

# 缺陷生成对抗网络：用于自动化缺陷检测的高保真缺陷合成

张功杰<sup>1</sup> 崔凯文<sup>1</sup> 洪子宜<sup>2</sup> 陆世健<sup>\*1</sup> <sup>1</sup>南洋理工大学 <sup>2</sup>新加坡德尔塔研究中心  
{gongjiezhang, shijian.lu}@ntu.edu.sg kaiwen001@e.ntu.edu.sg tzuyi.hung@deltaww.com

## 摘要

Automated defect inspection is critical for effective and efficient maintenance, repair, and operations in advanced manufacturing. On the other hand, automated defect inspection is often constrained by the lack of defect samples, especially when we adopt deep neural networks for this task. This paper presents Defect-GAN, an automated defect synthesis network that generates realistic and diverse defect samples for training accurate and robust defect inspection networks. Defect-GAN learns through defacement and restoration processes, where the defacement generates defects on normal surface images while the restoration removes defects to generate normal images. It employs a novel compositional layer-based architecture for generating realistic defects within various image backgrounds with different textures and appearances. It can also mimic the stochastic variations of defects and offer flexible control over the locations and categories of the generated defects within the image background. Extensive experiments show that Defect-GAN is capable of synthesizing various defects with superior diversity and fidelity. In addition, the synthesized defect samples demonstrate their effectiveness in training better defect inspection networks.

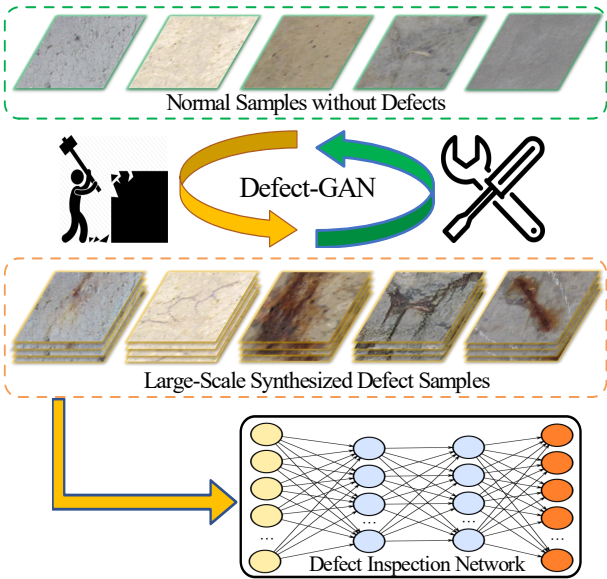


图1. Defect-GAN通过模拟易于获取的正常样本上的缺陷生成与修复过程，生成具有卓越保真度和多样性的大规模缺陷样本。生成的缺陷样本在训练精准且稳健的缺陷检测网络模型方面展现出显著成效。

## 1. 引言

自动化视觉缺陷检测旨在自动识别和判别各类图像缺陷，这一技术在制造业与建筑业等工业领域具有迫切需求。在制造业中，该技术是维护维修运行（MRO）体系的核心环节，能有效降低设备故障率并提升产能。同时，它在生产流水线各环节的质量控制中也发挥着重要作用，可用于精准定位异常问题。在建筑业领域，通过对建筑、桥梁等基础设施进行潜在风险识别，该技术为公共安全提供了关键保障。尽管自动化视觉缺陷检测技术已发展多年，它仍然面临着诸多挑战——

存在许多开放研究问题的任务。自动化视觉缺陷检测的一个关键挑战在于训练数据，这通常表现为两种不同形式。首先，收集大量带标签的缺陷样本往往成本高昂且耗时。由于缺陷样本的可重用性和可迁移性较差，这种情况变得更加严重——在处理各种新的缺陷检测任务时，我们常常需要重新收集并标注缺陷样本。其次，收集缺陷样本不仅关乎投入与成本。在许多情况下，缺陷样本本身就十分稀缺，现有数量远不能满足需求，尤其是在训练深度神经网络模型时。大规模缺陷样本的获取已成为有效设计和开发各类自动化缺陷检测系统的瓶颈之一。

\*Corresponding author.

缓解缺陷样本不足问题的一种直观方法是合成缺陷样本。尽管生成对抗网络（GAN）近年来在图像合成领域取得了卓越成就，但利用GAN合成缺陷样本仍面临诸多挑战： $\{v^*\}$  首先，现有GAN通常需要大规模训练数据，但在多数情况下难以获取大规模缺陷样本； $\{v^*\}$  其次，GAN本质上倾向于生成较简单的结构和模式[3]，因此不擅长合成具有复杂不规则形态及巨大随机变异特性的缺陷； $\{v^*\}$  再者，不同背景下的缺陷样本极难收集，导致GAN生成的缺陷样本往往与已收集的参考样本具有相似背景。这使得GAN合成的缺陷样本与参考样本具有相似的特征表示和分布，在面对各种新背景下的不同缺陷样本时难以提供有效帮助。

受[36]通过人工损伤正常工件表面来收集缺陷样本的启发，我们设计了Defect-GAN，旨在自动生成高保真度的缺陷样本，以训练准确且稳健的缺陷检测网络。Defect-GAN模拟了表面损伤和修复过程，通过利用通常易于获取的大规模正常样本，极大地缓解了缺陷样本不足的限制。我们设计了新颖的控制机制，使Defect-GAN能够灵活且逼真地在背景图像的不同位置生成各类缺陷。我们还在损伤过程中引入随机性以捕捉缺陷的随机变化，显著提升了生成缺陷样本的多样性。此外，我们设计了基于组合分层结构的网络架构，可在不同正常样本上生成缺陷，同时使正常样本的背景外观变化最小化。因此，使用此类生成的缺陷样本训练的模型能更好地处理具有各种不同背景的新缺陷样本。大量实验表明，Defect-GAN在应用于训练深度缺陷检测网络时，能够生成具有卓越保真度、多样性和有效性的大规模缺陷样本。

本工作的贡献可归纳为三个方面：*First*，我们设计了基于组合式分层的网络架构，在保留正常样本外观的同时从正常样本生成缺陷，通过模拟缺陷在不同正常样本上的呈现方式提升了缺陷多样性；*Second*，我们提出Defect-GAN，通过模拟损毁与修复过程来合成缺陷。该方法能对图像背景中生成缺陷的类别与空间位置实现更优的灵活性和控制力，并通过引入随机性实现了显著的缺陷多样性。

在生成过程中引入变化，能够通过正常样本的破坏与修复生成高保真度的缺陷样本。*Third*，大量实验表明Defect-GAN生成的缺陷样本有助于更有效地训练高精度缺陷检测网络。

## 2. 相关工作

图像合成。GANs [13] 是一种强大的生成模型，它同时训练生成器产生逼真的伪造图像，并训练判别器区分真实与伪造图像。早期研究 [13, 46, 2, 23, 4] 主要专注于无条件图像合成。近年来，越来越多工作涌现，致力于基于输入图像进行条件化图像合成，其应用范围涵盖风格迁移 [33, 20, 26, 76, 35, 28]、面部表情编辑 [7, 45, 6, 60, 59]、超分辨率重建 [30, 58, 48]、图像修复 [6, 6, 67, 44, 63] 等领域。另一趋势是多模态图像合成 [24, 25, 19, 8, 77]。然而现有方法在缺陷合成任务上泛化能力不足。我们提出的Defect-GAN通过模拟缺陷形成与修复过程，并引入随机性以模拟缺陷内部的随机变异，从而生成缺陷样本。此外，受 [64, 50, 42, 69] 启发，本方法将缺陷视作特殊前景，采用分层架构将缺陷合成至正常样本之上，既保留了正常样本的样式与外观特征，又实现了卓越的合成真实性与多样性。

从有限数据中学习。基于深度学习的技术[47, 71, 68]通常需要大量标注训练样本，但这往往难以实现。近期研究提出了许多解决数据不足问题的方法，主要可分为小样本学习与数据增强两大类。

少样本学习[51, 52, 12, 5, 31, 22, 62, 57, 70, 10, 55, 72]是指从未见过的类别中通过极少量训练样本（如1个或3个）进行学习的方法。然而其性能相当有限，远未达到实际应用水平。此外，少样本学习技术通常需要来自同一领域的大量样本，这并未缓解数据不足的限制。数据增强旨在从数量和多样性两个维度丰富训练数据集，从而训练出更优质的深度学习模型。近期若干研究尝试[1, 56, 61, 40]采用生成对抗网络作为数据增强方法，以合成逼真的训练样本。本文提出的Defect-GAN同样作为数据增强方法，通过合成具有卓越多样性和保真度的缺陷样本，来训练更优秀的缺陷检测网络。

