DOI:10.19651/j.cnki.emt.2519762

基于扩散模型的梯度引导图像修复算法研究*

张荣铠[1] 景明利[1] 焦 龙[2]
(1.西安石油大学电子工程学院 西安 710300; 2.西安石油大学化学化工学院 西安 710300)

摘 要: 去噪扩散概率模型(DDPM)展示了强大的图像生成能力,并成功应用于图像修复。近几年流行的方法通过引入结构先验来修复图像,尽管取得了良好的表现,但在结构还原的完整性与纹理细节的自然性方面仍存在不足,常出现断裂或不连续的修复区域。针对此不足之处,提出了一种基于扩散模型的梯度引导特征重建算法。首先对破损图像的梯度图进行修复,初步重建图像的纹理和结构;随后将修复后的梯度图作为生成引导,辅助原图进行修复;最后在噪声预测网络中引入高效通道注意力(ECA)模块,优化特征交互,进一步提升修复图像的一致性。实验结果表明,与主流算法相比,在 CelebA-HQ 数据集上,所提方法在峰值信噪比(PSNR)和结构相似性(SSIM)指标上分别提高 3.19% 和 2.74%,感知相似度(LPIPS)指标下降 8.82%;在 Places2 数据集上,PSNR 和 SSIM 分别提高 0.42% 和 2.81%,LPIPS 下降 2.75%,证明所提方法在图像修复任务中具备更强的结构还原与细节保持能力。

关键词: 图像修复;扩散模型;梯度图;通道注意力;噪声预测网络
中图分类号: TP391.41;TN0 **文献标识码**: A **国家标准学科分类代码**: 520.2060

Gradient-guided image inpainting algorithm based on diffusion models

Zhang Rongkai[1] Jing Mingli[1] Jiao Long[2]
(1. School of Electronic Engineering, Xi′an Shiyou University, Xi′an 710300, China;
2. School of Chemistry and Chemical Engineering, Xi′an Shiyou University, Xi′an 710300, China)

Abstract: Denoising diffusion probabilistic models(DDPM) have demonstrated powerful image generation capabilities and have been successfully applied to image inpainting. While many recent approaches introduce structural priors to assist the inpainting process and have achieved promising results, they still suffer from limitations in structural integrity and the naturalness of texture details, often resulting in fractured or discontinuous regions. To address these issues, this paper proposes a gradient-guided feature reconstruction algorithm based on diffusion models. Specifically, the gradient map of the corrupted image is first restored to preliminarily reconstruct its structure and texture. The restored gradient map is then used as generation guidance to assist the inpainting of the original image. Furthermore, an efficient channel attention (ECA) module is integrated into the noise prediction network to enhance feature interaction and improve the consistency of the reconstructed images. Experimental results show that compared with state-of-the-art methods, the proposed approach achieves improvements of 3.19% in PSNR and 2.74% in SSIM, and a reduction of 8.82% in LPIPS on the CelebA-HQ dataset. On the Places2 dataset, it achieves gains of 0.42% in PSNR and 2.81% in SSIM, with an 2.75% decrease in LPIPS, demonstrating superior capability in structural restoration and detail preservation for image inpainting tasks.

Keywords: image inpainting;diffusion model;gradient map;channel attention;noise prediction network

0 引 言

图像修复是计算机视觉领域的基础问题之一,其主要目标是对图像中因多种原因产生的缺失或破损区域进行有效补全,从而生成结构完整且符合人眼视觉的图像。

传统图像修复方法通过填充邻近像素块来修复小范围缺损,如杨竹青等[1]提出一种基于样本块匹配的传统算法,结合图像的梯度模值构建结构度量因子并引入方差调节策略来动态选择最佳样本块尺寸,以改善修复区域的纹理连续性和视觉效果。传统方法过于依赖局部信息,难以在处

收稿日期:2025-09-01
*基金项目:国家自然科学基金(22373075)项目资助

理大面积缺失时保持连贯性。随着计算硬件和能力的提升，基于深度学习的图像修复方法逐渐展现出优势。

近年来，图像修复领域出现了许多优秀的算法模型，显著推动了该领域的发展。早期的深度学习方法如：上下文编码器[2](context encoder，CE)和全局和局部一致的图像补全[3](globally and locally consistent image completion，GLCIC)，引入了全局与局部判别器，对生成图像的整体结构和细节一致性进行约束。生成式多列卷积神经网络[4](generative multi-column convolutional neural networks，GMCNN)，通过并行的多尺度卷积结构，有效增强了对图像特征的表达能力。具有上下文注意力的生成图像修复[5](generative image inpainting with contextual attention，GICA)引入自注意力机制来建模缺失区域与全局内容。EdgeConnect[5-6]通过边缘检测来引导图像修复，使生成图像更具结构性。开志强等[7]提出了一种结合全局语义学习和显著目标感知的激光干扰图像修复网络，提升目标轮廓清晰度与纹理一致性。李雪瑾等[8]通过结合语义损失与感知损失并改进 Sigmoid 激活函数，有效修复图像大面积缺失区域。上述方法大多基于编码器-解码器(encoder-decoder，ED)架构，并结合生成对抗网络[9](generative adversarial network，GAN)中的判别器模块，引入生成对抗损失减少生成图像与真实图像之间的差异。部分方法则在模型结构上进一步优化，Shift-Net[10]和 Deep Fusion Network[11]采用 U-Net 结构增强特征融合，而 Pyramid-Context Encoder Network[12]采用金字塔结构加强多尺度建模。随着 Transformer[13]在自然语言处理任务中的突破，研究者开始尝试将其引入图像修复任务，衍生出 TransFill[14]、MAE[15]、ZITS[16]等基于 Transformer 的模型，这些模型提升了修复的准确性、特征表示能力以及语义一致性；滕诗宇等[17]提出融合多尺度特征与全局-局部协同 Transformer，进一步降低了计算复杂度，提高了结构与细节的生成质量。

