GRIG：少样本生成式残差图像修复

Wanglong Lu、Xianta Jiang、Xiaogang Jin，*IEEE* 会员，Yong-Liang Yang、Minglun Gong、Tao Wang、Kaijie Shi 和 Hanli Zhao*

arXiv:2304.12035v1 [cs.CV] 24 Apr 2023

摘要—图像修复旨在用语义上有意义的内容填补图像中的缺失区域或掩码区域。近期方法在处理大规模缺失区域方面取得了显著进展。然而，这些方法通常需要大量训练数据集才能获得令人满意的结果，而针对如何使用少量样本训练这些模型的研究仍然有限。为此，我们提出一种新颖的少样本生成式残差图像修复方法，可生成高质量的修复结果。其核心思想是提出一种迭代残差推理方法，在生成对抗网络中结合卷积神经网络（CNNs）进行特征提取，并利用 Transformer 进行全局推理，同时采用图像级和图像块级判别器。我们还提出一种新颖的伪造图像块对抗训练策略，以生成真实纹理和细节外观。大量评估结果表明，在少样本图像修复任务中，无论从定量还是定性角度来看，我们的方法都优于以往方法。

索引词—图像修复、小样本学习、迭代推理、残差学习、生成对抗网络。

I. 引言

IMAGE 图像修复是计算机图形学 [1] 和计算机视觉 [2] 中的一项基础任务。它已应用于许多下游任务，例如受损现实场景的图像复原 [3], [4], [5], 和图像操控 [6]。近期提出的图像修复方法在真实图像和人脸图像上都取得了令人瞩目的成果 [2], [7] 。然而，这些方法存在一个被忽视的局限：它们需要大量数据来训练卷积神经网络（CNN）或 Transformer模型 [8]。当使用小型图像数据集训练这些模型时，极有可能出现过拟合和模型崩溃 [9]。实际上，如果只需少量训练图像，用户会方便得多。另一方面，某些领域（例如医疗、艺术和历史文物图像）的大型图像数据集要么收集成本过高，要么难以收集。

W. Lu 和 H. Zhao 任职于浙江省安全与应急智能信息学重点实验室、温州大学，中国温州 325035。W. Lu、X.Jiang 和 K. Shi 任职于纽芬兰纪念大学计算机科学系，加拿大纽芬兰省圣约翰斯 A1B 3X5。X. Jin 任职于浙江大学 CAD&CG国家重点实验室，中国杭州 310058。Y.-L. Yang 任职于巴斯大学计算机科学系，英国巴斯 BA2 7AY。M. Gong 任职于圭尔夫大学计算机科学学院，加拿大安大略省圭尔夫 N1G 2W1。T. Wang 任职于南京大学计算机科学与技术系，中国。

* 通讯作者。电子邮箱：hanlizhao@wzu.edu.cn

这严重限制了图像修复在现实场景中的应用。此外，提高模型训练的数据效率可能是加快其普及和应用的关键因素，从而提升那些受数据可用性限制的各个领域的可及性。

在少样本数据集上实现高质量图像修复仍然是一个具有挑战性且尚未解决的问题。现有的大多数方法 [2], [10], [11] 依赖单次推理，这可能会在待修复子区域生成含义模糊的结果。一些方法 [12], [13] 通过在渐进式方式下进行图像修复，重复利用早期细化阶段中先前修复的部分特征。然而，这些方法没有充分利用修复像素，将其作为下一次迭代的有用信息。此外，现有的图像修复方法并非专门针对小样本学习而设计，因此在训练样本有限的情况下可能无法取得良好效果。在图像修复中，可以采用领域相关先验知识 [14] 或轻量级生成模型 [15], [16]来缓解过拟合问题。然而，当两个任务之间存在较大的领域差距，或网络容量降低影响图像修复质量时，这些方法的表现并不理想。

为此，我们提出了一种新颖的少样本生成式残差图像修复框架（GRIG），能够在少样本数据集上实现高质量图像修复。为了有效优化修复结果，我们采用迭代推理来解决更准确且泛化能力更强的算法推理任务 [17] ，并结合残差学习 [18] 逐步改进先前的估计结果。我们探究将迭代推理和残差学习与卷积神经网络及 Transformer [8], ，以及解耦的图像级与图块级判别器相结合，能否构建出一种有前景的设计，从而打造更稳健、数据效率更高的方法来解决少样本图像修复任务。在该框架中，图像修复过程通过多次前向传播完成：每次都将前一轮的输出及对应掩码输入生成器。这使生成器能够得到充分训练，并专注于先前预测结果与真实值之间的残差信息，从而在少样本数据上实现高质量且泛化能力良好的图像修复。

我们将框架实现为三个组件：生成器、投影判别器和伪造图像块判别器。生成器使用卷积神经网络提取边缘和纹理的浅层特征，并利用Transformer模块捕获每个迭代步骤中特征上下文之间的全局交互。为加速网络收敛并减少过拟合，我们通过解耦图像分布学习，

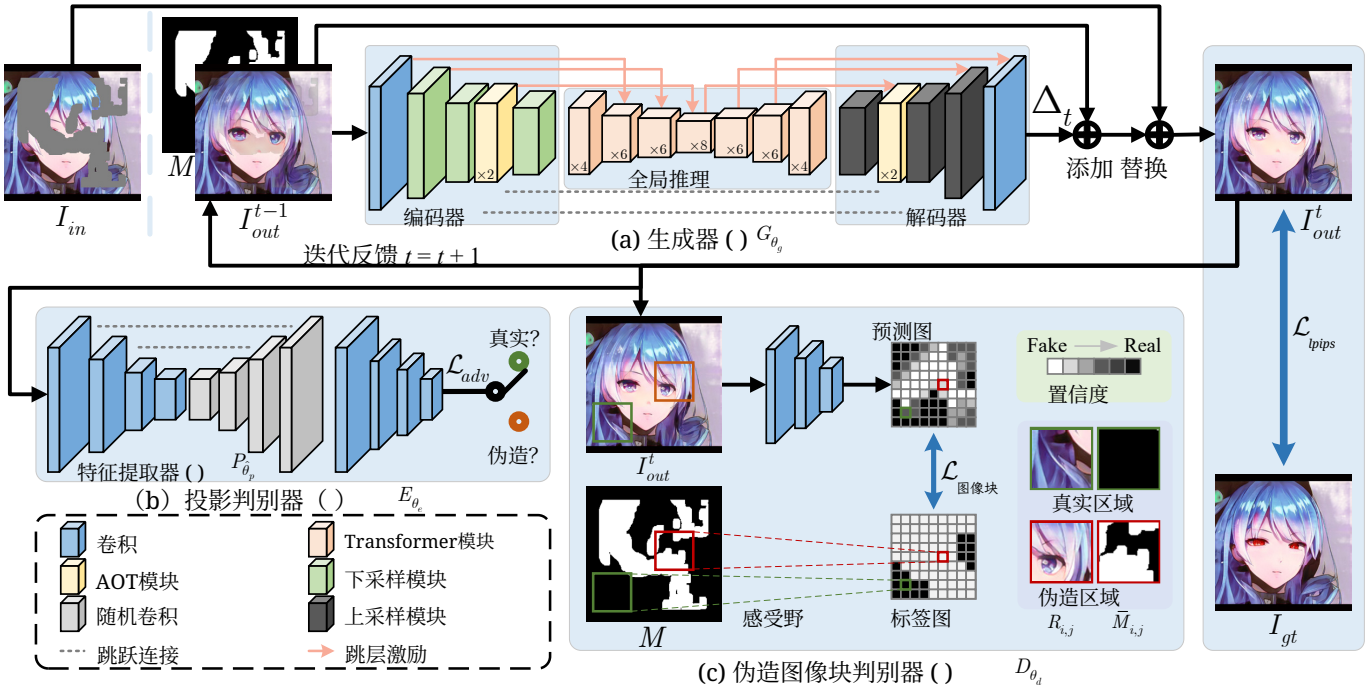

图1。少样本生成式残差图像修复框架（GRIG）的整体流程。对于第 t 个残差推理步骤，生成器 (a) 利用第 $(t-1)$ 个修复图像 I_{out}^{t-1}生成残差图像 Δ_t。在第一个残差推理步骤 $(t=1)$ 中，我们设定 $I_{out}^{0} = I_{in}$。利用修复图像 I_{out}^{t-1}和初始输入图像 I_{in}，执行相加与替换操作，得到用于下一次迭代细化的 I_{out}^{t}。在对抗训练期间，修复图像分别输入投影判别器 (b) 和伪造图像块判别器 (c)。在每个迭代推理步骤中，都会重新计算损失函数并执行相应的反向传播。在测试阶段，会执行类似的多步预测，但不使用损失函数，也不进行反向传播。为简洁起见，省略了下采样与上采样操作。

使用图像级和图块级判别器。我们首先构建投影判别器，以捕获完整的图像级分布。随后，我们提出一种伪造图像块判别器，用于增强生成图像的图块级细节，因为投影判别器难以捕获修复图像中的精细细节。在十个少样本数据集上的实验结果表明，就少样本和高质量图像修复而言，我们的方法优于最先进水平（SOTA）方法。

总而言之，本文的贡献如下：

- 提出一种新颖的少样本生成式残差图像修复框架，整合卷积神经网络和 Transformer，并采用解耦的图像级与图块级判别器，以提升在少样本数据上的性能。
- 提出一种伪造图像块判别器，辅助生成网络改善生成图像的精细细节，并防止少样本图像修复中的过拟合。
- 在十个内容和特征各异的小样本基准数据集上取得最先进的性能，涵盖人脸、照片级真实图像、动物、医学图像、卡通图像和艺术风格图像。

