研究与开发

DGSF-AOT：动态门控与自注意力融合增强的人脸图像修复

柏武贰[1,2]，张乾[1,2,3]，刘霜[1,2]，滕林[1,2]，杨思红[1,2]
（1. 贵州民族大学数据科学与信息工程学院，贵州 贵阳 550025；
2. 贵州省模式识别与智能系统重点实验室，贵州 贵阳 550025；
3. 贵州民族大学教务处，贵州 贵阳 550025）

摘　要：针对复杂背景下的人脸图像修复任务中普遍存在的细粒度纹理合成不足、结构修复断层和语义失谐的现象，提出了基于动态门控机制与自注意力模块融合增强的人脸图像修复网络。新算法通过构建多级膨胀卷积组捕获局部细节与长程上下文信息，并引入双重创新机制：（1）深度动态门控机制采用多层卷积与批归一化实现空间自适应的特征选择，取代传统残差连接的固定融合方式，显著提升了特征表达的灵活性和精准度；（2）自注意力机制显式建模全局像素依赖关系，有效解决了大范围缺损修复中的结构连贯性和细粒度纹理合成难题。实验结果表明，相对于较优对比算法SCAT，新算法在FFHQ、CelebA-HQ和LFW人脸数据集上的PSNR和SSIM指标平均提升了0.382 dB和0.004 1，FID平均改善了7.81%，尤其是在大面积遮挡（>50%）场景下，FID平均下降了2.153 4，显著提升了复杂背景下人脸图像修复质量，在生成逼真纹理、结构一致性方面有突出的修复优势。

关键词：上下文建模；动态特征融合；动态门控机制；自注意力机制
中图分类号：TN957.52；TP391.41
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2026016

DGSF-AOT: dynamic gating and self-attention fusion enhancement for face image restoration

Bai Wuer[1,2], Zhang Qian[1,2,3], Liu Shuang[1,2], Teng Lin[1,2], Yang Sihong[1,2]
1. School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China
2. Key Laboratory of Pattern Recognition and Intelligent Systems of Guizhou, Guiyang 550025, China
3. Academic Affairs Office, Guizhou Minzu University, Guiyang 550025, China

Abstract: Aiming at the phenomena of insufficient fine-grained texture synthesis, structural repair faults, and seman-

收稿日期：2025−06−09；**修回日期：**2025−07−04
通信作者：张乾，gzmuzq@gzmu.edu.cn
基金项目：贵州省高等学校大数据分析与智能计算重点实验室项目（No.黔教技[2023]012号）；贵州民族大学校级科研项目（No.GZMUZK [2021] YB23，No.GZMUZK [2023] QN10）
Foundation Items: Key Laboratory of Big Data Analysis and Intelligent Computing in Guizhou Higher Education Institutions (No.Guizhou Education Technology [2023] 012), School-level Scientific Research Projects of Guizhou Minzu University (No.GZMUZK [2021] YB23, No.GZMUZK [2023] QN10)

tic detuning, which are commonly found in face image restoration tasks in complex contexts, a face image restoration network based on the fusion enhancement of a dynamic gating mechanism with a self-attention module was proposed. The algorithm captured local details and long-range contextual information by constructing a multilevel dilated convolutional group, and introduced a dual innovative mechanism: (1) the deep dynamic gating mechanism adopted multilayer convolution with batch normalization to achieve spatially adaptive feature selection, replacing the fixed fusion of the traditional residual connection, which significantly enhanced the flexibility and accuracy of feature expression; (2) the self-attention mechanism explicitly modeled global pixel dependencies, which effectively solved the difficulties of structural coherence and fine-grained texture synthesis in large-scale defect repair. Experiments show that, compared with the better comparison algorithm SCAT, this new method improves PSNR and SSIM metrics by an average of 0.382 dB and 0.004 1, and improves FID by an average of 7.81% on three face datasets, namely, FFHQ, CelebA-HQ, and LFW, especially in the scene of large-area occlusion (>50%), the FID decreased by an average of 2.153 4, significantly improving the accuracy of face images in complex backgrounds. It improves the quality of face image restoration under complex backgrounds, especially in generating realistic textures and structural consistency, showing outstanding advantages.

Key words: contextual modeling, dynamic feature fusion, dynamic gating mechanism, self-attention mechanism

0　引言

图像修复[1]旨在基于图像的已知语义与纹理信息，生成视觉合理且结构连贯的缺失区域内容，是计算机视觉领域的关键任务之一。该技术在数字文物保护[2]、影视特效编辑[3]及自动驾驶场景补全[4]等领域具有重要应用价值。

早期图像修复主要依赖物理先验与手工设计特征，如文献[1]提出的基于扩散方程的图像补全方法，通过局部像素传播实现小范围缺失修复；文献[5]中改进的PatchMatch算法利用纹理合成技术实现内容填充。然而，这些传统方法在复杂语义推理与大规模缺失修复中存在显著局限性。

随着深度学习技术的发展，基于卷积神经网络（convolutional neural network，CNN）和生成对抗网络（generative adversarial network，GAN）的图像修复技术取得了突破性进展。例如，文献[6]提出的非局部神经网络通过计算特征图中任意两个位置之间的相似性，捕捉长距离依赖关系，在一定程度上提升了模型对远距离上下文的感知能力，但处理大规模图像时效率较低。文献[7]通过上下文注意力机制解决了远距离特征匹配问题，但受限于固定掩膜输入和传统卷积的刚性，又提出了部分卷积网络结合GAN架构[8]，通过动态自适应地更新卷积核的权重来处理自由形状缺失区域的信息，但复杂场景下的细节恢复不够细腻。文献[9]通过概率学习框架和注意力机制，解决了传统图像修复方法只能生成单一结果的问题，同时在修复效果和多样性之间取得平衡，但缺失区域与上下文关联较弱时，会出现细节模糊以及结构失真现象。文献[10]提出了动态特征选择机制，通过可变形卷积与区域机制结合的可变掩码卷积（variable mask convolution，VMC）模块，解决了传统卷积在修复过程中无效信息干扰导致的特征不稳定性问题，但在无纹理或低纹理区域（如纯色背景），动态选择过度依赖周围信息，导致纹理一致不足或色彩不匹配。文献[11]将小波变换深度整合到修复网络中，将输入图像分解为低频和高频子带，有效处理了不同尺度的结构信息和纹理细节，但在大面积遮挡场景中，低频子带缺乏足够的先验信息，导致修复区域的结构模糊。文献[12]使用聚合上下文变换（aggregated contextual-transformation，AOT）捕获信息丰富的远距离图像上下文和丰富的感兴趣区域上

下文推理模式，有效保持全局结构一致性，但在复杂纹理区域修复时，容易导致局部纹理模糊和结构扭曲。文献[13]提出了生成记忆引导的语义推理模式，有效解决了传统图像修复方法在处理大范围缺失区域时整体语义推理能力不足的问题，但在局部细节修复时不清晰或物体形状修复不准确。文献[14]提出了一种基于分割混淆对抗训练（segmentation confusion adversarial training，SCAT）和对比学习的新型图像修复方法，利用判别器的特征空间进行对比学习，以提高修复图像的真实性和一致性，但在复杂背景下的图像难以生成细粒度高的纹理，且结构连贯性合成能力不足。

为此，针对目前复杂背景下人脸图像修复细粒度纹理合成能力不足、结构断层和语义失谐的问题，本文提出一种基于动态门控与自注意力模块融合增强的人脸图像修复网络。该模型使用自注意力机制显式建模全局像素特征，并融合空间动态自适应局部特征实现修复任务。

1 本文方法

本文方法中，原始图像$\boldsymbol{I}_{\mathrm{gt}}\in\mathbf{R}^{H\times W\times 3}$（其中，$H$表示图像高度，$W$表示图像宽度，3表示RGB通道，$\mathbf{R}$为实数集）和掩码图像$\boldsymbol{M}\in\mathbf{R}^{H\times W\times 3}$经过Hadamard乘积后退化为受损图像$\boldsymbol{I}_{\mathrm{m}}$，表达式为：

$$\boldsymbol{I}_{\mathrm{m}}=\boldsymbol{I}_{\mathrm{gt}}\odot\boldsymbol{M} \tag{1}$$