在 Ho 等[18]于 2020 年提出去噪扩散概率模(denoising diffusion probabilistic models，DDPM)后，基于扩散模型的图像修复方法迅速发展，相较于 GAN，DDPM 具有显著的理论优势：通过变分下界推导的均方误差损失避免了模式崩塌问题，且链式生成特性支持细致的条件控制。研究者将 DDPM 引入图像修复任务，通过掩码引导的条件采样策略实现上下文重建。RePaint[19]通过预训练扩散模型和重采样策略提升修复质量与多样性；SmartBrush[20]通过文本和形状信息引导实现精确修复；StrDiffusion[21]通过结构指导纹理去噪，实现了一致的语义生成；张绪等[22]提出二阶段扩散模型框架，结合改进的 U-Net 和采样策略，有效缓解语义不一致问题。在添加结构约束后，扩散模型在图像修复中展现出较大的潜力，但仍存在一些不足之处：首先，未考虑结构与纹理的相互作用，导致生成图像的局部细节和全局结构不一致；其次，对于复杂图像，冗余信息过多，导致特征表达不完整和不准确，从而影响修复效果。

针对上述问题，提出梯度引导特征重建网络，由只携带结构和高频细节信息的梯度图作为先验信息，引导图像进行修复，并在去噪过程中，采用高效通道注意力模块强化通道选择，平衡不同尺度特征之间的信息流动。具体来说，对输入图像的梯度图进行修复，将其作为先验来引导原图进行修复；在去噪过程中，引入高效通道注意力模块，增强对重要特征的关注，保留结构信息并有效结合浅层和深层特征，减少冗余噪声，恢复精细纹理。实验结果表明，与其他图像修复方法对比，本文方法在 CelebA-HQ 数据集和 Places2 数据集上均展现了较好的修复性能。

1　本文方法

1.1　扩散模型

扩散模型[18]的核心思想在于将图像生成过程建模为一个逐步去噪的反向过程，其基础构成包括一个前向扩散过程和一个学习得到的逆向去噪过程。

前向扩散过程中，使用高斯噪声逐步破坏训练数据，设 $q(x_0)$ 是图像的数据密度，其中，x_0 代表未添加高斯噪声的原始图像数据。给定一个原始的训练样本 $x_0 \in p(x_0)$，根据下面的马尔科夫过程得到噪声样本 $x_1, x_2, \cdots, x_T$：

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t \boldsymbol{I}), \forall t \in \{1, \cdots, T\} \tag{1}$$

其中，T 是扩散步长，$\beta_1, \cdots, \beta_T \in [0,1)$ 是超参数，表示扩散步之间的方差，$\boldsymbol{I}$ 是与输入图像 x_0 维度相同的单位矩阵，而 $\mathcal{N}(x; \mu, \sigma)$ 表示用于生成 x 的均值为 μ、协方差为 σ 的正态分布。该递归公式支持对 x_t 进行直接采样；当 t 从均匀分布中采样时，即 $t \sim \mathcal{U}(\{1, \cdots, T\})$：

$$q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\overline{\alpha}_t} \cdot x_0, (1-\overline{\alpha}_t) \cdot \boldsymbol{I}), \tag{2}$$

其中，$\alpha_t = 1-\beta_t$，$\overline{\alpha} = \prod_{i=0}^{t} \overline{\alpha_i}$。对式(1)的递归方程进行 T 次迭代运算后，当施加于数据的累计噪声总量足够多时，目标分布 $p(x_T)$ 可以很好地逼近标准高斯分布，即 $p(x_t) \approx \mathcal{N}(0, \boldsymbol{I})$，因此可以通过从 $x_t \sim \mathcal{N}(0, \boldsymbol{I})$ 采样并按照逆向步骤 $p(x_{t-1} \mid x_t)$ 进行推导，从而重构出符合原始数据分布 $p(x_0)$ 的高质量新样本 x_0：

$$p(x_{t-1} \mid x_t, x_0) = \mathcal{N}(x_{t-1}; \mu_t(x_t, x_0), \sigma_t^2 \boldsymbol{I}), \tag{3}$$

其中，均值 $\mu_t(x_t, x_0)$，方差 σ_t^2 分别为：

$$\begin{cases} \mu_t(x_t, x_0) = \dfrac{1}{\sqrt{\alpha_t}}\left(x_t - \varepsilon \dfrac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\right) \\ \sigma_t^2 = \dfrac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t \end{cases} \tag{4}$$

其中，ε 表示 x_t 中的噪声，是逆向过程中唯一不确定的变量。扩散模型采用去噪网络 $\varepsilon_\theta(x_t, t)$ 来估计 ε。为了训练 $\varepsilon_\theta(x_t, t)$，给定一个加噪前的图像 x_0，扩散模型随机采样一个时间步长 t 和一个噪声 $\varepsilon \sim \mathcal{N}(0, \boldsymbol{I})$，根据

式(2)生成含有噪声的图像 x_t,随后扩散模型对 ε_θ 的网络参数 θ 用如下方法进行优化:

$$\nabla_\theta \| \varepsilon - \varepsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \varepsilon\sqrt{1-\bar{\alpha}_t}, t) \|_2^2 \quad (5)$$