II. 相关工作

早期图像修复技术高度依赖像素和图像块的低级特征。基于扩散的方法 [19]–[21]沿着边界将未受损的信息传播到孔洞中心。基于图像块的方法 [22]–[25] 会迭代地搜索并复制来自图像数据集或已知背景的相似外观。一些变体包括用于总结非平稳模式的基于 GPU 的并行方法 [26]，[27]，以及采用非局部纹理相似性进行图像修复的方法 [28]。由于缺乏对图像的语义理解，这些方法在处理小规模、狭窄的缺失区域时效果良好，但无法为较大的孔洞恢复有意义的内容。

基于深度学习的图像修复方法在语义补全方面取得了巨大成功。深度神经网络已被广泛用于提升图像修复的视觉质量 [29]。这些工作包括基于自动编码器的架构 [30] 及其变体架构 [31]–[33]。研究者开发了多种复杂模块或学习策略，以提升图像修复的效果，包括全局和局部判别器 [34]，、上下文注意力 [35]–[38] （用于改善语义理解）、处理不规则孔洞的方法 [10], [11], [39], ，以及利用辅助信息（例如草图 [6]，、前景轮廓 [40], 和结构 [41]）。近期研究还探讨了高分辨率 [2], [7], [12], [37], [42], [43]，、3 D 摄影 [1]，、多样化生成 [2], [10], [44]–[46],以及大孔洞填充 [2], [10], [13], [43], [45]–[49]等问题。上述方法旨在实现高语义质量的补全，但在样本数量较少的数据上训练时可能会过拟合。

基于渐进式的图像修复方法 [12], [13],[47]–[49]与我们的工作密切相关。这些方法主要以渐进方式从孔洞边界向中心修复像素 [12], [47]–[49] ，或采用多阶段

细化方案 [12], [13]。例如，Zeng 等人 [12]通过迭代预测置信度图及相应的中间结果，改进了高分辨率图像修复。这类方法仅重复利用部分预测信息，并且在下一次迭代图像修复中不会改变置信度较高的像素。循环特征推理（RFR） [13] 多次将嵌入的特征图输入其特征推理模块，以生成多个特征，用于自适应特征融合。另一方面，RFR 的最终图像修复结果由其解码器通过单次前向传播生成，这表明该模型无法在像素级重新调整结果以获得更精细的细节。在本文中，我们首次尝试使用少量样本进行图像修复训练。我们展示了框架如何通过迭代残差推理细化图像修复结果；该方法巧妙地结合了卷积神经网络和 Transformer，以及图像级和图像块级判别器。我们的方法能够高效地重复利用此前预测的全部像素和网络参数，从而为少样本图像修复任务带来有效的模型。

III. 方法

迭代推理通过反复将底层计算应用于先前推理步骤的输出，有望解决更准确且更具泛化能力的算法推理任务 [17], , 而残差学习 [18], [50] 则有助于逐步优化先前预测的结果。每个推理步骤中，我们的 GRIG 通过迭代预测残差偏移，并在同一生成器中重用先前预测的信息，从而动态学习在每一步细化输入图像。该方法避免了记忆输入与其真实图像之间的映射，因此能够防止过拟合，并提升小样本图像修复的视觉质量。

如图1所示，我们的框架由三个主要部分组成：生成器、投影判别器和伪造图像块判别器。给定一幅真实图像 $I_{gt} \in \mathbb{R}^{h\times w\times 3}$ 和一个二值掩码 $M \in \mathbb{R}^{h\times w\times 1}$ （其中 1 表示未知像素，0 表示已知像素），掩码图像 $I_{in} \in \mathbb{R}^{h\times w\times 3}$ 可表示为 $I_{in} = I_{gt} \odot (1 - M)$，其中 $\odot$表示 Hadamard积。GRIG 的目标是通过 $T > 1$ 个迭代推理步骤自动修复出一幅逼真的图像 $I_{out}^{T} \in \mathbb{R}^{h\times w\times 3}$ ，其中 $T > 1$ 表示训练期间的迭代推理步骤。对于第 t个迭代残差推理步骤，将先前的修复图像 I_{out}^{t-1}输入生成器以获得残差预测。在第一个残差推理步骤（$t = 1$）中，我们设定 $I_{out}^{0} = I_{in}$。随后执行加法和替换操作，生成新的图像补全结果 I_{out}^{t}。每个迭代步骤都会进行对抗训练，并通过反向传播相应地更新网络权重。

A. 网络架构

1) 生成器：生成器以前一次迭代的图像修复结果作为输入，旨在结合卷积神经网络和 Transformer [8] ，以在少样本图像修复中进行高效的迭代残差推理。生成器 G_{θ_g} 由编码器、由一系列 Restormer 的 Transformer模块组成的全局推理模块 [51],以及解码器构成（见图1a）。基于卷积神经网络的编码器和解码器在特征提取方面表现出色，而 Transformer模块则擅长动态注意力、全局上下文整合和泛化。这种组合有助于生成器在少量训练样本上实现有效泛化。为了提取特征并扩大感受野，以捕获具有信息量的远距离图像上下文和丰富的关注模式，我们首先堆叠一个卷积层、若干残差下采样模块 [15], 和 AOT模块 [42]。随后，将提取的特征输入 Restormer 的 Transformer模块堆栈，以进行全局上下文推理。同时，采用跳层激励模块（SLE） [15]实现快捷的梯度流，并通过跳跃连接收集解码器中的多分辨率特征图。随后，使用上采样模块 [15],、AOT模块 [42], 和一个卷积层构建解码器。解码器利用编码器和全局推理模块输出的多分辨率特征图生成中间预测结果 Δ_t 。为实现稳定的对抗训练，我们对网络的全部卷积层应用谱归一化 [52]。

在每个第 t 个迭代推理步骤中，将前一次迭代得到的修复图像 I_{out}^{t-1}及其对应的掩码M （见图1a）输入带有可学习网络参数 θ_g的生成网络G_{θ_g}。 G_{θ_g}生成中间残差图像修复结果 $\Delta_t = G_{\theta_g}(I_{out}^{t-1}, M) \in \mathbb{R}^{h\times w\times 3}$。随后，第 t 个修复图像 I_{out}^{t} 计算如下：

$$I_{out}^{t} = (I_{out}^{t-1} + \Delta_t) \odot M + I_{in} \odot (1 - M). \qquad (1)$$

2）投影判别器：为了稳定生成对抗网络训练并提高数据效率，我们利用预训练表示中的先验知识，训练紧凑型分类器，以学习少样本图像的全局分布。投影判别器（见图1b）学习为从真实图像中提取的特征图分配较高的置信度分数，同时为合成图像分配较低的分数。首先，使用带有预训练ˆ网络参数 θ_p 的类 U-Net 投影器 $P_{\hat{\theta}_p}$ ，从输入图像 I （即 I_{out}^{t} 或 I_{gt}）中提取特征图。$P_{\hat{\theta}_p}$ 由经过预训练的 EfficientNet-Lite1 [54] 实现，并采用跨通道混合和跨尺度混合机制 [16]。随后，带有可学习网络参数 θ_e 的投影判别器 E_{θ_e} 将提取的特征图映射为一个标量。这里，我们从 Projected GAN [16] 中选取具有最大特征投影尺度的判别器（即移除其他三个小尺度判别器）。

3）伪造图像块判别器：由于投影判别器主要专注于提取全局图像特征以实现鲁棒分类，因此这些投影特征中可能会忽略一些细节特征。为了帮助生成器生成逼真的细粒度纹理，并避免小样本训练中过拟合，我们提出了一种伪造图像块判别器，它根据判别器的感受野 [55] 学习识别真实图像块和修复补丁。

如图1c所示，伪造图像块判别网络 D_{θ_d} 利用可学习的网络参数 θ_d 学习

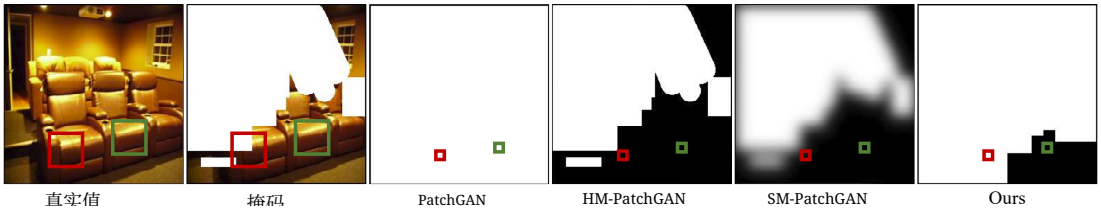

图2。PatchGAN、HM-PatchGAN、SM-PatchGAN 与我们的算法中判别器之间差异的示意图。方框表示判别器感受野大小的图像块（左侧两幅图像），以及这些图像块在结果标签图中的对应投影位置（右侧四幅图像），标签图覆盖（红色）掩码区域和（绿色）未掩码区域；标签图中的像素值表示伪造（白色）和真实（黑色）图像块的标签。

从给定图像 I（即 I^t_{out} 或 I_{gt}）中识别真实或伪造的图像块。判别器 D_{θ_d} 将 I 映射为预测图，其中每个单元都基于感受野指示相应图像块的置信度分数。在本文中，我们采用了来自 PatchGAN [53]的网络架构 D_{θ_d}。神经网络中的图像块级感受野也被研究用于交互式视频风格化 [56]中的过拟合规避，以及提升图像生成 [57]的多样性和泛化能力。