缺陷检测。表面缺陷检测是指基于机器视觉识别与定位表面缺陷的过程，这项重要任务在工业制造领域具有广泛的实际应用价值。

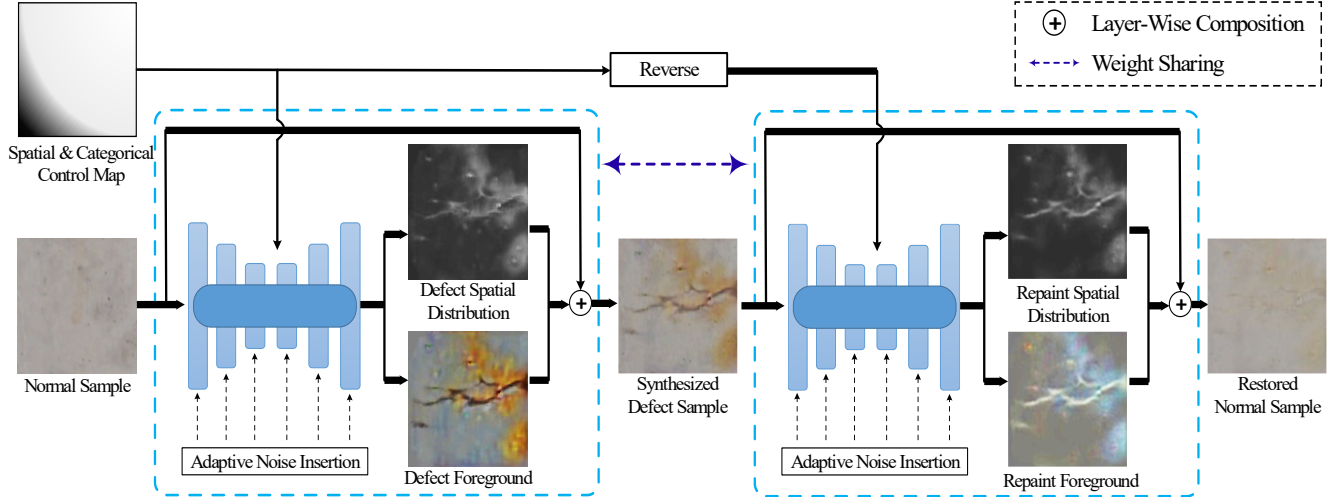


图2. 所提出Defect-GAN的生成流程：采用编码器-解码器结构，通过模拟损伤与修复过程来合成缺陷。从类别向量生成的 *Spatial & Categorical Control Map* 控制在所提供的正常样本中生成缺陷的位置与类型。*Adaptive Noise Injection* 通过引入随机变异来增强生成缺陷的多样性。此外，Defect-GAN采用 *Layer-Wise Composition* 策略，根据对应的空间分布图生成缺陷区域与修复前景。这有助于保留正常样本的风格与外观特征，从而实现卓越的缺陷合成真实感。

安全检测、建筑施工等。在深度学习时代之前，传统方法[39, 54, 29, 75, 53]需要设计手工特征提取器和启发式流程，这既需要专业知识又缺乏鲁棒性。进入深度学习时代后，许多研究[32, 9, 41, 38]采用基于卷积神经网络（CNNs）的模型进行缺陷检测，并取得了显著成效。

然而在实际场景中，缺陷样本数量有限始终是瓶颈问题。为缓解此类缺陷样本不足的问题，[36]通过人工损毁工件来收集缺陷样本；[37, 18]进一步采用计算机辅助设计（CAD）技术合成缺陷样本。但这类方法仅能处理简单情况。最新提出的SDGAN[40]采用生成对抗网络进行缺陷样本合成以实现数据增强。我们同样提出利用生成对抗网络合成缺陷样本，以训练更优的缺陷检测网络。通过采用分层组合策略模拟缺陷生成与修复过程，我们提出的Defect-GAN能够生成具有卓越真实性、多样性和灵活性的缺陷样本。该方法还能通过将学习到的缺陷模式迁移至未知表面，提供更优异的泛化能力。

### 3. 方法论

在本节中，我们将详细讨论所提出的方法。如图1所示，我们提出的方法包含两个部分：(1) 用于自动合成缺陷样本的Defect-GAN设计；(2) 利用合成缺陷样本进行缺陷检测。

#### 3.1. 用于缺陷合成的缺陷生成对抗网络

我们假设存在足够数量的正常样本，而缺陷样本数量有限，因为缺陷通常罕见且难以捕捉。基于这种数据可用性假设，我们提出采用非配对图像到图像转换[76, 7]范式进行缺陷合成，该方法通常需要较少训练数据并能获得更好的合成保真度。我们提出的Defect-GAN基于这样的直观认知：缺陷并非凭空产生，即始终存在对正常样本进行缺陷生成的损坏过程，同时也存在将缺陷样本恢复正常的复原过程。通过模拟上述损坏与复原过程，我们能够利用大量正常样本生成所需的缺陷样本。

Defect-GAN架构包含一个生成器 $G$ 和一个判别器 $D$ 。在训练阶段，Defect-GAN通过 $G$ 在两个循环中执行图像转换： $n \rightarrow d \rightarrow \hat{n}$ 和 $d \rightarrow n \rightarrow \hat{d}$ ，其中 $n \in R^{H \times W \times 3}$ 表示正常样本， $d \in R^{H \times W \times 3}$ 表示缺陷样本， $\hat{n}, \hat{d} \in R^{H \times W \times 3}$ 分别表示修复后的正常样本与缺陷样本。由于两个循环结构相同且同步执行，为简化说明，后续章节将仅以 $n \rightarrow d \rightarrow \hat{n}$ 循环为例进行阐述。

生成器 $G$ 的结构如图2所示，其采用编码器-解码器架构。 $G$ 的主体结构主要遵循常用的图像到图像转换网络[20, 76, 7]，这些网络首先对输入图像进行编码—

通过步长为4的年龄编码，然后解码回原始尺寸。为了提高缺陷生成的合成真实性和多样性，我们特别为 $G$ 设计了空间和分类控制、随机变异以及基于层的组合。 $D$ 的网络架构与StarGAN[7]相同，包含一个 $D_{src}$ （使用PatchGAN[21]区分真假样本）和一个 $D_{cls}$ （用于预测生成缺陷的类别）。

缺陷生成的空间与类别控制。不同类型的缺陷可能出现在正常样本的不同位置。为了对生成的缺陷提供更好的属性（空间与类别）控制，我们将属性控制映射 $A \in R^{H \times W \times C}$ 输入到 $G$ 中，从而在特定位置添加特定类型的缺陷。其中 $A_{x,y} \in R^C$ 表示对应位置是否存在缺陷， $C$ 表示缺陷类别数量。通过SPADE归一化[43]将 $A$ 嵌入网络，并输入到 $G$ 解码器的每个模块中。

请注意，由于我们仅假设图像级标注可用，在训练阶段，属性控制映射 $A$ 应对图像所有位置保持恒定，即通过空间维度重复目标缺陷标签 $c \in R^C$ 来获取 $A$ 。该限制可在推理阶段解除，这使得Defect-GAN能够以上下文兼容的方式在不同位置添加缺陷。

缺陷的随机性变异。与普通物体不同，缺陷通常具有复杂不规则的形态，其极高的随机变异特性使得生成对抗网络难以建模。为缓解这一问题，我们在编码器-解码器架构的每个模块中引入了自适应噪声注入机制，即在每个卷积块后的特征图中显式注入高斯噪声。每次噪声注入时会学习独立标量系数以调节噪声强度。通过显式模拟缺陷内部的随机变异，Defect-GAN能够生成更具真实性且多样性显著提升的缺陷样本。

逐层组合。如图2所示，Defect-GAN与现有图像到图像转换GAN[20, 76, 7, 40]的区别还在于：我们将最终生成视为两个图层的组合。具体而言，在缺陷生成过程中，最终缺陷样本是通过在正常样本上方叠加缺陷前景层生成的；同理，在修复过程中，最终修复样本是通过在缺陷样本上方叠加重绘前景层生成的。

污损过程可以表述为：

$$f_d, m_d = G(n, A_{n \rightarrow d}) \quad (1)$$

$$d = n \odot (1 - m_d) + f_d \odot m_d \quad (2)$$

其中 $\odot$ 表示空间乘法， $f_d$ 表示生成的缺陷前景， $m_d \in [0, 1]$ 表示

$f_d$ 对应的空间分布图。同样，还原过程可以表述为：

$$f_{\hat{n}}, m_{\hat{n}} = G(d, A_{d \rightarrow \hat{n}}) \quad (3)$$

$$\hat{n} = d \odot (1 - m_{\hat{n}}) + f_{\hat{n}} \odot m_{\hat{n}} \quad (4)$$

其中 $\hat{n}$ 表示无缺陷的恢复正常样本。

这种分层组合策略的直观思想是，缺陷可被视为叠加在背景（正常样本）上的一种特殊前景。同理，从背景中移除缺陷的修复过程也可视作覆盖缺陷区域的“重绘”过程。Defect-GAN并非直接生成合成图像，而是分别生成缺陷前景与对应的空间分布图，随后通过分层组合操作生成合成缺陷样本。