1.2 梯度图

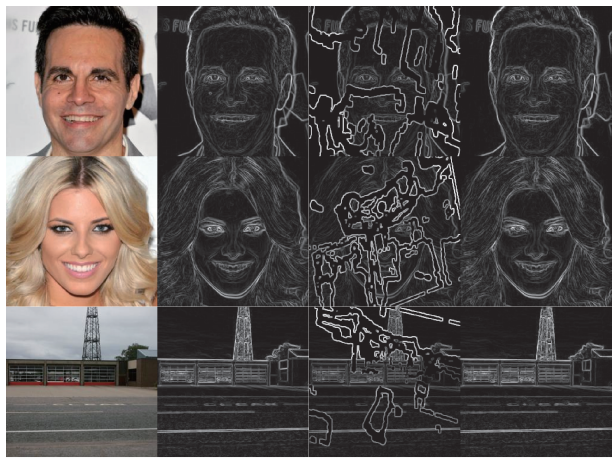
梯度图是量化图像局部变化的工具,广泛应用于图像分析与特征提取。它通过计算像素灰度值在空间方向上的变化速率与方向,不仅能够从宏观层面勾勒出图像的整体结构轮廓,同时也具备较强的局部细节刻画能力,能够在微观尺度上反映出如皮肤褶皱、发丝纹理等细节特征,在表达图像的结构信息与局部纹理方面具有明显优势;与此同时,梯度图的表达形式相对简洁,数据维度更低,使其在图像修复任务中更易于建模与优化。

具体而言,首先使用同一扩散框架对损坏图像的梯度图进行独立修复,得到结构完成的梯度先验。梯度修复模型与主扩散模型采用相同的去噪网络结构与扩散过程设定,但二者在训练阶段互相独立,分别针对梯度图与原图进行优化。在原图扩散模型的反向去噪阶段,修复完成的梯度图以条件输入的形式嵌入到去噪网络中,对原图修复施加约束,在保持纹理自然性的同时,更加准确地恢复图像的轮廓、边缘和几何细节。

整体上,梯度图负责重建图像的结构与框架,而主扩散模型在这种结构先验的指导下生成高保真度的像素细节。人脸和建筑的梯度图的修复效果如图 1 所示。

1.3 模型框架

问题定义:给定一个原图 $x_g \in \mathbf{R}^{1\times H\times W}$ 以及二值掩码 $x_m \in \{0,1\}^{1\times H\times W}$,其中,$x_m$ 为 0 代表着缺失区域,为 1 代表着已知区域。假设原图 x_g 沿着图像掩码 x_m 退化为一幅破损的图像 x_c,即 $x_c = x_g \odot x_m$。图像修复的目的是在破损图像中修复出缺失区域,使修复的区域应该与周围已知区域无缝融合,并且图片内容在逻辑上保持一致。



(a) 原图 (b) 梯度图 (c) 破损图像 (d) 修复图像

(a) Original image (b) Gradient image (c) Damaged image (d) Inpainted image

图 1 人脸和建筑的梯度图修复

Fig. 1 Inpainting results of faces and buildings after conversion to gradient maps

在 U-Net 网络的基础上,通过融合梯度图先验信息,并且增强跨层特征交互,提升复杂纹理区域的修复质量。模型框架如图 2 所示。

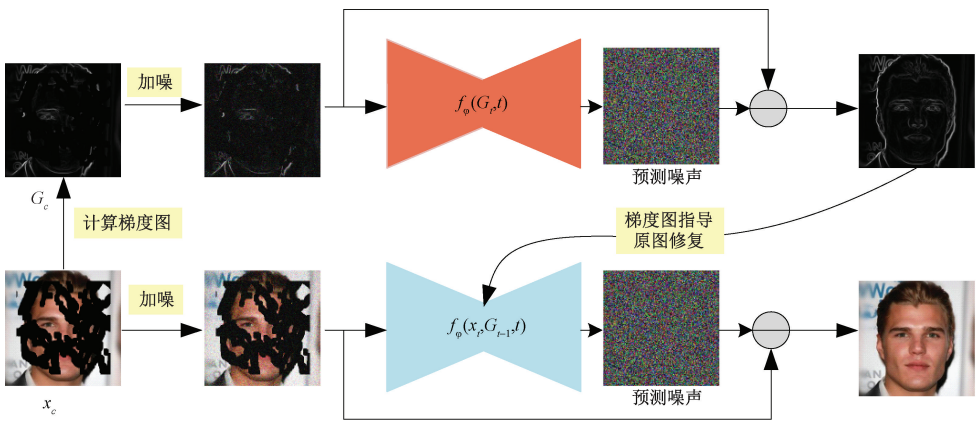


图 2 模型框架

Fig. 2 The framework of our method

梯度图承载了图像的边缘和纹理等高频信息,相比原始 RGB 图像,其维度更低、结构简洁且语义明确。通过对梯度图进行重建并作为先验引入修复过程,可以帮助模型更好地捕捉关键结构特征。在修复过程中,梯度图作为强约束信号,引导模型逐步恢复图像细节,提升结构一致性和感知质量。

在扩散模型在反向采样过程中,噪声图像 x_T 通过 U-Net 网络预测噪声不断去噪,在噪声中构建样本,直到生成想要的图像 x_0。当引入梯度图作为反向采样过程的条件时,t 时刻的梯度图 G_t、t 时刻的图像 x_t 以及 $t-1$ 时刻的图像 x_{t-1} 可以通过贝叶斯公式,将反向采样过程表示为:

$$q(x_{t-1}\mid x_t, G_{t-1}) = \frac{q(x_{t-1}\mid x_t)\,q(G_{t-1}\mid x_{t-1}, x_t)}{q(G_{t-1}\mid x_t)} = \frac{q(x_t\mid x_{t-1})\,q(x_{t-1})\,q(G_{t-1}\mid x_{t-1})}{q(G_{t-1}, x_t)} \quad (6)$$

其中，x_t、x_{t-1} 和 G_t 在扩散过程中实际上都是以他们的初始状态 x_0 和 G_0 为条件的。进而：$q(x_{t-1}) = q(x_{t-1}|x_0)$；以及 $q(G_{t-1},x_t) = q(G_{t-1})q(x_t|G_{t-1}) = q(G_{t-1}|G_0)q(x_t|G_{t-1},x_0)$，因此，式(6)可以简化为：

$$q(x_{t-1}|x_t,x_0,G_{t-1},G_0) = \frac{q(x_t|x_{t-1})q(G_{t-1}|x_{t-1})}{q(x_t|G_{t-1},x_0)} \times \frac{q(x_{t-1}|x_0)}{q(G_{t-1}|G_0)} \approx \frac{q(x_{t-1}|x_0)}{q(G_{t-1}|G_0)} \tag{7}$$

