



· 文化遗产数字化保护 ·

SN-CLPGAN: 基于谱归一化的中国传统山水画风格迁移方法

胡琦瑶^{1,2}, 刘乾龙³, 彭先霖⁴, 张翔¹, 彭盛霖¹, 范建平^{1,5}

(1. 西北大学 信息科学与技术学院, 陕西 西安 710127;

2. 生成式人工智能与混合现实陕西省高等学校重点实验室, 陕西 西安 710127;

3. 西北大学 网络与数据中心, 陕西 西安 710127; 4. 西北大学 艺术学院, 陕西 西安 710127;

5. 西北大学 新型网络智能信息服务国家地方联合工程研究中心, 陕西 西安 710127)

摘要 中国传统山水画的风格迁移为文化遗产数字化保护与传承提供了新的路径, 近年来, 深度学习技术已实现了不同图像间的风格迁移, 并展现出栩栩如生的效果。中国传统山水画的风格迁移旨在继承中国古代画家独特的绘画技巧, 但存在3个缺陷: ①缺乏高质量的中国传统山水画图像数据集; ②忽略了中国传统山水画特有的技法和笔墨细节; ③风格迁移效果与真实山水画有所差距。为了弥补上述缺陷, 首先, 创建了一个基于风格迁移的中国传统山水画数据集 STCLP, 包含4 281幅高质量的中国山水画以及自然景观图像, 并提出了一种基于谱归一化的中国山水画风格迁移方法 SN-CLPGAN。其次, 提出了在生成器中使用融合反射填充层的残差稠密块(residual-in-residual dense block, RRDB)学习中国山水画独特的笔触和技法。接着, 引入了多尺度结构相似性指数测量(multi-scale structural similarity index measure, MS-SSIM)损失以减少2幅图像之间的像素差异, 使生成图像更接近传统绘画的色彩和颜料。最后, 采用了融合谱归一化(spectral normalization, SN)的U-Net判别器增强图像纹理细节, 并确保模型训练过程的稳定性。大量实验验证了提出的方法在中国传统山水画风格迁移任务中的有效性和先进性。

关键词 风格迁移; 人工智能艺术; 中国传统山水画; 生成对抗网络

中图分类号: TP391.41 **DOI**: 10.16152/j.cnki.xdxbsr.2025-01-005

SN-CLPGAN: A method of style transfer in Chinese traditional landscape painting based on spectral normalization

HU Qiyao^{1,2}, LIU Qianlong³, PENG Xianlin⁴, ZHANG Xiang¹,
PENG Shenglin¹, FAN Jianping^{1,5}

(1. School of Information Science and Technology, Northwest University, Xi'an 710127, China;

2. Shaanxi Key Laboratory of Higher Education Institution of Generative Artificial Intelligence and Mixed Reality,

Xi'an 710127, China; 3. Network and Data Center, Northwest University, Xi'an 710127, China;

4. School of Art, Northwest University, Xi'an 710127, China; 5. State-Province Joint Engineering and Research Center of

收稿日期: 2024-05-20

基金项目: 国家自然科学基金(62471390、62306237); 陕西省重点研发计划(2024GX-YBXM-149); 西北大学研究生创新项目(CX2024204、CX2024206)。

第一作者: 胡琦瑶, 女, 讲师, 从事人工智能艺术、AIGC研究, huqiyao@nwu.edu.cn。

通信作者: 彭先霖, 男, 副教授, 从事文化遗产数字化保护研究, pxl@nwu.edu.cn。

Advanced Networking and Intelligent Information Services, Xi'an 710127, China)

Abstract The style transfer of Chinese landscape paintings offering new avenues for the digital preservation and inheritance of cultural heritage. In recent years, deep learning technologies have enabled style transfer between different images, achieving lifelike effects. Style transfer in Chinese landscape paintings aims to preserve the unique paintings skills of ancient Chinese painters, but faces three main challenges: ① The lack of high-quality datasets of traditional Chinese landscape paintings. ② The oversight of the unique techniques and ink details specific to traditional Chinese landscape paintings. ③ The gap between the style transfer outcomes and real landscape paintings. To address these deficiencies, this paper first introduces a Chinese landscape paintings dataset for style transfer, STCLP, which contains 4281 high-quality images of Chinese landscape paintings and natural landscapes. A generative adversarial network of style transfer in Chinese landscape painting based on spectral normalization is proposed, termed SN-CLPGAN. Additionally, it introduces the use of residual-in-residual dense blocks (RRDB) with reflect padding layers in the generator to learn the distinctive brushstrokes and techniques of Chinese landscape paintings. Furthermore, it employs the multi-scale structural similarity index measure (MS-SSIM) loss to minimize pixel-level differences between images, thereby producing images closer to traditional paintings in terms of color and pigmentation. Finally, the U-Net discriminator fused with SN is utilized to enhance the textural details of images, ensuring the stability of the model training process. Extensive experiments validate the effectiveness and advancement of the proposed method in the task of style transfer for Chinese landscape paintings.

