DOI：10.20189/j.cnki.CN/61-1527/E.202504012

一种语义引导的高效图像去摩尔纹框架研究

燕　松，李　敏，张雨森，施　昊
（火箭军工程大学，西安 710025）

摘　要：为提高模型对图像内容的理解能力、抑制模型在生成过程中的随机性，提出一种基于语义引导的图像去摩尔纹框架StableMoiré。首先，针对模型在生成过程中的语义漂移问题，将图像从像素空间经由变分自编码器和CLIP（contrastive language-image pre-training）图像编码器分别编码至隐空间和CLIP空间，并利用交叉注意力机制提取隐空间编码中与图像去摩尔纹相关的信息，使模型专注图像中与摩尔纹相关的部分，从而避免冗余信息的干扰。其次，针对使用语义信息对预训练文本生成图像模型进行引导时存在较大随机性的问题，提出一种基于残差注入的图像细节保留模块，避免模型在生成过程中偏离原图像的像素信息。结果表明：所提StableMoiré框架在摩尔纹消除方面有着更好的性能；与同类基线算法PromptIR相比较，其PSNR与SSIM分别提升了0.509 2和0.011 3，而FID与KID则分别降低了0.03和0.66。

关键词：深度学习；图像去摩尔纹；预训练文生图模型；扩散模型
中图分类号：TP391.4　　**文献标志码**：A　　**开放科学标识码**（OSID）：
文章编号：2097-2741（2025）04-114-10

听语音　聊科研　与作者互动

Study on an Efficient Semantics-Guided Image Demoiré Framework

YAN Song, LI Min, ZHANG Yusen, SHI Hao
（Rocket Force University of Engineering, Xi'an 710025, China）

ABSTRACT: StableMoiré, a semantics-guided framework for moiré removal, was proposed to enhance the model's understanding of the image content and reduce randomness in the generation process. First, to address semantic drifts during the generation process, images were encoded into latent and contrastive language-image pretraining (CLIP) spaces using a VAE and CLIP image encoder, respectively. Moiré-related information in the latent space was extracted using cross-attention mechanisms, allowing the model to focus on the parts of the images that were related to the moiré pattern, thereby avoiding the interference of redundant information. Subsequently, to address the significant randomness in semantic guidance for pretrained text-to-image models, a residual injection-based image detail preservation module was proposed, preventing it from deviating from the pixel information of the original image during the generation process. The results

收稿日期：2025-02-20　　**修回日期**：2025-03-17　　**录用日期**：2025-04-15
第一作者：燕松（2001—），男，博士研究生，主要从事生成式人工智能相关领域研究。E-mail：gary_144@outlook.com

引用格式：燕松，李敏，张雨森，等.一种语义引导的高效图像去摩尔纹框架研究[J].火箭军工程大学学报，2025，39（4）：114-123. *YAN Song, LI Min, ZHANG Yusen, et al. Stady on an efficient semantics-guided image demoiré framework[J]. Journal of Rocket Force University of Engineering, 2025, 39 (4): 114-123.*

show that StableMoiré performs better in terms of moiré removal. Compared with PromptIR, a similar baseline algorithm, StableMoiré improves its peak signal-to-noise ratio and structural similarity index measure by 0.509 2 and 0.011 3, respectively, while reducing the Fréchet inception distance and kernel inception distance by 0.03 and 0.66, respectively.

KEYWORDS: deep learning; image demoir; pretrained text-to-image model; diffusion model

随着信息时代的到来，人们越来越倾向于使用智能手机记录屏幕上的内容，但拍摄的图像中常常出现摩尔纹。从成像原理来看，摩尔纹是相机的色彩滤光阵列（color filter array，CFA）与屏幕的液晶显示（liquid crystal display，LCD）子像素之间的空间频率混叠导致的。摩尔纹会严重影响拍摄图像的质量，且难以与普通的图像噪声区分开来。

摩尔纹的尺度不尽相同、边界与形状模糊、颜色多样且频率变化无常，这些特征都会增加去摩尔纹算法的难度。随着深度学习的风靡，研究者提出了许多基于深度神经网络的去摩尔纹方法[1-3]。其中，最常见的手段是利用多尺度架构来循环消除各种尺寸的摩尔纹[1-4]。此外，一些研究者还主张利用摩尔纹中的频率成分来设计基于频率的去摩尔纹网络，从而实现更好的恢复结果[5]。这些方法直接使用成对的去摩尔纹数据集对模型进行训练，并取得了一系列成果，但面对复杂场景去摩尔纹任务时仍具有较大的局限性。主要是因为训练数据集覆盖的场景不全，导致模型在面对某些比较罕见的场景时表现不佳。

随着扩散模型的出现，许多方案将其引入图像去摩尔纹领域，希望借助其良好的数学定义以及强大的学习能力学习到摩尔纹的本质属性，从而更好地在复杂场景中提取与摩尔纹相关的特征[6]。此类方法缓解了数据集缺陷造成的性能下降，但数据本身引发的问题仍未得到有效解决。

近年来，随着大规模预训练模型的普及，许多基于预训练文本生成图像模型（文生图模型）的图像恢复方法被研究者提出，包括但不限于图像超分辨率重建[7-9]、图像去雾[10-11]、图像去雨[12]等。如图1所示，由于预训练文生图模型在超大规模数据集上进行了预训练，获得了来自大规模数据集的海量先验知识[13]，使其能够经过微调将预训练数据集上的知识迁移到不同任务上，实现高质量的生成。在图像恢复领域的诸多任务中，现有的方法主要使用ControlNet[14]将条件注入预训练模型之中，使模型实现可控的生成。但由于预训练模型的预训练与微调方式并不一致，传统方案直接使用预训练模型网络权重进行生成，存在着严重的语义漂移问题，即模型无法区分图像与文本指令之间的关系，导致模型的生成结果中出现人类不需要的内容，从而导致模型性能下降。而扩散模型内在的随机性与像素级引导的缺乏又将导致模型生成的结果与原图像在结构和内容上大相径庭。因此，如何将预训练文生图模型运用到图像去摩尔纹领域，利用大规模预训练数据集弥补去摩尔纹数据集的潜在缺陷，并有效抑制模型生成过程中的语义漂移与随机性问题，是一个亟待解决的难题。