其中，由于扩散过程中的图像 x_t 可以通过条件的辅助梯度图 G_{t-1} 以及初始图像 x_0 获得，基于此，本研究设定 $q(x_t|x_{t-1}) \approx q(x_t|G_{t-1},x_0)$，以及 $q(G_{t-1}|x_{t-1}) \approx 1$。基于文献[17]中的推导结果，将式(7)化简后，采用最小负对数似然求出 $\tilde{x}_{t-1}^*$，随后进行求梯度操作并令梯度为 0，得出图像理想逆向状态 $\tilde{x}_{t-1}^*$ 为：

$$\tilde{x}_{t-1}^* = \frac{1-e^{-2\bar{\theta}_{t-1}}}{1-e^{-2\bar{\delta}_t}} e^{-\delta_t'}(G_t-\mu_G) + \frac{(1-e^{-2\bar{\theta}_{t-1}})(e^{-2\bar{\delta}_t}-e^{-2\delta_t'})}{(1-e^{-2\bar{\theta}_{t-1}})(1-e^{-2\bar{\delta}_t})} e^{-\bar{\theta}_{t-1}}(G_0-\mu_G) + e^{-\bar{\theta}_{t-1}}(x_0-\mu_x)+\mu_x \tag{8}$$

其中，$\tilde{x}_{t-1}^*$ 的更新在过去时刻的预测与最终目标之间权衡，形成一种时间递进的“梯度图引导修复轨迹”，为原图修复提供参考。设定图像分量的均值 μ_x 是初始图片 x_0 的掩码版本，θ_t 是描述均值回归速度的时变参数；梯度图分量的均值 μ_G 为初始状态 G_0 的掩码版本，δ_t 是描述均值回归速度的时变参数。定义 $\bar{\theta}_{t-1} = \int_0^{t-1}\theta_z \mathrm{d}z$，$\bar{\delta}_{t-1} = \int_0^{t-1}\delta_z \mathrm{d}z$。

为提升模型在细节纹理恢复中的通道选择能力，在噪声预测网络中引入高效通道注意力模块[23]（efficient channel attention，ECA），以增强特征通道的自适应建模能力，其结构如图 3 所示。ECA 模块解决了传统通道注意力机制在建模通道间依赖关系时存在的效率低和信息损失问题，在图像修复任务中，这种信息丢失往往会削弱模型对纹理细节的感知能力，从而影响修复质量。

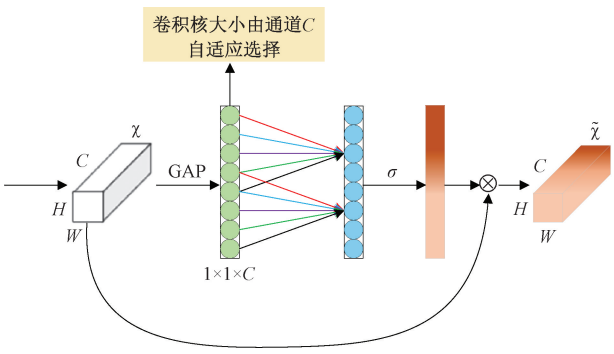


图 3　ECA 结构

Fig. 3　Structure of the ECA

传统的压缩激励网络（squeeze-and-excitation network，SENet）[24]采用“压缩—激励”策略，通过全局平均池化提取通道级全局上下文信息。在“激励”阶段，利用两层全连接网络对通道间依赖关系进行建模，从而生成通道注意力权重，实现通道特征的自适应重标定。这一过程包括显式的降维以及后续的升维操作。降维操作减少了参数量和计算成本，但可能导致信息丢失，从而削弱模型捕捉通道间复杂依赖关系的能力，影响细节恢复的准确性。

ECA 模块提出了一种高效简洁的解决方案，旨在提高通道特征依赖关系的表达能力，同时保持建模效率。其核心思想是通过局部跨通道交互替代全局建模，强调通道间的局部相关性。ECA 摒弃全连接层，采用一维卷积操作，仅在每个通道与相邻的 k 个通道间交互，从而避免冗余计算与信息损失。卷积核大小 k 由与通道维度相关的自适应函数动态生成，自动调整注意力范围。该机制有效保持通道特征的完整性，避免额外参数开销，并实现对关键通道依赖关系的高效捕捉。

ECA 模块继承了 SENet 的全局平均池化操作，将空间维度压缩为通道统计量，对于输入特征图 $\boldsymbol{X} \in \mathbf{R}^{H\times W\times C}$，$H$、$W$ 是空间维度，C 是通道数。首先对输入 $\boldsymbol{X}$ 进行全局平均池化，计算每个通道上的平均值：

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{X}_{i,j,C} \tag{9}$$

其中，$z_c \in \mathbf{R}^C$ 表示第 C 通道的池化值，整个特征图的全局平均池化特征表示为 $\boldsymbol{z} = [z_1,z_2,\cdots,z_C] \in \mathbf{R}^C$。随后 ECA 使用 1D 卷积来对 $\boldsymbol{z}$ 进行通道间的注意力计算，而不是使用全连接层，计算如式(10)所示。

$$\omega = \sigma(\mathrm{Conv1}D_k(z)) \tag{10}$$

其中，σ 是 Sigmoid 激活函数，将输出归一化到[0,1]之间，得到通道权重 ω，$\mathrm{Conv1}D_k(z)$ 表示 1D 卷积，其卷积核大小 k 由 C 自适应决定，即：

$$k = \left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{\mathrm{odd}} \tag{11}$$