B. 目标函数

GRIG 通过以下目标函数进行训练，以优化可学习的网络参数 θ_g、θ_e 和 θ_d。

1) LPIPS 损失：在每个迭代推理步骤中，我们使用学习感知图像块相似度（LPIPS）度量 [58]，以约束修复图像 I^t_{out} 与真实图像 I_{gt} 之间的感知相似度：

$$\mathcal{L}_{lpips}(\theta_g) = \|F(I^t_{out}) - F(I_{gt})\|_2, \tag{2}$$

其中，$F(\cdot)$ 是预训练的感知特征提取器；在我们的工作中，我们使用 VGG-16 [59]。这有助于我们的生成网络学习保持较高的视觉质量。

2) 投影对抗损失：为实现快速收敛，投影对抗损失利用预训练的分类模型提取先验知识（见图1b）。我们采用铰链损失 [16] 分别优化投影判别器 E_{θ_e} 和生成网络 G_{θ_g}。目标函数可表示为：

$$\begin{aligned}\mathcal{L}^E_{adv}(\theta_e) &= \mathbb{E}_{I_{gt}}[ReLU(1 - E_{\theta_e}(P_{\hat{\theta}_p}(I_{gt})))] \\ &\quad + \mathbb{E}_{I^t_{out}}[ReLU(1 + E_{\theta_e}(P_{\hat{\theta}_p}(I^t_{out})))], \\ \mathcal{L}^G_{adv}(\theta_g) &= -\mathbb{E}_{I^t_{out}}[E_{\theta_e}(P_{\hat{\theta}_p}(I^t_{out}))].\end{aligned} \tag{3}$$

投影判别器被约束为对修复图像赋予低分，对真实图像赋予高分；而生成器 G_{θ_g} 则由投影判别器监督，根据真实图像的分布修复掩码输入。

3）对抗式伪造图像块损失：如图1c所示，我们要求伪造图像块判别器区分给定图像中的伪造图像块和真实图像块。为此，我们构建相应的标签图 $X \in \mathbb{R}^{h' \times w'}$，以监督判别器。具体而言，我们根据伪造图像块判别器 D_{θ_d} 的感受野，将 I 和 M 划分为 $h' \times w'$ 对部分重叠的图像块（$R_{i,\,j}$ 和 $M_{i,\,j}$）。其中，$1 \le i \le h'$ 和 $1 \le j \le w'$ 是水平和垂直索引，$R_{i,\,j}$ 和 $M_{i,\,j}$ 的尺寸均等于感受野 $N \times N$。标签图表示如下：

$$X_{i,j} = \begin{cases} 0 & \text{if } \|\overline{M}_{i,j}\|_0 = 0; \\ 1 & \text{otherwise,} \end{cases} \tag{4}$$

其中，$\|\overline{M}_{i,\,j}\|_0$ 定义为子区域掩码 $M_{i,\,j}$的 L0范数。如果 $\|M_{i,\,j}\|_0$ 不为零，则说明该子区域掩码中存在一些掩码像素，且图像块$R_{i,\,j}$ 包含修复像素。因此，我们设定 $X_{i,\,j} = 1$，这意味着子区域$R_{i,\,j}$ 是一个伪造区域。否则，它就是真实区域。对抗性伪造区域损失的铰链版本表示为：

$$\begin{aligned}\mathcal{L}^D_{patch}(\theta_d) &= \mathbb{E}_{I_{gt}}[ReLU(1 - D_{\theta_d}(I_{gt}))] \\ &\quad + \mathbb{E}_{I^t_{out}}[ReLU(1 - D_{\theta_d}(I^t_{out})) \odot (1 - X)] \\ &\quad + \mathbb{E}_{I^t_{out}}[ReLU(1 + D_{\theta_d}(I^t_{out})) \odot X], \\ \mathcal{L}^G_{patch}(\theta_g) &= -\mathbb{E}_{I^t_{out}}[D_{\theta_d}(I^t_{out}) \odot X].\end{aligned} \tag{5}$$

图2展示了所提出的伪造图像块判别器与密切相关的判别器之间的差异。PatchGAN 的判别器 [53] 直接将修复图像中的所有图像块判定为伪造图像块；当提取的图像块不包含任何生成像素时，这可能会使判别器产生混淆。HM-PatchGAN 和 SM-PatchGAN [42] 旨在根据图像修复掩码分割缺失区域的合成图像块。由于必须先对图像修复掩码进行下采样，以使其与判别器输出的空间尺寸对齐，因此掩码边界周围的约束可能不够明确。例如，对图像修复掩码进行下采样会导致修复像素精确位置的信息丢失。SM-PatchGAN 试图区分生成图像块和真实图像块，而我们的判别器更进一步，考虑给定图像块中的生成像素是否与周围的真实像素一致。我们的判别器根据感受野构建标签图，并将包含任何修复像素的所有图像块视为伪造图像块，因此比 PatchGAN 和 SM-PatchGAN 提供了更多约束。

4) 总目标函数：生成器的总训练目标函数表示为：

$$\mathcal{L}^G_{total} = \lambda_{lpips}\mathcal{L}_{lpips} + \lambda_{adv}\mathcal{L}^G_{adv} + \lambda_{patch}\mathcal{L}^G_{patch}, \tag{6}$$

其中，λ_{lpips}、λ_{adv} 和 λ_{patch} 分别是相应损失的权重。在训练过程中，我们交替优化参数 θ_g、θ_e 和 θ_d。

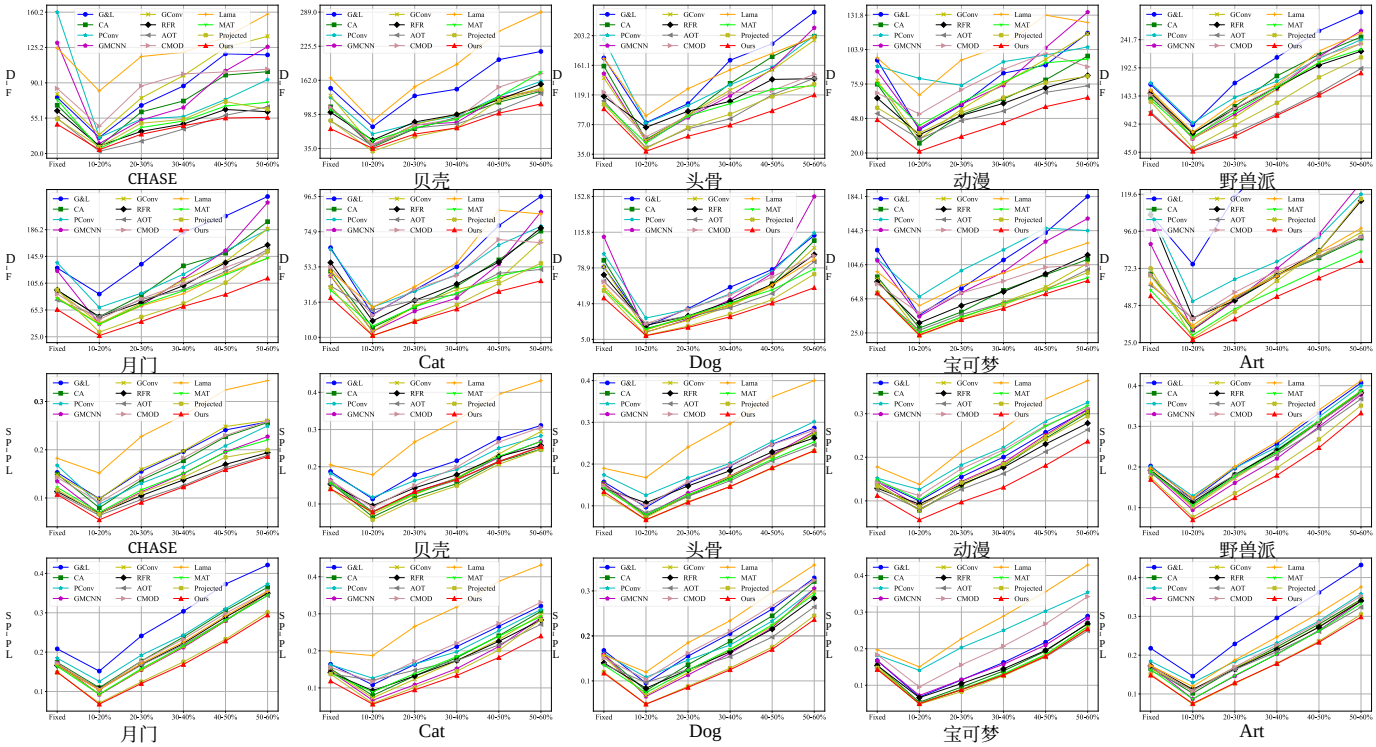

图3. GRIG 与最先进水平的图像修复方法在十个少样本数据集和两个评估指标上的定量比较（上两行为 FID，下两行为 LPIPS）。“固定”表示固定的中心 25% 矩形掩码。