其中，$\boldsymbol{M}$的元素取值为0或者1（0表示图像的缺失区域部分，1表示图像的有效区域），⊙表示Hadamard乘积。

修复网络将受损图像$\boldsymbol{I}_{\mathrm{m}}$和掩码图像$\boldsymbol{M}$作为双输入数据，在编解码的生成器$G$下，通过多尺度特征提取模块、注意力机制以及在掩码引导下生成具有空间感知的修复估计值$\boldsymbol{I}_{\mathrm{out}}\in\mathbf{R}^{H\times W\times 3}$。将受损图像$\boldsymbol{I}_{\mathrm{m}}$的有效区域和估计值$\boldsymbol{I}_{\mathrm{out}}$的掩码生成区域进行结合，生成修复结果$\boldsymbol{I}_{\mathrm{comp}}$，表达式为：

$$\boldsymbol{I}_{\mathrm{comp}}=(1-\boldsymbol{M})\odot\boldsymbol{I}_{\mathrm{out}}+\boldsymbol{M}\odot\boldsymbol{I}_{\mathrm{m}} \tag{2}$$

修复网络的生成器基于GAN框架，由判别器与生成器协同训练驱动。生成器主要由编码、深度堆叠的动态门控与自注意力融合增强AOT（dynamic gating and self-attention fusion enhanced AOT，DGSF-AOT）模块、解码模块组成。在训练策略方面，修复网络采用了文献[14]提出的分割混淆对抗训练损失$\mathcal{L}_{\mathrm{SCAT}}$、对比学习损失$\mathcal{L}_{\mathrm{contra}}$、全局对抗训练损失$\mathcal{L}_{\mathrm{adv}}$和$L1$重建损失$\mathcal{L}_{\mathrm{rec}}$进行联合训练。DGSF-AOT-GAN修复框架如图1所示。

1.1 DGSF-AOT模块

为了解决复杂背景下人脸图像修复出现细粒度纹理合成不足、结构断层和语义失谐的问题，本文在生成器中引入DGSF-AOT模块，其核心设计继承并改进了文献[12]的AOT模块架构。

DGSF-AOT模块在AOT的基础上，引入基于通道压缩的自注意力机制[15]，该机制通过建模特征空间的长程依赖，增强了对全局语义一致性的感知。同时，将静态的门控机制改为多层卷积

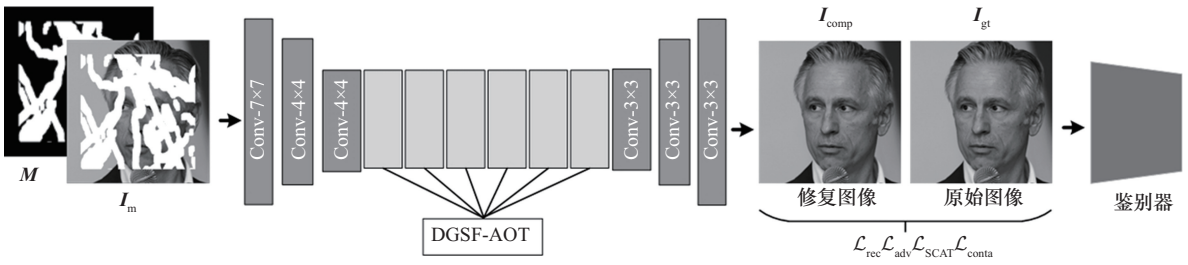


图1 DGSF-AOT-GAN修复框架

注：人脸数据来自FFHQ数据集。

嵌套批归一化的动态门控机制，以端到端方式学习像素级空间自适应融合权重，相较于原始AOT的单层门控结构，显著提高了特征融合的细粒度与精度。DGSF-AOT模块结构如图2所示。

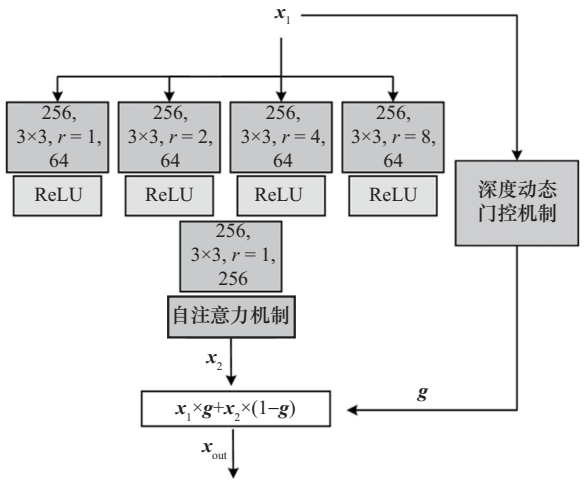


图2　DGSF-AOT模块结构

首先，受损图像$\boldsymbol{I}_{\mathrm{m}}$和掩码$\boldsymbol{M}$经过生成器$G$的编码器处理后得到空间压缩的图像$\boldsymbol{x}_1$，定义为：

$$\boldsymbol{x}_1=\mathrm{encoder}\left(\boldsymbol{I}_{\mathrm{m}},\boldsymbol{M}\right) \tag{3}$$

其中，encoder表示编码器。

$\boldsymbol{x}_1$作为DGSF-AOT模块的输入数据，经过并行多分支中集成差异化膨胀率（r=1,2,4,8）的空洞卷积，以解耦式捕获局部细节与多尺度上下文全局特征，并压缩通道以减少计算冗余，定义为：

$$\boldsymbol{x}_i'=\sigma\left(\mathrm{Conv}_{3\times3}^{i}\left(\boldsymbol{x}_1\right)\right)\quad i=1,2,4,8 \tag{4}$$

其中，σ表示ReLU函数，$\mathrm{Conv}_{3\times3}^{i}$表示膨胀率为$i$、卷积核为3×3的卷积。

将$\boldsymbol{x}_i'(i=1,2,4,8)$按照通道进行融合，并经过卷积和自注意力模块操作后，得到具有高级语义特征的$\boldsymbol{x}_2$。$\boldsymbol{x}_2$与深度动态门控值$\boldsymbol{g}$结合运算得到$\boldsymbol{x}_{\mathrm{out}}$，并将$\boldsymbol{x}_{\mathrm{out}}$作为下一个DGSF-AOT模块的输入，以此类推。$\boldsymbol{x}_2$定义为：

$$\boldsymbol{x}_2=\mathrm{att}\left(\mathrm{Conv}_{3\times3}^{1}\left[\boldsymbol{x}_1';\boldsymbol{x}_2';\boldsymbol{x}_4';\boldsymbol{x}_8'\right]\right) \tag{5}$$

$$\boldsymbol{g}=\mathrm{DG}\left(\boldsymbol{x}_1\right) \tag{6}$$

$$\boldsymbol{x}_{\mathrm{out}}=\boldsymbol{x}_1\times\boldsymbol{g}+\boldsymbol{x}_2\times\left(1-\boldsymbol{g}\right) \tag{7}$$

其中，[;;;]表示按通道拼接，att表示自注意力机制模块，DG表示深度动态门控机制，$\boldsymbol{g}$表示门控权重。

1.1.1　自注意力机制

为了增强模型对复杂背景下人脸图像的全局语义一致性、结构感知，更精准地利用图像中长距离的上下文信息，使结构和语义上的修复都更加连贯、自然，引入自注意力机制模块是一种有效的解决方案。自注意力机制模块如图3所示。

假设输入$\boldsymbol{x}_2''$的形状为（B,C,H,W），其中B表示批量大小，C表示通道数，在查询（Query）矩阵分支中，经过卷积操作，其通道压缩为原来的1/8，再经过展平和转置操作后，得到特征$\boldsymbol{Q}$，定义为：

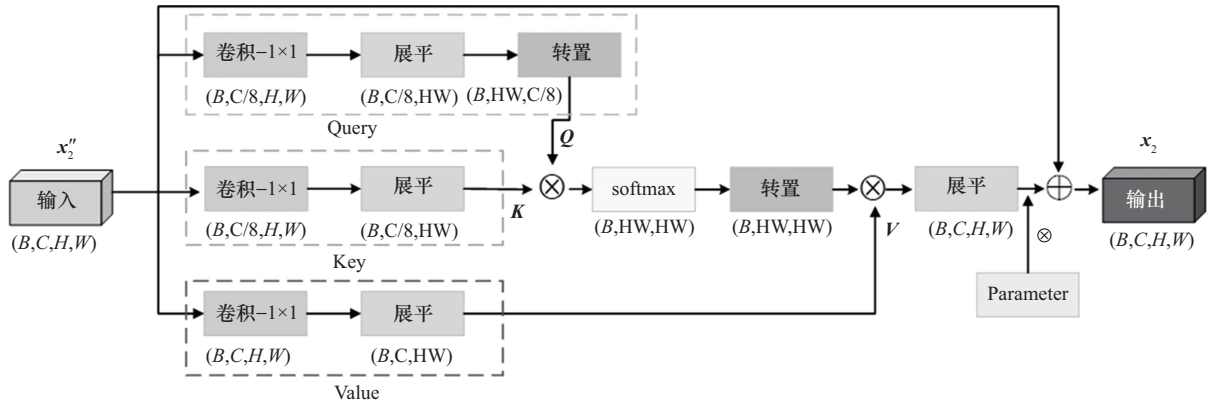


图3　自注意力机制模块

$$\boldsymbol{Q}=\varphi\left(\tau\left(\mathrm{Conv}_{1\times1}\left(\boldsymbol{x}_2''\right)\right)\right) \tag{8}$$