其中，γ 和 b 是超参数，$|\cdot|_{\mathrm{odd}}$ 表示取奇数操作，确保 k 为奇数，以保证对称性。相比全连接层，1D 卷积的参数量大幅下降，但较远通道之间的依赖性较弱，ECA 通过局部交互可以有效的增强重要通道的影响，削弱次要通道，即：

$$X_C' = \omega_C \cdot X_C \tag{12}$$

即输入 X 的每个通道 X_C 乘以注意力权重 ω_C，以捕获关键通道关系。

噪声预测网络如图 4 所示，在输入阶段，图像首先通过卷积层和两个残差块提取基础特征，随后利用线性自注意力模块建立全局依赖关系，加强特征间的长距离建模能力；接着通过梯度图引导模块引入条件信息，实现空间自适应归一化调制。在每一级特征下采样后，网络继续堆叠残差块与线性自注意力模块，并再次融合梯度图信息，引

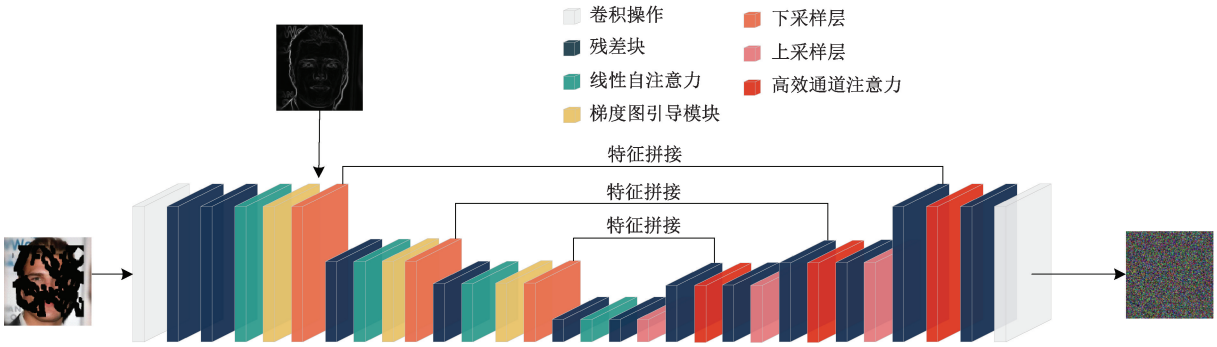


图 4 噪声预测网络

Fig. 4 Structure of the noise prediction network

导网络在压缩尺度下仍保留清晰的语义结构信息。

在解码阶段,U-Net 采用跳跃连接机制引入下采样的浅层特征,以弥补空间细节的损失,并减少噪声和冗余信息。ECA 模块集成于上采样后的残差块之间,首先对每个通道进行全局平均池化提取统计特征,随后通过轻量级的一维卷积建模通道之间的局部依赖关系,最终生成通道注意力权重用于加权输入特征。该模块无需显式压缩维度或增加大量参数,便可实现高效的通道建模,从而在不显著增加计算量的前提下,有效突出关键通道、抑制冗余特征,增强重建图像在结构上的全局一致性。

2 实验及分析

2.1 实验所用数据集

为全面评估所提出方法的性能,在 3 个主流的数据集 Paris Street View(PSV)[25]、CelebA-HQ[26]和 Places2[27]上进行实验。PSV 数据集由巴黎街景图像构成,包含 14 900 张用于训练的图像和 100 张用于验证的图像。CelebA-HQ 是一个高质量的人脸图像数据集,共包含 30 000 张图像,其中,26 000 张用于训练,4 000 张用于验证,可测试模型在细节丰富、结构复杂的人脸图像修复任务中的表现。Places2 是一个涵盖 400 多个类别的大规模自然场景图像数据集,原始规模超过 180 万张图像。从中选取 10 个典型类别,每个类别随机采样 3 000 张图像,最终构建了包含 25 000 张训练图像和 5 000 张测试图像的子集,用以评估模型在多样化自然场景下的泛化能力。所有图像均被统一裁剪并缩放至 256×256 的分辨率,以确保训练和测试阶段的一致性。此外,为模拟真实应用场景中图像缺损的复杂性,实验所采用的掩码数据来自不规则掩码数据集[28](irregular mask dataset),该数据集通过随机绘制线条、曲线、多边形及噪声区域生成,具有形状复杂、边界不规则、连通性随机等特征。其遮挡比例分布在 1%～60%,有效覆盖从小区域缺失到大面积遮挡的多种情况,使用该掩码数据集以提升模型对不同类型缺失区域的鲁棒性与泛化能力。

2.2 评价指标

为全面评估所提出方法的图像修复效果,采用 3 种主流的定量评价指标:峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性指数(structural similarity index, SSIM)以及感知相似度指标(learned perceptual image patch similarity, LPIPS)。

PSNR 和 SSIM 均属于客观评价指标。而 LPIPS 作为一种基于深度神经网络的感知指标,更加侧重于模拟人眼对图像语义和纹理的感知一致性。LPIPS 值越小,表示修复图像在人眼视觉感知上与原图越相似。

2.3 定量比较

为验证所提出方法在图像修复任务中的有效性,选取多个具有代表性的图像修复模型,并在 CelebA-HQ 与 Places2 两个数据集上开展对比实验。对比方法包括:RFR[29]、CMT[30]、TFill[31]、MEDFE[32]、LaMa[33] 和 CoordFill[34]。

为了全面评估不同方法在不同掩码复杂度下的表现,将不规则掩码按缺失比例划分为 10%～20%、20%～30%、30%～40%、40%～50%这 4 个区间,并在每个掩码比例设定下进行对比实验。

表 1 展示了在上述设置下,各方法在 PSNR、SSIM 和 LPIPS 指标上的性能对比结果,其中黑体代表最佳结果。可以观察到,所提方法在多数评价指标上均优于现有主流方法,且在低掩码比例的修复任务中表现出显著优势,说明所提方法在保持图像细节和结构一致性方面具有优势。