算法1 GRIG 的训练过程

当 G_{θ_g}、E_{θ_e} 和 D_{θ_d} 尚未收敛时**执行**
  从训练集中采样批量图像 $\mathcal{I}_{gt}$
  创建随机掩码 $\mathcal{M}$ 用于 $\mathcal{I}_{in}$
  获取输入 $\mathcal{I}_{in} \leftarrow \mathcal{I}_{gt} \odot (1-\mathcal{M})$
  设置输入 $\mathcal{I}_{out}^{0} \leftarrow \mathcal{I}_{in}$
  对于 迭代残差推理步骤 $t=1$ 至 T **执行**
    获取 $\Delta_t \leftarrow G_{\theta_g}(\mathcal{I}_{out}^{t-1}, \mathcal{M})$
    获取 $\mathcal{I}_{out}^{t} \leftarrow (I_{out}^{t-1}+\Delta_t) \odot \mathcal{M} + \mathcal{I}_{in} \odot (1-\mathcal{M})$
    更新G_{θ_g} 为 $\mathcal{L}_{total}^{G}$
    更新E_{θ_e} 为 $\mathcal{L}_{adv}^{E}$
    更新 D_{θ_d} 为 $\mathcal{L}_{patch}^{D}$
  结束 for 循环
结束 while 循环

C. 迭代残差推理

图像修复的迭代残差推理可以表述为对抗生成网络上的优化过程。这使生成器能够隐式地学习利用先前预测的结果，并专注于残差信息，从而实现高质量和更强的泛化能力。

我们引入生成网络 $G_{\theta_g}(I_{out}^{t-1})$ 作为预测残差信息的显式函数（见公式1）。在每个迭代推理步骤 t中，生成器 G_{θ_g} 经过训练，以最大化 $E_{\theta_e}(I_{out}^{t})$的置信度值以及 $D_{\theta_d}(I_{out}^{t})$预测图中的值，同时最小化 I_{out}^{t} 与 I_{gt}之间的感知相似度。因此， θ_g 通过 $\theta_g = \arg\min_{\theta_g} \mathcal{L}_{total}^{G}$求解。同时，$E_{\theta_e}$和 D_{θ_d}经过训练以分别区分图像（伪造或真实）和图像块，其中 $\theta_e = \arg\min_{\theta_e} \mathcal{L}_{adv}^{E}$ 以及 $\theta_d = \arg\min_{\theta_d} \mathcal{L}_{patch}^{D}$。生成网络 G_{θ_g} 直接预测残差信息，而其参数 θ_g 则由判别器以及I_{gt}进行监督。GRIG 训练过程的伪代码列于算法1中。

IV. 实验

A. 实验设置

我们使用 Python 和 PyTorch 构建了所提出的框架。除非另有说明，在训练阶段和测试阶段的所有实验中，我们均设置 $\lambda_{lpips}=1.5$、$\lambda_{adv}=1$、$\lambda_{patch}=1$ 和 $T=3$。我们使用 Adam优化器，其第一动量系数为 $\beta_1=0.5$，第二动量系数为 $\beta_2=0.999$，学习率为 0.0002。我们的掩码使用 CMOD掩码生成算法 [10] 创建。我们的生成器包含 31.76M 个参数，在 NVIDIA GeForce RTX 2080 GPU（8 GB）上，每个残差推理步骤的处理速度约为 21 FPS。

我们将 GRIG 与最先进水平的图像修复方法进行了比较，包括 Globally&Locally（G&L） [34], 上下文注意力（CA） [38], 部分卷积（PConv） [39], GMCNN [31],门控卷积（GConv） [11], 循环特征推理（RFR） [13], AOT-GAN（AOT） [42], Co-mod-GAN（CMOD） [10], Lama [7], 以及 MAT [2]。我们还将 GRIG 与一种基于轻量级 Projected GAN [16] 的图像修复模型（Projected）进行了比较，以进一步证明 GRIG 在小样本图像修复方面的优越性。公开可用的 MMEditing框架 [60], 是一个开源图像

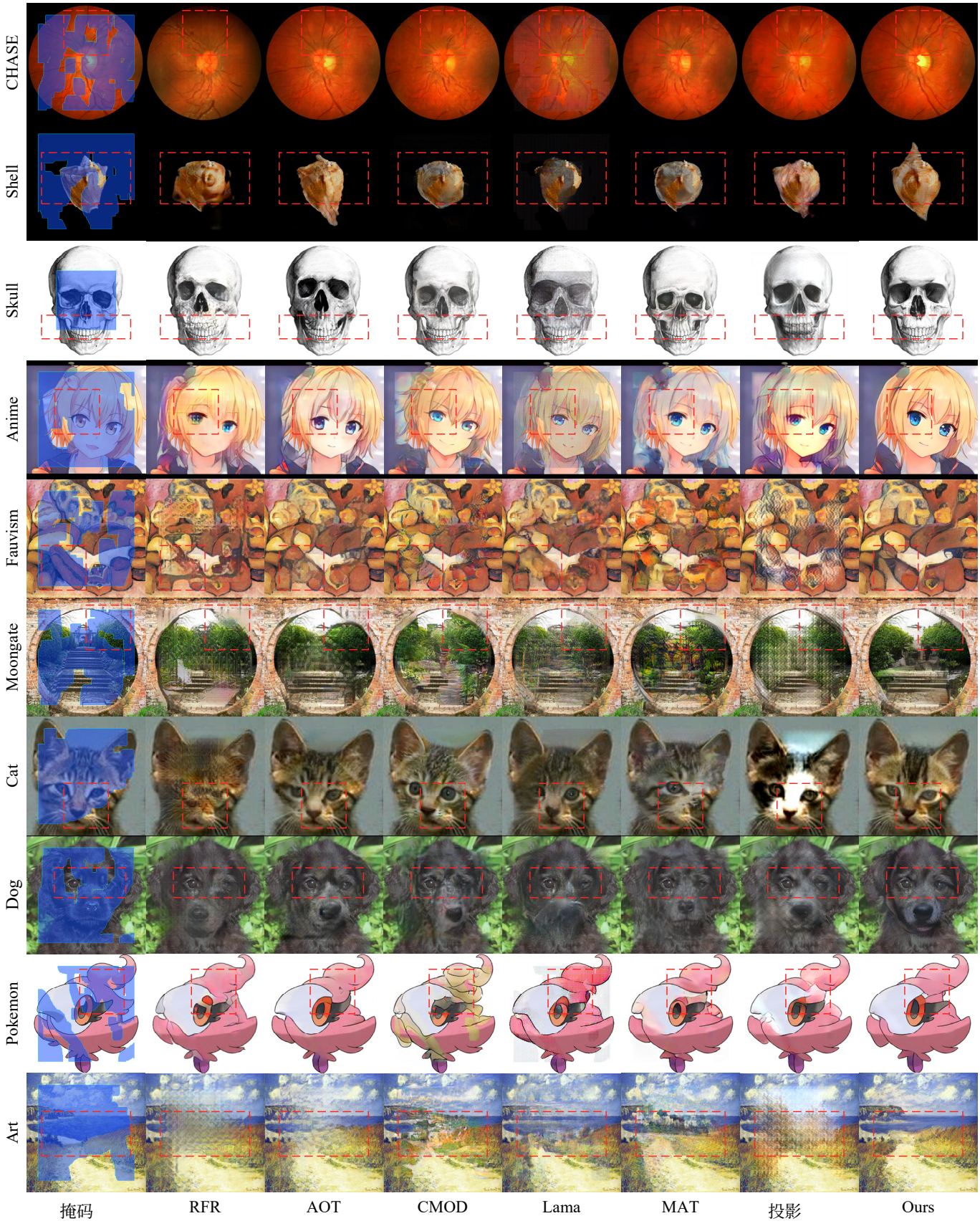

图4。GRIG 与最先进水平的图像修复方法在少样本数据集上的视觉比较。

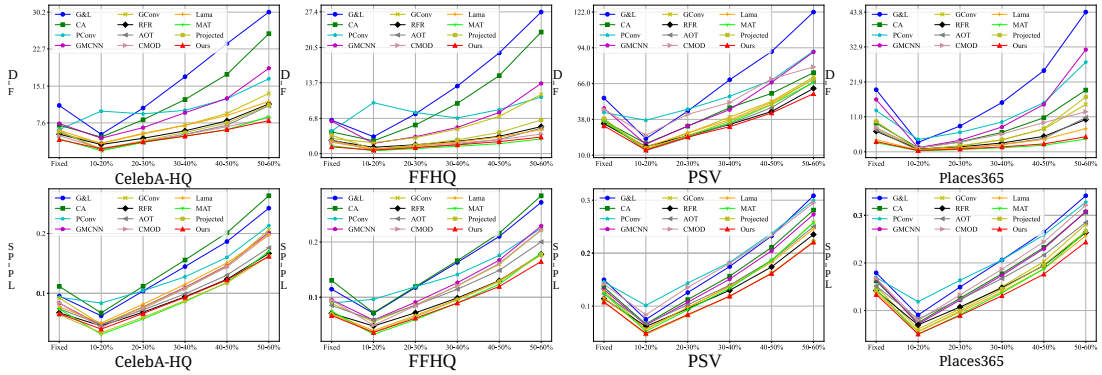

图 5. GRIG 与最先进的图像修复方法在四个大型数据集及两项评估指标（FID 和 LPIPS）上的定量比较。“固定”表示固定的中心 25% 矩形掩码。

表 I 十个实验用少样本图像数据集的详细信息。

数据集	# 训练集	# 测试集
CHASE [62]	18	10
贝壳 [15]	48	16
头骨 [15]	72	25
动漫 [15]	90	30
野兽派 [15]	94	30
月门 [15]	106	30
猫 [63]	120	40
狗 [63]	309	80
宝可梦 (pokemon.com)	633	200
艺术（wikiart.org）	750	250