基于分层组合的新型合成方法能够在真实性和多样性两方面显著提升缺陷合成效果。这主要得益于我们的模型以正常样本为背景，隐式专注于缺陷生成而无需考虑背景生成。这一特性使模型具备生成更逼真缺陷样本的能力。此外，缺陷可能存在于各种背景之上。由于缺陷样本的稀缺性，我们只能在极有限数量的背景上采集特定缺陷。因此，典型图像合成方法缺乏缺陷可迁移性，即只能在受限场景下合成缺陷样本。我们提出的分层组合策略可缓解这一问题，因为它能充分保留背景的固有特征（外观、风格等），迫使模型模拟缺陷与给定背景之间的真实交互作用。这显著提升了缺陷可迁移性，意味着我们的模型能够在各种不同背景中生成新的缺陷样本。

训练目标。为了生成视觉上逼真的图像，我们采用对抗性损失，使生成的缺陷 $d$ 与真实缺陷样本 $d_{real}$ 难以区分。

$$\mathcal{L}_{adv} = \min_G \max_{D_{src}} \mathbb{E}_{d_{real}} [\log D_{src}(d_{real})] + \mathbb{E}_d [\log(1 - D_{src}(d))] \quad (5)$$

我们的Defect-GAN旨在根据目标缺陷标签 $c \in R^C$ 生成缺陷。为使生成的缺陷与目标类别对齐，我们采用了包含两项的类别分类损失： $\mathcal{L}_{cls}^r$ 通过将真实缺陷样本 $d_{real}$ 分类至对应类别 $c'$ 来优化 $D$ ，以及 $\mathcal{L}_{cls}^f$ 通过生成目标类别 $c$ 的缺陷样本来优化 $G$ 。

$$\mathcal{L}_{cls}^r = \mathbb{E}_{d_{real}, c'} [-\log(D_{cls}(c' | d_{real}))] \quad (6)$$

$$\mathcal{L}_{cls}^f = \mathbb{E}_{d,c}[-\log(D_{cls}(c|d))] \quad (7)$$

此外，我们引入了重建损失 $\mathcal{L}_{rec}$ ，以尽可能保留输入图像的内容。我们采用L1损失作为重建损失函数。

$$\mathcal{L}_{rec} = \mathbb{E}_{n,\hat{n}}[||n - \hat{n}||_1] \quad (8)$$

分层组合策略将在破坏和修复过程中生成空间分布图，以指导最终合成。我们通过引入两个额外的空间约束（超越空间分布图）进一步改进合成效果，即循环一致性损失和区域约束损失。

为了将生成的缺陷样本精确恢复到正常样本，重绘空间分布图应与缺陷空间分布图保持一致。因此，我们设计了缺陷空间分布图与重绘空间分布图之间的空间分布循环一致性损失 $\mathcal{L}_{sd-cyc}$

$$\mathcal{L}_{sd-cyc} = \mathbb{E}_{m_n, m_{\hat{n}}} [||m_n - m_{\hat{n}}||_1] \quad (9)$$

与此同时，为避免缺陷前景和重绘前景占据整个图像区域，我们引入区域约束损失 $\mathcal{L}_{sd-con}$ 来惩罚过大的缺陷和前景分布图：

$$\mathcal{L}_{sd-con} = \mathbb{E}_{m_n, m_{\hat{n}}} [||m_n - 0||_1 + ||m_{\hat{n}} - 0||_1] \quad (10)$$

G和D的整体训练目标为：

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}^r \mathcal{L}_{cls}^r \quad (11)$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}^f \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec} + \lambda_{con} \mathcal{L}_{sd-cyc} + \lambda_c \mathcal{L}_{sd-con} \quad (12)$$

其中 $\lambda_{cls}^r$ 、 $\lambda_{cls}^f$ 、 $\lambda_{rec}$ 、 $\lambda_{sd-cyc}$ 、 $\lambda_{sd-con}$ 为超参数，根据经验分别设置为2.0、5.0、5.0、5.0、1.0。

### 3.2. 提升缺陷检测性能

前述Defect-GAN生成的大量缺陷样本可进一步用于训练最先进的视觉识别模型，以进行缺陷检测。我们采用最常用的图像识别模型ResNet[15]和DenseNet[17]执行缺陷检测任务。生成的缺陷样本将与原始数据集混合，共同用于训练识别模型。

然而，我们注意到尽管缺陷生成对抗网络能够合成逼真的缺陷样本，但生成样本与原始样本之间仍存在领域差距。若直接在增强数据上训练识别模型，将导致模型学会区分

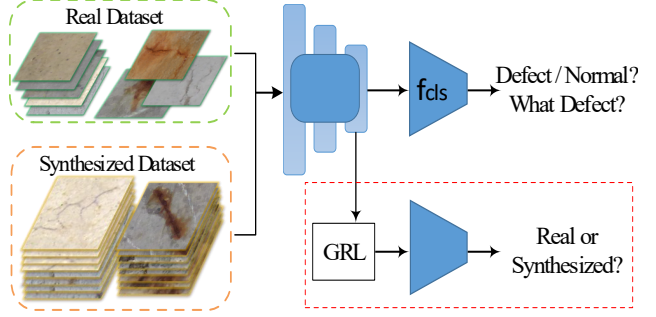


图3. 我们引入了一个源分类器 $f_{src}$ （通过梯度反转层）与网络主干 $f_\theta$ 相连，用于显式区分合成样本与真实样本。这使得缺陷检测网络不会针对此类任务进行非预期学习。

这两个领域不理想地。我们附加一个额外的源分类器 $f_{src}$ 来明确区分合成样本与真实样本，并通过梯度反转层（GRL）[11]将该领域分类器连接到网络骨干，如图3所示。因此，由 $f_\theta$ 提取的合成样本与真实样本特征之间将不存在可区分的差异，这确保了所有训练数据都能被有效学习。

## 4. 实验

本节展示了我们方法的实验。我们首先评估Defect-GAN的缺陷合成性能，然后证明其作为数据增强方法在提升缺陷检测性能方面的能力。

**数据集。**我们在CODEBRIM<sup>1</sup>[38]上评估Defect-GAN——这是一个混凝土桥梁缺陷检测数据集，具有六个互不排斥的类别：裂缝、剥落、风化、钢筋裸露、腐蚀和正常样本。该数据集不仅提供用于多标签分类的图像块，还包含裁剪这些图像块的原始全分辨率图像。与现有的缺陷检测开源数据集[49,65,34]相比，据我们所知CODEBRIM是最具挑战性和复杂性的数据集，能更好地反映实际场景。

### 4.1. 缺陷合成

**实现细节。**我们使用分类数据集集中的所有图像来训练Defect-GAN。此外，通过简单裁剪原始全分辨率图像，我们额外收集了50,000个正常图像块。所有图像在训练时均调整为128×128尺寸。为稳定训练并生成更优质图像，我们将公式5替换为带梯度惩罚的Wasserstein GAN目标函数[2,14]，并每进行五次判别器更新执行一次生成器更新。我们采用Adam优化器[27]，参数设置为 $\beta_1 = 0.5$ 和 $\beta_2 = 0.999$ 。

<sup>1</sup>Dataset available at <https://doi.org/10.5281/zenodo.2620293>

Methods	FID Scores ↓
StackGAN++ [73]	111.1
Conditional StackGAN++ [73]	132.1
StyleGAN v2 [25]	148.2
StyleGAN v2 [25] + DiffAug [74]	142.4
CycleGAN [76]	94.5
StarGAN [7]	295.1
StarGAN [7] + SPADE [43]	103.0
Defect-GAN (Ours)	<b>65.6</b>
Ideal Defect Synthesizer	25.0

表 1. Defect-GAN 与现有图像合成方法在 Fréchet Inception 距离 (FID) 上的定量比较。

训练Defect-GAN时，学习率从 $2 \times 10^{-4}$ 开始，逐渐降低至 $1 \times 10^{-6}$ 。我们将批次大小设为4，总训练迭代次数设为50万次。在单张NVIDIA 2080Ti GPU上训练耗时约一天。

评估指标。我们采用常用的Fréchet Inception Distance (FID) [16] 来评估合成缺陷样本的真实性。较低的FID分数表示更好的合成真实性。

定量实验结果。表1展示了关于缺陷合成保真度的定量实验结果，其中第一模块为直接合成方法（从随机采样的潜代码生成图像），第二模块为图像到图像转换方法。我们在第三模块通过将真实缺陷样本随机分为两组并计算其间的FID分数，展示了理想缺陷合成器的FID基准值。如表1所示，由于缺乏缺陷训练样本及捕捉复杂不规则缺陷模式的能力有限，直接合成方法普遍表现欠佳。相比之下，通过模拟Defect-GAN的损毁与修复流程，现有图像到图像转换方法能生成质量显著更优的缺陷图像。这是因为此类方法以更多信息作为输入，通常具有更高的数据利用率，且能利用大量正常样本进行训练。另一方面，Defect-GAN在合成FID指标上显著优于其他方法，证明了其在缺陷合成领域的优越性。有趣的是，具有类别控制功能的模型在FID分数上往往反而不如无类别控制的模型。我们认为引入额外的类别控制可能会限制模型的合成真实感。但即使存在这种约束，Defect-GAN仍保持了最佳性能。