其中，τ和φ分别表示展平和转置操作。

同理，在键（Key）矩阵和值（Value）矩阵分支中，$\boldsymbol{x}_2''$经过1×1卷积层和展平操作后，得到$\boldsymbol{K}$和$\boldsymbol{V}$，定义为：

$$\boldsymbol{K}=\tau\left(\mathrm{Conv}_{1\times1}\left(\boldsymbol{x}_2''\right)\right) \tag{9}$$

$$\boldsymbol{V}=\tau\left(\mathrm{Conv}_{1\times1}\left(\boldsymbol{x}_2''\right)\right) \tag{10}$$

将查询矩阵$\boldsymbol{Q}$和键矩阵$\boldsymbol{K}$进行矩阵乘积，动态生成注意力权重$\boldsymbol{E}$，并对$\boldsymbol{E}$的最后一个维度进行归一化处理，确保每个位置的注意力权重和为1，定义为：

$$\boldsymbol{E}=\boldsymbol{Q}\cdot\boldsymbol{K} \tag{11}$$

$$\boldsymbol{A}=\delta\left(\boldsymbol{E}\right) \tag{12}$$

其中，δ表示softmax函数，$\boldsymbol{A}$表示注意力权重。

将注意力权重$\boldsymbol{A}$矩阵转置后与值矩阵$\boldsymbol{V}$相乘，得到加权聚合后的特征$\boldsymbol{O}$，定义为：

$$\boldsymbol{O}=\boldsymbol{V}\cdot\boldsymbol{A}^{\mathrm{T}} \tag{13}$$

最后，通过γ控制注意力结果的贡献进行残差连接，得到最终的结果$\boldsymbol{x}_2$，定义为：

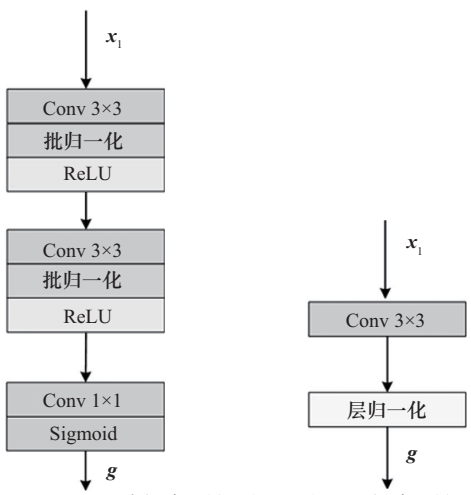
$$\boldsymbol{x}_2=\gamma\boldsymbol{O}+\boldsymbol{x}_2'' \tag{14}$$

其中，γ表示Parameter函数，设置γ初始值为0，随着训练的进行，γ逐渐学习增大。

1.1.2　深度动态门控机制

为解决复杂背景下的人脸图像修复中出现细粒度纹理合成不足、局部语义不合理的问题，本文引入深度动态门控机制网络。该网络通过多尺度空洞卷积提取局部–全局上下文特征，结合自注意力建模长程语义关联，并利用动态门控机制生成空间自适应的权重特征图，实时调节原始特征与多尺度增强特征的融合比例，提升特征融合精度与语义敏感性，最终实现纹理真实性、视觉语义一致性上的协同优化。门控机制网络如图4所示。

DGSF-AOT动态门控机制网络如图4（a）所示。深度动态门控机制网络由卷积、批归一化和ReLU激活函数构成，能动态适应不同区域的修复需求（如区分边缘和平滑区域），并学习更复杂的门控权重分布，增强对多尺度特征的精细化融合能力，适合处理复杂背景的纹理和结构修复。



（a）DGSF-AOT动态门控机制网络　（b）AOT门控机制网络

图4　门控机制网络

在受损图像$\boldsymbol{I}_{\mathrm{m}}$和掩码$\boldsymbol{M}$经过生成器$G$的编码器处理后得到$\boldsymbol{x}_1$特征图，将其作为深度动态门控机制的输入数据。首先，$\boldsymbol{x}_1$经过3×3的卷积提取初级空间特征，定义为：

$$\boldsymbol{g}_1=\mathrm{Conv}_{3\times3}\left(\boldsymbol{x}_1\right) \tag{15}$$

为了使提取的特征$\boldsymbol{g}_1$能够稳定训练，并减少内部协变量偏移，将其每个通道的数据进行批归一化处理，定义为：

$$\boldsymbol{g}_2=\xi(\boldsymbol{g}_1) \tag{16}$$

其中，ξ表示批归一化函数。

为了增加模型的表达能力和学习特征间的非线性关系，引入ReLU激活函数，该激活函数能够将所有负的像素值设置0，定义为：

$$\boldsymbol{g}_3=\sigma(\boldsymbol{g}_2) \tag{17}$$

模型训练的人脸图像数据具有复杂的背景，因此，提取高级语义关系权重尤为重要。本文在式（17）的基础上，再次引入3×3的卷积提取高级空间语义特征，同时引入ξ和σ函数以稳定模

型训练和增强模型学习非线性关系，定义为：

$$\boldsymbol{g}_4=\sigma\left(\xi\left(\mathrm{Conv}_{3\times3}\left(\boldsymbol{g}_3\right)\right)\right) \tag{18}$$

最后，使用 1×1 的卷积进行跨通道信息整合，并经过 Sigmoid 函数处理后映射到[0,1]，生成相应像素位置的修复权重，定义为：

$$\boldsymbol{g}=\varsigma\left(\mathrm{Conv}_{1\times1}\left(\boldsymbol{g}_4\right)\right) \tag{19}$$

其中，$\boldsymbol{g}_i$ 表示中间门控权重，ς 表示 Sigmoid 函数。

原始 AOT 门控机制网络如图 4（b）所示，其采用单层卷积和层归一化的轻量设计，适合简单修复任务，但难以捕捉复杂区域的动态权重关系，对具有复杂背景人脸图像特征的捕获效果较差。

1.2　损失函数

本文算法在训练阶段遵循文献[14]的损失函数，即分割混淆对抗训练损失 $\mathcal{L}_{\mathrm{SCAT}}$、对比学习损失 $\mathcal{L}_{\mathrm{contra}}$、全局对抗训练损失 $\mathcal{L}_{\mathrm{adv}}$ 和 $L1$ 重建损失 $\mathcal{L}_{\mathrm{rec}}$。

分割混淆对抗训练损失由分割网络 S 和生成器网络 G 组成，表达式分别为：

$$\mathcal{L}_{\mathrm{SCAT}}(S)=-\mathbb{E}\left[\begin{array}{l}\dfrac{1}{HW}\displaystyle\sum_{i=1}^{HW}\left[\boldsymbol{m}_i\ln S(\bar{\boldsymbol{x}})_i+(\boldsymbol{1}-\boldsymbol{m}_i)\ln\left(\boldsymbol{1}-S(\bar{\boldsymbol{x}})_i\right)\right]+\\ \dfrac{1}{HW}\displaystyle\sum_{i=1}^{HW}\left[\bar{\boldsymbol{m}}_i\ln S(\boldsymbol{x})_i+(\boldsymbol{1}-\bar{\boldsymbol{m}}_i)\ln\left(\boldsymbol{1}-S(\boldsymbol{x})_i\right)\right]\end{array}\right] \tag{20}$$

$$\mathcal{L}_{\mathrm{SCAT}}(G)=-\mathbb{E}\left[\frac{1}{HW}\sum_{i=1}^{HW}\left[\bar{\boldsymbol{m}}_i\ln S(\bar{\boldsymbol{x}})_i+(1-\bar{\boldsymbol{m}}_i)\ln\left(1-S(\bar{\boldsymbol{x}})_i\right)\right]\right] \tag{21}$$