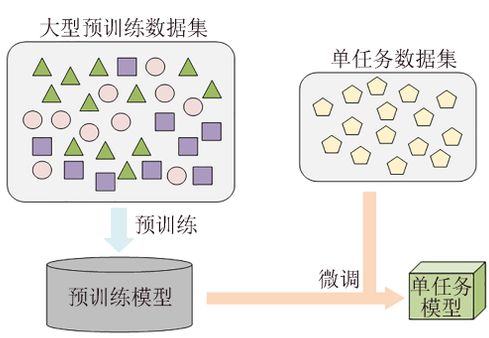


图1 使用大型预训练模型微调得到单任务模型的流程

Fig. 1 Process of obtaining a single-task model through fine-tuning of a large pretrained model

针对图像去摩尔纹任务中存在的上述问题，提出了全新的框架StableMoiré。StableMoiré对基于预训练的Stable Diffusion[15]进行了微调，一方面，将contrastive language-image pretraining（CLIP）空间的图像语义进行针对去摩尔纹任务的优化并注入模型，使模型通过该语义信息理解图像中摩尔纹的本质特征，从而抑制模型由于语义空间的不匹配而产生的语义漂移现象，并实现了基于语义引导的多场景泛化；另一方面，通过将语义信息与图像细节信息融合，并以残差的形式注入模型，显著抑制了生成过程的随机性，使生成图像的结构与原图像保持一致。最终，通过框架中的各部分协同工作，既解决了模型生成过程中的语义漂移与随机性问题，又有效利用了预训练模

型中的先验知识，提升模型对复杂场景的适应能力，从而实现了高质量的摩尔纹去除。最后，将本文所提算法在LCDMoiré数据集上与4个基线算法进行比较，在4个指标上均实现了相对最优。

1 StableMoiré整体架构及相关工作

1.1 预训练文生图模型

文生图模型通过理解人类指令而生成与给定文本描述相匹配的图像，对于图像编辑[16]、虚拟现实[17]和艺术创作[18-19]等多种下游任务具有重要意义。近年来，随着扩散模型在图像生成领域崭露头角，预训练文生图模型，如DALL-E[20]和Stable Diffusion[15]等开始涌现出来。DALL-E利用了Transformer[21]架构，能够生成与文本描述高度一致的图像。Stable Diffusion[15]则通过隐空间扩散模型（latent diffusion model，LDM）来生成图像，在生成质量、效率和多样性上都有所提升。这些预训练模型不仅推动了文生图技术的进步，也为未来的研究提供了强大的基座模型。

1.2 预训练文生图模型微调策略

预训练文生图模型的生成过程具有较大的随机性[15]。为了使模型的生成结果更好地满足人们的实际需求，研究者们提出了许多微调这些模型的方法来进行自定义生成。具体来说，为了使文生图模型按照自定义的条件进行输出，ControlNet[14]于2023年被提出，将超网络特定层的特征注入文生图模型来达到条件控制的目的。为了向预训练模型中注入新的知识以实现基于主体的自定义输出，DreamBooth[22]于2022年被提出，该方法通过将主体与占位符绑定的微调策略，让模型获得区分特定主体的能力。而在优化微调算力消耗方面，Hu等[23]提出的低秩近似策略大大缩减了微调过程中需要训练的模型参数规模。这些微调策略为预训练文生图模型适应更多下游任务做出了巨大贡献，也进一步推动了这类模型在商用领域的落地。

1.3 基于扩散模型的图像恢复方法

基于扩散模型的图像恢复方法分为零样本（zero-shot）图像恢复与任务导向型图像恢复2类。零样本图像恢复方法采用预先训练好的扩散模型[24]来进行线性[25-26]或非线性[27-28]图像修复，但在面对真实世界的数据时，这些模型的表现较差。相较之下，任务导向型图像恢复方法通过微调模型以适应特定任务，如超分辨率[29-30]、JPEG压缩恢复[31]、人脸修复[32-33]以及低光照增强[34]。StableSR[35]和DiffBIR[36]等方法采用条件扩散模型作为核心技术，针对不同的图像恢复任务的特点设计对应的模型架构，实现了高质量的生成。然而，在图像去摩尔纹领域，由于摩尔纹的噪声复杂多样，在特征上与传统噪声有着较大区别，导致此类方法的研究仍处于起步阶段。

1.4 StableMoiré模型整体架构

受文生图模型使用文本命令作为引导条件生成图像的启发，本文提出了StableMoiré。一方面，为了更好地引导模型进行生成，StableMoiré将图像中与摩尔纹相关的语义信息传递到预训练文生图模型中，起到与文本命令类似的作用，从而引导模型理解待处理图像中与摩尔纹相关的特征，并排除无关信息的干扰。另一方面，为了抑制模型在生成过程中的随机性，StableMoiré在去噪过程中不断向模型注入基于语义引导的图像细节信息，从而驱动模型保留除了摩尔纹以外的像素特征，抑制模型在生成过程中的随机性。通过这种结构设计，StableMoiré有效解决了模型在生成过程中的语义漂移与强随机性问题，从而实现了高质量的图像去摩尔纹效果。