以及基于 PyTorch 的视频编辑工具箱，实现了 G&L [34], CA [38], 部分卷积 [39],和 GConv [11]模型。我们使用了 GMCNN [31], RFR [13],AOT [42], Lama [7], 和 MAT [2] 的官方代码。我们使用官方的基于 TensorFlow 的代码创建了基于 PyTorch 的 CMOD [10]版本。为了实现 Projected 模型，我们为 Projected GAN [16]添加了带有跳跃连接的镜像编码器以及感知相似度 $\mathcal{L}_{lpips}$。该 Projected 模型使用 PyTorch 创建，并采用了与 GRIG 相同的超参数 $\lambda_{lpips} = 1.5$。

为确保公平性，所有实验均采用相同的训练集/测试集划分。所有图像均调整为 256×256 的分辨率。除非另有说明，所有对比模型均在论文中提到的数据集上重新训练，批量大小设为 8。所有模型均在 NVIDIA V100 GPU（32 GB）上进行训练和测试。在测试期间，所有实验均使用具有不同掩码比例 [39]的不规则掩码以及固定中心 25%（128×128）的矩形掩码，以模拟不同情况。

由于L1距离、峰值信噪比和结构相似性指标都严重偏向于模糊结果 [10], ，因此我们遵循近期文献中的既有做法，使用 Fr´echet初始距离（FID） [61] 和LPIPS指标进行定量评估 [7]。

B. 少样本数据集上的比较结果

实验在十个少样本图像数据集上进行。数据集的详细信息如表 I 所示。全部数据集均通过随机采样进行划分。所有模型均使用 400，000 个图像批次进行训练。

图3展示了GRIG与最先进水平的图像修复方法在十个少样本数据集上的定量比较结果。为比较图像修复性能，我们使用了具有不同比例掩码的各种不规则掩码 [39] ，以及固定中心25%（128×128）的矩形掩码，以模拟各种场景。在少样本设置下，从训练阶段采样图像时，所有参与比较的方法都应用了可微数据增强 [64]。如图3所示，在大多数基准数据集上，针对各种类型的掩码，GRIG在FID和LPIPS指标方面均以较大优势超过了所有基线方法。对于大多数数据集，我们的方法都取得了显著提升。值得注意的是，在50%至60%的掩码比例下，与排名第二的方法相比，GRIG在FID方面分别实现了10.03%（CHASE）、16.26%（贝壳）、11.03%（头骨）、14.49%（动漫）、4.24%（野兽派）、26.48%（月门）、15.61%（猫）、23.86%（狗）、3.70%（宝可梦）和7.31%（艺术）的相对提升。

图4展示了各种对比方法的图像修复结果。结果表明，对于训练样本少于100个的数据集（例如 CHASE、贝壳、头骨和动漫），大多数方法都会因过度拟合少量样本中的特征和模式而无法生成合理的内容。在包含500多个样本的数据集（例如宝可梦和艺术）上训练时，部分方法能够在掩码区域填充更多语义内容。然而，仔细观察仍能发现伪影。当掩码区域较大时，RFR [13] 容易在修复区域生成重复图像块。AOT [42]和 Lama [7]能够修复缺失区域中的结构，但会在精细细节处留下伪影。CMOD [10] 和 MAT [2] 由于可学习参数数量庞大，往往会过拟合训练数据。Projected GANs [16] 能够处理语义结构，但可能会在掩码边界附近造成颜色不一致。图3和图4表明，尽管我们的方法拥有比 GConv [11] （4.0M）更多的可学习参数（31.76M），且仅使用有限样本进行训练，GRIG 仍能在定量指标和视觉质量方面取得更好的表现。GRIG 在各种具有不同训练样本数量的少样本数据集上展现出强大的泛化能力，并

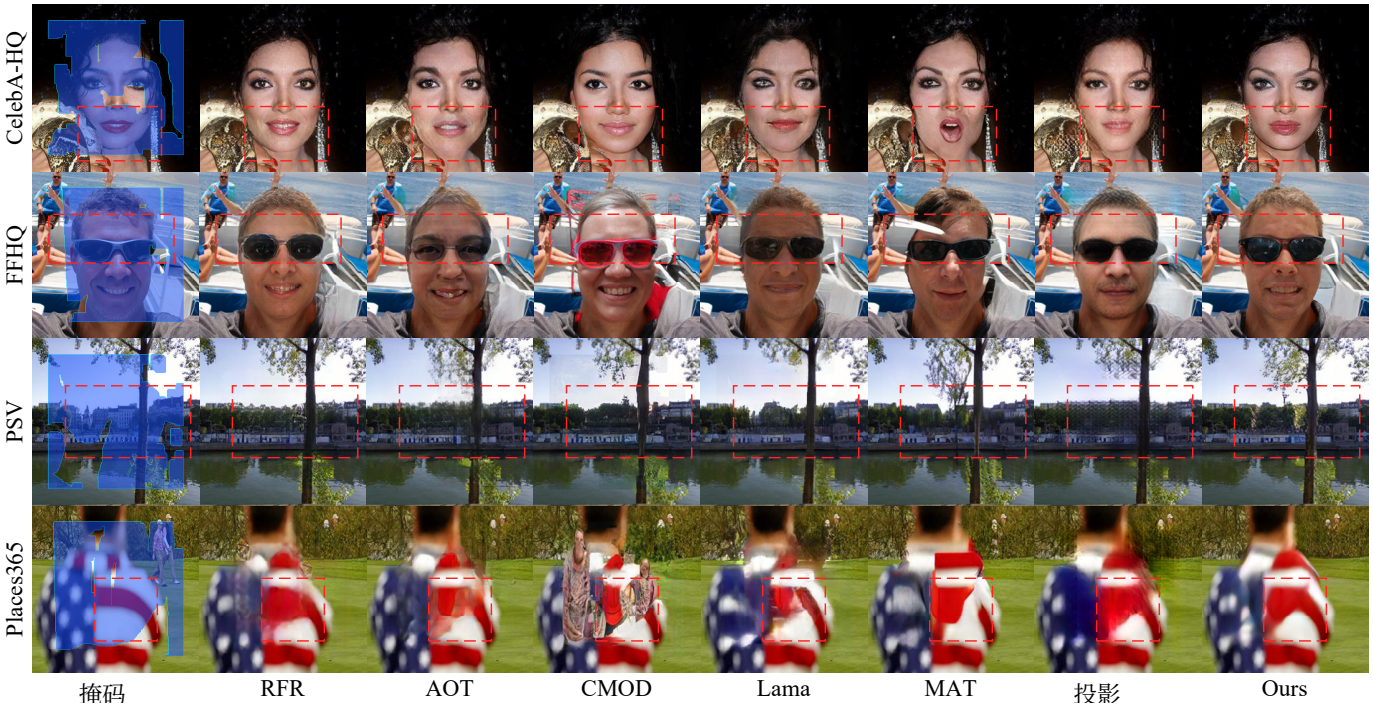

图6. GRIG与最先进水平的图像修复方法在大规模数据集上的视觉比较。

生成视觉质量更高的图像。

我们认为，在小样本图像修复中能够实现更好的泛化性能和图像修复质量有三个原因。首先，我们的迭代残差推理策略使生成器能够利用前几次迭代中学习到的信息，同时增加输入的多样性，从而提升性能。其次，Transformer中的自注意力机制 [8]在利用现有信息进行进一步上下文推理方面具有优势。在我们的生成器中，编码器和解码器用于提取局部特征，而Restormer Transformer模块 [51] 用于全局上下文推理。第三，投影判别器和伪造图像块判别器分别包含2.829M和2.765M个可学习参数，有助于提升我们方法的泛化能力。投影判别器基于预训练特征的通用性，聚焦于语义层面的图像。伪造图像块判别器则专注于学习图像块级一致性，以捕捉图像块统计信息，并区分真实图像块和修复补丁，从而避免记忆整幅图像的需要，防止过拟合。

C. 大规模数据集上的比较结果

我们还在四个大规模数据集上将我们的方法与最先进水平的图像修复方法进行了比较：CelebA-HQ [65]（训练集 28K，测试集 2K）、FFHQ [66]（训练集 60K，测试集 10K）、巴黎街景 [67]（PSV，训练集 14.9K，测试集 100）以及 Places365 [68]（训练集 1.8M，测试集 36.5K）。对于 PSV 和 Places365 数据集，我们直接使用了原始的训练集和测试集划分，而其他数据集则采用随机抽样进行划分。为确保公平，所有方法均使用其默认设置进行训练比较。我们的模型分别在 CelebA-HQ、FFHQ 和 PSV 上使用 1，000，000 个图像批次进行训练，在 Places365 上使用 2，000，000 个图像批次进行训练。

图 5 中的定量结果表明，在大规模数据集上，就 FID 和 LPIPS 指标而言，GRIG 的表现优于大多数最先进水平的图像修复方法。具体而言，GRIG 在 PSV 上取得了最佳 FID 分数，并在 PSV 和 Places365 上取得了最佳 LPIPS 分数。MAT [2] 在 FFHQ 和 Places365 上取得了最佳 FID 分数。总体而言，在其他大规模数据集上，GRIG 的表现与 MAT 相当，但其可学习权重少得多（31.76M），而 MAT 为 62.0M。我们的迭代残差学习有效地帮助网络将图像修复过程分解为多个推理步骤，并逐步细化图像修复结果。此外，通过投影判别器和伪造图像块判别器，将图像分布学习解耦为图像级和块级约束，有助于我们的 GRIG 模型在少样本场景和大规模数据集上均取得出色表现。