其中，$\bar{\boldsymbol{m}}$ 表示一个全 1 填充的掩码，$\bar{\boldsymbol{x}}$ 和 $\boldsymbol{x}$ 分别表示生成图像和真实图像数据，$\mathbb{E}$ 表示信息熵的期望损失。

对比学习损失主要由纹理对比学习损失 $\mathcal{L}_{\mathrm{contra}}^{\mathrm{text}}$ 和语义对比学习损失 $\mathcal{L}_{\mathrm{contra}}^{\mathrm{sem}}$ 组成。纹理对比学习损失定义为：

$$\mathcal{L}_{\mathrm{contra}}^{\mathrm{text}}=\mathbb{E}\sum_{i=1}^{N}\frac{d\left(\boldsymbol{D}_i(\bar{\boldsymbol{x}}),\boldsymbol{D}_i(\boldsymbol{x})\right)}{d\left(\boldsymbol{D}_i(\bar{\boldsymbol{x}}),\boldsymbol{D}_i(\tilde{\boldsymbol{x}})\right)} \tag{22}$$

其中，$d(\cdot,\cdot)$、$\boldsymbol{D}_i(\cdot)$ 和 N 分别表示距离度量、判别器 D 的第 i 层输出特征映射和使用的浅层总数，$\bar{\boldsymbol{x}}$ 表示锚点样本，$\boldsymbol{x}$ 表示正样本，$\tilde{\boldsymbol{x}}$ 表示负样本。

语义对比学习损失定义为：

$$\mathcal{L}_{\mathrm{contra}}^{\mathrm{sem}}=-\mathbb{E}\left[\ln\left(\frac{\exp\left(\boldsymbol{D}(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{D}(\boldsymbol{x})/t\right)}{\exp\left(\boldsymbol{D}(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{D}(\boldsymbol{x})/t\right)+\displaystyle\sum_{j=1}^{M}\exp\left(\boldsymbol{D}(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{D}\left(\tilde{\boldsymbol{x}}_j\right)/t\right)}\right)\right] \tag{23}$$

其中，M 表示负样本数量，t 表示温度参数，用于调节对比学习的难度和特征的分布。

对比学习损失定义为：

$$\mathcal{L}_{\mathrm{contra}}=\lambda_{\mathrm{text}}\mathcal{L}_{\mathrm{contra}}^{\mathrm{text}}+\lambda_{\mathrm{sem}}\mathcal{L}_{\mathrm{contra}}^{\mathrm{sem}} \tag{24}$$

其中，λ_{text} 和 λ_{sem} 分别表示平衡两个相应损失的权值。

全局对抗训练损失和 $L1$ 重建损失函数分别定义为：

$$\mathcal{L}_{\mathrm{adv}}=\min_{G}\max_{D}\mathbb{E}_x\left[\ln\boldsymbol{D}(\boldsymbol{x})\right]+\mathbb{E}_{\hat{x}}\left[\ln\left(1-\boldsymbol{D}(\hat{\boldsymbol{x}})\right)\right] \tag{25}$$

$$\mathcal{L}_{\mathrm{rec}}=\mathbb{E}\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_1 \tag{26}$$

其中，$\hat{\boldsymbol{x}}$ 为生成图像样本数据。

总体训练损失函数为：

$$\mathcal{L}_{\mathrm{total}}=\lambda_{\mathrm{adv}}\left(\mathcal{L}_{\mathrm{adv}}+\mathcal{L}_{\mathrm{SCAT}}\right)+\mathcal{L}_{\mathrm{contra}}+\lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} \tag{27}$$

其中，λ_{adv} 和 λ_{rec} 分别表示控制相应损失的权重。

2　实验及结果分析

2.1　实验设置

实验在 GTX3090 单块图形处理单元（graphics processing unit，GPU）上基于 PyTorch 1.13 的学习框架进行训练和测试。训练中采用 Adam 优

化器优化模型，批次大小设置为4，学习率设置为1×10^{-4}，共计迭代35万次。

2.2 数据集

实验在FFHQ[16]、LFW[17]和CelebA-HQ[18]人脸数据集上进行。其中，FFHQ数据集包含7万张1 024×1 024像素的PNG格式高分辨率人脸图像，经严格筛选和预处理后具备丰富的多样性和真实性，涵盖不同年龄、性别、种族、表情及配饰（如眼镜、帽子等），并遵循伦理规范，实验从数据集中随机抽取30 000张作为训练集，2 000张作为测试集，并统一裁剪为256×256尺寸。LFW数据集包含了13 233张人脸数据，该数据取自互联网上公开的面部图像，包含了明星、普通人、政治家等不同身份来源的人脸图像，涵盖不同光照条件、姿态变化、表情变化、年龄、种族等，具备一定的多样性和复杂性，实验取11 910张图像作为训练集，剩余的1 323张图像作为测试集。CelebA-HQ数据集包含了30 000张高清人脸图像，每张图像都附带人脸关键点和属性信息，实验取28 000张图像作为训练集，2 000张图像作为测试集。掩码数据采用不规则掩码数据集[19]，包含12 000张随机生成的不规则掩码图像。根据掩码覆盖面积比例划分为1%～10%、11%～20%、21%～30%、31%～40%、41%～50%、51%～60%共6个区间，而每个区间包含2 000张随机生成的掩码数据，用于不同比例的遮挡测试。

2.3 对比实验

为了全面且严谨地评估所提方法性能的优劣性，本文选取了一系列具有代表性的先进对比算法，采用定量评估指标与定性视觉比较相结合的方式展开深入探究。算法包括AOT-GAN[12]、DSNet[10]、CTSDG[20]、PICMM[21]、MISF[22]、DCDPI[23]和SCAT[14]等前沿模型。针对图像修复效果评估指标，本文选取峰值信噪比（peak signal-to-noise ratio，PSNR）、结构相似性指数（structural similarity index，SSIM）、$L1$损失、弗雷歇初始距离（Frechet inception distance，FID）以及学习感知图像块相似度（learned perceptual image patch similarity，LPIPS）这五大核心指标，细致入微地对各算法进行对比分析。

2.3.1 定量对比实验分析

不同算法在FFHQ数据集上的对比结果见表1，其中加粗的为最优值。本文算法在PSNR、SSIM和$L1$指标上均显著优于对比算法。在大掩码比例（41%～60%）下，本文算法的PSNR和SSIM指标相较次优模型SCAT算法，平均分别提升了1.26%和0.998%，$L1$和FID指标平均分别降低了3.79%和11.25%。然而，当掩码比例大于50%时，相较于对比算法DCDPI，LPIPS值上升了5.79%，这反映了超大缺失区域的长程依赖建模瓶颈，但在常规遮挡条件（≤50%掩膜）下仍保持LPIPS领先。在小中掩码比例（1%～40%）下，除本文算法在掩码1%～10%下FID略高于SCAT算法外，其余指标明显优于其他算法，充分表明了本文算法的实用性和优越性。

不同算法在LFW数据集上的对比结果见表2。在给定的掩码范围下，本文算法的SSIM和$L1$指标均优于对比算法。在PSNR、FID和LPIPS指标中，本文算法除了在51%～60%的掩码比例下略低于或高于个别对比算法，其余掩码比例下均为最优值。例如，相较于DCDPI算法，本文算法在51%～60%的掩码下PSNR指标低于该算法0.107 2 dB，但本文算法在其他掩码下PSNR指标均高于该算法，并且平均提升了0.981 8 dB。对比MISF算法的FID指标，本文算法在51%～60%掩码下高于该算法0.614，但本文算法在FID上平均降低了0.898 2。相较于DSNet算法，本文算法在51%～60%下LPIPS指标略高于该算法0.003 4，但在所有掩码比例下平均降低了0.008 1。