StableMoiré的整体架构如图2所示。首先，给

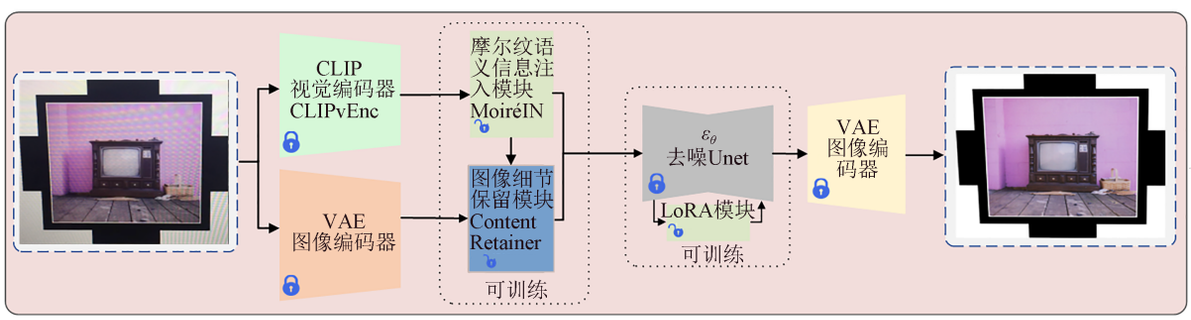


图2 StableMoiré模型架构

Fig. 2 Schematic of the StableMoiré model

定一张有摩尔纹的待处理图像，由CLIP视觉编码器与VAE图像编码器分别将图像由像素空间编码到CLIP语义空间与VAE隐空间中。接下来，摩尔纹语义信息注入模块MoiréIN，将图像中与摩尔纹相关的语义信息注入主干预训练扩散模型；而基于残差注入的图像细节保留模块Content Re-tainer则在像素层面引导模型保留原图像空间信息，抑制模型生成过程中的随机性，最终精准地去除摩尔纹，从而获得干净的图像。在模型训练的过程中，也一并使用LoRA对去噪网络进行微调，从而使主干去噪网络更好地适应微调后的语义空间。

2 StableMoiré各模块设计

2.1 摩尔纹语义信息注入模块

传统的文本生成图像模型大多经由CLIP空间上的大规模预训练得到，其对图像与文本的理解能力主要来源于预训练CLIP的感知能力。然而，CLIP在预训练过程中并未单独对摩尔纹这一特征进行优化，导致模型在理解图像时主要关注图像主体的内容信息，而非摩尔纹相关的信息。直接使用这一信息对模型的生成过程进行引导容易导致模型出现语义漂移的现象。因此，为了更好地帮助模型在去摩尔纹任务中关注图像中与摩尔纹相关的信息，就需要对原CLIP图像语义进行调整，使其微调后能够更专注于图像中的摩尔纹噪声，从而精准地去除摩尔纹。

基于以上思考，提出全新的摩尔纹语义信息注入模块MoiréIN，以提取图像中与摩尔纹相关的语义信息，并将其编码为类似文本命令的形式输入模型，指导模型关注图像中的摩尔纹噪声部分，并排除无关信息的干扰。MoiréIN是一个单分支结构，由3层带有层归一化和LeakyReLU激活函数的MLP层组成。首先，预训练好的CLIP视觉编码器将等待处理的图像编码成语义向量，用于提取图像整体上的语义信息；随后，编码得到的向量被输入MoiréIN模块中进行进一步的特征提取，得到无干扰信息的摩尔纹语义。具体来说，对于1张输入图像，CLIP图像编码器CLIPvEnc先将其编码到CLIP空间中

$$\boldsymbol{I}_{\mathrm{CLIP}} = \mathrm{CLIPvEnc}(\boldsymbol{I}) \tag{1}$$

式中：$\boldsymbol{I}_{\mathrm{CLIP}}$为经过CLIP图像编码器编码后得到的语义向量。紧接着，$\boldsymbol{I}_{\mathrm{CLIP}}$经由MoiréIN模块处理，得到更加纯粹的摩尔纹语义信息$\boldsymbol{I}_{\mathrm{M}}$，再传入主干扩散模型，引导模型进行摩尔纹去除

$$\boldsymbol{I}_{\mathrm{M}} = \mathrm{MoiréIN}(\boldsymbol{I}_{\mathrm{CLIP}}) \tag{2}$$

经过损失函数的引导，MoiréIN模块所提取到的$\boldsymbol{I}_{\mathrm{M}}$将与主干扩散模型同步进行优化，直到$\boldsymbol{I}_{\mathrm{M}}$精准引导模型关注摩尔纹信息。因此，图像在CLIP空间中复杂的语义信息经过MoiréIN模块的处理，得到更加适合图像去摩尔纹任务的语义信息，从而辅助模型实现更高质量的去摩尔纹结果。其次，模块的信息提取过程是在训练中不断优化的，从而避免了任何额外的人工干预，极大降低了提示工程的难度，使模型的工作流程更加简洁高效、用户界面更加友好。

2.2 基于残差注入的图像细节保留模块

虽然语义信息对于模型理解摩尔纹相关特征至关重要，但难以完整保留图像在像素空间中的诸多细节。如果缺乏图像细节信息的注入，模型推理时将依赖于没有图像内容信息的随机噪声以及单纯的语义引导信息，从而增加了模型生成结果的随机性。尽管这种随机性在文生图任务中是被提倡的，但对于图像恢复任务而言，更期望恢复结果能保留原图像的像素信息。因此，不仅需要对模型进行语义上的引导，还需要在像素空间进行有效的引导，才能达到理想的图像恢复效果。在这样的目标下，通过提供包含图像内容信息的初始噪声并进行DDIM反演是至关重要的步骤。

