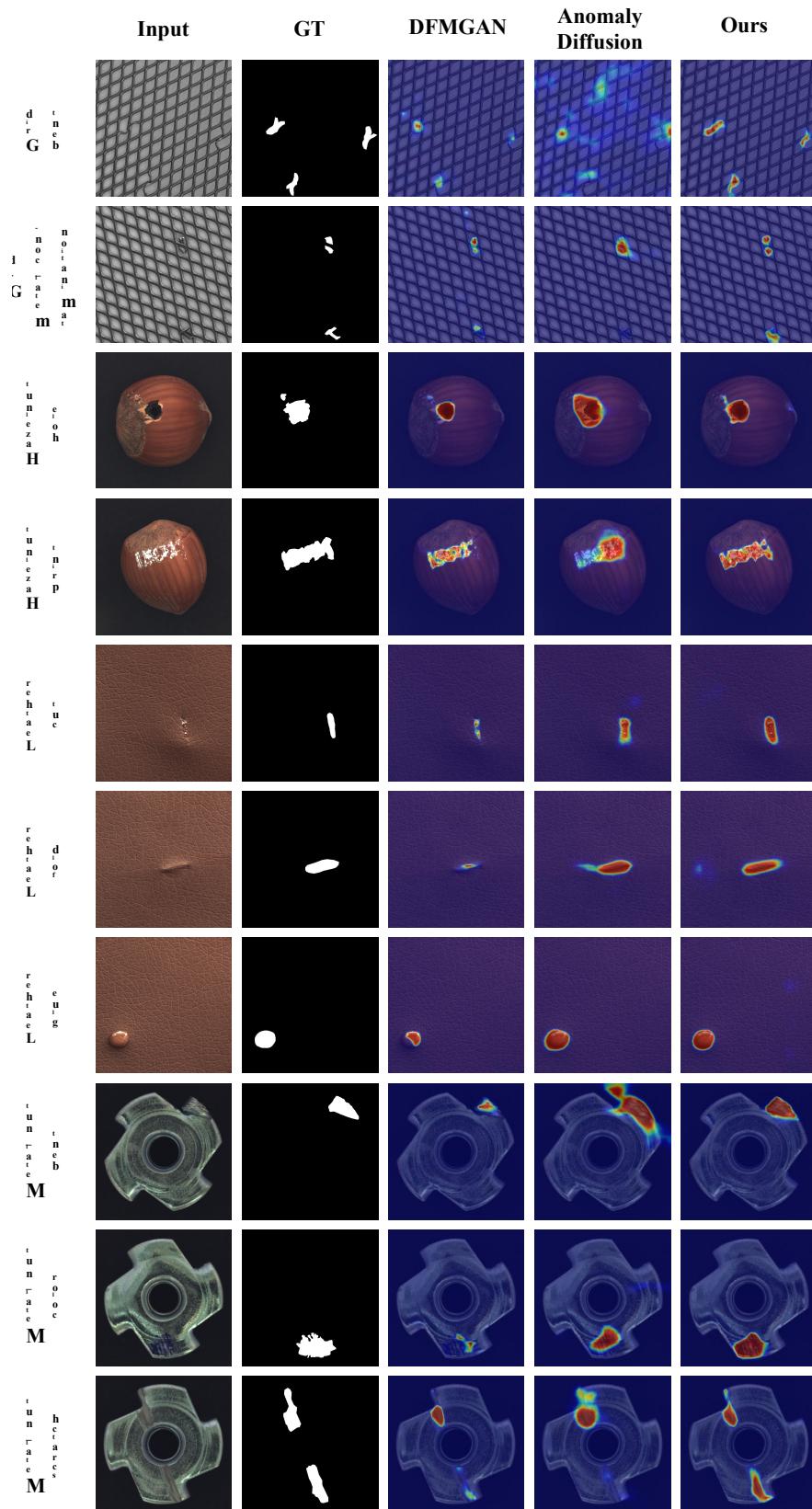


图19. 在MVTec AD上与异常监督分割模型BiSeNet V2的对比结果。图中从上至下依次为{v*}、*hazelnut*、*leather*和*metal_nut*类别的结果。