我们通过在表2中展示定量消融实验，进一步证明了Defect-GAN中所提出设计的有效性。若未采用我们设计的组件，

Design Choices				FID Scores ↓
SCC	ANI	LWC	SC	
×	×	×	×	295.1
✓	×	×	×	103.0
✓	✓	×	×	99.7
✓	×	✓	×	76.8
✓	×	✓	✓	69.5
✓	✓	✓	✓	<b>65.6</b>

表2. 所提出Defect-GAN的消融研究：我们设计的空间与分类控制(SCC)、自适应噪声注入(ANI)、分层组合(LWC)以及附加空间约束(SC)具有互补性，共同提升合成缺陷的质量。

Defect-GAN 会退化为 StarGAN [7]——一种广泛使用的多领域图像转换模型。但该模型在此任务中无法收敛，也无法合成任何类缺陷图案。通过引入空间与分类控制 (SCC)，模型能够收敛并生成与现有方法质量相当的缺陷样本。在此基础上，分层组合 (LWC) 能显著提升合成逼真度。我们认为这源于双重因素：（1）通过让网络完全专注于缺陷生成，突破了缺陷数据不足的限制；（2）能生成上下文更自然的缺陷。此外，自适应噪声注入 (ANI) 和训练中增加的空间约束 (SC) 也能提升缺陷合成性能。这些组件被证明具有互补性，使 Defect-GAN 实现了最先进的缺陷合成质量。

定性实验结果。图4展示了Defect-GAN的定性结果以及与其他合成方法的对比。第1-2行展示了当前最先进的直接合成方法：StackGAN++[73]和采用DiffAug[74]的StyleGAN v2[25]的合成效果。可以看出许多生成样本未包含清晰缺陷，部分样本在视觉上不够自然。这验证了前述GAN在缺陷合成方面的局限性。对于图像转换方法，我们选择搭载SPADE[43]的StarGAN[7]作为对比方法，因为它与Defect-GAN同样提供类别控制。而CycleGAN[76]和SDGAN[40]等方法产生的视觉效果相似。如第4-5行所示，搭载SPADE的StarGAN和Defect-GAN都能基于正常样本生成视觉逼真且多样化的缺陷样本。虽然搭载SPADE的StarGAN生成的缺陷样本与Defect-GAN效果相当，但其倾向于改变背景特征，而Defect-GAN通过分层组合策略能保留正常样本的外观与风格。另一方面，搭载SPADE的StarGAN完全无法将学习到的缺陷模式迁移到新的

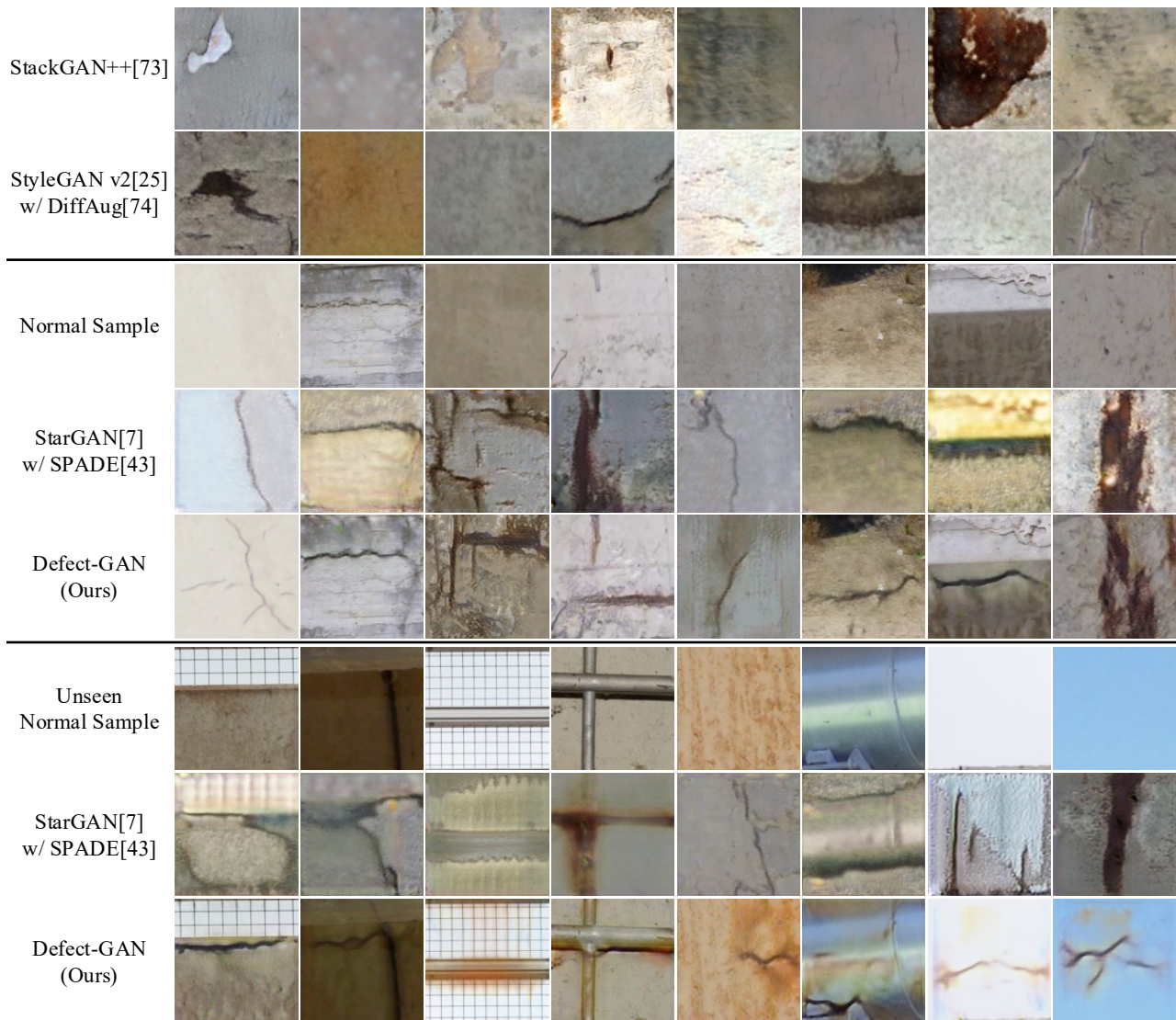


图4. Defect-GAN与前沿图像合成方法的定性对比：第1-2行展示了两种最新图像合成方法通过随机噪声直接生成缺陷的效果。第4-5行分别通过StarGAN with SPADE和我们的Defect-GAN，对第3行中用于网络训练的*Normal Samples*进行缺陷生成对比；而第7-8行则分别通过StarGAN with SPADE和我们的Defect-GAN，对第6行中未参与网络训练的*Unseen Normal Samples*进行缺陷生成对比。

在训练过程中未见的背景，而Defect-GAN则展现出卓越的缺陷迁移能力和合成真实感（如第7-8行所示）。这一特性对于向训练数据引入新信息至关重要。

此外，我们在图5中展示了Defect-GAN在缺陷生成方面的类别控制能力——基于同一正常图像可生成不同类型的缺陷。图6则展示了Defect-GAN在缺陷生成中的空间控制能力，其中红色框标注了预设的缺陷生成区域。Defect-GAN能在特定位置生成缺陷的同时保持上下文环境的自然真实。

## 4.2. 缺陷检测

实现细节。我们使用训练集和额外的50,000张正常图像来训练Defect-GAN。随后，Defect-GAN通过合成50,000个缺陷样本来扩充训练数据。生成的缺陷样本与原始训练数据混合用于训练缺陷检测网络，同时加入修复后的正常样本以避免数据不平衡。所有图像尺寸调整为 $224 \times 224$ 。我们采用学习率为 $1 \times 10^{-3}$ 的SGD优化器训练网络直至收敛，批量大小设置为16。使用验证集选择最佳性能模型，并在测试集上报告性能指标。