然而，反演过程不稳定且耗时。为了缓解这种情况，以ControlNet为灵感来源，提出了基于残差注入的图像细节保留模块Content Retainer，以有效地向模型中注入图像细节信息。Content Retainer与ControlNet的主要区别在于模型微调过程中，Content Retainer融合了提取到的摩尔纹语义信息$\boldsymbol{I}_{\mathrm{M}}$，并进行训练，从而帮助模块将更好的降噪信息注入主干网络。如图3所示，Content Retainer包括2个部分。首先，时间步信息t经由3层MLP编码后形成时间编码$\boldsymbol{I}_t$，而输入图像经由VAE图像编码器被编码到隐空间之中，得到输入图像在隐空间中的编码$\boldsymbol{I}_{\mathrm{VAE}}$。紧接着，这2部分信息与$\boldsymbol{I}_{\mathrm{M}}$一同被输入Content Retainer模块进行信息提取，得到残差注入信息$\boldsymbol{I}_{\mathrm{res}}$为

$$\boldsymbol{I}_{\mathrm{res}} = \mathrm{Att}(\mathrm{Res}(\boldsymbol{I}_t, \boldsymbol{I}_{\mathrm{VAE}}), \boldsymbol{I}_t, \boldsymbol{I}_{\mathrm{M}}) \tag{3}$$

式中：Res和Att模块分别为标准残差网络和交叉注意力Transformer模块。Res的输出被用作查询特征，而$\boldsymbol{I}_{\mathrm{M}}$在交叉注意力层中既作为键特征又作为值特征。经过Content Retainer的处理，$\boldsymbol{I}_{\mathrm{res}}$以残

差的形式被输入到预训练文生图模型，从而实现图像细节信息的注入，以更好地指导模型的反演过程，有效地抑制模型在生成过程中的随机性。

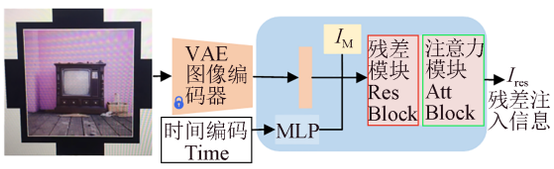


图 3 基于残差注入的图像细节保留模块结构

Fig. 3 Structure of image detail retention module based on residual injection

为了训练 StableMoiré，采用与 Stable Diffusion 类似的去噪损失函数，将 $\boldsymbol{I}_{\text{res}}$ 和 $\boldsymbol{I}_{\text{M}}$ 引入扩散去噪训练过程

$$\mathcal{L}=E_{z_0,t,I_{\text{M}},I_{\text{res}},\epsilon\sim N(0,1)}\left[\left\|\epsilon-\epsilon_\theta(z_t,t,\boldsymbol{I}_{\text{M}},\boldsymbol{I}_{\text{res}})\right\|_2^2\right] \quad (4)$$

式中：$\mathcal{L}$ 为损失函数；E 为数学期望；z_0 为参考图像 $\hat{I}$ 的隐空间表示，经由 VAE 图像编码器编码得到；$z_t=\sqrt{\bar{\alpha}}\,z_0+\sqrt{1-\alpha_t}\,\epsilon_\theta$，为时间步为 t 时图像噪声的隐空间表示；α_t 为时间步为 t 时的衰减系数；$\bar{\alpha}$ 为时间步 α_t 累积乘积；ϵ 为服从正态分布的噪声；ϵ_θ 为去噪神经网络。该损失函数能确保 StableMoiré 得到稳定且高效的训练，以结合特定于图像的摩尔纹语义信息与像素信息来指导模型的扩散过程。

3 实验结果与分析

3.1 数据集介绍

使用 LCDMoiré 数据集对 StableMoiré 模型进行训练和评估，该数据集包含 10 200 对合成生成的图像，每对图像包括带有摩尔纹的图像和 1 张干净的参考图像。数据集分为 10 000 对图像的训练集、100 对图像的验证集和 100 对图像的测试集，每张图像的分辨率均为 1 024×1 024 像素。数据生成过程模拟用户使用智能手机拍摄 LCD 屏幕时的场景。具体步骤包括：将输入的 RGB 图像重新采样为 RGB 子像素马赛克；应用随机投影变换，以模拟显示设备和相机的不同相对位置和角度；通过拜耳色彩滤波阵列重新采样，以模拟 RAW 数据；添加高斯噪声以模拟传感器噪声，以及应用简单的图像信号处理流程，包括去马赛克和去噪。

3.2 实验环境

本次实验使用的硬件平台配置为 8 张 NVIDIA A100 显卡服务器平台，单张显卡显存为 80 G。服务器软件配置为 Ubuntu20.04 操作系统，Python3.8 语言，Cuda11.8 显卡驱动和 Pycharm 集成开发环境。训练过程采用随机梯度下降法（stochastic gradient descent，SGD）进行优化，初始学习率为 0.000 1，Batchsize 设置为 64。

3.3 评估指标

为了客观评估模型的生成效果，采用文献[37]中使用的评估方法。具体来说，所使用的 4 个指标分别为 FID（Fréchet inception distance）、KID（kernel inception distance）、PSNR（peak signal-to-noise ratio）和 SSIM（structural similarity index measure）。

FID 是一种用于评估图像生成质量的指标，通过计算 2 个图像数据集的特征分布之间的距离来衡量他们的相似性。使用预训练的 Inception 模型提取图像的特征向量，然后计算生成图像和真实图像特征向量的均值和协方差矩阵之间的差异。FID 值越小，表示生成的图像与真实图像越相似。