Keywords style transfer; artificial intelligence art; Chinese landscape paintings; generative adversarial network

中国传统山水画作为中国古代文化与艺术的璀璨珍宝,其声望不仅源于其反映自然物象的光影变幻与表象细节,更在于其深刻表达了画家的主观情绪和精神境界。近年来,随着深度学习和图像处理技术的迅猛发展^[1-2],风格迁移——即图像的风格化或非现实渲染,作为深度学习在图像处理领域的关键技术之一,已经成为计算机视觉与传统绘画融合的一个创新领域。特别是将这项技术应用于中国传统山水画,不仅能在视觉上复现其独特的艺术韵味,更重要的是,这种结合开辟了中国文化遗产数字化保护和传承的新路径。风格迁移技术的发展不仅体现了深度学习在图像处理领域的应用成就,也彰显了图像处理技术与中国传统绘画相结合的新可能,为保护和传播中华优秀传统文化提供了新的技术手段。

在此背景下,生成对抗网络(generative adversarial network, GAN)^[3]及其后续发展的技术,如 CycleGAN^[4]和 Pix2pix^[5]模型为风格迁移提供了强大的工具。具体地,Goodfellow 等人提出 GAN 模型^[3]对图像风格迁移领域产生了革命性的影响,标志着风格迁移任务进入深度学习的时代;紧接着,Zhu 等人提出了 CycleGAN 框架^[4],旨在基于非配对数据集训练出满足目标风格的图像,从

而实现风格迁移;Zheng 等人提出了一种 P2-GAN 模型^[6],能有效地从单个风格图像中学习笔画风格,并实现高质量的图像风格迁移;Wang 等人提出了一种美学风格注意力(Aes-SA)模块^[7],将美学特征纳入风格迁移过程,以增强生成图像的视觉效果。

尽管基于神经网络的风格迁移技术在自然图像处理领域取得了非凡的成就,但针对中国传统绘画这一特定类别的风格迁移任务绝非易事。为了兼顾图像生成的质量并保留中国传统山水画独特技法,需要解决以下 3 个问题:①如何获取高质量的中国传统山水画图像数据?②如何保留中国传统山水画独特的技法和笔墨细节?③如何使风格迁移自然生动,确保高质量的中国传统山水生成效果?为解决上述问题,我们利用 CycleGAN 模型设计了一个基于谱归一化的中国传统山水画风格迁移模型,命名为 SN-CLPGAN。为了获得高质量的图像数据,从多个网站和文化遗产机构收集了 4 281 幅高质量的自然景观图像以及中国历代传统山水画。此外,还进行了图像增强处理,以确保原始图像的质量。本文创建的数据集以中国传统山水画为特色侧重于山、石、树等具体物象的轮廓细节,以及中国传统绘画独特的笔墨技巧。

首先,在生成器模块中引入了残差稠密块(residual-in-residual dense block, RRDB)^[8],可以进一步学习山、石、树木等物象的绘画技法 and 笔触。此外,添加了反射填充层,以避免伪影和光斑,进一步提高生成图像的质量。其次,采用融合谱归一化(spectral normalization, SN)^[9]技术的 U-Net 判别器^[10],在稳定训练过程的同时改善图像恢复的纹理细节信息。最后,引入 MS-SSIM (multi-scale structural similarity index measure) 损失^[11-12],让模型专注于在亮度、对比度和结构方面约束风格迁移后的图像,从而使风格迁移生成的图像更符合人类的视觉感知,更接近中国传统山水画所独有的画面色彩。借助这些方法,我们成功保留了中国传统山水画中独特的笔墨韵味与细节,从而生成高质量且富有韵味的山水画。