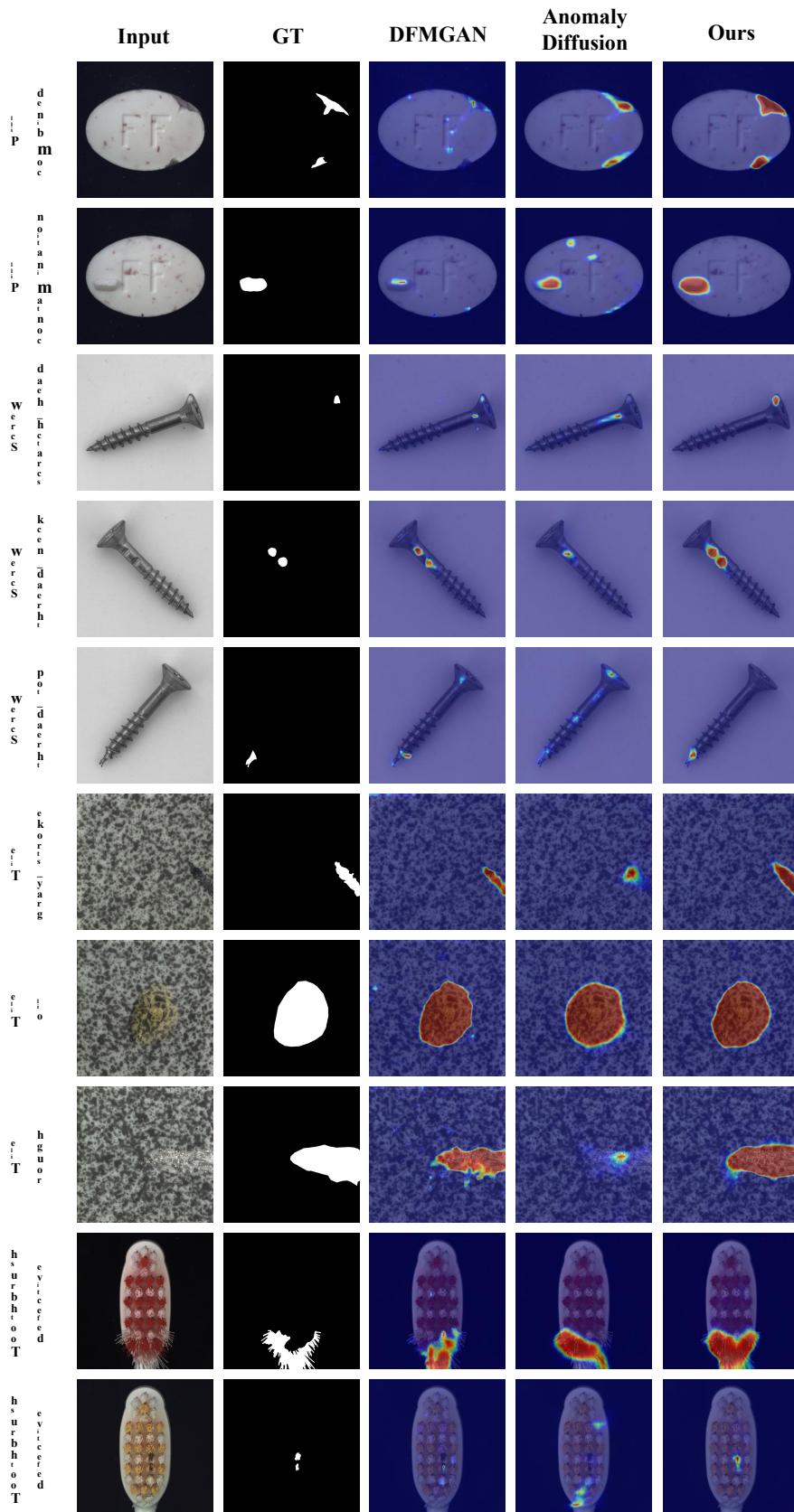


图20. 在MVTec AD上与异常监督分割模型BiSeNet V2的对比结果。图中从上到下依次为pill、screw、tile和toothbrush类别的结果。