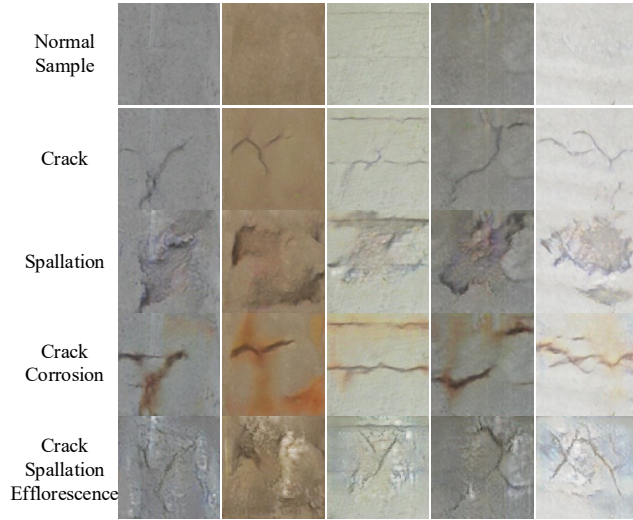


图5. Defect-GAN在缺陷生成中的分类控制示意图：对于第1行中的每个正常样本，第2-3行和第4-5行分别展示了基于单个目标类别和多个目标类别生成的缺陷样本。

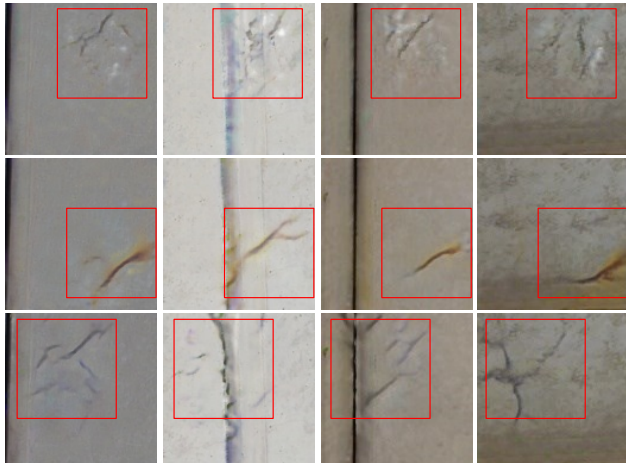


图6. Defect-GAN在缺陷生成中的空间控制示意图：每行展示使用不同正常样本但相同空间控制生成的缺陷样本，每列展示使用相同正常样本但不同空间控制生成的缺陷样本。

定量实验结果。由于CODEBRIM具有多标签分类任务的特点，我们只能采用类别控制的方法来扩展训练样本。缺陷检测的结果如表3所示。每个模块的第一行显示仅使用原始训练数据的缺陷检测性能，每个模块的其余三行显示使用原始训练数据和通过不同合成方法生成的增强样本的缺陷检测性能。为了公平比较，

Networks	Augmentation Methods	Accuracy(%)
ResNet34 [15]	None	70.25
	Conditional StackGAN++[73]	62.59
	StarGAN[7]+SPADE[43]	71.90
	Defect-GAN (Ours)	<b>75.48</b>
DenseNet121 [17]	None	70.77
	Conditional StackGAN++[73]	58.68
	StarGAN[7]+SPADE[43]	72.61
	Defect-GAN (Ours)	<b>75.79</b>

表3. 缺陷检测的定量实验结果。

为所有合成方法增广了50,000个合成缺陷样本。结果显示，Conditional StackGAN++[73]生成的缺陷样本会大幅降低缺陷检测性能。这是因为StackGAN++由于其在缺陷建模方面的能力有限，甚至无法生成逼真的缺陷样本。StackGAN++生成的缺陷样本对网络训练有害。另一方面，StarGAN[7]+SPADE[43]生成的样本可以略微提升检测性能。而我们提出的Defect-GAN则能进一步显著提高训练后的缺陷检测网络精度。尽管两种方法都能生成具有良好视觉真实感的缺陷样本，但Defect-GAN能够模拟训练期间未见背景上的学习缺陷。这一特性使得Defect-GAN生成的样本更加多样化，从而能为训练数据引入新信息，显著提升训练模型的性能。结果也证明了Defect-GAN在保真度、多样性和可迁移性方面对于缺陷合成的优越性。

## 5. 结论

本文提出了一种新型Defect-GAN，通过模拟缺陷形成与修复过程来生成缺陷样本。该模型能够捕捉缺陷内部的随机性变化，并灵活控制生成缺陷的位置与类别。此外，借助新颖的分层组合架构，它能在保留背景风格与外观的前提下生成缺陷。所提出的Defect-GAN可生成具有卓越保真度与多样性的缺陷样本，能显著提升缺陷检测网络的性能。

致谢：本工作由台达-南洋理工网络物理系统企业实验室开展，并获得了台达电子股份有限公司及新加坡国家研究基金会的资助，资助项目为企业实验室@大学计划（项目编号：DELTA-NTU CORP-SMA-RP15）。

## 参考文献

- [1] Antreas Antoniou, Amos Storkey, 与 Harrison Edwards. 数据增强生成对抗网络. *arXiv preprint arXiv:1711.04340*, 2017年.
- [2] Martin Arjovsky, Soumith Chintala, 与 Léon Bottou. Wasserstein生成对抗网络. 见*ICML*, 第214–223页, 2017年.
- [3] David Bau, 朱俊彦, Jonas Wulff, William Peebles, Hendrik Strobelt, 周博磊, 与 Antonio Torralba. 洞察GAN无法生成的内容. 见*ICCV*, 第4502–4511页, 2019年.
- [4] Andrew Brock, Jeff Donahue, 与 Karen Simonyan. 面向高保真自然图像合成的大规模GAN训练. 见*ICLR*, 2019年.
- [5] 陈威宇, 刘彦承, Zsolt Kira, 王昱祺, 与黄佳彬. 细观小样本分类. 见*ICLR*, 2019年.
- [6] 陈英聪, 沈晓晖, 林哲, 鲁鑫, I Pao, 贾佳亚等. 面向人脸属性操控的语义成分分解. 见*CVPR*, 第9859–9867页, 2019年.
- [7] 崔允宰, 崔旻宰, 金文英, 河正宇, 金成勋, 与 朱宰旭. StarGAN: 统一式多领域图像到图像转换生成对抗网络. 见*CVPR*, 第8789–8797页, 2018年.
- [8] 崔允宰, Uh Youngjung, 柳在俊, 与 河正宇. StarGAN v2: 多领域多样化图像合成. 见*CVPR*, 第8188–8197页, 2020年.
- [9] Shahrzad Faghieh-Roohi, Siamak Hajizadeh, Alfredo Núñez, Robert Babuska, 与 Bart De Schutter. 基于深度卷积神经网络的钢轨表面缺陷检测. 见*International joint conference on neural networks (IJCNN)*, 第2584–2589页, IEEE, 2016年.
- [10] 范奇, 卓炜, 汤凯腾, 与 戴宇荣. 基于注意力RPN与多关系检测器的小样本目标检测. 见*CVPR*, 2020年.
- [11] Yaroslav Ganin 与 Victor Lempitsky. 基于反向传播的无监督域自适应. 见*ICML*, 第1180–1189页, 2015年.
- [12] Spyros Gidaris 与 Nikos Komodakis. 动态小样本视觉学习不忘记. 见*CVPR 2018*, 第4367–4375页, 2018年.
- [13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, 徐冰, David Warde-Farley, Sherjil Ozair, Aaron Courville, 与 Yoshua Bengio. 生成对抗网络. 见*NeurIPS*, 第2672–2680页, 2014年.
- [14] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, 与 Aaron C Courville. Wasserstein GANs的改进训练方法. 见*NeurIPS*, 第5767–5777页, 2017年.
- [15] 何恺明, 张祥雨, 任少卿, 与 孙剑. 面向图像识别的深度残差学习. 见*CVPR*, 第770–778页, 2016年.
- [16] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, 与 Sep Hochreiter. 采用双时间尺度更新规则训练的GAN收敛至局部纳什均衡. 见*NeurIPS*, 第6626–6637页, 2017年.
- [17] 高黄, 庄刘, 劳伦斯·范德马滕和基利安·Q·温伯格. 密集连接卷积网络. 见*CVPR*, 第4700–4708页, 2017年.
- [18] 黄倩, 吴元, John Baruch, 蒋平和彭永红. 一种用于评估X射线无损检测的缺陷模拟模板模型. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(2):466–475, 2009年.
- [19] 黄勋, 刘明宇, Serge Belongie和Jan Kautz. 多模态无监督图像到图像翻译. 见*ECCV*, 2018年.
- [20] Phillip Isola, 朱俊彦, 周廷辉和Alexei A. Efros. 基于条件对抗网络的图像到图像翻译. 见*CVPR*, 第1125–1134页, 2017年.
- [21] Phillip Isola, 朱俊彦, 周廷辉和Alexei A. Efros. 基于条件对抗网络的图像到图像翻译. 见*CVPR*, 第1125–1134页, 2017年.
- [22] 康秉翼, 刘壮, 王欣, Fisher Yu, 冯家石和Trevor Darrell. 通过特征重加权的小样本目标检测. 见*ICCV*, 第8420–8429页, 2019年.
- [23] Tero Karras, Timo Aila, Samuli Laine和Jaakko Lehtinen. 渐进式增长GAN以提升质量、稳定性和多样性. 见*ICLR*, 2018年.
- [24] Tero Karras, Samuli Laine和Timo Aila. 一种基于风格的生成器架构用于生成对抗网络. 见*CVPR*, 第4401–4410页, 2019年.
- [25] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen和Timo Aila. StyleGAN图像质量分析与改进. 见*CVPR*, 第8110–8119页, 2020年.
- [26] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee和Jiwon Kim. 学习用生成对抗网络发现跨领域关系. *arXiv preprint arXiv:1703.05192*, 2017年.
- [27] Diederik P. Kingma和Jimmy Ba. Adam: 一种随机优化方法. 见*ICLR*, 2015年.
- [28] Ali Koksul和卢世健. RF-GAN: 一种轻量可重构网络用于非配对图像到图像翻译. 见*ACCV*, 2020年.
- [29] 郭重峰, 许建同, 刘宗贤和吴汉城. 使用两阶段反向传播神经网络的LED芯片自动检测系统. *Journal of Intelligent Manufacturing*, 25(6):1235–1243, 2014年.
- [30] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, 王泽翰等. 使用生成对抗网络的照片级真实感单图像超分辨率. 见*CVPR*, 第4681–4690页, 2017年.
- [31] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran和Stefano Soatto. 基于可微凸优化的元学习. 见*CVPR*, 2019年.
- [32] 李运东, 赵伟刚和潘佳浩. 基于Fisher准则深度学习的可变形图案织物缺陷检测. *IEEE Transactions on Automation Science and Engineering*, 14(2):1256–1264, 2016年.