具体而言,所提方法在 CelebA-HQ 与 Places2 两个数据集的大多数评估指标中均表现出色,特别是在 CelebA-HQ 数据集上,在所有掩码率下的 PSNR、SSIM 和 LPIPS 这 3 个指标上均取得最优结果,充分体现了所提方法在结构还原与细节重建方面的优势。在 Places2 数据集中,虽然在部分掩码比例和个别指标上略逊于部分方法,但整体仍保持较强竞争力,特别是在感知指标 LPIPS 方面依然达到最优或次优水平,说明本文所提方法在不同场景和掩码条件下具有良好的泛化能力和鲁棒性。

表 1　在 CelebA-HQ 和 Places2 数据集上不同方法定量比较

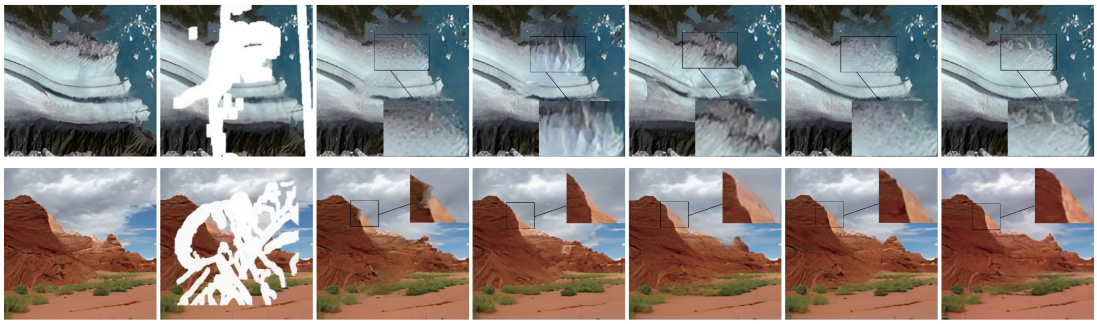
Table 1　Quantitative comparison of different methods on the CelebA-HQ and Places2 datasets

指标	方法	CelebA-HQ				Places2			
		10%～20%	20%～30%	30%～40%	40%～50%	10%～20%	20%～30%	30%～40%	40%～50%
PSNR↑	MEDFE	30.97	27.75	25.36	23.47	29.05	25.92	23.78	22.07
	TFill	31.35	28.32	26.02	24.21	28.87	25.76	23.65	21.54
	RFR	30.93	26.05	27.31	24.28	27.36	24.95	22.83	21.25
	LaMa	32.55	29.31	26.82	25.10	27.45	24.91	22.84	21.31
	CMT	33.44	30.05	27.32	25.41	29.46	26.22	23.87	**21.76**
	CoordFill	33.31	30.12	27.35	25.21	**29.53**	26.31	24.03	21.25
	本文	**34.37**	**30.54**	**27.70**	**25.44**	29.47	**26.31**	**24.13**	21.69
SSIM↑	MEDFE	0.971	0.943	0.908	0.865	0.941	0.888	0.825	0.752
	TFill	0.975	0.951	0.925	0.912	0.957	0.916	0.865	0.804
	RFR	0.944	0.906	0.861	0.812	0.929	0.892	0.831	0.756
	LaMa	0.976	0.962	0.899	0.850	0.943	0.902	0.848	0.784
	CMT	0.983	0.966	0.950	0.913	0.962	**0.953**	0.875	0.809
	CoordFill	0.981	0.964	0.942	0.911	0.963	0.924	0.879	0.818
	本文	**0.986**	**0.973**	**0.957**	**0.938**	**0.972**	0.927	**0.887**	**0.841**
LPIPS↓	MEDFE	0.032 5	0.055 2	0.080 9	0.111 2	0.063 1	0.105 3	0.150 5	0.201 2
	TFill	0.028 1	0.058 5	0.095 1	0.125 4	0.071 1	0.080 5	0.121 6	0.160 5
	RFR	0.037 6	0.086 3	0.105 4	0.139 2	0.052 4	0.083 4	0.114 5	0.145 6
	LaMa	0.047 1	0.065 2	0.102 9	0.136 9	0.045 2	0.082 5	0.124 5	0.168 9
	CMT	0.028 3	0.054 0	0.083 7	0.127 2	**0.035 6**	0.068 5	0.106 2	0.152 0
	CoordFill	0.027 9	0.053 8	0.088 4	0.120 3	0.049 5	0.067 4	**0.103 8**	0.145 6
	本文	**0.027 5**	**0.052 1**	**0.080 6**	**0.113 8**	0.039 1	**0.065 4**	0.104 2	**0.141 6**

2.4　定性比较

尽管评价指标能够为模型的性能提供乐观的量化参考，但难以全面反映人类的视觉感知和主观差异。因此，下面本文引入定性比较，作为对提出方法性能分析的补充。

由于篇幅限制，图 5 展示了 MEDFE、TFill、CMT、CoordFill 和本文方法在 CelebA-HQ 数据集和 Places2 数据集上受损图像的修复效果，其中，图 5(a)为原始图像，图 5(b)为增加掩码后的破损图像，图 5(c)～(g)为对比模型的修复效果和本文的修复效果。前 3 行是在 CelebA-HQ 数据集上得到的结果，最后两行是在 Places2 数据集上得到的结果。与其他模型相对比，在 CelebA-HQ 数据集上，本文所提出的方法在第 1 行在牙齿以及眼睛的修复效果上更自然且接近真实图像；第 2 和第 3 行所生成的头发细节上更接近真实纹理。在 Places2 数据集上，本文所提方法更合理的修复了背景纹理，没有明显的伪影。



(a) 原图 (a) Original image　(b) 破损图像 (b) Damaged image　(c) MEDFE　(d) TFill　(e) CMT　(f) CoordFill　(g) 本文方法 (g) Ours