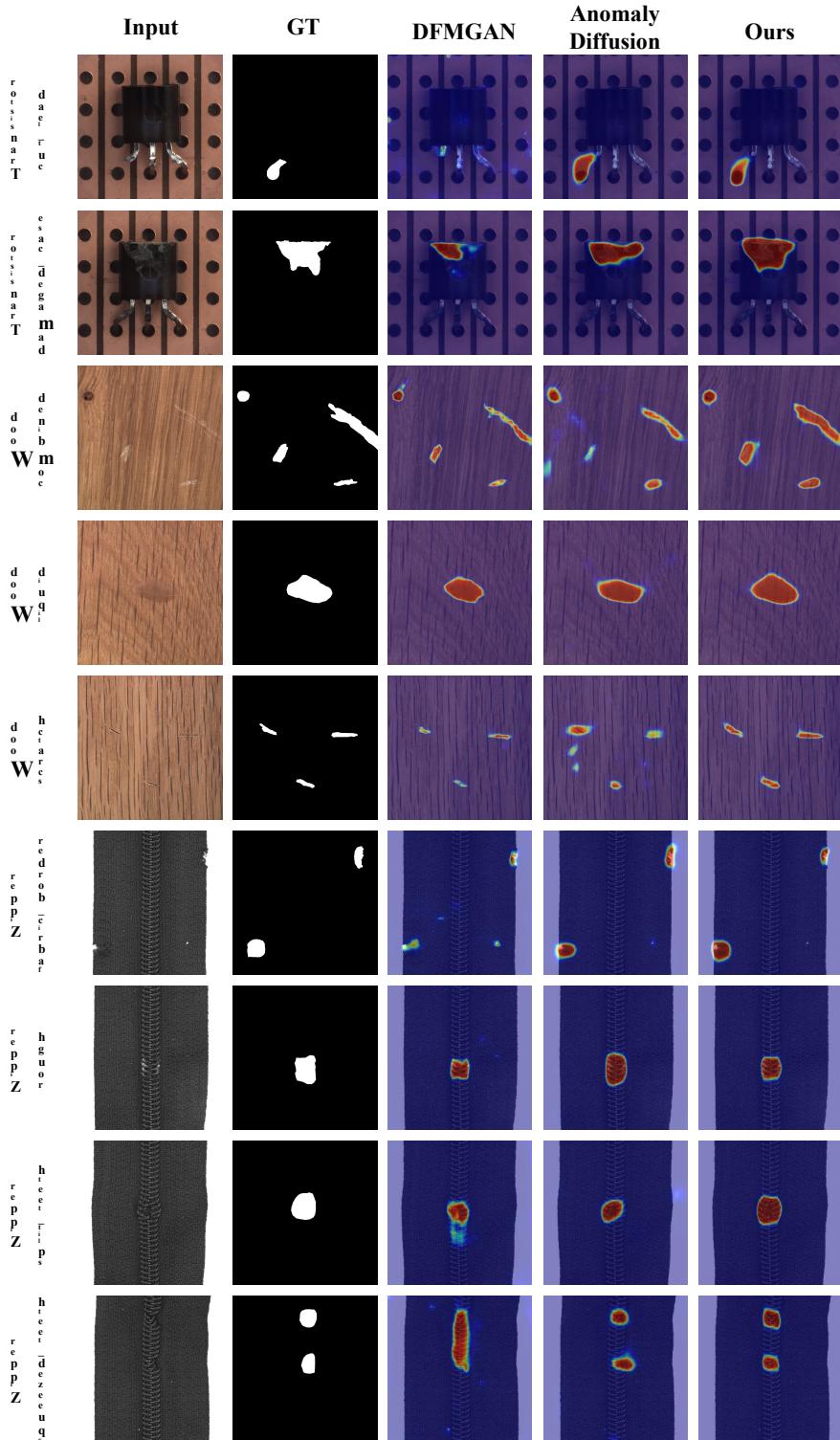


图21. 在MVTec AD上与异常监督分割模型BiSeNet V2的对比结果。图中从上到下依次为transistor、wood和zipper类别的结果。

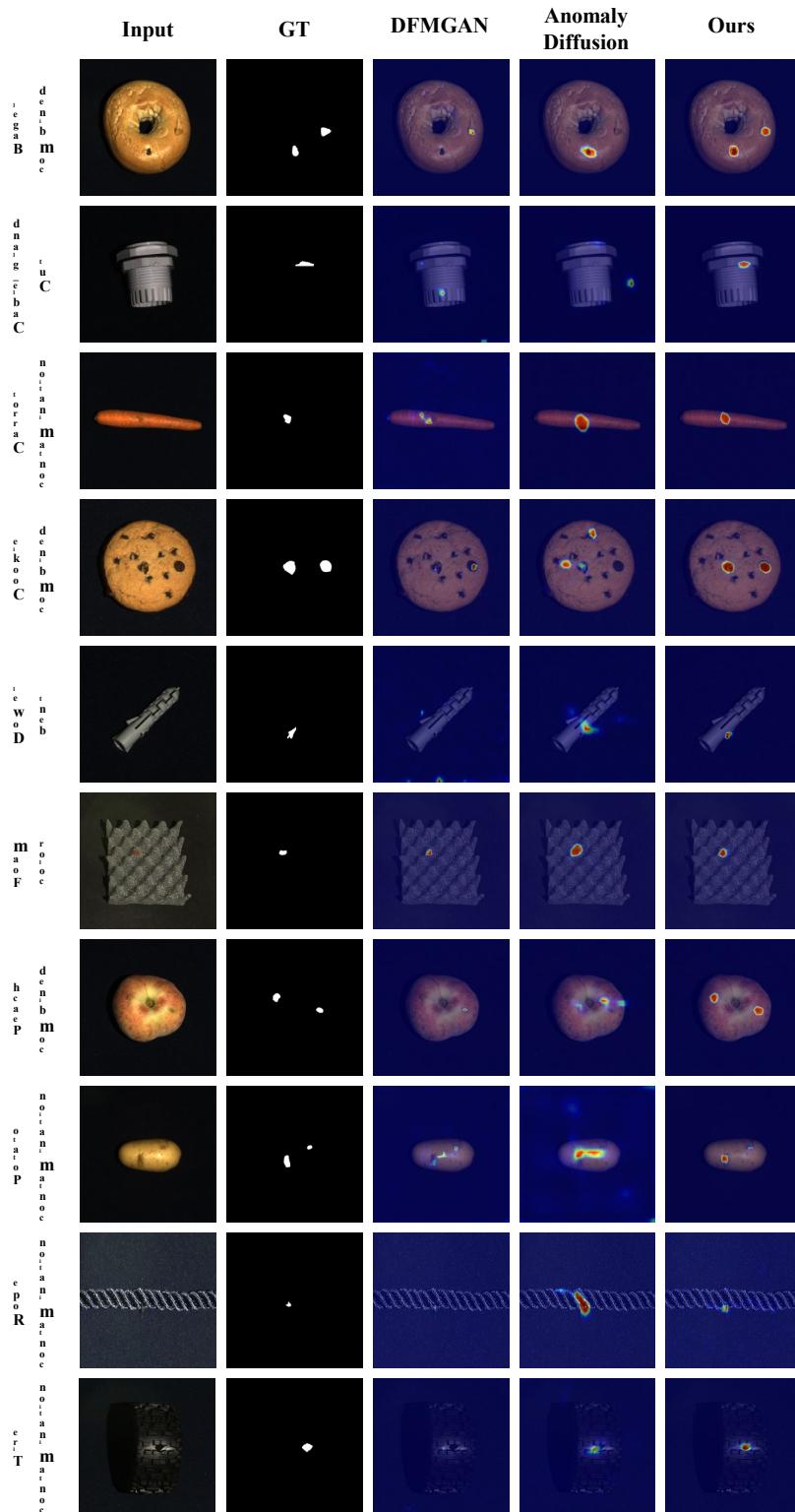


图22. 使用BiSeNet V2在MVTec 3D AD数据集上的定性监督异常分割结果。

我们在MVTec AD数据集上报告了SeaS每个类别的详细分割结果，与DFMGAN [11]和AnomalyDiffusion [17]进行了比较，具体数据从表22到表27呈现。

A.8. 基于SeaS生成的图像-掩码对训练的不同监督分割模型的更多定性比较结果

在本节中，我们基于MVTec AD和MVTec 3D AD数据集，提供了不同监督分割模型的进一步定性结果。我们选择了三种参数量级不同的模型（BiSeNet V2 [40]: 3.341M, UPerNet [38]: 64.042M, LFD [45]: 0.936M），并展示了SeaS针对各类别中不同类型异常的分割结果。相关结果见图23至图27。

表22. 基于BiSeNet V2的监督式异常分割对比。

Category	DFMGAN				AnomalyDiffusion				Ours			
	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU
bottle	89.34	64.67	62.78	44.71	99.00	88.02	80.53	68.25	99.46	93.43	85.59	75.86
cable	93.87	67.98	64.74	44.02	92.84	69.86	66.32	46.49	89.85	72.07	71.58	53.24
capsule	74.88	16.43	23.01	29.97	92.71	38.11	40.67	19.44	86.33	24.64	30.54	39.70
carpet	94.53	42.53	47.44	39.88	98.65	73.10	65.83	43.25	99.61	82.30	72.94	55.52
grid	96.86	24.40	37.40	29.93	80.59	8.08	16.79	14.26	99.36	37.91	42.50	39.80
hazelnut	99.87	96.75	90.07	71.68	97.71	63.34	59.87	43.12	97.82	78.55	73.09	68.47
leather	97.50	51.10	52.26	50.67	99.30	57.49	59.62	43.94	98.91	59.84	58.62	45.82
metal_nut	99.39	97.59	92.52	70.40	99.03	95.67	88.69	58.8	99.69	98.29	93.23	74.40
pill	97.09	83.98	79.26	36.39	99.44	93.16	86.62	41.18	98.31	76.97	68.00	55.43
screw	97.94	37.10	41.01	31.63	94.08	17.95	25.90	20.00	97.64	40.20	45.35	38.43
tile	99.65	97.08	91.16	75.94	97.79	85.58	78.28	60.46	99.67	97.29	91.48	75.75
toothbrush	97.70	51.32	54.05	23.38	98.43	49.64	54.08	26.53	97.15	46.09	49.02	28.56
transistor	84.31	45.34	46.07	30.00	98.85	85.27	77.95	49.83	96.75	69.52	66.11	57.24
wood	98.32	64.82	63.11	58.99	96.78	63.38	60.31	45.73	98.38	80.81	74.03	56.22
zipper	97.29	65.18	63.24	49.93	98.81	78.89	72.66	62.03	99.23	80.27	73.41	64.80
Average	94.57	60.42	60.54	45.83	96.27	64.5	62.27	42.89	97.21	69.21	66.37	55.28