1 相关工作

传统的风格迁移方法主要依靠纹理分析和合成技术识别应用源图像中的风格特征。随着深度学习技术的迅猛发展,风格迁移迎来了技术革新。相较于传统方法主要关注纹理转换,深度学习使得风格迁移技术不仅能够捕捉到更细腻的风格特征,还能在更广泛的艺术领域内进行应用^[13]。当前,基于深度学习的风格迁移主要分为2大类:基于图像迭代和基于模型迭代的方法。这些方法通过不断优化过程,有效地将一幅图像的风格迁移到另一幅图像上。Gatys 等人首次提出了基于图像迭代的图像风格迁移算法,它是基于深度学习的图像风格迁移技术的基本理论基础^[14];受 Gatys 等人开创性方法的影响,结合 GAN 和成对学习思想,Zhu 等人提出了 CycleGAN 这一无监督风格迁移方法^[4];Chen 等人提出一种可以有效捕捉纹理和风格等局部高频特征的风格迁移方法 NiceGAN,但在对全局分布建模时效果不佳^[15];Yi 等人提出 DualGAN,结合了 GAN 和对偶学习策略,使用多个卷积层构建编码器,使用多个转置卷积层构建解码器,然后使用一个能够在两者之间产生变化的深度网络作为特征转换器^[16];U-GAT-IT^[17]则引入注意力机制,并提出了正则化方法 AdaLIN,从而构建了一种跨域风格迁移模型。

虽然上述工作取得了一定进展,但也存在提取的图像内容忽略了细节特征、输出图像存在伪影等问题。针对目前风格迁移算法中存在的图像

失真等质量问题,张美玉等人通过将2种特征进行拉普拉斯算子滤波后计算2者的L1误差,约束了图像的变化,抑制了伪影生成^[18];茹超等人提出了一种颜色迁移算法,重点考虑了边缘所存在的弱化现象,增强了生成图像的视觉效果^[19];Peng 等人提出了一种轮廓增强 CycleGAN 框架,用于从风景照片到中国传统山水画的风格迁移^[20];Zhang 等人在 CycleGAN 的基础上进一步设计了 CycleGAN-AdaIN 框架,旨在将实景照片转换为中国水墨画^[21]。尽管基于 GAN 的风格迁移方法已经取得了令人瞩目的成果,但很少有风格迁移技术应用于中国传统山水画这一特定领域的相关工作。此外,现有方法很少考虑中国传统绘画所具有的独特笔墨技法,因此,本文提出的中国传统山水画风格迁移方法将着力解决上述问题。

2 研究方法

2.1 网络框架

CycleGAN^[4]模型为本文所提出的中国山水画图像风格迁移工作奠定了坚实的理论基础。CycleGAN 通过引入循环一致性损失机制,使得模型能够在没有一对一配对训练数据的情况下自行学习图像间的对应关系,从而有效地处理包括复杂风格迁移在内的图像到图像转换任务。基于此,本文选择 CycleGAN 模型作为本文中国山水画风格迁移模型的骨干网络。

图1展示了 SN-CLPGAN 模型的网络框架图,SN-CLPGAN 模型由2个镜像对称的 GAN 组成循环结构,其中,每个生成器负责将图像从 X 域转换到 Y 域或从 Y 域转换到 X 域,而相应的判别器则负责区分转换图像与真实图像,每个方向的转换由一个改进的生成器和一个改进判别器组成。具体的,改进的生成器由初始卷积块、下采样块、RRDB 块、上采样块以及输出卷积块5部分组成[见图1(b)];改进的判别器由融合谱归一化的 U-Net 判别器组成[见图1(c)]。每个转换方向的生成器不仅受到来自其对应判别器的对抗性损失的约束,还受到循环一致性损失的约束。这种循环一致性损失确保从 X 域到 Y 域或从 Y 域到 X 域的转换后,再转换回原始域时,能够重建接近原始图像的结果。这意味着,即便是在缺乏一对一配对训练样本的情况下,网络也能学习到2个域之间具有明确意义的映射关系。