KID 是一种与 FID 类似的生成质量度量指标，用于评估生成图像与真实图像的分布差异。通过计算 2 个数据集的特征向量在再生核希尔伯特空间（reproducing kernel Hilbert space，RKHS）中的最大均值差异来衡量他们的相似性。KID 值越小，表示生成的图像与真实图像越相似。

PSNR 是一种传统的图像质量评估指标，基于均方误差（mean squared error，MSE）衡量 2 幅图像之间的差异。通过比较原始图像和处理后图像的像素值差异来计算，值越高表示图像质量越好。PSNR 计算简单，物理意义清晰，被广泛应用于图像压缩和重建等领域，但其主要基于像素差异，可能无法完全反映人眼的视觉感受。

SSIM 是一种更符合人眼视觉感知的图像质量评估指标，通过比较图像的亮度、对比度和结构信息来衡量图像的相似性。SSIM 不仅考虑了像素值的差异，还考虑了图像的结构信息，能够更好地反映人眼对图像质量的主观感受，被广泛应用于图像处理和计算机视觉领域。SSIM 值越接近 1，表示 2 幅图像越相似。

3.4 对比实验

本文将 StableMoiré 与图像去摩尔纹领域最先进的模型进行比较，包括 DMCNN[38]、Air-Net[39]、PromptIR[40]、Diff-Plugin[6]。这些方法的设计分别基于不同的思想。其中，DMCNN 是传

统的基于CNN的单任务去摩尔纹模型，而AirNet、PromptIR和Diff-Plugin则能够应对更加丰富的摩尔纹噪声类型，且Diff-Plugin的设计是基于扩散模型，相较于其他方法具有更加优异的表现。

3.4.1 定性实验结果

StableMoiré与其他模型的生成结果在视觉上的比较如图4所示。从可视化结果分析可知，StableMoiré无论是从像素空间的恢复质量、还是从语义信息的保留程度来看，都有最佳表现。

与Diff-Plugin的恢复结果对比，StableMoiré恢复后得到的图像更加纯净，不存在Diff-Plugin恢复结果中所出现的明显阴影斑块，且恢复得到的图像曝光度与对比度得到了更好的平衡，而Diff-Plugin的恢复结果则存在明显的过曝情况。

与DMCNN的恢复结果对比，StableMoiré在所有的去摩尔纹场景下都取得了更好的恢复结果，证明该模型能够利用预训练过程中获得的一系列先验知识，实现去摩尔纹知识在多场景下的泛化，使模型在特殊场景下的表现更加优异，而DMCNN则在部分场景下无法完全消除摩尔纹，泛化性较差。

AirNet和PromptIR虽然能在部分场景的摩尔纹去除上获得较好的结果，但与StableMoiré对比，在某些场景下的去摩尔纹表现并不稳定。具体来说，这2种方法在部分场景下，恢复得到的图像依旧存在肉眼可见的噪声纹理或是人眼能轻松感知到的曝光不稳定现象。因此，这2种方法在稳定性上不如StableMoiré。

由此可见，StableMoiré无论是在摩尔纹的去除程度、还是在图像恢复质量上，都有更优的表现，且实现了不同场景下复杂摩尔纹特征的稳定去除，进一步体现了其鲁棒性和泛化性。

3.4.2 定量实验结果

不同去摩尔纹算法的定量评估结果如表1所示。显然，StableMoiré在所有方法中取得了最佳的定量评估结果，表现最为优秀。具体来说，在FID与KID指标上，比排名第2的PromptIR算法分别减少了0.03和0.66；而在PSNR与SSIM指标上，分别提升了0.509 2和0.011 3，从定量的角度证明了StableMoiré的优越性。

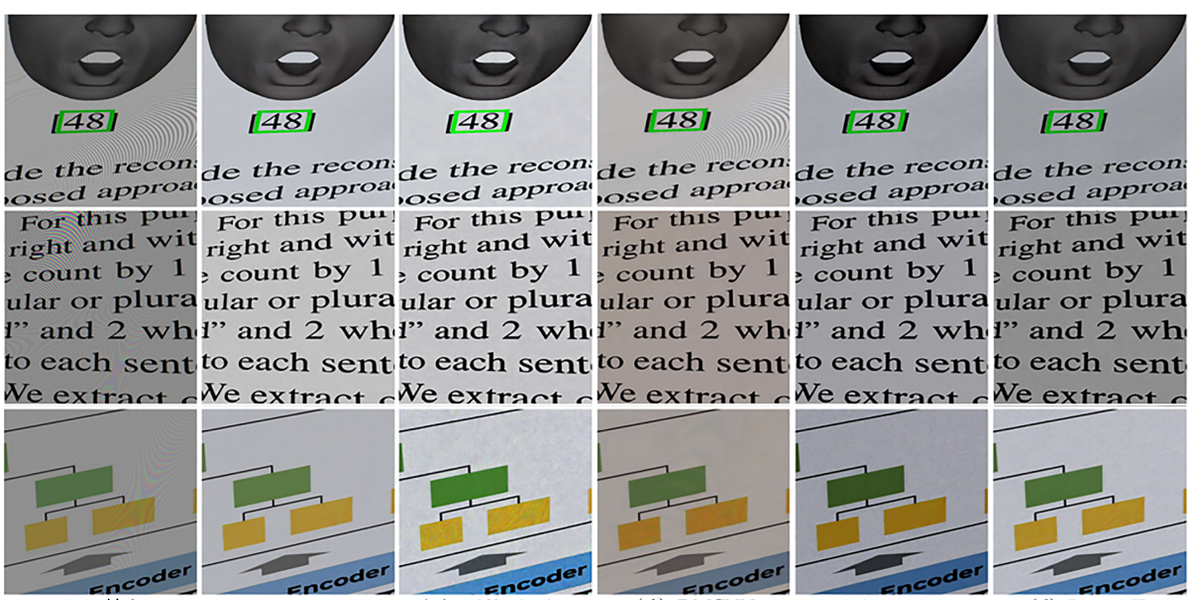
表1 不同模型对比实验定量评估结果

Table 1 Quantitative evaluation results of comparative experiments of different models