表23. BiSeNe图像级异常检测对比
t V2。

Category	DFMGAN			AnomalyDiffusion			Ours		
	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max
bottle	96.74	98.75	95.35	98.14	99.34	97.67	100.00	100.00	100.00
cable	79.47	85.00	74.13	95.37	96.71	92.91	94.61	96.39	89.83
capsule	85.51	95.16	89.82	84.06	95.01	89.74	88.81	96.92	89.21
carpet	91.42	96.29	88.89	90.55	96.41	90.32	98.16	99.31	97.56
grid	99.64	99.82	97.56	81.19	89.92	83.95	99.17	99.63	98.73
hazelnut	100.00	100.00	100.00	93.39	95.74	90.91	100.00	100.00	100.00
leather	98.31	99.23	95.24	100.00	100.00	100.00	95.83	98.38	95.93
metal_nut	97.37	99.16	94.66	99.01	99.66	97.71	100.00	100.00	100.00
pill	84.86	95.27	91.00	90.38	97.43	91.35	96.59	99.12	95.24
screw	74.95	85.50	80.72	58.18	75.32	81.25	77.24	89.55	80.60
tile	99.47	99.74	99.12	98.78	99.44	97.39	100.00	100.00	100.00
toothbrush	78.33	87.73	83.72	78.33	89.26	79.17	90.42	94.49	89.47
transistor	79.52	75.77	69.57	94.40	94.68	94.34	99.23	98.39	94.92
wood	98.87	99.46	97.67	90.48	94.12	93.33	100.00	100.00	100.00
zipper	98.97	99.64	97.56	98.89	99.62	97.56	100.00	100.00	100.00
Average	90.90	94.43	90.33	90.08	94.84	91.84	96.00	98.14	95.43

表24. UPerNet上有监督异常分割的对比。

Category	DFMGAN				AnomalyDiffusion				Ours			
	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU
bottle	87.94	56.89	56.56	45.41	99.54	93.01	85.94	75.31	99.28	91.73	84.53	78.73
cable	87.52	64.30	65.61	41.02	91.00	68.12	67.49	51.84	91.08	76.25	74.63	59.00
capsule	67.92	12.31	20.32	30.47	97.64	51.90	51.66	37.00	92.09	39.60	43.89	50.18
carpet	95.85	36.05	34.52	48.10	99.45	82.13	72.55	53.17	99.67	82.01	73.53	60.60
grid	97.49	29.67	36.15	31.37	94.22	28.97	38.50	32.93	99.18	44.94	48.28	44.21
hazelnut	99.36	79.76	71.10	72.90	97.77	70.48	67.93	54.47	99.54	81.84	75.48	73.30
leather	80.97	17.60	26.21	30.17	99.48	63.46	60.54	48.70	99.42	68.26	65.52	57.01
metal_nut	98.44	95.64	91.48	64.92	98.62	95.11	88.62	61.31	99.70	98.33	92.90	76.07
pill	97.58	83.74	80.02	42.33	99.33	95.04	88.77	49.18	98.59	81.16	74.26	62.62
screw	97.49	53.83	53.02	42.05	93.89	36.60	42.68	34.08	98.97	52.02	51.65	46.61
tile	99.79	97.29	91.11	77.46	94.70	73.34	67.79	58.54	99.67	95.89	90.71	77.89
toothbrush	97.42	51.09	59.23	28.33	97.52	60.67	59.46	33.98	98.50	63.62	63.07	42.09
transistor	82.07	36.31	39.48	27.44	94.26	73.68	69.50	53.64	93.88	70.37	68.12	56.98
wood	97.90	69.02	62.21	63.10	96.09	70.10	64.38	51.44	99.28	85.28	76.28	65.09
zipper	97.28	71.60	66.64	54.54	99.54	86.18	78.50	66.47	99.17	85.01	77.57	68.21
Average	92.33	57.01	56.91	46.64	96.87	69.92	66.95	50.80	97.87	74.42	70.70	61.24

表25. UPerNe上图像级异常检测的对比

 t_o

Category	DFMGAN			AnomalyDiffusion			Ours		
	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max
bottle	94.19	97.86	93.18	100.00	100.00	100.00	100.00	100.00	100.00
cable	85.64	90.03	80.33	95.58	97.06	92.56	94.40	96.38	92.44
capsule	81.04	94.26	87.01	96.00	98.77	95.48	94.43	98.44	92.21
carpet	96.72	98.58	93.75	98.68	99.53	98.36	99.94	99.97	99.20
grid	98.33	99.13	96.30	96.67	98.73	97.44	99.76	99.88	98.73
hazelnut	99.84	99.87	97.96	99.17	99.43	97.87	100.00	100.00	100.00
leather	79.91	90.70	81.75	100.00	100.00	100.00	100.00	100.00	100.00
metal_nut	98.30	99.38	97.71	98.65	99.62	98.41	99.72	99.91	99.21
pill	88.54	96.56	92.39	91.23	97.78	90.91	98.28	99.58	97.92
screw	89.01	94.54	88.24	85.06	93.87	85.33	93.47	97.07	90.45
tile	99.68	99.81	99.13	99.68	99.81	99.13	100.00	100.00	100.00
toothbrush	75.00	86.99	80.00	90.00	95.13	90.00	95.00	97.65	94.74
transistor	83.04	73.59	74.19	100.00	100.00	100.00	99.52	99.16	96.43
wood	93.36	95.60	95.45	98.62	99.49	97.62	99.87	99.94	98.82
zipper	98.48	99.51	98.14	100.00	100.00	100.00	100.00	100.00	100.00
Average	90.74	94.43	90.37	96.62	98.61	96.21	98.29	99.20	97.34

表26. LFD上有监督异常分割的比较。

Category	DFMGAN				AnomalyDiffusion				Ours			
	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU
bottle	90.41	61.51	58.49	40.19	98.71	89.64	81.55	67.10	99.28	92.65	84.86	73.82
cable	96.49	79.40	75.25	53.47	97.89	79.85	72.75	53.69	94.53	75.41	72.70	55.98
capsule	91.82	56.11	58.56	32.50	95.80	38.17	48.92	32.04	91.80	49.76	53.69	41.14
carpet	89.10	48.04	49.89	39.46	94.83	53.15	51.79	42.21	99.10	82.74	74.51	57.56
grid	89.18	34.89	41.21	19.21	85.19	24.32	34.76	18.22	98.78	62.24	58.44	41.69
hazelnut	99.36	95.16	89.80	76.43	98.54	77.39	70.42	45.97	98.97	88.00	81.77	73.39
leather	97.82	51.86	52.25	48.09	98.99	65.73	62.85	42.65	99.11	76.49	69.30	56.51
metal_nut	98.16	95.16	90.99	63.02	99.38	97.34	91.63	64.59	99.23	96.66	91.42	75.15
pill	95.80	75.90	70.31	31.73	98.96	92.51	85.35	50.04	98.11	79.63	72.54	56.73
screw	93.96	38.00	41.69	30.88	92.68	44.64	49.17	34.08	98.27	52.40	52.32	41.02
tile	97.37	88.79	82.05	66.30	92.98	79.59	73.52	55.08	99.38	96.24	89.90	75.50
toothbrush	95.17	55.21	53.95	28.83	98.31	68.60	66.14	29.67	96.97	54.84	53.19	27.91
transistor	97.68	89.68	84.18	46.98	98.20	83.97	75.84	44.22	98.80	84.32	77.02	55.57
wood	97.47	77.72	70.91	58.77	95.68	67.54	63.06	42.78	98.60	88.57	81.46	62.94
zipper	93.80	58.43	56.82	46.44	98.42	84.05	77.08	64.14	99.15	86.67	79.09	69.37
Average	94.91	67.06	65.09	45.49	96.30	69.77	66.99	45.77	98.01	77.77	72.81	57.62