- [33] 刘明宇、Thomas Breuel与Jan Kautz。无监督图像到图像转换网络。发表于*NeurIPS*, 第700–708页, 2017年。[34] Marc Maguire、Sattar Dorafshan与Robert J. Thomas。SDNET2018: 面向机器学习应用的混凝土裂缝图像数据集。2018年。[35] Youssef Alami Mejjati、Christian Richardt、James Tompkin、Darren Cosker与Kwang In Kim。无监督注意力引导的图像到图像转换。发表于*NeurIPS*, 第3693–3703页, 2018年。[36] Domingo Mery与Dieter Filbert。基于潜在缺陷在射线图像序列中追踪的铝铸件自动缺陷检测。《*IEEE Transactions on Robotics and Automation*》, 18(6):890–901, 2002年。[37] D Mery、D Hahn与N Hirschfeld。使用缺陷CAD模型与真实X射线图像的铝铸件缺陷仿真。《*Insight-Non-Destructive Testing and Condition Monitoring*》, 47(10):618–624, 2005年。[38] Martin Mundt、Sagnik Majumder、Sreenivas Murali、Panagiotis Panetsos与Visvanathan Ramesh。基于元学习的混凝土缺陷多目标分类卷积神经网络架构——基于混凝土缺陷桥梁图像数据集。发表于*CVPR*, 第11196–11205页, 2019年。[39] Henry YT Ngan、Grantham KH Pang与Nelson HC Yung。织物缺陷自动检测技术综述。《*Image and vision computing*》, 29(7):442–458, 2011年。[40] 牛双龙、李彬、王兴刚与林辉。基于GAN的缺陷图像样本生成方法及其在缺陷识别中的改进。《*IEEE Transactions on Automation Science and Engineering*》, 2020年。[41] 牛双龙、林辉、牛同质、李彬与王兴刚。Defect GAN: 基于生成对抗网络的弱监督缺陷检测。发表于*CASE*, 第127–132页, 2019年。[42] Dim P Papadopoulos、Youssef Tamaazousti、Ferda Ofli、Ingmar Weber与Antonio Torralba。披萨制作指南: 基于分层组合结构的生成对抗网络模型研究。发表于*CVPR*, 第8002–8011页, 2019年。[43] 朴泰成、刘明宇、王廷春与朱俊彦。基于空间自适应归一化的语义图像合成。发表于*CVPR*, 第2337–2346页, 2019年。[44] Deepak Pathak、Philipp Krahenbuhl、Jeff Donahue、Trevor Darrell与Alexei A Efros。上下文编码器: 通过图像修复实现特征学习。发表于*CVPR*, 第2536–2544页, 2016年。[45] Albert Pumarola、Antonio Agudo、Aleix M Martinez、Alberto Sanfeliu与Francesc Moreno-Noguer。GANimation: 基于单张图像的解剖结构感知面部动画。发表于*ECCV*, 第818–833页, 2018年。[46] Alec Radford、Luke Metz与Soumith Chintala。基于深度卷积生成对抗网络的无监督表示学习。《*arXiv preprint arXiv:1511.06434*》, 2015年。[47] 任少卿、何恺明、Ross Girshick与孙剑。Faster R-CNN: 基于区域提议网络的实时目标检测算法。发表于*NeurIPS*, 第91–99页, 2015年。[48] Mehdi SM Sajjadi、Bernhard Scholkopf与Michael Hirsch。Enhancenet: 通过自动化纹理合成。于*ICCV*, 第4491–4500页, 2017年。[49] 石勇、崔立蒙、齐志全、孟凡、陈振松。基于随机结构森林的道路裂缝自动检测。《*IEEE Transactions on Intelligent Transportation Systems*》, 17(12):3434–3445, 2016年。[50] Krishna Kumar Singh、Utkarsh Ojha、李永杰。FineGAN: 细粒度物体生成与发现的无监督分层解缠。收录于*CVPR*, 第6490–6499页, 2019年。[51] Jake Snell、Kevin Swersky、Richard Zemel。原型网络在小样本学习中的应用。收录于*NeurIPS 2017*, 第4077–4087页, 2017年。[52] 宋 Flood、杨永新、张莉、汤涛、Philip HS Torr、Timothy M Hospedales。学习比较: 小样本学习中的关系网络。收录于*CVPR 2018*, 第1199–1208页, 2018年。[53] AS Tolba、Hazem M Raafat。用于薄膜缺陷检测的多尺度图像质量度量。《*The International Journal of Advanced Manufacturing Technology*》, 79(1-4):113–122, 2015年。[54] 蔡德明、谢志勇。方向性纹理的自动表面检测。《*Image and Vision computing*》, 18(1):49–62, 1999年。[55] 王欣、Thomas E. Huang、Trevor Darrell、Joseph E Gonzalez、喻飞。令人沮丧的简单小样本目标检测。2020年。[56] 王宇雄、Ross Girshick、Martial Hebert、Bharath Hariharan。基于虚构数据的低样本学习。收录于*CVPR*, 第7278–7286页, 2018年。[57] 王宇雄、Deva Ramanan、Martial Hebert。元学习检测稀有目标。《*ICCV*》, 第9924–9933页, 2019年。[58] 王铮、叶芒、杨帆、白翔、佐藤信一。基于级联SR-GAN的自适应低分辨率行人重识别。收录于*IJCAI*, 第3891–3897页, 2018年。[59] 吴荣亮、卢世健。LEED: 通过解缠实现无标签表情编辑。收录于*ECCV*, 第781–798页, 2020年。[60] 吴荣亮、张功杰、卢世健、陈涛。级联EF-GAN: 局部聚焦的渐进式面部表情编辑。收录于*CVPR*, 第5021–5030页, 2020年。[61] 鲜永勤、Saurabh Sharma、Bernt Schiele、Zeynep Akata。f-VAEGAN-D2: 面向任意样本学习的特征生成框架。收录于*CVPR*, 第10275–10284页, 2019年。[62] 颜小鹏、陈子良、徐安霓、王晓丹、梁晓丹、林亮。Meta R-CNN: 实例级低样本学习的通用求解器。收录于*ICCV*, 2019年。[63] 闫兆义、李小明、李穆、左旺孟、山世光。Shift-net: 通过深度特征重排的图像修复。收录于*ECCV*, 第1–17页, 2018年。[64] 杨建伟、Anitha Kannan、Dhruv Batra、Devi Parikh。LR-GAN: 分层递归生成对抗网络用于图像生成。收录于*ICLR*, 2017年。[65] 杨亮、李兵、李伟、刘兆明、杨国勇、肖继忠。使用无人机进行混凝土深层检测并构建CSSC数据库。收录于*IEEE/RSJ International Conference on Intelligent Robots and Systems*。IEEE, 2017年。