图 6 展示了相应的定性性能。结果表明，对于大掩码，语义图像修复对大多数图像修复方法来说仍然很困难。RFR [13] 在修复区域生成重复图像块，因为在特征空间中进行迭代细化可能会忽略图像空间中的精细细节。AOT [42] 和 CMOD [10] 在这些数据集上表现良好。然而，面对复杂背景时，它们在某些情况下难以处理较大的掩码区域。由于一次性推理无法重新调整图像修复结果，复杂背景可能会对 MAT 的图像修复质量产生负面影响。由于投影特征中的精细细节很容易被忽略，基于投影的模型 [16]往往会生成模糊的结果。总体而言，所提出的方法，

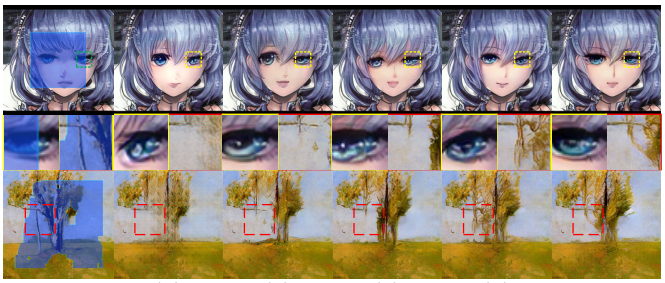

图7. 消融研究的视觉示例：(A) 不含伪造图像块判别器的 GRIG；(B) 不含投影判别器的 GRIG；(C) 将伪造图像块判别器替换为 PatchGAN 判别器的 GRIG；(D) 将伪造图像块判别器替换为 SM-PatchGAN 判别器的 GRIG；以及我们的方法 GRIG。

表II
消融研究的定量结果（FID）：(A) 不含伪造图像块判别器的 GRIG，(B) 不含投影判别器的 GRIG，(C) 将 GRIG 的伪造图像块判别器替换为 PatchGAN 的判别器，(D) 将 GRIG 的伪造图像块判别器替换为 SM-PatchGAN 的判别器，以及 GRIG（我们的方法）。结果在 $50-60\%$ 掩码比例下进行评估。**Bold**：数量指标排名第一。

数据集	(A)	(B)	(C)	(D)	Ours
CHASE	73.96	57.00	56.89	64.17	**55.84**
动漫	77.49	68.56	66.03	71.12	**65.05**
Dog	65.33	62.92	61.17	59.83	**58.49**
Art	96.84	79.19	79.16	78.83	**77.32**
CelebA-HQ	10.14	8.92	8.41	8.96	**8.06**
PSV	60.53	61.76	59.62	61.07	**58.08**

Lama [7], 和 MAT [2] 能够在高掩码比例下修复复杂结构中的合理内容。

D. 消融研究

1) 判别器的消融研究：我们进一步分析了 GRIG 中判别器的作用，分别移除每个组件，并将我们的伪造图像块判别器分别替换为 PatchGAN [53]和 SM-PatchGAN [42], 。所有参与比较的判别器均采用相同的网络架构，即 70×70大小的 PatchGAN。我们评估了它们的图像修复性能，以展示这些变化的影响。

表 II 展示了所比较变体的定量结果。在不同的少样本和大规模数据集上，就 FID 分数而言，GRIG 均优于全部变体。当移除投影判别器或伪造图像块判别器中的任意一个时，FID 分数都会大幅增加。将我们的伪造图像块判别器替换为其他判别器也会导致更高的 FID 分数。这些结果表明，移除我们的判别器或替换所提出的伪造图像块判别器，会导致图像修复性能迅速下降。GRIG 在各种数据集上取得的最佳 FID 分数验证了我们的伪造图像块判别器在提升性能以及缓解少样本图像修复中过拟合方面的有效性。

图7展示了相应的视觉比较。当移除伪造图像块判别器后，如图7A所示，图像修复结果在掩码边界附近出现明显的伪影，生成的纹理也较为模糊。当

表III 使用不同迭代推理步骤训练的模型的定量结果（FID） T。在$50-60\%$掩码比例下对结果进行了评估。**Bold**：排名第1的最佳结果。

数据集	$T=1$	$T=3$	$T=5$	$T=7$	$T=9$
CHASE	62.76	55.84	56.58	**53.39**	57.84
动漫	69.55	**65.05**	68.05	66.50	69.20
Dog	63.64	**58.49**	59.66	62.09	61.47
Art	78.40	**77.32**	78.04	78.29	78.23

移除投影判别器后，定量性能和视觉质量都会下降。在这种情况下，输出结果更难保持语义结构，例如图7B所示的不对称动漫人脸。当我们将伪造图像块判别器替换为PatchGAN判别器时，生成像素与已知像素之间的对齐可能会受到影响，如图7C所示。当我们将伪造图像块判别器替换为SM-PatchGAN判别器时，虽然可以生成合理的内容，但与已知区域的一致性较差，如图7D所示。相比之下，GRIG在定量指标和定性指标方面都表现最佳。

2）迭代推理步骤的消融研究：为评估迭代推理步骤 T 的有效性，我们针对具有不同迭代推理步骤 T的 GRIG 进行了消融研究，并分别在 $50-60\%$掩码比例下测试了 FID 分数，如表III所示。每次测试均采用与相应训练阶段相同的迭代推理步骤。与使用 $T=1$训练的模型相比，使用 $T>1$ 训练的模型取得了显著的性能提升。例如，在 CHASE 数据集上，使用 $T=3$ 训练的模型的 FID 分数比使用 $T=1$ 训练的模型低 6.92（55.84 对 62.76）。当 $T>5$时，性能提升趋于饱和，或在一定程度上有所下降，但在大多数情况下，图像修复性能仍优于 $T=1$ 。结果表明，GRIG 在早期步骤即可生成令人满意的图像修复结果，而残差偏移可能在后续步骤中发生波动，从而可能导致图像修复质量出现变化。然而，GRIG 有效地平衡了步骤数量与图像修复质量的提升，在少样本图像修复任务中实现了更优性能。本文采用 $T=3$ ，以在计算成本与视觉质量之间取得平衡。图8可视化了使用$T=3$训练的模型在每个步骤 t 的残差输出 Δ_t 。掩码图像逐渐得到修复。模型在早期步骤优先处理语义特征，在后期步骤则处理精细细节。

如图9所示，我们分别在使用$T=1$、3、5、7、9进行训练的模型上，评估了GRIG在固定中心 25% 矩形掩码下的表现。Anime和Dog数据集上的FID分数表明，使用更高迭代推理步骤 T 训练的模型比使用较低 T训练的模型收敛更快，而使用 $T=1$ 训练的模型难以收敛。具体而言，使用 $T>1$ 训练的模型在大约10、000个图像批次后收敛，而使用 $T=1$训练的模型距离收敛仍很远，即使经过10、000个图像批次后，其结果仍剧烈波动。这表明我们的框架能够有效帮助网络更快收敛。

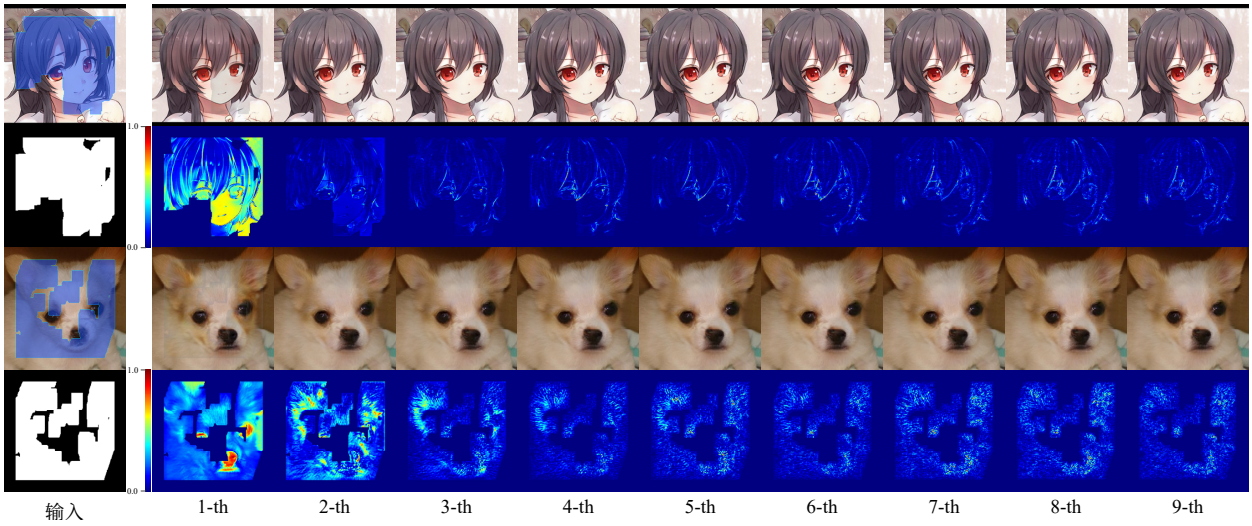

图8。可视化每个迭代推理步骤（每组）的图像修复性能：（左上）掩码图像，（左下）输入二值掩码，（右上）修复图像，以及（右下）残差输出的热力图 Δ_t。热力图中的红色/蓝色分别表示对应像素的高变化/低变化。