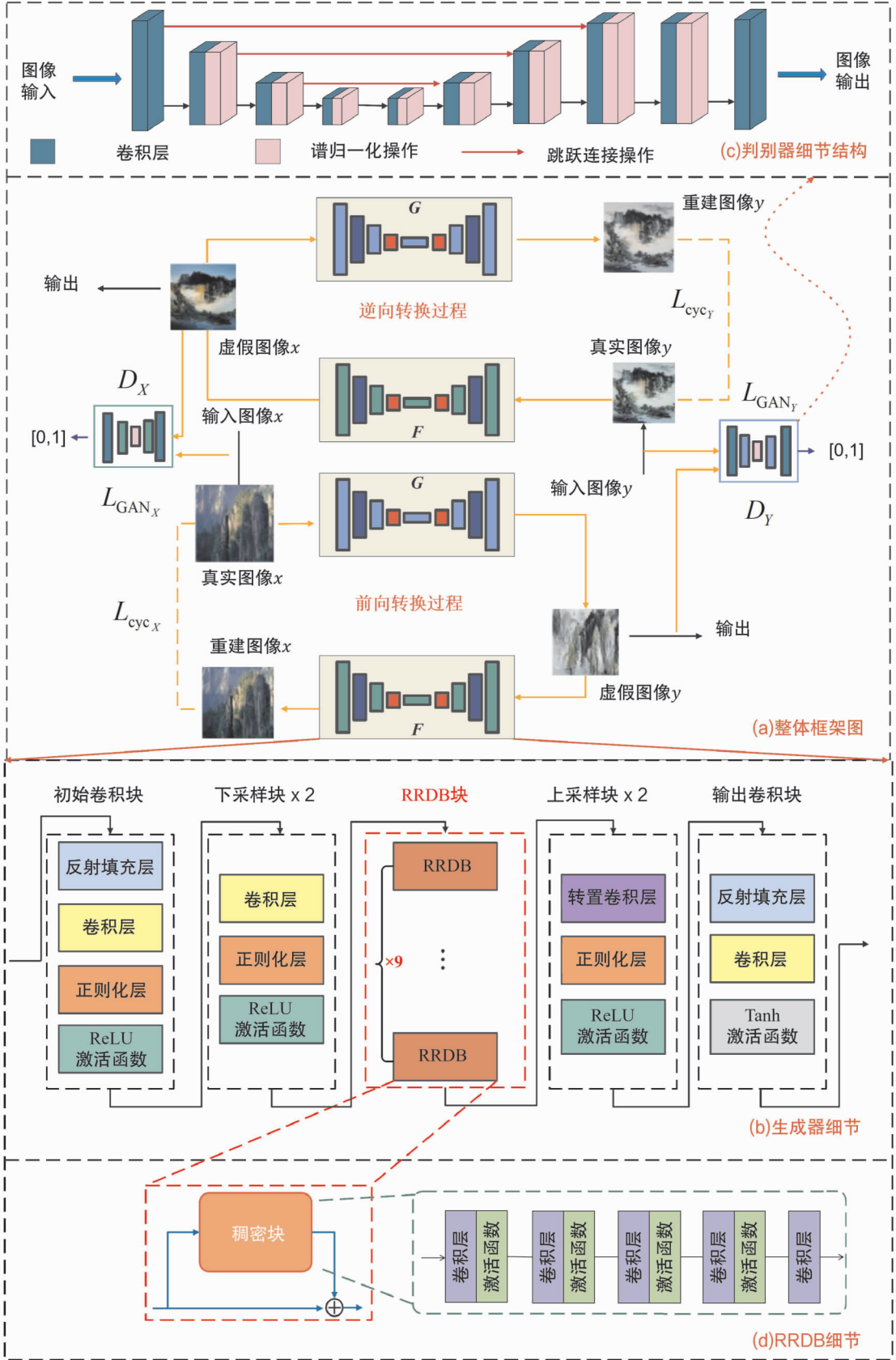


图 1 SN-CLPGAN 模型整体框架图

Fig.1 SN-CLPGAN model overall framework

此外,引入 MS-SSIM 损失 (multi-scale structural similarity index measure loss, MS-SSIM Loss), 以减少 2 幅图像之间的像素差异,可以做到生成图像更接近传统绘画的色彩和颜料,从而使得生成的山水画风格迁移图像更加生动且自然。

2.2 生成器细节

图 1(b) 显示了 SN-CLPGAN 网络模型的生成器结构,该生成器由 5 个基本模块组成:①初始卷积块,由反射填充层、 7×7 卷积层、实例归一化层和 ReLU 激活层组成,用于初始特征提取和消除伪影;②下采样块,由 3×3 卷积层、实例归一化层和 ReLU 激活层组成,用于降低图像的空间分辨率,同时增加特征图的深度;③RRDB 块,由 9 个具有残差密集块组成,并移除 BN 层,以进一步细化和增强特征;④上采样块,由 3×3 转置卷积层、实例归一化层和 ReLU 激活层组成,用于恢复图像的空间分辨率,生成清晰的输出图像;⑤输出卷积块将上采样后的特征图转换为最终的输出图像,并利用 Tanh 激活函数使得输出图像的像素值映射到 $[-1, 1]$ 之间的合理范围,此外,还负责将多维特征映射转换成与目标域兼容的图像。