[66] Raymond A Yeh、陈辰、Teck Yian Lim、Alexander G Sc hwing、Mark Hasegawa-Johnson与Minh N Do。基于深度生成模型的语义图像修复。见*CVPR*，第5485–5493页，2017年。[67] 余嘉慧、林哲、杨继梅、沈晓晖、卢欣与黄托马斯。基于上下文注意力的生成式图像修复。见*CVPR*，第5505–5514页，2018年。[68] 詹方能、吕世坚与薛楚辉。面向场景文本精准检测与识别的逼真图像合成。2018年。[69] 詹方能、朱洪源与吕世坚。面向图像合成的空间融合生成对抗网络。见*CVPR*，第3653–3662页，2019年。[70] 张功杰、崔凯文、吴荣亮、吕世坚与田永鸿。PNPDet：通过即插即用子网络实现高效且免遗忘的小样本检测。见*WACV*，2021年。[71] 张功杰、吕世坚与张伟。CAD-Net：面向遥感图像目标检测的上下文感知网络。《*IEEE Transactions on Geoscience and Remote Sensing*》，57(12): 10015–10024，2019年。[72] 张功杰、罗志鹏、崔凯文与吕世坚。Meta-DETR：通过统一图像级元学习的小样本目标检测。《*ArXiv*》，abs/2103.11731，2021年。[73] 张晗、徐韬、李洪生、张少亭、王孝刚、黄晓雷与Dimitris N Metaxas。StackGAN++：基于堆叠式生成对抗网络的逼真图像合成。《*IEEE transactions on pattern analysis and machine intelligence*》，41(8): 1947–1962，2018年。[74] 赵晟宇、刘志健、林济、朱俊彦与韩松。面向数据高效GAN训练的可微分增强。《*arXiv preprint arXiv:2006.10738*》，2020年。[75] 周文举、费敏瑞、周慧瑜与李康。基于稀疏表示的瓶盖表面缺陷快速检测方法。《*Neurocomputing*》，123: 406–414，2014年。[76] 朱俊彦、朴泰成、Phillip Isola与Alexei A Efros。基于循环一致对抗网络的无配对图像转换。见*ICCV*，第2223–2232页，2017年。[77] 朱振、徐志亮、游安生与白翔。语义多模态图像合成。见*CVPR*，第5467–5476页，2020年。

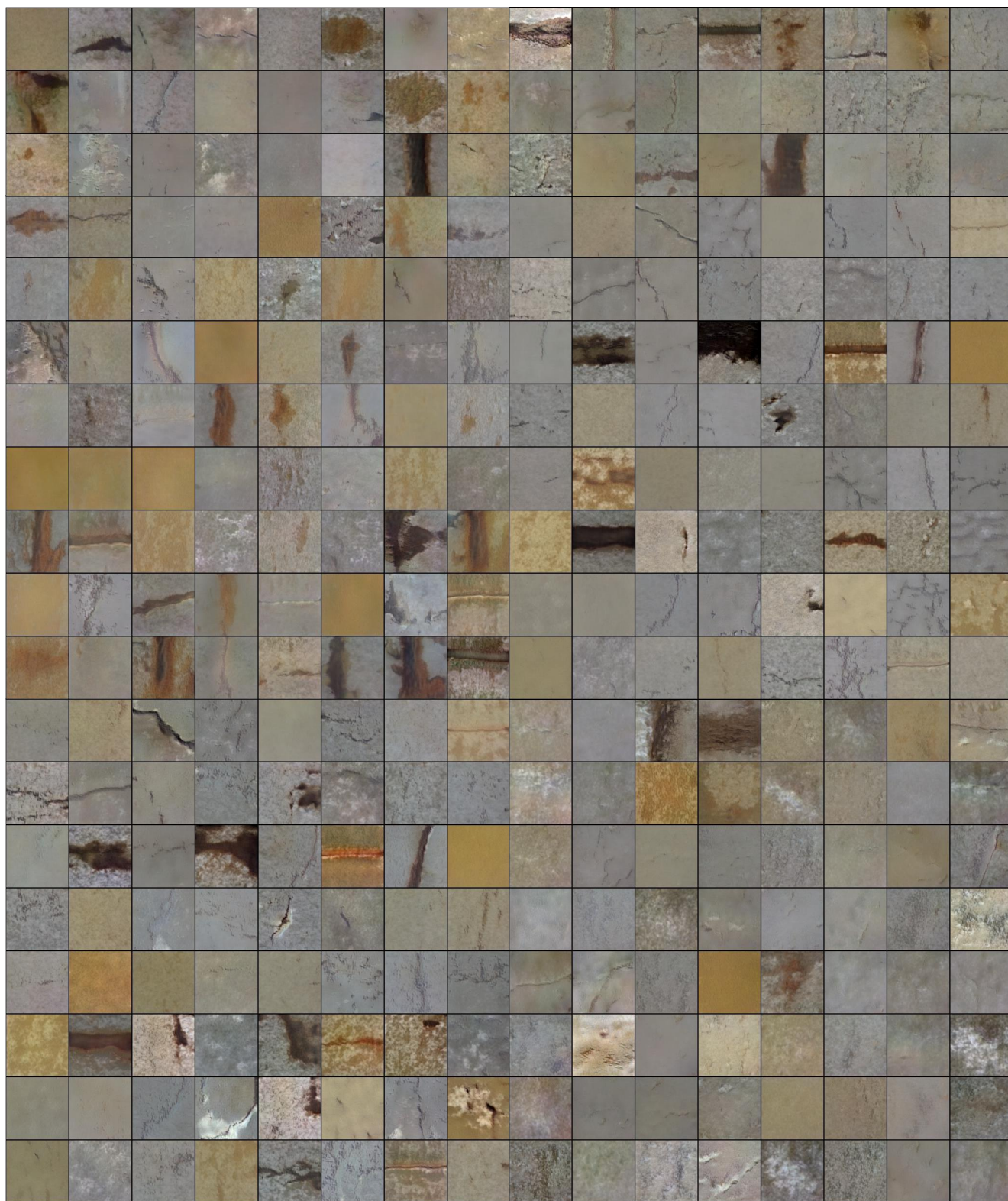


图7. 经StyleGAN v2 w/ DiffAug生成的随机缺陷样本：虽然能够生成部分视觉真实的缺陷样本，但大量生成样本仍不包含任何缺陷，这验证了模型在捕捉和建模复杂不规则缺陷纹理方面存在能力局限。