表27. LFD图像级异常检测对比。

Category	DFMGAN			AnomalyDiffusion			Ours		
	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max
bottle	96.98	98.76	95.35	100.00	100.00	100.00	100.00	100.00	100.00
cable	90.98	94.21	88.14	99.52	99.55	97.71	92.05	94.95	88.70
capsule	86.32	95.99	88.46	83.25	94.62	89.44	93.80	98.19	93.42
carpet	88.02	95.33	87.60	86.00	93.42	87.22	97.98	99.22	96.67
grid	85.48	92.61	85.71	93.69	97.08	91.14	96.79	98.76	96.10
hazelnut	99.90	99.91	98.97	98.28	98.60	95.83	100.00	100.00	100.00
leather	95.93	98.15	93.65	99.90	99.95	99.20	100.00	100.00	100.00
metal_nut	96.16	98.57	96.18	99.01	99.65	98.46	98.58	99.54	97.64
pill	82.85	94.40	92.00	94.15	98.42	94.47	98.16	99.50	96.84
screw	82.60	92.15	82.22	81.54	91.32	82.05	87.83	94.39	85.54
tile	98.94	99.43	96.55	98.25	99.13	95.65	99.36	99.69	99.12
toothbrush	77.08	87.68	80.95	100.00	100.00	100.00	87.92	94.08	87.80
transistor	88.04	85.06	77.78	97.38	96.57	92.86	98.10	96.90	94.55
wood	99.87	99.94	98.82	97.24	98.70	96.47	100.00	100.00	100.00
zipper	97.07	98.78	96.25	99.01	99.71	99.39	100.00	100.00	100.00
Average	91.08	95.40	90.58	95.15	97.78	94.66	96.70	98.35	95.76

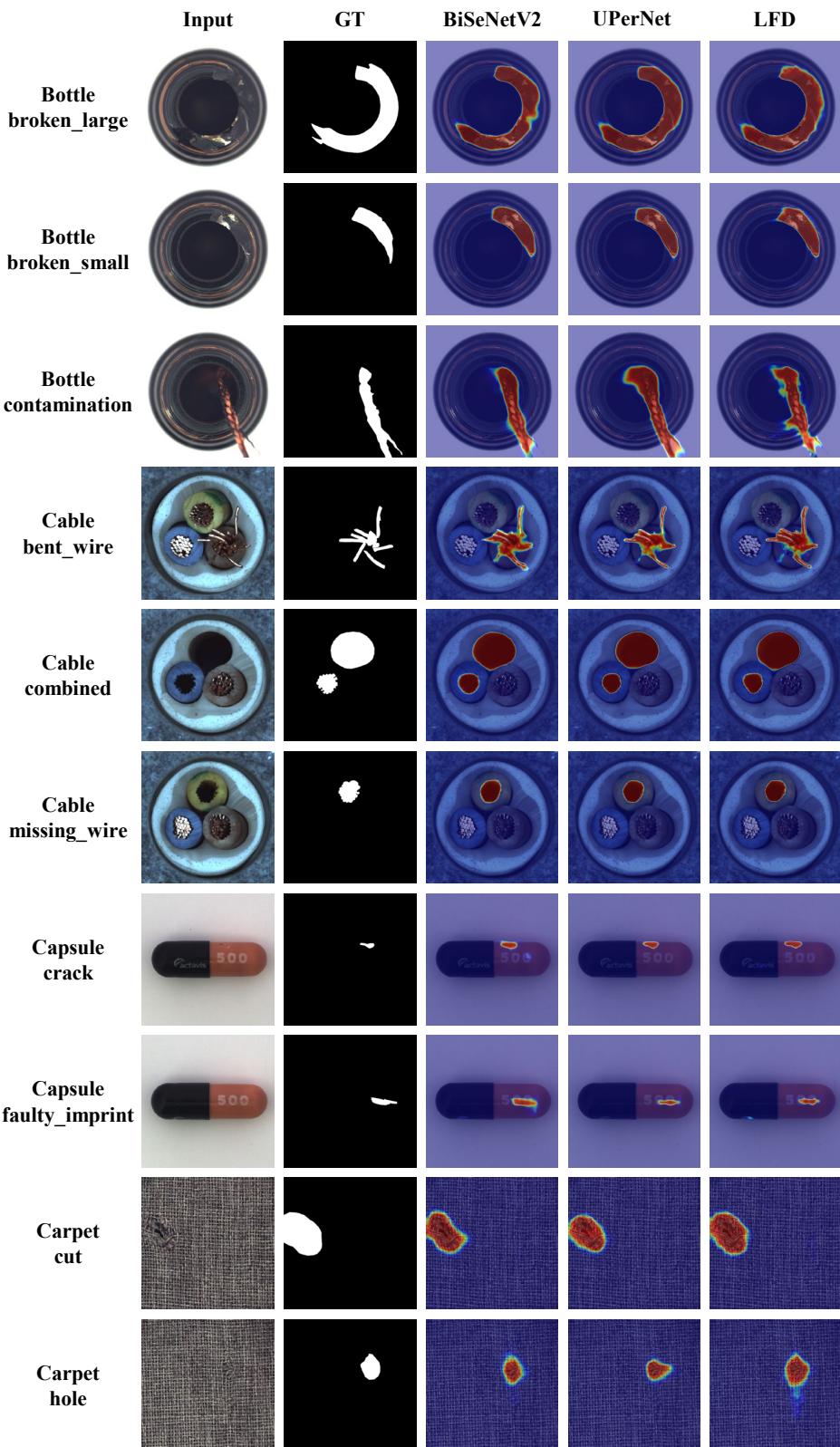


图23. 在MVTec AD上与监督分割模型的定性比较结果。图中从上至下依次为 *bottle*、*cable*、*capsule*和*carpet*类别的结果。