已有研究实验表明,使用 RRDB 块作为基本块更有利于模型的表示,并能获得令人满意的性能^[8]。图 1(d) 显示了 RRDB 块的具体细节,RRDB 块结合了残差网络 (ResNet) 和稠密连接网络 (DenseNet) 的概念,由若干层卷积和激活函数组成,每一层都与前面所有层直接连接。值得注意的是,RRDB 块移除了批归一化层 (batch normalization, BN)。通过这种结合,既提高了模型的特征学习能力,又保持了网络的深度和效率。

关于图像生成任务,多项研究证实引入批归一化层的性能往往不能令人满意^[22],例如图像超分辨率^[23]和图像去模糊^[24]的任务。因此,移除批归一化层有助于提高泛化能力,降低计算复杂度和内存使用量。

2.3 判别器细节

原始的 CycleGAN 网络采用补丁大小为 70×70 像素的 PatchGAN 作为判别器网络。与全图像判别器相比,这种补丁级判别器架构的参数较少。然而,由于中国山水画中存在细粒度的特征信息和丰富的细节区域,对中国传统山水画的风格迁移任务提出了新的要求,例如需要处理足够大的退化空间、理解图像中复杂的关系、保留细粒度的特征信息以及更稳健的训练过程等。鉴于此,本

文采用了融合谱归一化的 U-Net 判别器稳定训练动态 [见图 1(c)]。判别器的激活函数是 LeakyReLU,网络结构采用独特的跳跃连接。判别器包含 10 个卷积层,通过这些卷积层,模型能够逐渐抽取并处理输入图像的细粒度特征,第一层和最后一层没有进行谱归一化操作,这种设计目的是为了在输入层允许更多的灵活性和在输出层保持网络的判别能力,同时,在中间层通过谱归一化确保稳定性和有效性。谱归一化有助于减轻 GAN 训练阶段所产生的伪影,从而带来更生动、真实的图像细节信息。通过这种设计有效地训练 SN-CLPGAN,并在去除伪影的图像增强效果方面实现局部-全局细节的良好平衡。

2.4 损失函数

对抗损失 (adversarial loss) 沿用原始 GAN 的对抗损失方法^[3],并将其应用于 2 个对称映射函数。例如生成器 G 有如下映射关系,即 $G: x \rightarrow y$, 对应的判别器为 D_Y 。目标损失如式(1)所示。

$$L_{\text{GAN}}(G, D_Y, X, Y) = E_{y \sim p(y)} [\log D_Y(y)] + E_{x \sim p(x)} [\log (1 - D_Y(G(x)))] \quad (1)$$

式中: G 是生成与 Y 域中的图像所相似的图像; D_Y 是判断生成的图像 $G(x)$ 与真实图像 y 之间的差异性; E 是取得样本的平均值。

至于另一个生成器 F (即 $F: y \rightarrow x$) 及其对应的判别器 D_X , 目标函数 $L_{\text{GAN}}(F, D_X, Y, X)$ 的构建过程完全是镜像对称的,不再赘述。

循环一致性损失 (cycle consistency loss) 对抗训练可以学习映射 G 和 F , 并生成与目标域 Y 和 X 分布相似的输出。然而,仅依赖对抗损失无法保证学习到的映射函数能够精确地将每个输入 X_i 映射到一个确切的、预期的输出 Y_i , 为了弥补这一点,引入了循环一致性损失。 X 域中的每张图片 x 都应能在循环平移中回到原点,反之亦然,即 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ 。因此,循环一致性损失函数可以表示为式(2)。

$$L_{\text{cyc}}(G, F) = E_{x \sim p(x)} [\|F(G(x)) - x\|_1] + E_{y \sim p(y)} [\|G(F(y)) - y\|_1] \quad (2)$$

本体映射损失 (identity loss) 旨在保持输入图像与输出图像之间的颜色一致性。通过这种损失,可以约束生成器 G 和 F , 使其在不需要进行域转换时,避免对输入图像的色彩进行自由改变。具体而言,当一张属于目标 Y 域的图像通过映射 G 进行处理时,或当一张属于源域 X 的图像通过映射 F 处理时,期望输出图像与输入图像保持一致,