不同算法在CelebA-HQ数据集上的对比结果见表3。本文算法除了在51%～60%掩码比例下

表 1　不同算法在 FFHQ 数据集上的对比结果

指标	掩码覆盖面积比例	AOT-GAN	DSNet	CTSDG	PICMM	MISF	DCDPI	SCAT	本文算法
PSNR/dB ↑	1%～10%	36.964 3	36.922 5	38.907 6	34.279 8	37.777 9	38.286 7	40.382 9	**40.594 9**
	11%～20%	30.391 4	31.234 9	32.721 5	28.619 0	31.988 6	32.545 7	33.962 1	**34.235 8**
	21%～30%	26.350 7	28.001 8	29.218 5	25.350 9	28.634 4	29.174 2	30.179 3	**30.489 5**
	31%～40%	22.989 4	25.654 4	26.546 2	22.752 3	26.123 8	26.635 7	27.288 0	**27.608 3**
	41%～50%	19.985 4	23.778 2	24.526 9	20.681 0	24.136 3	24.699 1	25.083 8	**25.404 9**
	51%～60%	14.840 3	21.144 8	21.513 2	17.455 4	21.156 2	21.773 6	21.574 4	**21.841 3**
SSIM ↑	1%～10%	0.980 4	0.978 5	0.984 2	0.968 6	0.981 7	0.983 1	0.988 2	**0.988 6**
	11%～20%	0.945 7	0.944 0	0.955 5	0.922 0	0.951 2	0.954 8	0.965 9	**0.967 1**
	21%～30%	0.895 5	0.901 1	0.916 4	0.864 6	0.911 5	0.918 0	0.933 7	**0.936 3**
	31%～40%	0.833 0	0.855 2	0.871 8	0.800 8	0.867 2	0.877 0	0.894 5	**0.899 1**
	41%～50%	0.752 5	0.804 4	0.822 0	0.729 5	0.817 0	0.830 9	0.848 4	**0.855 3**
	51%～60%	0.609 3	0.732 1	0.743 4	0.623 5	0.737 1	0.758 6	0.765 2	**0.774 4**
*L*1 ↓	1%～10%	0.003 7	0.002 7	0.002	0.004 9	0.002 5	0.002 2	**0.001 7**	**0.001 7**
	11%～20%	0.009 0	0.007 3	0.005 7	0.011 4	0.006 6	0.006 1	0.004 9	**0.004 8**
	21%～30%	0.017 3	0.013 2	0.010 8	0.019 9	0.012 2	0.011 2	0.009 6	**0.009 2**
	31%～40%	0.029 4	0.019 9	0.017 0	0.030 6	0.018 7	0.017 4	0.015 5	**0.014 9**
	41%～50%	0.048 4	0.028 0	0.024 5	0.043 8	0.026 8	0.024 6	0.022 9	**0.022 0**
	51%～60%	0.104 1	0.043 2	0.040 5	0.073 3	0.043 4	0.039 6	0.040 4	**0.038 9**
FID ↓	1%～10%	1.911 0	2.237 5	1.495 6	4.294 5	1.617 6	2.054 0	**1.244 5**	1.246 0
	11%～20%	5.049 0	4.831 8	3.907 1	9.138 1	3.660 2	4.501 0	3.193 3	**3.138 4**
	21%～30%	9.764 5	7.472 6	6.764 5	13.461 9	5.925 4	7.170 8	5.653 6	**5.517 7**
	31%～40%	17.871 6	10.610 1	10.381 2	18.911 1	8.693 4	10.042 5	8.459 7	**8.083 2**
	41%～50%	34.176 5	14.109 8	14.547 9	26.000 3	11.638 4	13.543 8	11.665 2	**11.018 6**
	51%～60%	72.086 5	19.306 8	22.598 1	37.212 1	**17.565 8**	19.273 6	20.797 8	17.794 0
LPIPS ↓	1%～10%	0.010 0	0.009 6	0.007 5	0.019 2	0.007 8	0.008 3	0.006 4	**0.006 2**
	11%～20%	0.030 5	0.025 4	0.022 0	0.047 2	0.021 1	0.021 9	0.018 6	**0.018 2**
	21%～30%	0.063 6	0.046 1	0.042 9	0.082 0	0.039 3	0.040 4	0.036 3	**0.035 8**
	31%～40%	0.111 0	0.069 5	0.068 9	0.122 8	0.061 2	0.062 4	0.058 3	**0.057 4**
	41%～50%	0.183 1	0.096 9	0.099 3	0.169 4	0.087 6	0.088 1	0.085 3	**0.083 8**
	51%～60%	0.319 2	0.142 1	0.156 0	0.249 6	0.138 3	**0.136 5**	0.149 5	0.144 4

注：↑表示越高越好，↓表示越小越好。

略低于或高于个别对比算法，其余掩码比例下均为最优值。相较于 DCDPI 算法，尽管本文在 PSNR 和 SSIM 指标上略低于该算法，但本文算法在 PSNR 和 SSIM 上平均高于该算法 0.880 6 dB 和 0.007 9。同理，相较于 MISF 算法，尽管本文算法在 FID 和 LPIPS 指标上略高于该算法，但在所有掩码比例下本文算法在 3 个指标上平均降低了 0.002 2、0.031 7 和 0.002 1。

2.3.2　定性对比实验分析

为了更直观地对比不同方法的修复效果，本文在 3 个数据集上进行定性对比分析。不同算法在 FFHQ、LFW、CelebA-HQ 数据集上的定性对比效果分别如图 5、图 6、图 7 所示。图 5 中第 1 行～第 4 行分别表示掩码比例区间从 21%～30% 到 51%～60% 下不同方法的修复效果。图 6 和图 7 中第 1 行～第 3 行分别表示掩码 31%～40%、41%～50% 和 51%～60% 区间下的对比修复效果。图 5

表2　不同算法在LFW数据集上的对比结果

指标	掩码覆盖面积比例	AOT-GAN	DSNet	CTSDG	PICMM	MISF	DCDPI	SCAT	本文算法
PSNR/dB ↑	1%～10%	38.935 4	38.726 9	41.151 5	35.027 3	39.734 5	40.145 7	42.039 3	**42.402 6**
	11%～20%	32.034 6	32.112 7	33.936 2	28.695 2	33.022 0	33.581 3	34.719 2	**35.047 1**
	21%～30%	27.668 2	28.351 6	29.730 4	24.940 4	29.131 5	29.655 4	30.360 4	**30.708 1**
	31%～40%	24.369 7	25.718 7	26.678 2	22.142 4	26.279 0	26.798 8	27.117 4	**27.555 4**
	41%～50%	21.704 0	23.591 7	24.403 7	19.831 6	23.978 5	24.582 9	24.573 3	**25.048 8**
	51%～60%	16.919 8	20.534 5	20.920 2	16.409 9	20.563 7	**21.217 2**	20.685 8	21.110 0
SSIM ↑	1%～10%	0.983 7	0.983 5	0.989 0	0.971 9	0.986 3	0.987 3	0.990 7	**0.991 1**
	11%～20%	0.955 3	0.954 9	0.967 1	0.928 0	0.961 4	0.964 2	0.971 7	**0.972 9**
	21%～30%	0.912 2	0.915 5	0.933 3	0.869 6	0.925 3	0.930 2	0.941 5	**0.944 2**
	31%～40%	0.860 3	0.872 4	0.892 7	0.803 6	0.883 5	0.890 3	0.902 8	**0.908 1**
	41%～50%	0.798 6	0.822 9	0.846 0	0.725 6	0.833 7	0.844 1	0.855 1	**0.864 0**
	51%～60%	0.690 5	0.747 0	0.765 0	0.607 8	0.747 5	0.763 6	0.761 7	**0.776 5**
*L*1 ↓	1%～10%	0.003 3	0.002 3	0.001 6	0.004 9	0.002 0	0.001 9	0.001 5	**0.001 4**
	11%～20%	0.007 5	0.006 5	0.004 9	0.011 5	0.005 8	0.005 5	0.004 5	**0.004 4**
	21%～30%	0.014 3	0.012 3	0.009 9	0.020 7	0.011 1	0.010 7	0.009 3	**0.008 9**
	31%～40%	0.023 6	0.019 2	0.016 3	0.032 4	0.017 8	0.017 2	0.015 6	**0.014 7**
	41%～50%	0.036 4	0.027 7	0.024 0	0.047 5	0.026 3	0.025 0	0.023 6	**0.022 2**
	51%～60%	0.074 6	0.044 6	0.041 8	0.081 9	0.044 7	0.042 3	0.043 3	**0.040 8**
FID ↓	1%～10%	3.434 4	3.350 5	2.712 2	5.516 9	2.878 3	3.136 8	2.511 2	**2.424 9**
	11%～20%	7.810 9	6.824 2	5.636 3	11.401 4	5.685 6	6.186 9	5.024 3	**4.809 9**
	21%～30%	14.550 7	10.829 1	9.535 8	19.162 6	8.945 4	10.003 7	8.000 6	**7.627 3**
	31%～40%	24.546 8	14.957 1	14.335 8	27.848 5	12.670 6	14.119 6	11.884 0	**11.131 7**
	41%～50%	39.429 5	19.619 5	20.212 0	39.555 1	16.959 2	19.098 3	16.140 7	**15.141 8**
	51%～60%	74.742 7	26.164 0	30.768 6	58.827 9	**24.351 6**	27.528 7	26.831 5	24.965 6
LPIPS ↓	1%～10%	0.009 6	0.008 8	0.006 5	0.019 1	0.007 1	0.008 8	0.005 9	**0.005 6**
	11%～20%	0.030 1	0.024 7	0.020 1	0.050 3	0.020 3	0.023 9	0.017 8	**0.017 0**
	21%～30%	0.064 3	0.046 6	0.041 3	0.092 4	0.039 7	0.044 9	0.036 2	**0.034 8**
	31%～40%	0.109 7	0.072 0	0.068 1	0.144 2	0.064 0	0.069 9	0.060 4	**0.057 8**
	41%～50%	0.168 0	0.102 4	0.100 6	0.207 0	0.094 3	0.099 5	0.091 3	**0.087 6**
	51%～60%	0.289 3	**0.155 5**	0.166 7	0.307 9	0.156 5	0.159 4	0.163 0	0.158 9