方法	FID	KID	PSNR	SSIM
DMCNN（ICIP 2018）	29.13	1.62	14.872 3	0.865 3
AirNet（CVPR 2022）	33.05	4.27	12.864 5	0.811 7
PromptIR（NeurIPS 2023）	29.01	1.56	14.981 7	0.871 3
Diff-Plugin（CVPR 2024）	29.77	1.75	14.401 6	0.852 4
StableMoiré	28.98	0.90	15.490 9	0.882 6

3.4.3 不同方法所需训练参数量对比

不同去摩尔纹算法训练时所需要优化的参数



（a）输入 （b）StableMoiré （c）Diff-Plugin （d）DMCNN （e）AirNet （f）PromptIR

图4 对比实验结果

Fig. 4 Results of comparative experiments

量如表2所示。从数量上看，DMCNN所使用的参数量最少，但去摩尔纹性能表现欠佳。AirNet与PromptIR虽然在部分场景的去摩尔纹任务中取得了不错的结果，但需要庞大的参数量，较小的性能提升却耗费了较大的计算资源。而StableMoiré在平衡参数量的同时，实现了定量与定性角度上相对最优的去摩尔纹结果，其对计算资源的低需求与性能上的高表现进一步彰显了本文理论及模型设计上的优越性。

表 2　不同模型所需训练参数量

Table 2　Numbers of training parameters required for different models

方法	参数量/M
DMCNN（ICIP 2018）	1.43
AirNet（CVPR 2022）	28.70
PromptIR（NeurIPS 2023）	33.00
Diff-Plugin（CVPR 2024）	3.35
StableMoiré	2.83

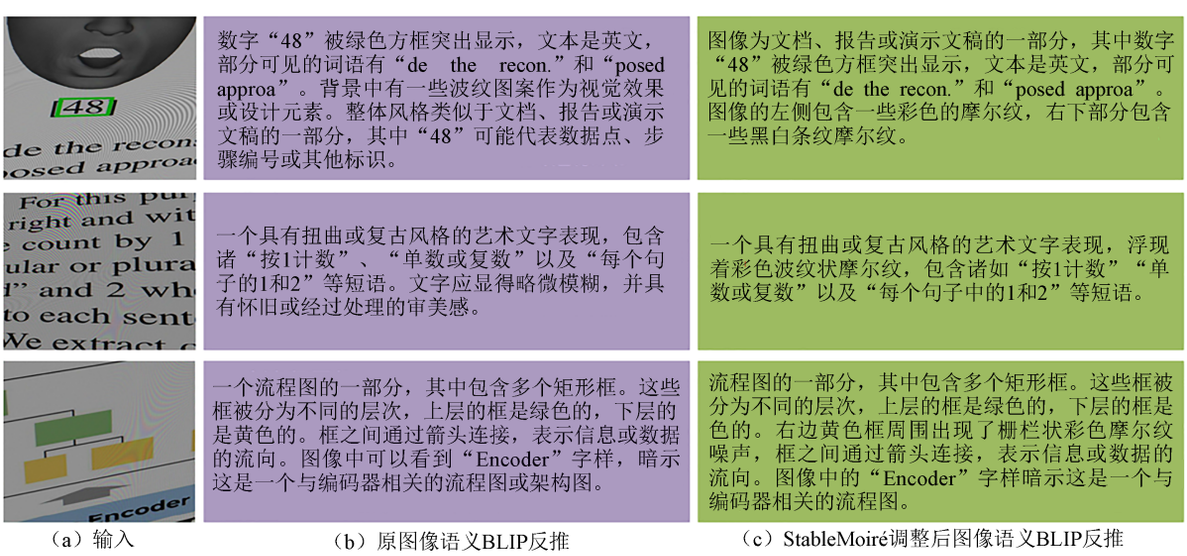
3.4.4　BLIP[41]反推文本提示词实验结果

为了更好地对比分析经StableMoiré微调前后语义空间之间的差别，将微调前后的图像语义特征单独抽取出来，并使用图像反推文本提示词方法BLIP分别对微调前后的语义特征进行提示词反演，结果如图5所示。从实验结果来看，经StableMoiré微调前的图像语义信息主要关注图像的整体像素特征，并未区分图像内容与噪声之间的区别。对部分图像而言，摩尔纹噪声更是被理解成了“波纹状图案装饰”，导致模型在这部分摩尔纹的去除上出现了偏差，最终导致摩尔纹去除不完整、或是本应为空白的区域却出现阴影的情况。而经过StableMoiré微调后的语义空间不仅保留了图像的大部分描述，还重点识别了图像中的摩尔纹区域，使得模型能够更好地区分摩尔纹噪声与图像内容，从而提高模型的恢复性能。

3.5　消融实验

为了评估StableMoiré各模块及其协同工作的有效性，从以下几个角度出发，设计了不同的消融实验：1）为了更好地评估MoiréIN模块，验证本文针对摩尔纹语义信息所提出的相关理论的有效性，采取将MoiréIN模块去除但保留Content Retainer模块的实验设置，记为WioMIN；2）为了更好地评估Content Retainer是否在图像像素细节信息保留与模型生成随机性抑制上起到了作用，采取将Content Retainer模块去除但保留MoiréIN模块的实验设置，记为WioCR；3）为了更好地评估StableMoiré整体架构设计的先进性，将MoiréIN与Content Retainer模块同时去除，以评估StableMoiré整体架构的有效性，并将其记为WioBoth。