即 $G(Y) \approx Y$ 和 $F(X) \approx X$, 具体如式(3)所示。

$$L_{\text{identity}}(G, F) = E_{y \sim p(y)} [\|G(y) - y\|_1] + E_{x \sim p(x)} [\|F(x) - x\|_1] \quad (3)$$

L_1 、 L_2 损失函数具体如式(4)、(5)所示。

$$L_1 = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)| \quad (4)$$

$$L_2 = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (5)$$

式中: x_i 表示输入; $f(\cdot)$ 表示网络; y_i 表示实际的观测值。

本文采用了损失函数 L_{SSIM} , 在考虑的所有指标中展示了其卓越的性能。 L_{SSIM} 是根据亮度、对比度和结构的 SSIM 指标制定的, 考虑到了人类的视觉感知^[25], 其广泛应用于图像生成和图像增强任务中, 目的是最小化生成图像与原始图像之间的结构差异, 使生成图像更接近原始图像。 L_{SSIM} 如式(6)所示。

$$L_{\text{SSIM}}(P) = 1 - \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \times \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}(\bar{p}) \quad (6)$$

式中: $\bar{p}$ 是补丁 P 的中心像素; μ_x, μ_y 表示 x, y 的均值; σ_x, σ_y 表示 x 和 y 的方差; σ_{xy} 表示 x 和 y 之间的协方差; C_1, C_2 是维持稳定的常数。

多尺度结构相似性 (MS-SSIM) 是一种评估图像质量的指标, 广泛应用于图像重建和超分辨率等领域。MS-SSIM 的核心思想是计算图像在多个尺度上的相似性, 具体做法是对原始图像进行多个下采样和模糊处理, 每个处理产生不同尺度的图像。对每个尺度计算 SSIM 分数, 然后将这些分数合并, 从而得到最终的 MS-SSIM 值, 类似于式(6), 具体的损失函数 $L_{\text{MS-SSIM}_{\text{raw}}}$ 如式(7)所示。

$$L_{\text{MS-SSIM}_{\text{raw}}}(P) = 1 - \text{MS-SSIM}(\bar{p}) \quad (7)$$

无论局部结构如何, L_1 损失都能以相同的误差权重保留图像的色彩和亮度。为了捕捉 2 种误差函数的最佳特性, 使迁移后的图像在风格上更加自然, 并结合来自不同分辨率的多尺度图像细节, 采用了一种结合 $L_{\text{MS-SSIM}_{\text{raw}}}$ 和 L_1 损失的优化方法。因此, 可以得到更新后的 $L_{\text{MS-SSIM}}$ 如式(8)所示。

$$L_{\text{MS-SSIM}}(G, F) = \alpha(1 - \text{MS-SSIM}_X) + (1 - \alpha)L_{1_x} + \alpha(1 - \text{MS-SSIM}_Y) + (1 - \alpha)L_{1_y} \quad (8)$$

式中: α 是 $L_{\text{MS-SSIM}_{\text{raw}}}$ 的权重系数。根据多次实验结

果和超参数灵敏度分析, 本文将 α 的最佳值设为 0.16, 从而使模型达到最佳性能。

综上所述, 可以得到总体目标损失函数 L_{total} 如式(9)所示。

$$L_{\text{total}} = L_{\text{GAN}}(G, D_Y, X, Y) + L_{\text{GAN}}(F, D_X, Y, X) + \lambda_1 L_{\text{cyc}}(G, F) + \lambda_2 L_{\text{identity}}(G, F) + \lambda_3 L_{\text{MS-SSIM}}(G, F) \quad (9)$$

式中: λ_1, λ_2 和 λ_3 为控制不同损失函数相对重要性的权重系数。 $\lambda_1 = 10, \lambda_2 = 5$, 与 CycleGAN 中的权重设置一致, 通过反复调参并对结果可视化分析, 设置 $\lambda_3 = 1$ 。

3 实验与分析

3.1 数据集和实验环境

本文自建的风景画图像数据集 STCLP 包括风景照片画和未配对的自然图像, 该数据集共有 4 281 幅图像, 包括 2 175 幅自然风景照片和 2 106 幅中国山水画。用于训练和测试的数据分别为 1 675 幅和 500 幅自然景观照片以及 1 545 幅和 561 幅中国山水画。在训练过程中, 所有图像的大小均调整为 256×256 像素。此外, 为了提高模型的训练效率与生成图像效果, 还进行了旋转、裁剪等必要的图像增强处理, 以确保输入模型的图像数据更加均匀和规范化。