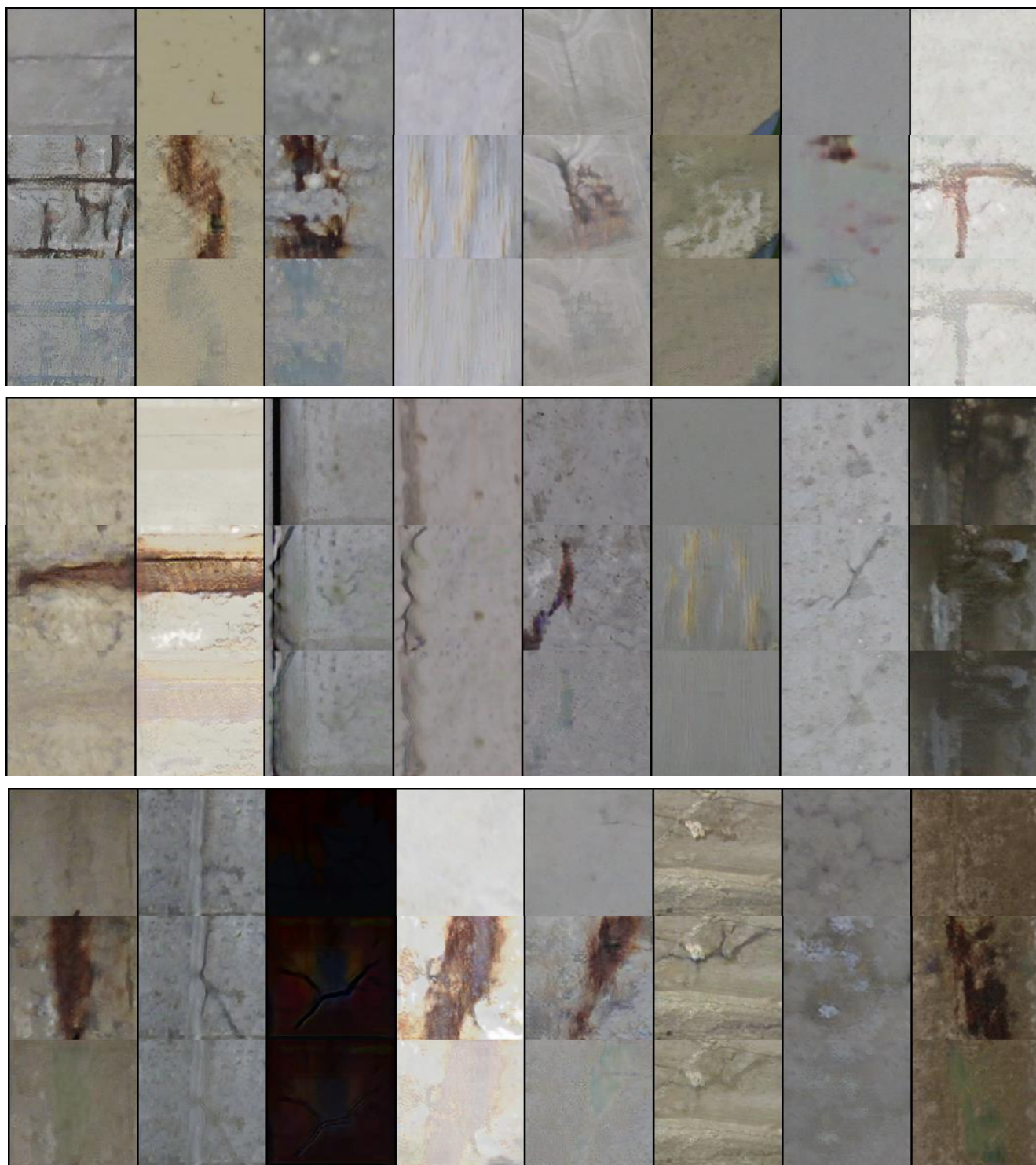


图8. 所提出的Defect-GAN模拟缺损与修复过程的补充结果：每个图块中，第一行为真实正常样本；第二行与第三行分别为生成的缺损样本与修复后的正常样本。



图9. 所提出的Defect-GAN模拟缺损与修复过程的补充结果：每个图块中，第一行显示真实正常样本；第二行与第三行分别展示生成的缺损样本及修复后的正常样本。

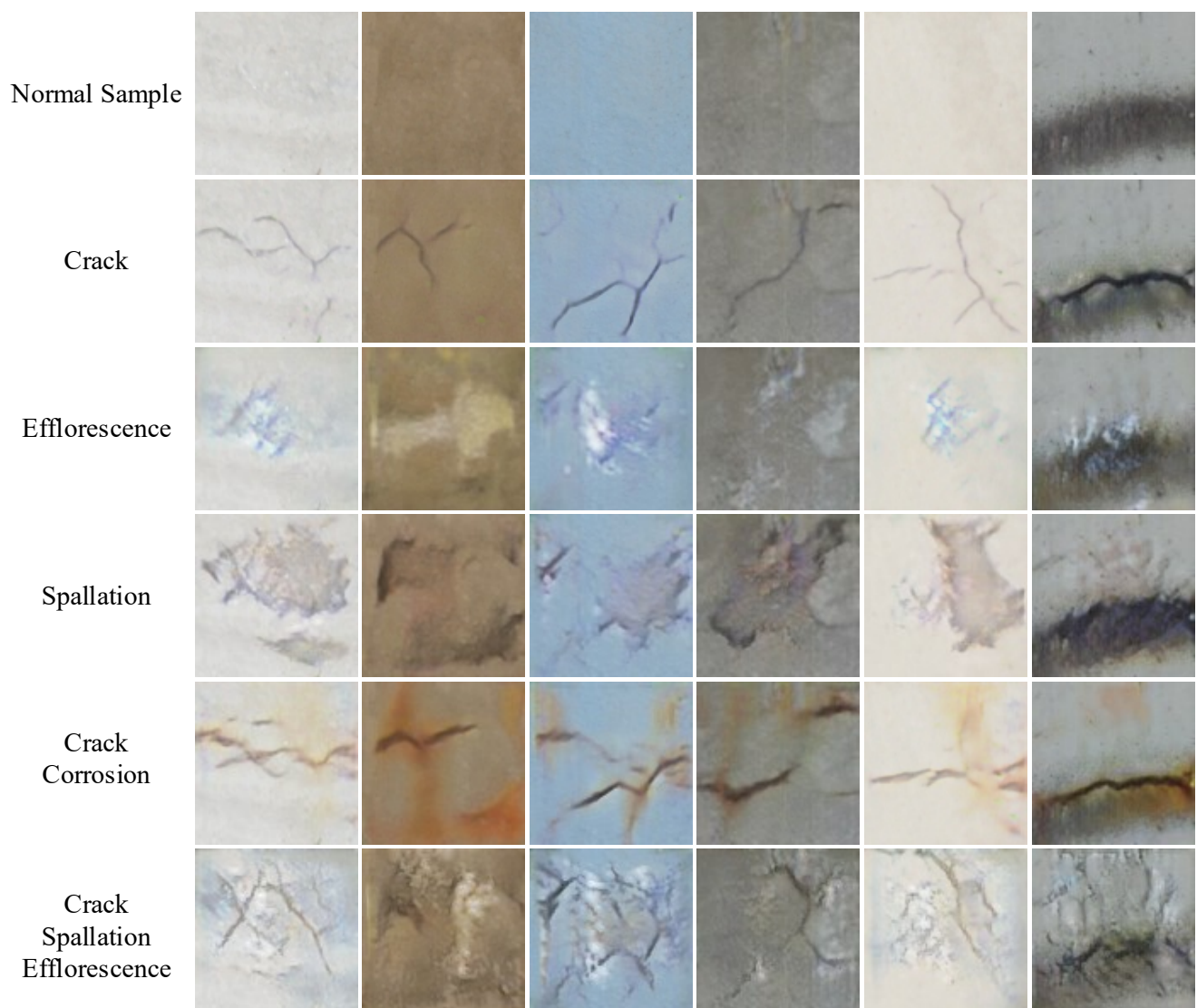


图10. Defect-GAN在缺陷生成中的类别控制示意图：第一行中的每个正常样本分别对应第二至第六行中基于目标类别生成的缺陷样本。

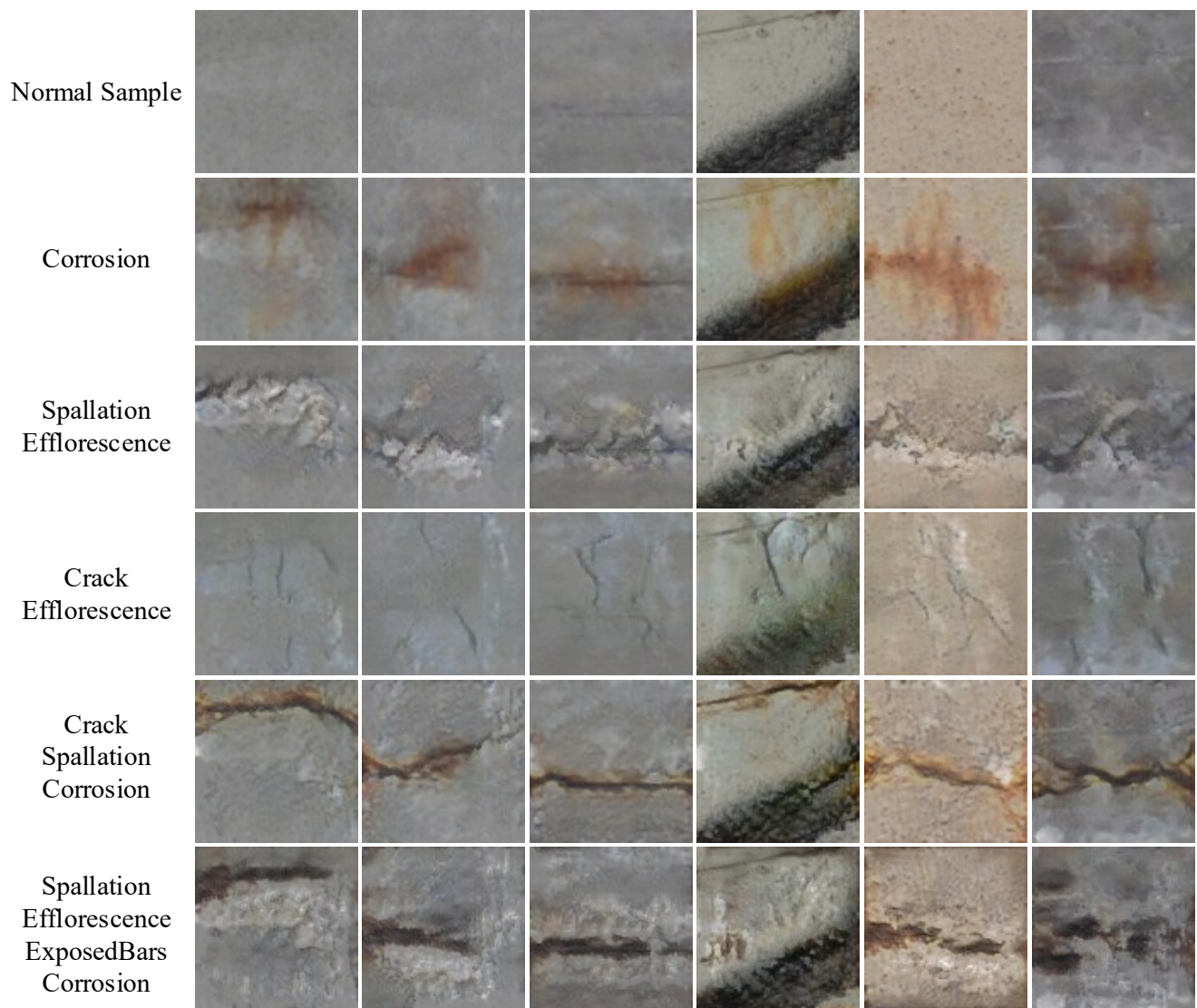


图11. Defect-GAN在缺陷生成中的类别控制示意图：第一行中的每个正常样本分别对应第二至第六行中基于目标类别生成的缺陷样本。