注：↑表示越高越好，↓表示越小越好。

～图7中的第3列～第9列表示各种对比算法。

从图5～图7可以看出，相较于其他对比算法，本文算法修复的图像在整体视觉观感上与真实图像的契合度更高。在中度遮挡的场景下（21%～40%掩码比例），本文算法在结构构建、纹理与语义生成层面表现更优。以图5第1行、第2行和图6第1行、图7第1行的图像为例，针对耳钉、下颚线、眼眶、瞳孔等细节的修复，本文算法不仅能合理还原物体形态，还实现了与周边区域的自然融合，更贴近真实图像特征。在面对大面积遮挡图像时（41%～60%掩码比例），本文算法的局部纹理细节修复和结构合成的优势进一步凸显，如图5第3行人脸修复中，可精准还原瞳孔的纹理细节；图6第2行、第3行的眼眶周围和耳朵修复时，相较于对比算法，其产生的伪影现象得到了较为有效的抑制，修复区域与周边内容的衔接更为自然、流畅；图5中第4行人物牙套的金属质感、眉毛的毛发走向细节能够被细

表3　不同算法在CelebA-HQ数据集上的对比结果

指标	掩码覆盖面积比例	AOT-GAN	DSNet	CTSDG	PICMM	MISF	DCDPI	SCAT	本文算法
PSNR/dB ↑	1%～10%	37.349 6	37.424 2	39.630 6	35.266 6	38.309 1	38.754 8	40.229 7	**40.856 7**
	11%～20%	31.086 4	31.636 5	33.293 2	29.816 6	32.420 0	32.963 9	33.826 7	**34.305 6**
	21%～30%	27.309 2	28.454 2	29.743 7	26.782 7	29.107 3	29.600 7	30.137 9	**30.566 2**
	31%～40%	24.476 9	26.188 7	27.185 0	24.477 3	26.680 4	27.176 9	27.412 3	**27.845 6**
	41%～50%	22.113 6	24.357 4	25.166 1	22.557 9	24.756 6	25.243 0	25.225 8	**25.644 1**
	51%～60%	17.958 3	21.824 6	22.157 5	19.505 7	21.793 8	**22.316 6**	21.711 9	22.121 1
SSIM ↑	1%～10%	0.979 4	0.979 3	0.985 5	0.970 9	0.982 4	0.984 2	0.987 9	**0.988 3**
	11%～20%	0.943 8	0.945 1	0.958 3	0.929 3	0.952 3	0.957 0	0.964 7	**0.965 5**
	21%～30%	0.894 7	0.903 7	0.921 4	0.880 1	0.914 1	0.921 9	0.932 0	**0.933 5**
	31%～40%	0.838 0	0.859 7	0.879 1	0.826 9	0.871 5	0.882 5	0.892 3	**0.895 0**
	41%～50%	0.773 1	0.811 4	0.831 6	0.766 8	0.823 8	0.838 8	0.846 1	**0.850 2**
	51%～60%	0.673 6	0.743 8	0.755 8	0.672 2	0.747 4	**0.769 4**	0.762 9	0.768 6
*L*1 ↓	1%～10%	0.003 7	0.002 6	0.001 9	0.004 4	0.002 3	0.002 2	0.001 8	**0.001 6**
	11%～20%	0.008 3	0.007 0	0.005 3	0.009 7	0.006 3	0.005 9	0.005 1	**0.004 7**
	21%～30%	0.015 2	0.012 7	0.010 1	0.016 4	0.011 5	0.010 8	0.009 9	**0.009 2**
	31%～40%	0.024 1	0.019 0	0.015 8	0.024 3	0.017 5	0.016 6	0.015 8	**0.014 6**
	41%～50%	0.036 1	0.026 4	0.022 6	0.033 9	0.024 8	0.023 4	0.023 0	**0.021 3**
	51%～60%	0.069 3	0.039 6	**0.036 9**	0.054 9	0.039 6	0.037 2	0.040 3	0.037 2
FID ↓	1%～10%	1.830 8	1.687 8	1.171 2	2.785 3	1.275 8	1.605 4	1.088 3	**0.983 8**
	11%～20%	5.022 5	3.749 6	3.056 3	5.652 0	2.841 2	3.399 9	2.741 4	**2.408 8**
	21%～30%	10.731 9	6.106 8	5.433 0	9.368 0	4.723 7	5.542 1	4.774 3	**4.257 5**
	31%～40%	19.828 3	8.781 8	8.289 0	13.546 5	6.597 0	7.692 9	6.874 3	**6.212 7**
	41%～50%	33.514 4	12.157 1	11.912 6	18.705 3	9.044 5	10.548 0	9.794 6	**8.793 6**
	51%～60%	66.464 2	16.488 9	17.408 0	29.256 8	**12.884 0**	15.204 2	16.110 4	14.520 0
LPIPS ↓	1%～10%	0.010 8	0.009 3	0.007 1	0.016 3	0.007 5	0.008 5	0.006 7	**0.005 8**
	11%～20%	0.032 9	0.024 7	0.020 8	0.038 6	0.020 4	0.022 1	0.019 5	**0.017 0**
	21%～30%	0.067 0	0.044 1	0.040 7	0.065 0	0.037 8	0.039 9	0.037 6	**0.033 3**
	31%～40%	0.112 3	0.065 8	0.065 2	0.094 8	0.058 2	0.060 4	0.059 6	**0.053 2**
	41%～50%	0.172 4	0.091 0	0.094 5	0.129 5	0.082 6	0.085 1	0.086 7	**0.078 4**
	51%～60%	0.297 0	0.131 7	0.148 7	0.193 1	**0.129 6**	0.130 9	0.148 8	0.136 1