本实验基于 Ubuntu 21.10 操作系统, 该系统配备了 24 GiB 显存的 RTX-3090 显卡, 实验代码基于 Pytorch 和 TensorFlow 框架实现, 模型使用 Adam 优化器^[26], 默认学习率设置为 0.000 2。

3.2 定性分析

将本文提出的方法与经典的图像风格迁移模型进行了定性分析与比较, 包括原始 CycleGAN^[4]、NICEGAN^[15]、DualGAN^[16] 和 U-GAT-IT^[17], 可视化结果如图 2 所示。NICEGAN 方法生成的图像存在结构失真和内容丢失等严重问题; U-GAT-IT 方法生成的图像内容损失、图像模糊和色彩缺失; DualGAN 方法在具体内容保留方面效果显著, 但在图像纹理风格迁移方面效果不甚理想; 原始的 CycleGAN 方法可视化结果不存在内容结构缺失的问题, 但在保持色彩多样性方面效果不佳。从图 2 的第 3 行可以看出, 本文方法能够较好地学习到山体的细节和边缘。原始 CycleGAN 模型对山、石、树木等具体物象信息重构效果较差, 本文方法引入了 RRDB 模块, 可

以进一步学习山、石、树木等物象信息的独特绘画技法和笔触。相比之下,本文提出的方法可以做

到最大程度保留原始图像的语义和结构信息,风格迁移效果令人满意。



图 2 不同方法风格迁移可视化效果

Fig. 2 Visualization of different methods of style transfer.

3.3 定量分析

为评估本文生成的风格迁移图像质量,拟采用 4 种评价指标进行比较分析,包括结构相似性指数 (SSIM)、峰值信噪比 (PSNR)、Inception 分数

(IS)^[26-27]和神经图像评估^[28](NIMA)。研究表明,PSNR 和 SSIM 这 2 个传统指标与人类对视觉质量的评价存在较低的相关性,基于此,本文引入 NIMA 作为补充评价指标以进一步评估图像的美学质量,指标测试结果如表 1 所示。

表 1 与基线方法的性能对比

对比方法	CycleGAN ^[4]	NICEGAN ^[15]	DualGAN ^[16]	U-GAT-IT ^[17]	本文方法
SSIM	0.733	0.645	0.670	0.685	0.749
PSNR	9.359	9.006	8.889	13.105	9.482
IS	3.724	3.415	3.751	3.632	3.969
NIMA	4.732	4.563	4.740	4.697	4.887

图 3 是 500 张测试图像的 NIMA 分数直方图,显示了 NIMA 分数的平均值或 NIMA 分数的差值。在中国山水像进行风格迁移的实验中,本文提出的方法在 SSIM 和 IS 评分上优于对比方法。这表明本方法在保持图像结构相似性及生成图像的多样性和清晰度方面具有显著优势。虽然 U-GAT-IT 在 PSNR 指标上表现最佳,反映了其在保持生成图像与目标图像像素级差异方面的能力,但在保持图像结构、多样性和清晰度方面表现不如本文方法。综上所述,相比传统风格迁移方法,本文方法展现出显著的性能优势。