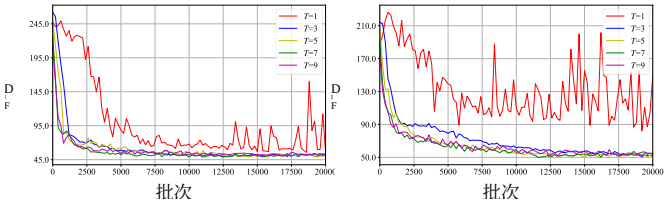

图9。Anime（左）和 Dog（右）数据集上每次训练迭代的 FID 分数对比。结果均在固定的中心 25% 矩形掩码上进行评估。

3）少样本设置下的消融研究： 我们在CHASE、Anime、Cat、Dog和Art数据集上测试了GRIG对不同少样本设置的敏感性。表IV展示了FID分数的定量结果。随着训练样本数量的增加（例如，50样本图像），FID分数会降低，这表明更多训练样本能够提升图像修复质量。

图10展示了相应的修复图像示例。当模型使用较少样本进行训练时，修复图像的质量会迅速下降。例如，使用 5 样本图像训练的模型无法修复掩码区域内的语义结构；而使用 10样本和 30样本图像训练的模型能够修复出更合理的内容，但部分输出结果在掩码边界周围仍表现出明显的颜色不一致。图4中 CMOD 和 MAT 上也出现了类似现象。相比之下，在 50样本设置下训练的模型能够生成更清晰的结果，并呈现更复杂的纹理和更丰富的色彩。我们可以发现，模型使用的训练样本越多，其在定量评估和定性评估中的性能就越好。

表IV 不同n样本设置下训练的模型的定量结果（FID）。在 $50-60\%$ 掩码比例下对结果进行评估。“n样本”表示我们使用 n 张图像进行训练，而“全部”表示本文提及的训练集。**BOLD**：最佳结果排名第一。

数据集	5样本	10样本	30样本	50样本	All
CHASE	70.26	64.39	-	-	**55.84**
动漫	98.50	90.02	77.58	77.32	**65.05**
Cat	120.42	73.10	58.93	55.38	**44.73**
Dog	138.12	126.37	98.83	88.63	**58.49**
Art	113.27	102.02	96.17	91.31	**77.32**
CelebA-HQ	52.46	31.78	20.06	15.53	**8.06**
PSV	122.23	105.09	97.33	90.23	**58.08**

五、结论、局限性与未来工作

本文迈出了探索少样本图像修复解决方案的第一步。通过引入图像级与图像块级判别器解耦的迭代残差推理，我们提出了一种新颖的少样本生成式残差图像修复框架。所提出的生成器有效利用卷积神经网络进行特征提取，并利用 Transformer 进行全局推理。为帮助生成网络学习图像细节，我们引入了伪造图像块判别器。此外，我们在多个少样本数据集上取得了新的最先进性能，广泛的实验也证明了所提方法的有效性。

我们的方法存在一些局限性。该方法能够在少样本数据集上有效执行高保真图像修复。然而，GRIG 无法直接利用条件信息来进行基于引导的图像修复。引入更复杂的方案或模块来引导图像修复过程，将有助于实现可控少样本图像补全。此外，GRIG 并不专用于多样化图像修复。使用映射网络将随机风格编码嵌入生成器，可能是提高少样本图像修复多样性的有效方案。

参考文献

[1] S. Mori、O. Erat、W. Broll、H. Saito、D. Schmalstieg 和 D. Kalkofen，“Inpaintfusion：面向三维场景的增量式 RGB-D 修复”，*TVCG*，第 26 卷第 10 期，第 2994–3007 页，2020 年。

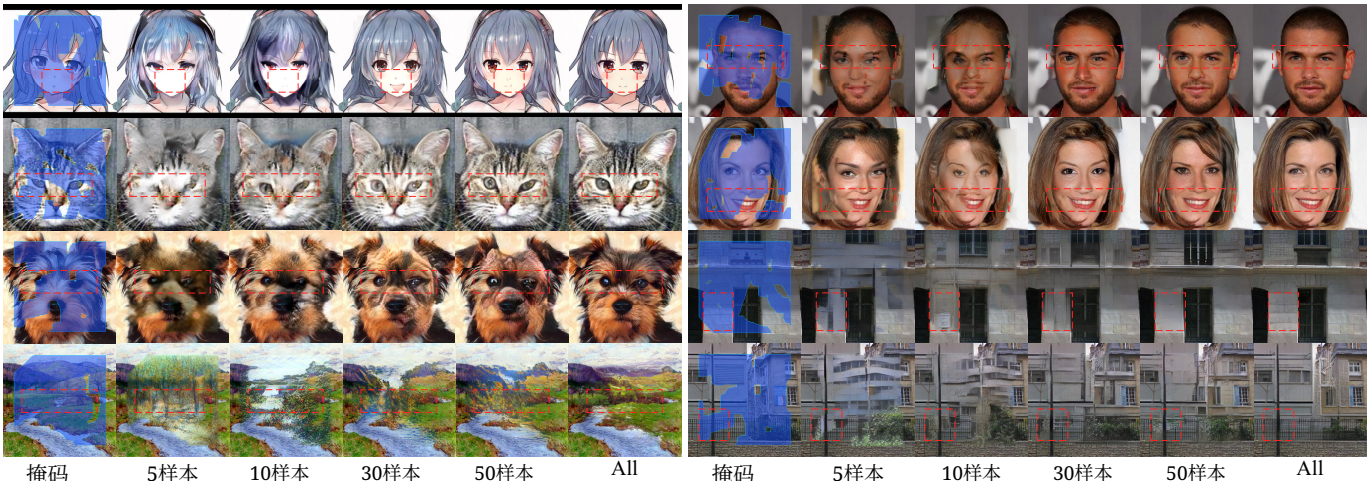

图10。不同　　样本设置下训练的模型之间的视觉比较。“全部”表示论文中提到的训练集　　er.

[2] W. Li、Z. Lin、K. Zhou、L. Qi、Y. Wang 和 J. Jia，“MAT：面向大孔洞图像修复的掩码感知 Transformer”，载于 CVPR，2022 年，第 10 748–10 758 页。[3] N. Kawai、T. Sato 和 N. Yokoya，“基于图像修复并考虑背景几何结构的减弱现实”，TVCG，第 22 卷第 3 期，第 1236–1247 页，2016 年。[4] S. Mori、D. Schmalstieg 和 D. Kalkofen，“用于图像修复的优质关键帧”，TVCG，第 1–1 页，2022 年。[5] Z. Wan、B. Zhang、D. Chen、P. Zhang、D. Chen、J. Liao 和 F. Wen，“让老照片重焕生机”，载于 CVPR，2020 年，第 2747–2757 页。[6] Y. Jo 和 J. Park，“SC-FEGAN：结合用户草图与颜色的人脸编辑生成对抗网络”，载于 ICCV，2019 年，第 1745–1753 页。[7] R. Suvorov、E. Logacheva、A. Mashikhin、A. Remizova、A. Ashukha、A. Silvestrov、N. Kong、H. Goka、K. Park 和 V. Lempitsky，“ 利用傅里叶卷积实现具有分辨率鲁棒性的超大掩码图像修复”，载于 WACV，2022 年，第 3172–3182 页。[8] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser 和 I. Polosukhin，“注意力机制就是你所需要的一切”，载于 NeurIPS，第 30 卷，2017 年。[9] M. Arjovsky 和 L. Bottou，“面向生成对抗网络训练的原则性方法”，载于 ICLR，2017 年。[10] S. Zhao、J. Cui、Y. Sheng、Y. Dong、X. Liang、E. I. Chang 和 Y. Xu，“通过协同调制生成对抗网络实现大规模图像补全”，载于 ICLR，2021 年。[11] J. Yu、Z. Lin、J. Yang、X. Shen、X. Lu 和 T. Huang，“利用门控卷积进行自由形状图像修复”，载于 ICCV，2019 年，第 4470–4479 页。[12] Y. Zeng、Z. Lin、J. Yang、J. Zhang、E. Shechtman 和 H. Lu，“结合迭代置信度反馈与引导式上采样的高分辨率图像修复”，载于 ECCV，2020 年，第 1–17 页。[13] J. Li、N. Wang、L. Zhang、B. Du 和 D. Tao，“用于图像修复的循环特征推理”，载于 CVPR，2020 年。[14] U. Ojha、Y. Li、C. Lu、A. A. Efros、Y. J. Lee、E. Shechtman 和 R. Zhang，“通过跨域对应关系实现少样本图像生成”，载于 CVPR，2021 年，第 10 738–10 747 页。[15] B. Liu、Y. Zhu、K. Song 和 A. Elgammal，“面向更快速、更稳定的高保真少样本图像合成 GAN 训练”，载于 ICLR，2021 年。[16] A. Sauer、K. Chitta、J. M¨uller 和 A. Geiger，“投影 GAN 收敛更快”，载于 NeurIPS，第 34 卷，2021 年，第 17 480–17 492 页。[17] Y. Du、S. Li、B. J. Tenenbaum 和 I. Mordatch，“通过能量最小化学习迭代推理”，载于 ICML，2022 年。[18] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，载于 CVPR，2016 年，第 770–778 页。[19] C. Ballester 、 M. Bertalmio 、 V. Caselles 、 G. Sapiro 和 J. Verdera，“通过联合插值向量场与灰度值进行填充”，TIP，第 10 卷第 8 期，第 1200–1211 页，2001 年。[20] M. Bertalmio、G. Sapiro、V. Caselles 和 C. Ballester，“图像修复”，载于 SIGGRAPH，2000 年，第 417–424 页。[21] Levin、Zomet 和 Weiss，“从全局图像统计信息中学习如何进行图像修复”，载于 ICCV，第 1 卷，2003 年，第 305–312 页。[22] A. Telea，“一种基于快速行进法的图像修复技术”，Journal of Graphics Tools，第 9 卷第 1 期，第 23–34 页，2004 年。