注：↑表示越高越好，↓表示越小越好。

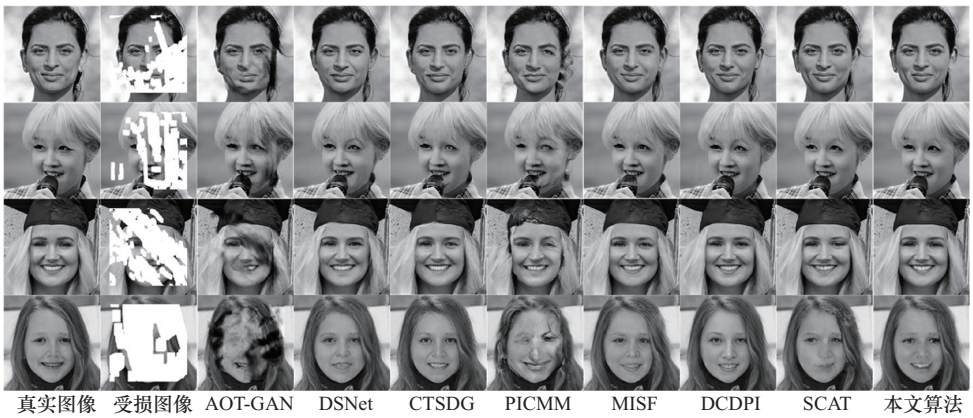


图5　不同算法在FFHQ数据集上的定性对比效果

注：人脸数据来自FFHQ数据集。

图6　不同算法在LFW数据集上的定性对比效果

注：人脸数据来自LFW数据集。

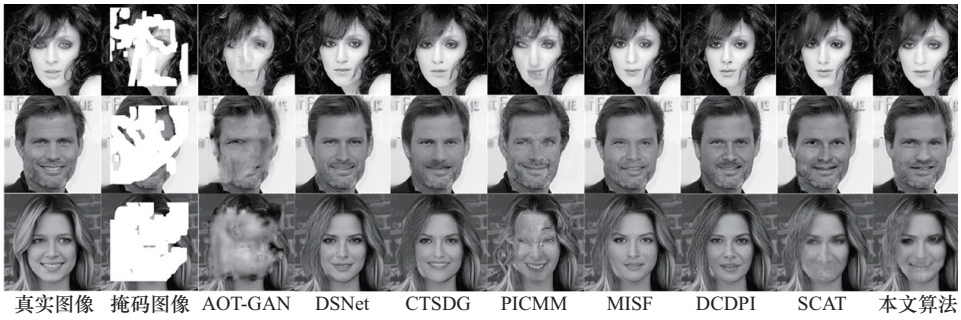


图7　不同算法在CelebA-HQ数据集上的定性对比效果

注：人脸数据来自CelebA-HQ数据集。

腻刻画。然而，从图7第3行中可以明显看出，在51%～60%掩码下本文算法在视觉效果与结构完整性方面较MISF算法存在明显劣势，但在21%～50%的掩码下，反观其余对比算法，在同类场景中普遍存在修复瑕疵，或因局部细节模糊而特征失真，或出现纹理错误（如瞳孔纹理偏移、牙套形态扭曲、眉毛纹理紊乱等），严重影响了图像真实性。综上，本文算法能够实现视觉语义观感与细粒度纹理还原的双重高质量修复效果。

2.4　消融实验

为了验证模块的有效性，本文在FFHQ数据集上开展消融实验分析。基准模型为未添加任何模块的原始网络；Net1模型是在基准模型基础上引入自注意力机制的网络；Net2模型是在基准模型上增添深度动态门控机制的网络；Net3模型是同时添加自注意力机制与深度动态门控机制后的网络。模块消融实验客观指标对比见表4。

表4　模块消融实验客观指标对比

模型	自注意力机制	深度动态门控机制
基准	×	×
Net1	√	×
Net2	×	√
Net3	√	√

注：×表示未添加该模块，√表示添加该模块。

2.4.1　定量消融实验

定量消融实验结果见表5。由表5可知，Net3网络在多项指标上展现出显著优势。具体而言，相较于基准模型，Net3在感知质量指标方面取得显著提升：PSNR值平均提高了1.098 dB，SSIM平均提升了0.012，同时$L1$损失降低了0.002。这些数据表明，引入自注意力机制和动态门控机制能够有效改善图像生成质量，促使模型在纹理生成层面更注重感知一致性，而非简单的像素级匹配，从而产生更细腻的修复效果。

在生成图像真实性评估方面，FID指标呈现出掩码比例相关的特性变化。在小掩码1%～20%下，

表5　定量消融实验结果

指标	模型	掩码1%～10%	掩码11%～20%	掩码21%～30%	掩码31%～40%	掩码41%～50%	掩码51%～60%
PSNR/dB ↑	基准	39.217 7	32.995 0	29.339 8	26.568 0	24.475 7	20.992 5
	Net1	39.453 2	33.193 7	29.537 0	26.726 5	24.569 4	21.106 5
	Net2	40.401 0	33.936 5	30.144 3	27.265 9	25.076 9	21.521 5
	Net3	**40.594 9**	**34.235 8**	**30.489 5**	**27.608 3**	**25.404 9**	**21.841 3**
SSIM ↑	基准	0.986 3	0.961 3	0.926 3	0.885 1	0.837 7	0.751 5
	Net1	0.986 7	0.962 1	0.927 7	0.886 6	0.838 9	0.752 4
	Net2	0.988 2	0.965 8	0.933 6	0.894 8	0.849 3	0.765 5
	Net3	**0.988 6**	**0.967 1**	**0.936 3**	**0.899 1**	**0.855 3**	**0.774 4**
*L*1 ↓	基准	0.002 0	0.005 6	0.010 7	0.017 1	0.024 9	0.043 7
	Net1	0.001 9	0.005 4	0.010 4	0.016 8	0.024 6	0.043 2
	Net2	**0.001 7**	0.004 9	0.009 6	0.015 4	0.022 7	0.040 3
	Net3	**0.001 7**	**0.004 8**	**0.009 2**	**0.014 9**	**0.022 0**	**0.038 9**
FID ↓	基准	1.523 1	3.748 9	6.552 7	9.745 1	13.526 2	23.195 9
	Net1	1.468 2	3.676 6	6.429 0	9.650 9	13.441 5	23.105 5
	Net2	**1.197 3**	**3.110 6**	5.523 4	8.301 5	11.424 9	20.663 2
	Net3	1.246 0	3.138 4	**5.517 7**	**8.083 2**	**11.018 6**	**17.794 0**
LPIPS ↓	基准	0.007 5	0.021 6	0.041 3	0.065 5	0.094 7	0.159 7
	Net1	0.007 5	0.021 5	0.041 1	0.065 1	0.094 5	0.159 8
	Net2	**0.006 1**	**0.018 0**	**0.035 5**	**0.057 4**	0.084 3	0.147 1
	Net3	0.006 2	0.018 2	0.035 8	**0.057 4**	**0.083 8**	**0.144 4**

注：↑表示越高越好，↓表示越小越好。

Net3略低于Net2，但随着掩码比例超过21%，FID明显低于其他网络，这印证了自注意力机制的长程依赖建模能力随遮挡范围扩大而优势凸显。进一步分析LPIPS指标，在Net2架构中引入自注意力机制模块后，对31%以上受损图像的修复效果提升较为显著，证明该模块通过全局上下文感知有效增强了复杂背景下的语义连贯性。值得注意的是，动态门控机制与自注意力机制的协同作用在31%以上遮挡情况下表现出更强的鲁棒性，纹理细节修复更加自然。

2.4.2　定性消融实验

为了更加直观地对比不同网络修复性能的差异，本文在21%～60%掩码比例下进行定性对比分析。不同模块消融对比效果如图8所示，图8第1行到第4行表示从掩码区间21%～30%到51%～60%的修复效果。通过4组渐进式遮挡实验可以看出，Net3在大面积缺损修复中展现出显著的视觉优势。

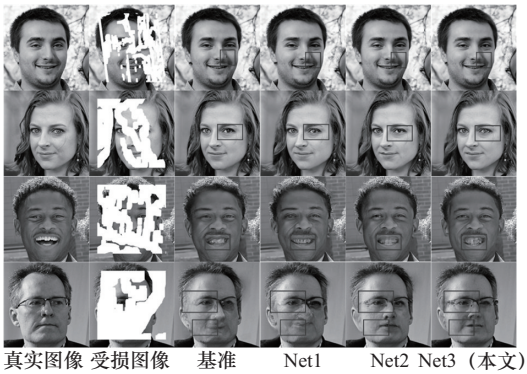


图8　不同模块消融对比效果

注：人脸数据来自FFHQ数据集。

在21%～30%中等遮挡情况下（图8第1行），Net3成功重建了鼻梁背光面的平滑渐变光影，准确捕捉到环境光与面部曲面法线的关系。在处理31%～40%遮挡时（图8第2行），Net3在眼部瞳孔修复中表现出色，精确恢复了虹膜的放射状纹理。相比之下，Net2因缺乏自注意力机制，瞳孔伪影较重。在大掩码51%～60%下，

Net3修复更加细腻，如图8第4行的眼镜框、皱纹线纹理等。

2.4.3 DGSF-AOT层数消融实验

本文通过消融实验探索DGSF-AOT模块堆叠层数对模型性能的影响机制。为确定最优堆叠深度，本文在21%～60%掩码范围内对模块数量进行1～10层的控制变量实验，重点关注参数量与性能指标的平衡关系。DGSF-AOT层数消融实验结果见表6。