图 5　不同方法的定性比较

Fig. 5　Qualitative comparison of different methods

2.5　消融实验

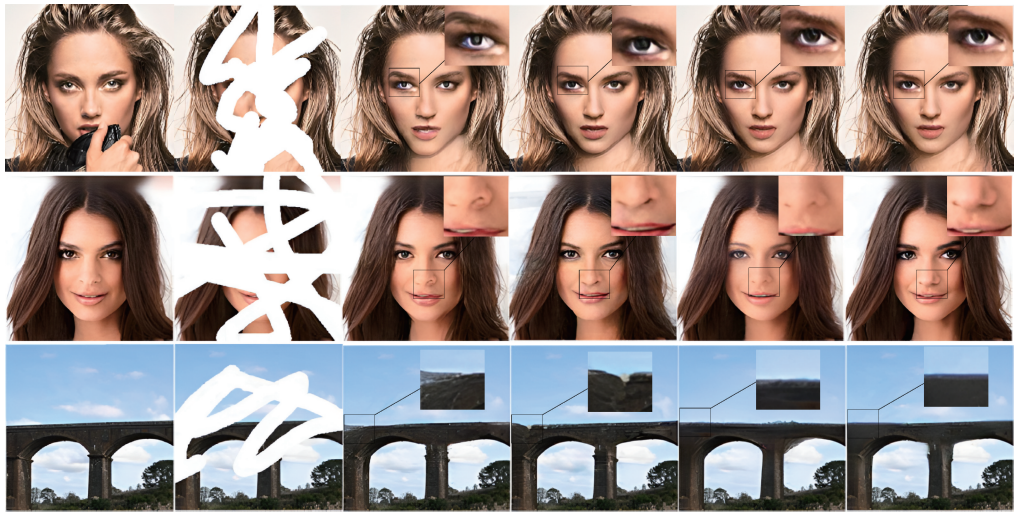
为验证本文提出的梯度图引导机制与 ECA 模块的有效性,在相同实验配置下于 CelebA-HQ 和 Places2 数据集上进行消融实验,采用不同比例的随机不规则掩码,其中 10%～30%划为简单掩码,30%～50%划为复杂掩码。实验共设计 4 种方案:1)仅使用基线网络模型方案 A;2)在基线模型上引入梯度图引导机制方案 B;3)在基线模型上加入 ECA 模块以增强通道注意力方案 C;4)同时引入梯度图引导与 ECA 模块方案 D。4 种方案在简单掩码和复杂掩码下的实验结果如表 2 所示,4 种方案的人脸和自然场景修复效果图如图 6 所示,其中图 6(a)为原始图像,图 6(b)为带掩码的破损图像,图 6(c)～(f)为 4 种方案。

从表 2 中实验结果可以看出,在 CelebA-HQ 和 Places2 两个数据集上,方案 B 和 C 均较基线方案 A 取得了明显的性能提升,说明二者分别在结构还原与通道特征强化方面均发挥了积极作用。进一步地,方案 D 在同时引入梯度图与 ECA 模块后表现最优,其中在 CelebA-HQ

表 2　消融研究

Table 2　Ablation study

指标	方法	CelebA-HQ		Places2	
		简单掩码	复杂掩码	简单掩码	复杂掩码
PSNR ↑	方案 A	30.37	25.03	25.02	21.34
	方案 B	32.15	25.89	26.53	22.13
	方案 C	31.92	25.65	26.65	21.92
	方案 D	32.64	26.45	27.33	22.89
SSIM ↑	方案 A	0.963	0.935	0.902	0.836
	方案 B	0.975	0.943	0.935	0.860
	方案 C	0.978	0.939	0.928	0.857
	方案 D	0.981	0.947	0.945	0.863
LPIPS ↓	方案 A	0.053 6	0.113 5	0.052 5	0.138 6
	方案 B	0.040 2	0.104 5	0.049 6	0.133 2
	方案 C	0.041 6	0.107 2	0.050 7	0.136 5
	方案 D	0.038 5	0.101 3	0.048 7	0.130 5



(a) 原图 (a) Original image　(b) 破损图像 (b) Damaged image　(c) 方案A (c) Ablation A　(d) 方案B (d) Ablation B　(e) 方案C (e) Ablation C　(f) 方案D (f) Ablation D

图 6　消融实验的定性比较

Fig. 6　Qualitative comparison of ablation experiments

数据集上，简单掩码下 PSNR 达到 32.64 dB、SSIM 达到 0.981、LPIPS 降低至 0.038 5；复杂掩码下 PSNR 为 26.45 dB、SSIM 为 0.947、LPIPS 为 0.107 2。在 Places2 数据集上亦取得相同趋势，方案 D 在两种掩码条件下均显著优于其他方案。综合来看，梯度引导修复对整体性能的提升更为显著，而 ECA 模块在进一步增强细节特征辨别方面起到了重要辅助作用。

3　结　　论

针对当前图像修复方法在结构还原不完整、细节感知能力不足的问题，本研究提出了一种融合梯度图引导与通道注意力机制的扩散模型图像修复方法。通过引入梯度图作为结构先验，引导模型更准确地还原图像的全局结构与细节信息；同时，ECA 模块增强了通道维度上的关键特征建模能力，有效提升了修复图像在纹理细节与语义一致性方面的表现。实验结果表明，所提方法在多个公开数据集上取得了优于主流方法的性能提升，验证了其在结构完整性、细节保留以及整体视觉质量方面的综合优势。未来的工作将尝试将该方法拓展至多模态修复场景，以实现更高层次的语义控制与多样性生成。

参考文献

[1] 杨竹青，谢宏. 基于方差调节策略耦合结构特征的图像修复算法[J]. 电子测量与仪器学报，2020，34(10)：25-32.
YANG ZH Q，XIE H. Image inpainting algorithm based on variance adjustment strategy coupling structural feature [J]. Journal of Electronic Measurement and Instrumentation，2020，34(10)：25-32.

[2] PATHAK D，KRÄHENBUHL P，DONAGUE J，et al. Context encoders：Feature learning by inpainting[C]. IEEE Conference on Computer Vision and Pattern Recognition，2016：2536-2544.