[23] V. Kwatra, I. Essa, A. Bobick 和 N. Kwatra，《用于基于示例合成的纹理优化》，TOG，第 24 卷，第 3 期，第 795–802 页，2005 年。[24] Q. Guo, S. Gao, X. Zhang, Y. Yin 和 C. Zhang，《通过两阶段低秩近似实现基于图像块的图像修复》，TVCG，第 24 卷，第 6 期，第 2023–2036 页，2018 年。[25] Z. Zhu, H.-Z. Huang, Z.-P. Tan, K. Xu 和 S.-M. Hu，《利用互联网图像忠实补全风景地标图像》，TVCG，第 22 卷，第 8 期，第 1945–1958 页，2016 年。[26] C. Barnes, E. Shechtman, A. Finkelstein 和 D. B. Goldman，《Patch-match：一种用于结构图像编辑的随机对应算法》，TOG，第 28 卷，第 3 期，第 24 篇文章，2009 年。[27] D. Simakov, Y. Caspi, E. Shechtman 和 M. Irani，《使用双向相似性总结视觉数据》，载于 CVPR，2008 年，第 1–8 页。[28] D. Ding, S. Ram 和 J. J. Rodr´ıguez，《使用非局部纹理匹配和非线性滤波进行图像修复》，TIP，第 28 卷，第 4 期，第 1705–1719 页，2019 年。[29] J. S. Ren, L. Xu, Q. Yan 和 W. Su n，《Shepard卷积神经网络》，载于 NeurIPS，第 28 卷，2015 年，第 901–909 页。[30] D. Pathak, P. Kr¨ahenb¨uhl, J. Donahue, T. Darrell 和 A. Efros，《上下文编码器：通过图像修复进行特征学习》，载于 CVPR，2016 年。[31] Y. Wang, X. Tao, X. Qi, X. Shen 和 J. Jia，《通过生成式多列卷积神经网络进行图像修复》，载于 NeurIPS，2018 年，第 331–340 页。[32] Z. Yan, X. Li, M. Li, W. Zuo 和 S. Shan，《Shift-net：通过深度特征重排进行图像修复》，载于 ECCV，2018 年。[33] Y. Zeng, J. Fu, H. Chao 和 B. Guo，《学习金字塔上下文编码器网络以实现高质量图像修复》，载于 CVPR，2019 年，第 1486–1494 页。[34] S. Iizuka, E. Simo-Serra 和 H. Ishikawa，《全局和局部一致的图像补全》，TOG，第 36 卷，第 4 期，第 107 篇文章，2017 年。[35] H. Liu, B. Jiang, Y. Xiao 和 C. Yang，《用于图像修复的连贯语义注意力》，载于 ICCV，2019 年，第 4169–4178 页。[36] C. Xie, S. Liu, C. Li, M.-M. Cheng, W. Zuo, X. Liu, S. Wen 和 E. Ding，《使用可学习的双向注意力图进行图像修复》，载于 ICCV，2019 年，第 8857–8866 页。[37] Z. Yi, Q. Tang, S. Azizi, D. Jang 和 Z. Xu，《用于超高分辨率图像修复的上下文残差聚合》，载于 CVPR，2020 年，第 7508–7517 页。[38] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu 和 T. S. Huang，《通过上下文注意力实现生成式图像修复》，载于 CVPR，2018 年，第 5505–5514 页。[39] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao 和 B. Catanzaro，《使用部分卷积修复不规则孔洞》，载于 ECCV，2018 年，第 89–105 页。[40] W. Xiong, J. Yu, Z. Lin, J. Yang, X. Lu, C. Barnes 和 J. Luo，《前景感知图像修复》，载于 CVPR，2019 年，第 5833–5841 页。[41] Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu 和 G. Li，《Structureflow：通过结构感知外观流进行图像修复》，载于 ICCV，2019 年，第 181–190 页。[42] Y. Zeng, J. Fu, H. Chao 和 B. Guo，《用于高分辨率图像修复的聚合上下文变换》，TVCG，第 1–1 页，2022 年。

[43] C. Zheng, T.-J. Cham, J. Cai 和 D. Phung，《衔接全局上下文交互以实现高保真图像补全》，载于 CVPR，2022 年，第 11 512–11 522 页。[44] L. Zhao, Q. Mo, S. Lin, Z. Wang, Z. Zuo, H. Chen, W. Xing 和 D. Lu，《Uctgan：基于无监督跨空间转换的多样化图像修复》，载于 CVPR，2020 年，第 5740–5749 页。[45] Z. Wan, J. Zhang, D. Chen 和 J. Liao，《利用 Transformer 实现高保真多样化图像补全》，载于 ICCV，2021 年，第 4672–4681 页。[46] Q. Liu, Z. Tan, D. Chen, Q. Chu, X. Dai, Y. Chen, M. Liu, L. Yuan 和 N. Yu，《减少 Transformer 中的信息丢失以实现多样化图像修复》，载于 CVPR，2022 年，第 11 347–11 357 页。[47] H. Zhang, Z. Hu, C. Luo, W. Zuo 和 M. Wang，《利用渐进式生成网络进行语义图像修复》，载于 ACM MM，2018 年，第 1939–1947 页。[48] J. Li, F. He, L. Zhang, B. Du 和 D. Tao，《面向图像修复的视觉结构渐进式重建》，载于 ICCV，2019 年。[49] Z. Guo, Z. Chen, T. Yu, J. Chen 和 S. Liu，《使用全分辨率残差网络进行渐进式图像修复》，载于 ACM MM，2019 年，第 2496–2504 页。[50] J. Hur 和 S. Roth，《用于联合光流与遮挡估计的迭代残差细化》，载于 CVPR，2019 年，第 5747–5756 页。[51] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan 和 M.-H. Yang，《Restormer：用于高分辨率图像复原的高效 Transformer》，载于 CVPR，2022 年。[52] T. Miyato, T. Kataoka, M. Koyama 和 Y. Yoshida，《生成对抗网络的谱归一化》，载于 ICLR，2018 年。[53] P. Isola, J.-Y. Zhu, T. Zhou 和 A. A. Efros，《使用条件对抗网络进行图像到图像转换》，载于 CVPR，2017 年，第 5967–5976 页。[54] M. Tan 和 Q. Le，《EfficientNet：重新思考卷积神经网络的模型缩放》，载于 ICML，第 97 卷，2019 年，第 6105–6114 页。[55] F. Yu 和 V. Koltun，《通过空洞卷积进行多尺度上下文聚合》，载于 ICLR，2016 年。[56] O. Texler, D. Futschik, M. Kuˇcera, O. Jamriˇska, ˇS´arka Sochorov´a, M. Chai, S. Tulyakov 和 D. S´ykora，《使用少样本基于图像块的训练进行交互式视频风格化》，TOG，第 39 卷，第 4 期，第 73 页，2020 年。[57] W. Wang, J. Bao, W. Zhou, D. Chen, D. Chen, L. Yuan 和 H. Li，《Sindiffusion：从单幅自然图像中学习扩散模型》，arXiv 预印本 arXiv:2211.12445，2022 年。[58] R. Zhang, P. Isola, A. A. Efros, E. Shechtman 和 O. Wang，《深度特征作为感知度量具有惊人效用》，载于 CVPR，2018 年，第 586–595 页。[59] K. Simonyan 和 A. Zisserman，《用于大规模图像识别的超深卷积网络》，载于 ICLR，2015 年。[60] MMEditing 贡献者，《MMEditing：OpenMMLab 图像与视频编辑工具箱》，https://github.com/open-mmlab/mmediting，2022 年 9 月。[61] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler 和 S. Hochreiter，《通过双时间尺度更新规则训练的 GAN 收敛于局部纳什均衡》，载于 NeurIPS，第 30 卷，2017 年，第 6629–6640 页。[62] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen 和 S. A. Barman，《一种应用于视网膜血管分割的集成分类方法》，IEEE 生物医学工程汇刊，第 59 卷，第 9 期，第 2538–2548 页，2012 年。[63] Z. Si 和 S.-C. Zhu，《通过信息投影学习混合图像模板（HIT）》，TPAMI，第 34 卷，第 7 期，第 1354–1367 页，2012 年。[64] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu 和 S. Han，《面向数据高效 GAN 训练的可微增强》，载于 NeurIPS，第 33 卷，2020 年，第 7559–7570 页。[65] T. Karras, T. Aila, S. Laine 和 J. Lehtinen，《用于提升质量、稳定性和多样性的 GAN 渐进式增长》，载于 ICLR，2018 年。[66] T. Karras, S. Laine 和 T. Aila，《用于生成对抗网络的基于风格的生成器架构》，TPAMI，第 43 卷，第 12 期，第 4217–4228 页，2021 年。[67] C. Doersch, S. Singh, A. Gupta, J. Sivic 和 A. A. Efros，《是什么让巴黎看起来像巴黎？》，载于 SIGGRAPH，2012 年，第 101:1–101:9 页。[68] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva 和 A. Torralba，《Places：一个包含 1000 万张图像的场景识别数据库》，TPAMI，第 40 卷，第 6 期，第 1452–1464 页，2018 年。