由表6可知，当DGSF-AOT堆叠至8层时，对指标PSNR和SSIM的提升较为有利；当堆叠层数达到9层及以上时，尽管部分性能指标仍存在边际增益（如$L1$），但$L1$和LPIPS指标在6层结构处时仅有略微的下降，且模型参数数量也会增加，不同层数参数量对比见表7。每增加一层DGSF-AOT，会增加2.51×10^6的参数量。在保证参数量最小化的前提下，通过综合评估各项性能指标，最终确定层数6为最优方案。相较于层数8，该模型在参数量减少5.02×10^6的情况下，仍保持优异的性能表现：在21%～50%掩码范围内，PSNR指标仅存在0.04的微弱差异，其余指标的下滑幅度均控制在合理阈值内。

表7 不同层数参数量对比

层数	参数量/M×(10^6)
1	3.55
2	6.06
3	8.57
4	11.08
5	13.59
6	16.10
7	18.60
8	21.11
9	23.62
10	26.13

表6 DGSF-AOT层数消融实验结果

指标	掩码覆盖面积比例	1层	2层	3层	4层	5层	6层	7层	8层	9层	10层
PSNR↑	21%～30%	28.865 2	29.267 0	29.914 6	30.183 4	30.170 1	30.489 5	30.334 7	**30.492 0**	30.451 2	30.346 2
	31%～40%	25.975 2	26.452 3	27.023 8	27.272 0	27.242 0	27.608 3	27.489 9	**27.613 9**	27.601 2	27.512 8
	41%～50%	23.796 6	24.293 1	24.812 4	25.080 9	25.048 2	25.404 9	25.290 1	**25.439 5**	25.437 1	25.350 7
	51%～60%	20.313 0	20.825 2	21.227 0	21.530 7	21.445 5	21.841 3	21.800 3	21.941 5	**21.948 3**	21.902 4
SSIM↓	21%～30%	0.921 9	0.926 5	0.932 1	0.934 4	0.935 7	0.936 3	0.935 9	**0.937 2**	0.936 4	0.934 5
	31%～40%	0.877 4	0.884 0	0.892 2	0.896 0	0.898 2	0.899 1	0.898 5	**0.900 3**	0.899 7	0.896 6
	41%～50%	0.826 7	0.834 5	0.845 4	0.851 2	0.854 6	0.855 3	0.854 6	**0.857 1**	0.856 8	0.852 3
	51%～60%	0.739 6	0.746 1	0.759 5	0.768 4	0.774 9	0.774 4	0.774 1	0.777 3	**0.778 4**	0.771 4
$L1$↓	21%～30%	0.011 6	0.010 9	0.009 8	0.009 6	0.009 6	**0.009 2**	0.009 3	**0.009 2**	**0.009 2**	0.009 4
	31%～40%	0.018 7	0.017 5	0.015 8	0.015 5	0.015 5	0.014 9	0.014 9	0.014 8	**0.014 7**	0.015 0
	41%～50%	0.027 5	0.025 7	0.023 4	0.022 8	0.022 9	0.022 0	0.022 0	0.021 8	**0.021 7**	0.022 0
	51%～60%	0.048 3	0.044 9	0.041 7	0.040 1	0.040 7	0.038 9	0.038 6	0.038 3	**0.038 1**	0.038 4
FID↓	21%～30%	6.896 6	5.905 8	5.602 5	5.502 8	5.518 0	5.517 7	5.358 1	5.385 1	5.346 8	**5.306 8**
	31%～40%	10.656 9	8.960 2	8.505 3	8.190 7	8.241 8	8.083 2	**7.920 5**	8.018 2	8.089 2	7.992 3
	41%～50%	15.240 2	12.369 1	11.456 3	10.996 5	11.157 1	11.018 6	**10.636 7**	10.931 1	10.864 3	10.916 7
	51%～60%	28.968 4	23.071 5	19.038 8	18.493 8	17.647 9	17.794 0	17.108 9	**16.950 4**	17.043 6	18.281 2
LPIPS↓	21%～30%	0.043 5	0.037 6	0.036 3	0.035 5	0.035 9	0.035 8	0.034 8	0.034 7	0.035 1	**0.033 1**
	31%～40%	0.071 4	0.061 3	0.058 8	0.057 7	0.058 4	0.057 4	0.055 7	0.055 9	0.056 5	**0.053 4**
	41%～50%	0.106 9	0.090 7	0.086 6	0.084 5	0.085 7	0.083 8	0.081 7	0.081 6	0.082 6	**0.078 5**
	51%～60%	0.193 7	0.160 4	0.151 2	0.146 2	0.148 9	0.144 4	0.140 4	0.139 0	0.142 1	**0.136 7**

注：↑表示越高越好，↓表示越小越好。

3　结束语

本文针对复杂背景下人脸图像修复任务中出现细粒度纹理修复不足、结构修复断层和语义失谐的问题，提出一种基于动态门控与自注意力融合增强的图像修复网络。该网络通过引入动态门控网络实现多尺度特征的自适应融合权重生成，并结合自注意力机制显式建模长程像素依赖关系。实验表明，该方法在 FFHQ、LFW 和 CelebA-HQ 人脸数据集上显著提升了细粒度纹理细腻修复、视觉语义合理性和结构连贯性合成的效果。

参考文献：

[1] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques-SIGGRAPH '00. New York: ACM Press, 2000: 417-424.

[2] Criminisi A, PéRez P, Toyama K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[3] Brooks T, Holynski A, Efros A A. InstructPix2Pix: learning to follow image editing instructions[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 18392-18402.

[4] Cao A Q, Dai A, De Charette R. Pasco: urban 3D panoptic scene completion with uncertainty awareness[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 14554-14564.

[5] Criminisi A, Perez P, Toyama K. Object removal by exemplar-based inpainting[C]//Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2003: II.

[6] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7794-7803.

[7] Yu J H, Lin Z, Yang J M, et al. Generative image inpainting with contextual attention[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5505-5514.

[8] Yu J H, Lin Z, Yang J M, et al. Free-form image inpainting with gated convolution[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 4470-4479.

[9] Cai W W, Wei Z G. PiiGAN: generative adversarial networks for pluralistic image inpainting[J]. IEEE Access, 2020, 8: 48451-48463.

[10] Wang N, Zhang Y P, Zhang L F. Dynamic selection network for image inpainting[J]. IEEE Transactions on Image Processing, 2021, 30: 1784-1798.

[11] Yu Y C, Zhan F N, Lu S J, et al. WaveFill: a wavelet-based generation network for image inpainting[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 14094-14103.

[12] Zeng Y H, Fu J L, Chao H Y, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7): 3266-3280.

[13] Feng X, Pei W J, Li F J, et al. Generative memory-guided semantic reasoning model for image inpainting[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(11): 7432-7447.

[14] Zuo Z W, Zhao L, Li A L, et al. Generative image inpainting with segmentation confusion adversarial training and contrastive learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(3): 3888-3896.

[15] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[C]//Proceedings of International Conference on Machine Learning. Maastricht: PMLR, 2019: 7354-7363.

[16] Karras T, Laine S, Aila T M. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 4396-4405.

[17] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: a database forstudying face recognition in unconstrained environments[J]. Computer Science, 2008: 1-11.

[18] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation[PP]. arXiv (2018-

02-26)[2025-03-11] arXiv: arXiv. 1710.10196.

[19] Liu G L, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Computer Vision-ECCV 2018. Cham: Springer, 2018: 89-105.

[20] Guo X F, Yang H Y, Huang D. Image inpainting via conditional texture and structure dual generation[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 14114-14123.

[21] Xia X B, Yang W H, Ren J, et al. Pluralistic image completion with Gaussian mixture models[C]//Proceedings of the Neural Information Processing Systems (NeurlPS 2022). Piscataway: IEEE Press, 2015: 1-14.

[22] Li X G, Guo Q, Lin D, et al. MISF: multi-level interactive Siamese filtering for high-fidelity image inpainting[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 1859-1868.

[23] Wang Z, Li K, Peng J. Dynamic context-driven progressive image inpainting with auxiliary generative units[J]. The Visual Computer, 2024, 40(5): 3457-3472.

[作者简介]

柏武贰（1995– ），男，贵州民族大学数据科学与信息工程学院硕士生，主要研究方向为图像修复。

张乾（1984– ），男，博士，贵州民族大学教务处教授，主要研究方向为机器学习、模式识别、计算机视觉。

刘霜（2001– ），女，贵州民族大学数据科学与信息工程学院硕士生，主要研究方向为图像修复。

滕林（1996– ），男，贵州民族大学数据科学与信息工程学院硕士生，主要研究方向为图像修复。

杨思红（1996– ），女，贵州民族大学数据科学与信息工程学院硕士生，主要研究方向为图像修复。