[3] IIZUKA S，SIMO-SERRA E，ISHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics(ToG)，2017，36(4)：1-14.

[4] WANG Y，TAO X，QI X J，et al. Image inpainting via generative multi-column convolutional neural networks[J]. ArXiv preprint arXiv：1810.08771，2018.

[5] YU J H，LIN ZH，YANG J M，et al. Generative image inpainting with contextual attention[C]. IEEE Conference on Computer Vision and Pattern Recognition，2018：5505-5514.

[6] NAZERI K，NG E，JOSEPH T，et al. EdgeConnect：Structure guided image inpainting using edge prediction [C]. IEEE International Conference on Computer Vision，2019：3265-3274.

[7] 开志强，苗锡奎，马天磊，等. 基于全局语义学习和显著目标感知的激光干扰图像修复[J]. 仪器仪表学报，2024，45(7)：38-51.
KAI ZH Q，MIAO X K，MA T L，et al. Laser jamming image inpainting based on global semantic learning and salient target awareness [J]. Chinese Journal of Scientific Instrument，2024，45(7)：38-51.

[8] 李雪瑾，李昕，徐艳杰. 基于生成对抗网络的数字图像修复技术[J]. 电子测量与仪器学报，2019，33(1)：40-46.
LI X J，LI X，XU Y J. Digital image restoration technology based on generative adversarial networks[J]. Journal of Electronic Measurement and Instrumentation，2019，33(1)：40-46.

[9] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial nets[J]. ArXiv preprint arXiv：1406.2661，2014.

[10] YAN ZH Y，LI X M，LI M，et al. Shift-net：Image inpainting via deep feature rearrangement [C]. European Conference on Computer Vision(ECCV)，2018：3-19.

[11] HONG X，XIONG P F，JI R H，et al. Deep fusion network for image completion [C]. 27th ACM International Conference on Multimedia，2019：2033-2042.

[12] ZENG Y H，FU J L，CHAO H Y，et al. Learning pyramid-context encoder network for high-quality image inpainting [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition，2019：1486-1494.

[13] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[J]. ArXiv preprint arXiv：1706.03762，2017.

[14] ZHOU Y Q，BARNES C，SHECHTMAN E，et al. Transfill：Reference-guided image inpainting by merging multiple color and spatial transformations[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：2266-2276.

[15] HE K M，CHEN X L，XIE S N，et al. Masked autoencoders are scalable vision learners[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：16000-16009.

[16] DONG Q L，CAO CH J，FU Y W. Incremental transformer structure enhanced image inpainting with masking positional encoding[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：11358-11368.

[17] 滕诗宇,何丽君.融合多特征与全局-局部 Transformer 的图像修复算法[J].电子测量技术,2025,48(6):121-129.
TENG SH Y, HE L J. Fusion of multi-features and global-local Transformer forimage inpainting [J]. Electronic Measurement Technology,2025,48(6):121-129.

[18] HO J, JAIN A N, ABBEEL P. Denoising diffusion probabilistic models [J]. Advances in Neural Information Processing Systems,2020, 33: 6840-6851.

[19] LUGMAYR A, DANELLJAN M, ROMERO A, et al. Repaint: Inpainting using denoising diffusion probabilistic models[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11461-11471.

[20] XIE SH AN, ZHANG ZH F, LIN ZH, et al. Smartbrush: Text and shape guided object inpainting with diffusion model[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 22428-22437.

[21] LIU H P, WANG Y, QIAN B, et al. Structure matters: Tackling the semantic discrepancy in diffusion models for image inpainting[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 8038-8047.

[22] 张绪,胡峻峰.基于扩散模型的二阶段细化图像修复模型[J].电子测量技术,2025,48(17):142-150.
ZHANG X, HU J F. Diff-2sIR: Diffusion-based refinement two-stage image restoration model [J]. Electronic Measurement Technology, 2025, 48(17): 142-150.

[23] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[24] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition,2020,42(8):2011-2023.

[25] DOERSCH C, SINGH S, GUPTA A, et al. What makes paris look like paris? [J]. ACM Transactions on Graphics,2012, 31(4): 1-9.

[26] KARRAS T, AILA T, LAINE S, et al. Progressive growing of gans for improved quality, stability, and variation[C]. Proceedings of International Conference on Learning Representations(ICLR), 2018.

[27] ZHOU B L, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million image database for scene recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017, 40(6): 1452-1464.

[28] LIU G L, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]. European Conference on Computer Vision (ECCV), 2018, 11215: 89-105.

[29] LI J Y, WANG N, ZHANG L F, et al. Recurrent feature reasoning for image inpainting[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7757-7765.

[30] KO K, KIM C S. Continuously masked transformer for image inpainting [C]. IEEE/CVF International Conference on Computer Vision,2023: 13123-13132.

[31] ZHENG CH X, CHAM T J, CAI J F, et al. Bridging global context interactions for high-fidelity image completion[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition,2022: 11512-11522.

[32] LIU H Y, JIANG B, SONG Y B, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations [C]. Computer Vision-ECCV 2020,2020,12347: 725-741.

[33] SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with fourier convolutions[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2022: 2149-2159.

[34] LIU W H, CUN X D, PUN C M, et al. Coordfill: Efficient high-resolution image inpainting via parameterized coordinate querying [C]. AAAI Conference on Artificial Intelligence, 2023, 37(2): 1746-1754.

作者简介

张荣铠,硕士研究生,主要研究方向为图像修复。
E-mail:23212030465@stumail.xsyu.edu.cn

景明利(通信作者),副教授,博士,主要研究方向为压缩c中的稀疏低秩恢复、数值优化算法、图像处理、机器学习。
E-mail:mljingsy@xsyu.edu.cn

焦龙,教授,博士,主要研究方向为化学计量学、模式识别。
E-mail:mop@xsyu.edu.cn