StableMoiré、WioMIN、WioCR和WioBoth的对比实验定量评估结果如表3所示。从结果分析可知，本文所提出的MoiréIN与Content Retainer模块在协同工作时能够在所有的实验设置中取得最佳的去摩尔纹结果。



（a）输入　（b）原图像语义BLIP反推　（c）StableMoiré调整后图像语义BLIP反推

图 5　调整前后图像语义经过BLIP文本提示词反推实验结果

Fig. 5　Image semantics inferred from BLIP text prompt pre- and post-adjustment

表3 消融实验定量评估结果

Table 3 Quantitative evaluation results of ablation experiments

方法	FID	KID	PSNR	SSIM
WioMIN	31.07	1.89	13.563 3	0.774 9
WioCR	35.27	2.68	11.263 8	0.621 3
WioBoth	42.17	4.38	9.803 1	0.541 2
StableMoiré	28.98	0.90	15.490 9	0.882 6

具体来看，当去除MoiréIN模块时，模型的性能在FID和KID指标上分别提升了2.09和0.99，在PSNR和SSIM指标上分别下降了1.927 6和0.107 7，证实了MoiréIN模块的有效性，说明摩尔纹相关语义信息的注入能够辅助模型更好地理解图像中与摩尔纹相关的特征，并对其进行去除；当去除Content Retainer模块时，模型的性能有了更明显的下降，在FID和KID指标上分别提升了6.29和1.78，在PSNR和SSIM指标上分别下降了4.227 1和0.261 3，证实了Content Retainer模块在图像细节保留中的重要性。如果没有图像细节的注入，将难以有效抑制模型生成过程的随机性，这将导致模型的生成结果变得不可控；当将2部分模块均去除时，模型将退化为传统的Stable Diffusion，此时的生成过程仅依赖带有图像信息的初始噪声，显然无法对图像中的摩尔纹进行有效去除，且无法抑制生成过程的随机性，在所有实验中的表现最差。

消融实验的结果充分证实了StableMoiré中各模块的有效性以及本文理论分析的正确性，说明在各部分的协同工作下，模型才能取得相对最佳的摩尔纹去除效果。

4 结论

本文基于预训练文生图模型提出了一种图像去摩尔纹算法StableMoiré。一方面，通过摩尔纹相关语义信息的引导，提升了模型对摩尔纹信息的理解能力，解决了传统方法存在的语义漂移问题；另一方面，通过基于残差的图像细节的注入，实现了图像去摩尔纹过程中像素层级的引导，有效抑制了扩散模型在生成过程中的随机性。多模块的协同工作有效利用了预训练模型中的先验知识，提升了模型对复杂场景的适应能力，从而实现了高质量的摩尔纹去除。

通过对比实验证明，StableMoiré无论是在视觉效果上、还是定量指标上，相比现有方法中表现最优的PromptIR，FID和KID指标分别降低了0.03和0.66，PSNR与SSIM指标分别提升了0.509 2和0.011 3。在训练完成后，StableMoiré可作为插件集成于用户端，实现用户拍摄过程中的无缝后处理。然而，图像去摩尔纹缺乏真实的数据集，使用合成数据集可能无法完全覆盖真实世界中的复杂场景。因此，模型在真实场景中的表现还有待进一步测试。未来的研究方向主要集中在2个方面：一是要进一步加强图像去摩尔纹在真实数据上的支撑；二是要将本文提出的方法进一步应用于图像去雨、去雾等其他底层视觉任务，以拓展其在图像恢复领域的应用范围。

参考文献

[1] SUN Y J，YU Y Z，WANG W P. Moiré photo restoration using multiresolution convolutional neural networks[J]. IEEE Transactions on Image Processing，2018：4160-4172.

[2] CHENG X，FU Z Y，YANG J. Multi-scale dynamic feature encoding network for image demoiréing[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop（ICCVW）. Piscataway，NJ，USA：IEEE，2019：3486-3493.

[3] LIU Y，ZHU Z，BAI X. WDNet：watermarkdecomposition network for visible watermark removal[C]//2021 IEEE Winter Conference on Applications of Computer Vision（WACV）. Piscataway，NJ，USA：IEEE，2021：3684-3692.

[4] LIU S，LI C H，NAN N，et al. MMDM：multi-frame and multi-scale for image demoiréing[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. Piscataway，NJ，USA：IEEE，2020：1751-1759.

[5] LIU C M，WANG Y B，ZHANG N H，et al. Learning moiré pattern elimination in both frequency and spatial domains for image demoiréing[J]. Sensors，2022，22(21)：8322.

[6] LIU Y H，KE Z H，LIU F，et al. Diff-Plugin：revitalizing details for diffusion-based low-level tasks[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Piscataway，NJ，USA：IEEE，2024：4197-4208.

[7] WANG Y F，YANG W H，CHEN X Y，et al. SinSR：diffusion-based image super-resolution in a single step[C]//2024 IEEE/CVF Conference on

Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 25796-25805.

[8] WANG J Y, YUE Z S, ZHOU S C, et al. Exploiting diffusion prior for real-world image super-resolution[J]. International Journal of Computer Vision, 2024, 132(12): 5929-5949.

[9] NOROOZI M, HADJI I, MARTINEZ B, et al. You only need one step: fast super-resolution with stable diffusion via scale distillation[C]//Computer Vision-ECCV 2024. Cham: Springer Nature, 2025: 145-146.

[10] YANG Y M, ZOU D S, SONG X Y, et al. DehazeDM: image dehazing via patch autoencoder based on diffusion models[C]//2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway, NJ, USA: IEEE, 2023: 3783-3788.