重新

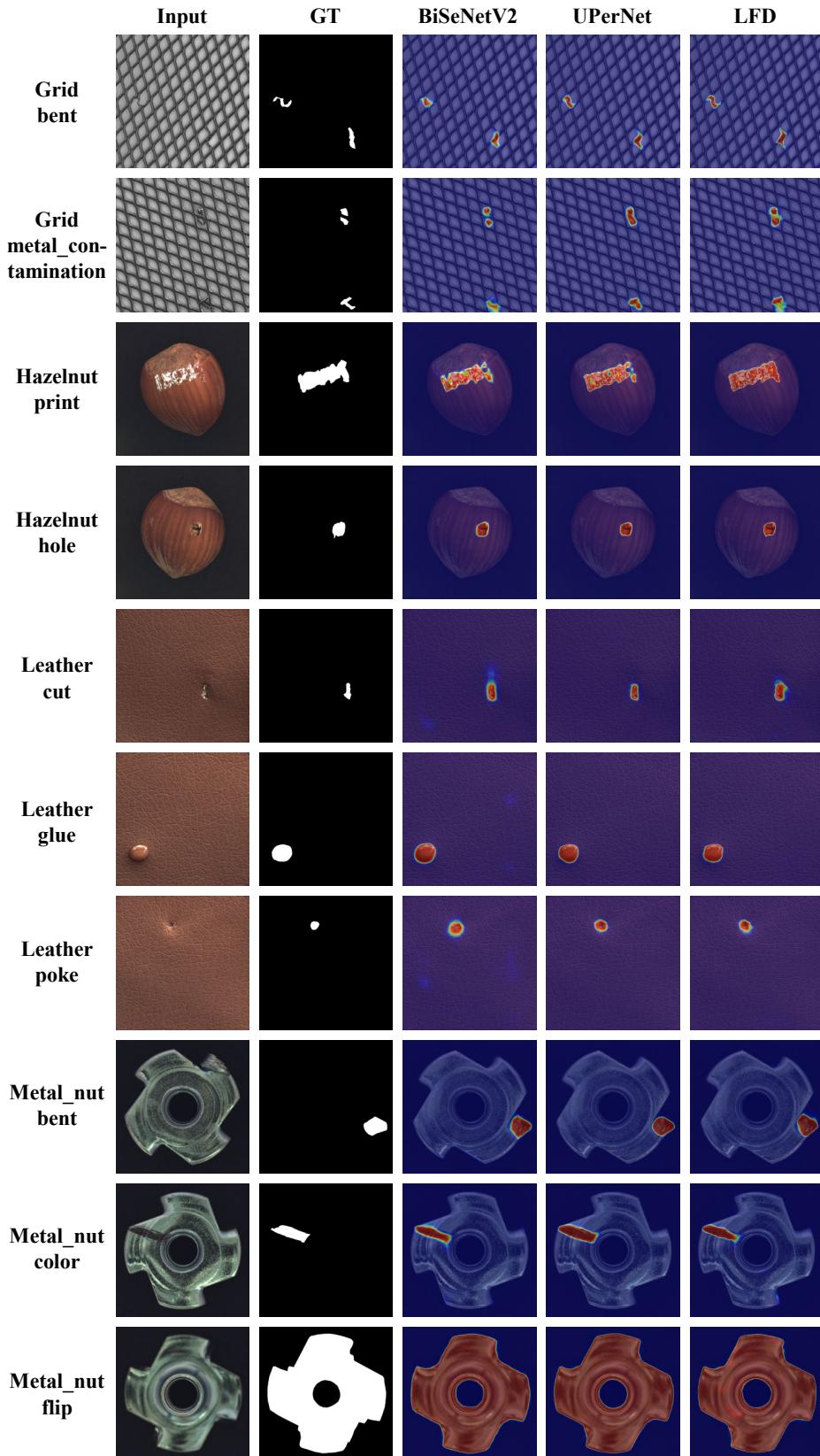


图24. 在MVTec AD数据集上与监督分割模型的定性比较结果。图中从上到下依次为grid、hazelnut、leather和metal_nut类别的结果。

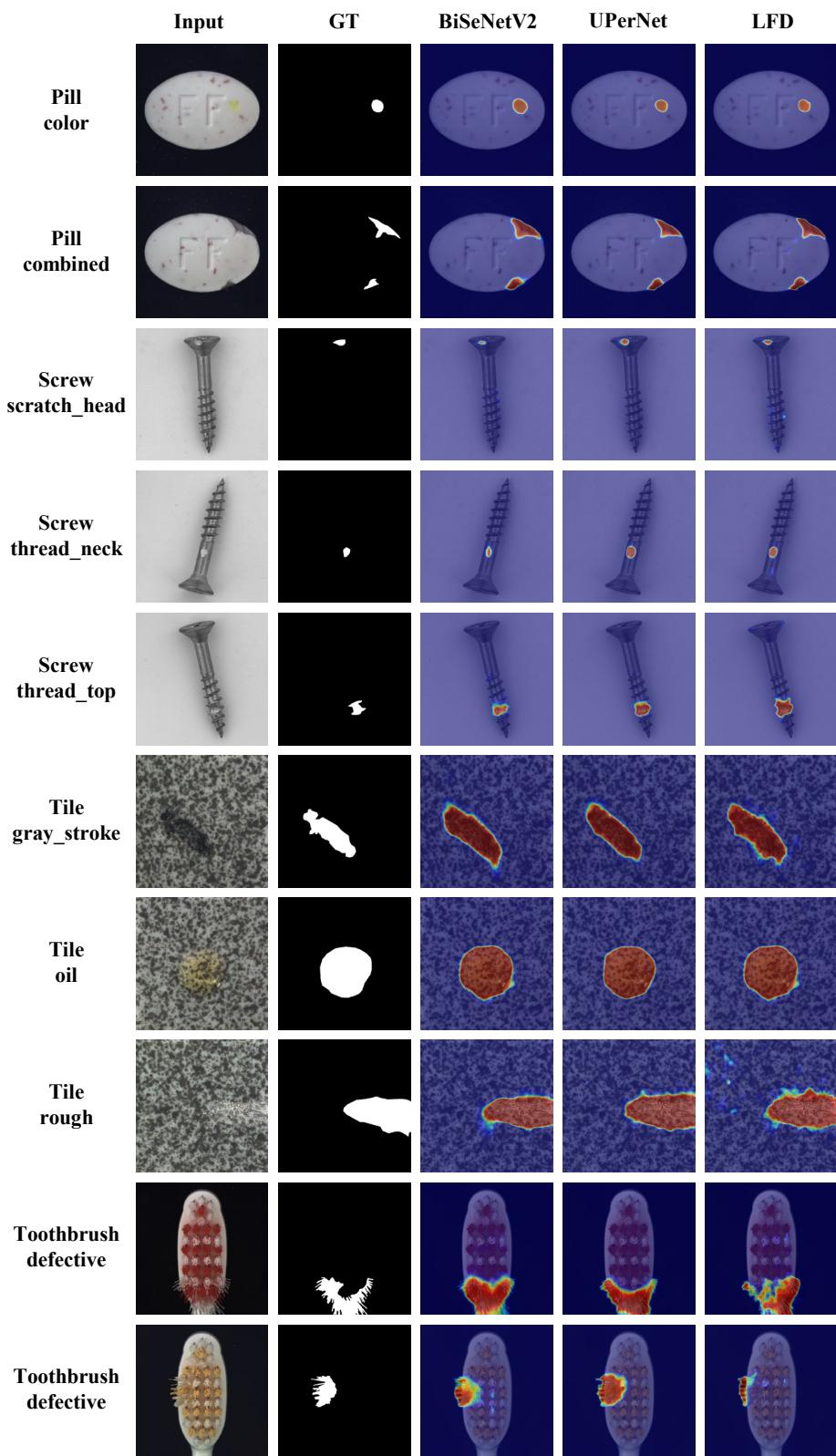


图25. 在MVTec AD数据集上与监督分割模型的定性比较结果。图中从上到下依次为pill、screw、tile和toothbrush类别的结果。