[11] GUAN F X, LAI H T, ZANG H Y, et al. Image dehazing method based on diffusion model[C]// 2024 IEEE International Conference on Mechatronics and Automation (ICMA). Piscataway, NJ, USA: IEEE, 2024: 477-484.

[12] SHEN Y Y, WEI M Q, WANG Y Z, et al. Rethinking real-world image deraining via an unpaired degradation-conditioned diffusion model[EB/OL]. (2024-05-01)[2025-03-17]. https://arxiv.org/pdf/2301.09430.pdf.

[13] LU S, BIGOULAEVA I, SACHDEVA R, et al. Are emergent abilities in large language models just in-context learning?[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2024: 5098-5139.

[14] ZHANG L M, RAO A Y, AGRAWALA M. Adding conditional control to text-to-image diffusion models[C]//2023 IEEE / CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 3813-3824.

[15] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 10674-10685.

[16] HUANG Y, HUANG J C, LIU Y F, et al. Diffusion model-based image editing: a survey[EB/OL].(2025-03-11)[2025-03-17]. https://arxiv.org/abs/2402.17525v4.

[17] ZHANG C, WU Q Y, GAMBARDELLA C C, et al. Taming stable diffusion for text to 360° panorama image generation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 6347-6357.

[18] WANG B Y, CHEN Q F, WANG Z Y. Diffusion-based visual art creation: a survey and new perspectives[EB/OL].(2024-08-22) [2025-03-17]. https://arxiv.org/abs/2408.12128.

[19] BOUROU A, GENOVESIO A, MEZGER V. GANs conditioning methods: a survey[EB/OL]. (2024-08-28)[2025-03-17]. https://arxiv.org/pdf/2408.15640.

[20] RAMESH A, PAVLOV M, GOH G, et al. Zero-shot text-to-image generation[EB/OL].(2021-02-26) [2025-03-17]. https://arxiv.org/abs/2102.12092.

[21] VASWANI ASHISH, SHAZEER NOAM, PARMAR NIKI, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, LA, USA: Curran Associates Inc., 2017: 6000-6010.

[22] RUIZ N, LI Y Z, JAMPANI V, et al. DreamBooth: fine tuning text-to-image diffusion models for subject-driven generation[C]//2023 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 22500-22510.

[23] HU E, SHEN Y L, WALLIS P, et al. LoRA: low-rank adaptation of large language models [EB/OL].(2021-10-16)[2025-03-17]. https://arxiv.org/abs/2106.09685.

[24] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[EB/OL].(2020-12-16) [2025-03-17]. https://arxiv.org/abs/2006.11239.

[25] KAWAR B, ELAD M, ERMON S, et al. Denoising diffusion restoration models[EB/OL].(2022-10-12)[2025-03-17]. https://arxiv.org/abs/2201.11793.

[26] WANG Y H, YU J W, ZHANG J. Zero-shot image restoration using denoising diffusion nullspace model[EB/OL]. (2022-12-07)[2025-03-17]. https://arxiv.org/abs/2212.00490.

[27] CHUNG H, KIM J, MCCANN M T, et al. Diffusion posterior sampling for general noisy inverse problems[EB/OL].(2022-09-29)[2025-03-17]. https://arxiv.org/pdf/2209.14687.

[28] EI B, LYU Z Y, PAN L, et al. Generative diffusion prior for unified image restoration and en-

hancement[C]//2023 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 9935-9946.

[29] SAHARIA C, HO J, CHAN W, et al. Image super-resolution via iterative refinement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4713-4726.

[30] XIA B, ZHANG Y L, WANG S Y, et al. DiffIR: efficient diffusion model for image restoration [C]//2023 IEEE / CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 13049-13059.

[31] SAHARIA C, CHAN W, CHANG H W, et al. Palette: image-to-image diffusion models[C]//Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings. New York, USA: ACM, 2022: 1-10.

[32] WANG Z X, ZHANG Z Y, ZHANG X Y, et al. DR2: Diffusion-based robust degradation remover for blind face restoration[C] //2023 IEEE /CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2023: 1704-1713.

[33] ZHAO Y, HOU T B, SU Y C, et al. Towards authentic face restoration with iterative diffusion models and beyond[C]//2023 IEEE / CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 7278-7288.

[34] JIANG H, LUO A, FAN H Q, et al. Low-light image enhancement with wavelet-based diffusion models[J]. ACM Transactions on Graphics, 2023, 42(6): 1-14.

[35] WANG J Y, YUE Z S, ZHOU S C, et al. Exploiting diffusion prior for real-world image super-resolution[J]. International Journal of Computer Vision, 2024, 132(12): 5929-5949.

[36] LIN X Q, HE J W, CHEN Z Y, et al. DiffBIR: toward blind image restoration with generative diffusion prior[M]//Computer Vision ECCV 2024. Cham: Springer Nature, 2024: 430-448.

[37] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 10674-10685.

[38] ZHANG X S, YANG W H, HU Y Y, et al. Dmcnn: dual-domain multi-scale convolutional neural network for compression artifacts removal[C]// 2018 25th IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2018: 390-394.

[39] LI B Y, LIU X, HU P, et al. All-in-one image restoration for unknown corruption[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 17431-17441.

[40] POTLAPALLI V, ZAMIR S W, KHAN S H, et al. PromptIR: prompting for all-in-one image restoration[C]//Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023 (NeurIPS 2023). New Orleans, LA, USA: NeurIPS Foundation, 2023: 1-19.

[41] LI J N, LI D X, SAVARESE S, et al. BLIP2: bootstrapping language-image pre-training with frozen image encoders and large language models[EB/OL]. (2023-01-30) [2025-03-17]. https://arxiv.org/pdf/2301.12597.

（编辑：夏菁）