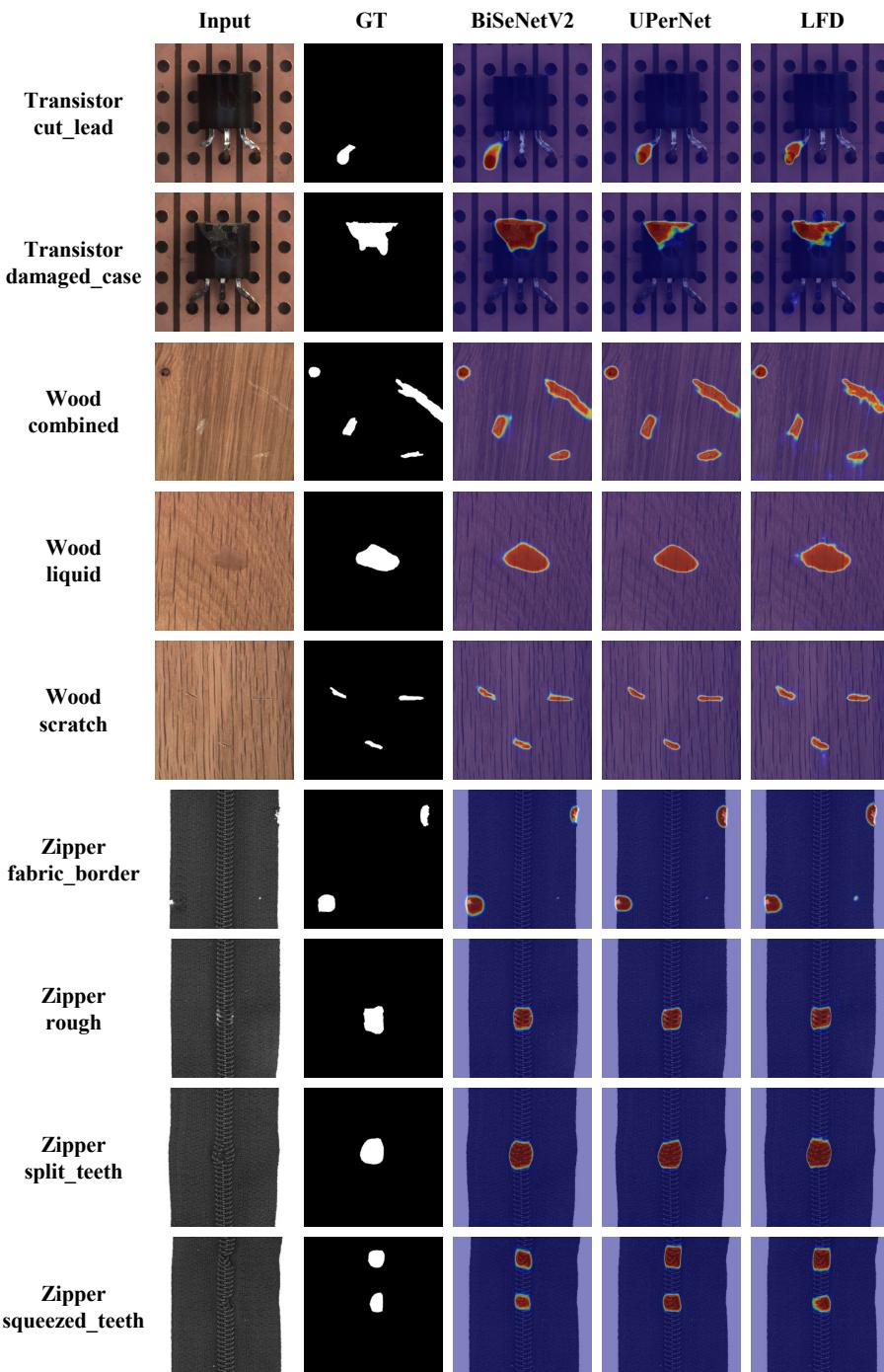


图26. 在MVTec AD上与监督分割模型的定性比较结果。图中从上至下依次为transistor、wood和zipper类别的结果。

重新

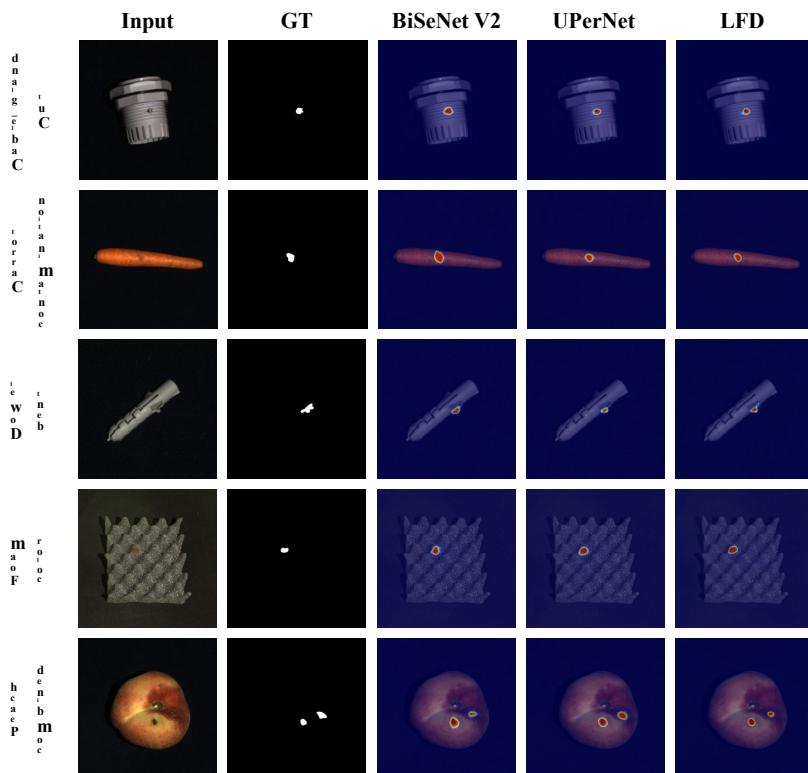


图27. 在MVTec 3D AD数据集上与有监督异常分割模型的定性对比结果。图中从上至下依次为cable_gland、carrot、dowel、foam和peach类别的结果。

A.9. 与文本反转的对比

我们进行了仅使用文本反转 (TI) [12]方法来学习产品的实验，生成的图像如图28所示。由于可学习参数数量有限，TI方法难以生成与真实产品相似的图像。相比之下，在AIG方法中，产品在局部细节上虽有微小变化但满足全局一致性，而异常则具有随机性，因此生成的产品应与真实产品保持全局一致。因此，与仅凭TI方法即可满足异常生成需求的AG方法AnomalyDiffusion[17]不同，我们通过微调U-Net来确保生成产品的全局一致性。

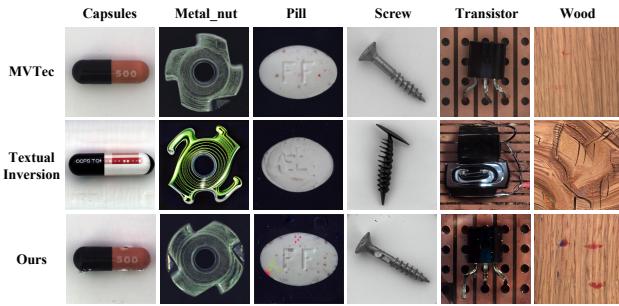


图28. 与Textual Inversion生成结果的定性比较。

A.10. 更多关于光照条件的实验

我们从MVTec3D数据集中的产品桃子中选取了一个缺陷类别进行实验，该类别在光照条件和背景上存在显著差异。强光照条件下的图像呈现桃子的顶部，而弱光照条件下的图像则显示桃子底部。因此，图像中的背景——无论是桃子顶部还是底部——也有所不同。我们选择了三种不同光照条件下的训练集进行实验：1) 仅包含强光照条件下顶部图像；2) 仅包含弱光照条件下底部图像；3) 一半为强光照条件下的顶部图像，另一半为弱光照条件下的底部图像。不同设置下生成的图像如图29所示。可以看出，SeaS对光照条件和背景变化具有鲁棒性。

A.11. 关于小缺陷生成的更多结果。

SeaS能够在小尺度异常中保留细粒度细节，如图30所示。然而，由于潜在空间的分辨率有限，生成极其细微的异常可能具有挑战性。我们将在未来的工作中探讨这一点。

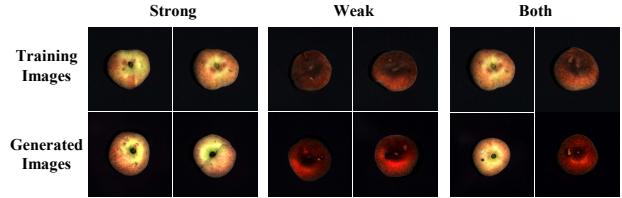


图29. 在不同光照条件和背景下的MVTec3D AD生成结果可视化。图中第一行为训练图像，第二行为生成图像。

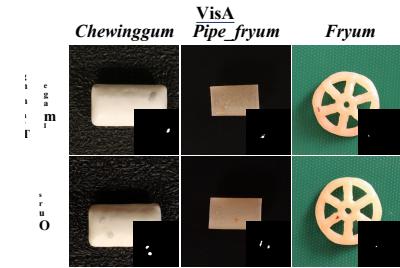


图30. 小规模异常生成结果。

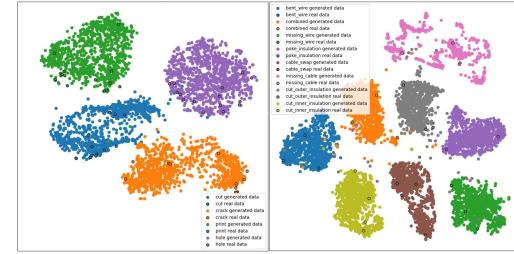


图31. 同一产品在真实数据与生成数据中不同异常类型的T-SNE可视化。

A.12. 关于未见异常类型生成的更多分析。

SeaS能够在已知异常类型内生成多样化的未见异常，如附录A.2所分析。然而，生成真正未见过的异常类型仍然具有挑战性。图31中的t-SNE可视化显示，同一产品的不同类型异常会形成紧密的簇。虽然可以实现簇内变异，但缺乏先验知识限制了跨簇泛化的能力。我们认为泛化至未见异常类型具有重要意义，并将在未来工作中对此进行探索。

A.13. 与DRAEM对比的更多实验。

如表28所示，使用与SeaS相同的异常图像训练DRAEM [41]相比仅使用无异常图像能获得更好的结果。然而，DRAEM + SeaS实现了进一步的提升，这表明性能增益不仅来自真实异常，也源于SeaS生成的多样且逼真的异常样本。

表28. 在多个数据集上结合生成异常与基于合成的异常检测方法的比较。

Segmentation Models	MVTec AD									VisA									MVTec 3D AD								
	Image-level			Pixel-level			Image-level			Pixel-level			Image-level			Pixel-level			Image-level			Pixel-level			Image-level		
	AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU	AUROC	AP	F_1 -max	IoU
DRAEM	98.00	98.45	96.34	97.90	67.89	66.04	60.30	86.28	85.30	81.66	92.92	17.15	22.95	13.57	79.16	90.90	89.78	86.73	14.02	17.00	12.42						
DRAEM + training data	97.43	98.84	97.84	96.41	74.42	71.86	59.84	83.74	86.00	82.75	94.63	39.22	43.06	29.02	73.86	88.46	86.69	82.43	19.36	25.05	17.01						
DRAEM + SeaS	98.64	99.40	97.89	98.11	76.55	72.70	58.87	88.12	87.04	83.04	98.45	49.05	48.62	35.00	85.45	93.58	90.85	95.43	20.09	26.10	17.07						