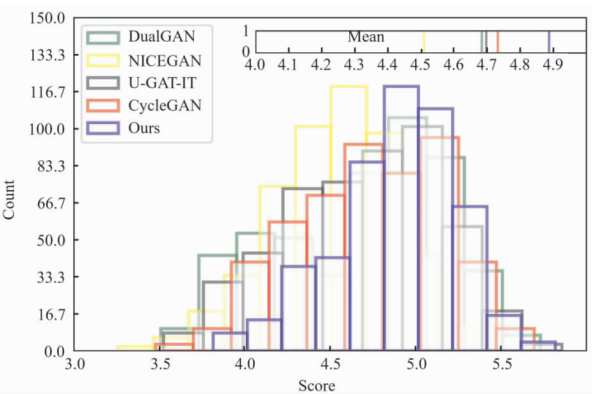


图 3 500 张测试图像的 NIMA 分数直方图

Fig. 3 Histogram of NIMA scores for 500 test images

图 4 显示了不同方法生成的风格迁移图像的 NIMA 分数与可视化效果图,分数越高说明生成

的风格迁移图像质量越好。

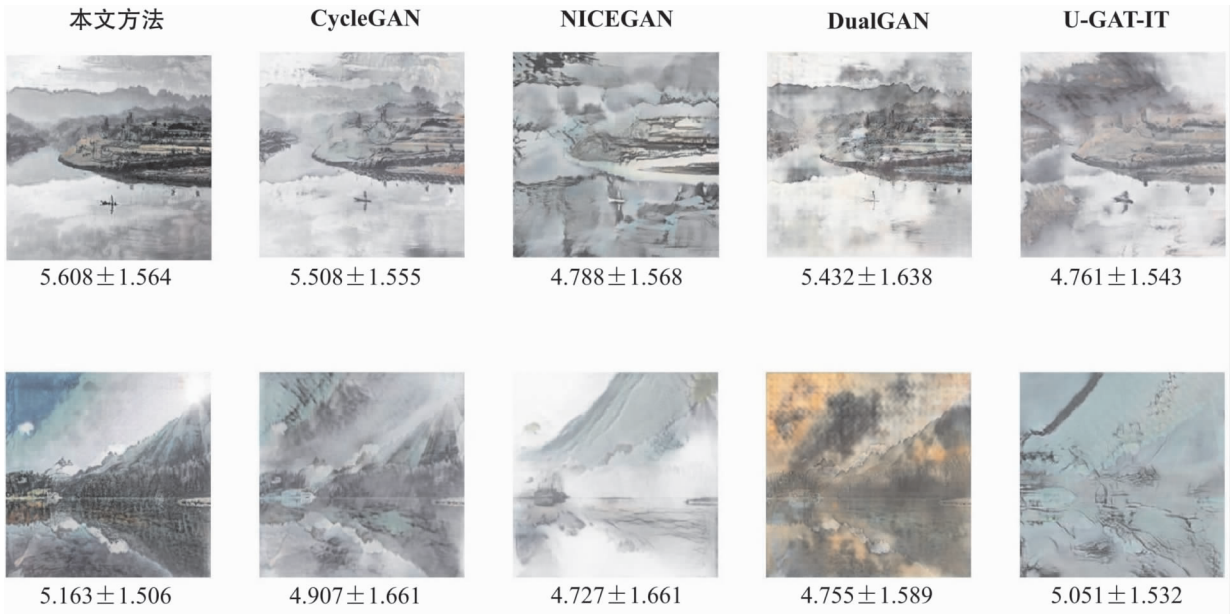


图 4 不同对比方法的 NIMA 分数与可视化效果图

Fig. 4 NIMA scores and visualization of different comparison methods

3.4 消融研究

为了进一步评估本文算法在风格迁移方面的性能,并证明每个模块的有效性,构建了 5 个变体模型:① SN-CLPGAN w/o RRDB& MS-SSIM;② SN-CLPGAN w/o SN&RRDB;③ SN-CLPGAN w/o SN&MS-SSIM;④ SN-CLPGAN w/o RRDB;⑤ SN-CLPGAN w/o MS-SSIM。表 2 统计了 5 个模型在 SSIM、PSNR、IS 和 NIMA 这 4 个指标上的评分情况,粗体显示最佳实验结果。可以看出,本文模型在 SSIM、IS 和 NIMA 3 项指标上都取得了最佳

结果,而本文方法在移除 RRDB 模块的情况下,PSNR 指标取得了最佳结果,说明此种情况下添加 RRDB 模块会对模型的准确表达造成一定负面影响。然而,从图 5 消融研究的可视化效果展示图可以看出,尽管第 6 列(SN-CLPGAN w/o RRDB 模型)的 PSNR 分值最高,但存在明显伪影和亮点。第 4 列(SN-CLPGAN w/o SN&RRDB 模型)中的 PSNR 值是次优的。实验结果表明,PSNR 分数与人眼看到的视觉质量并不完全一致,较高的 PSNR 可能比较低的 PSNR 看起来更糟糕。

表 2 消融实验结果

Tab. 2 Ablation study results

模型方法	SSIM	PSNR	IS	NIMA
SN-CLPGAN w/o RRDB& MS-SSIM	0. 738	9. 191	3. 897	4. 835
SN-CLPGAN w/o SN&RRDB	0. 736	9. 510	3. 651	4. 666
SN-CLPGAN w/o SN&MS-SSIM	0. 707	9. 341	3. 343	4. 694
SN-CLPGAN w/o RRDB	0. 815	9. 532	3. 917	4. 899
SN-CLPGAN w/o MS-SSIM	0. 743	9. 473	3. 934	4. 949
本文方法	0. 889	9. 423	4. 169	5. 187

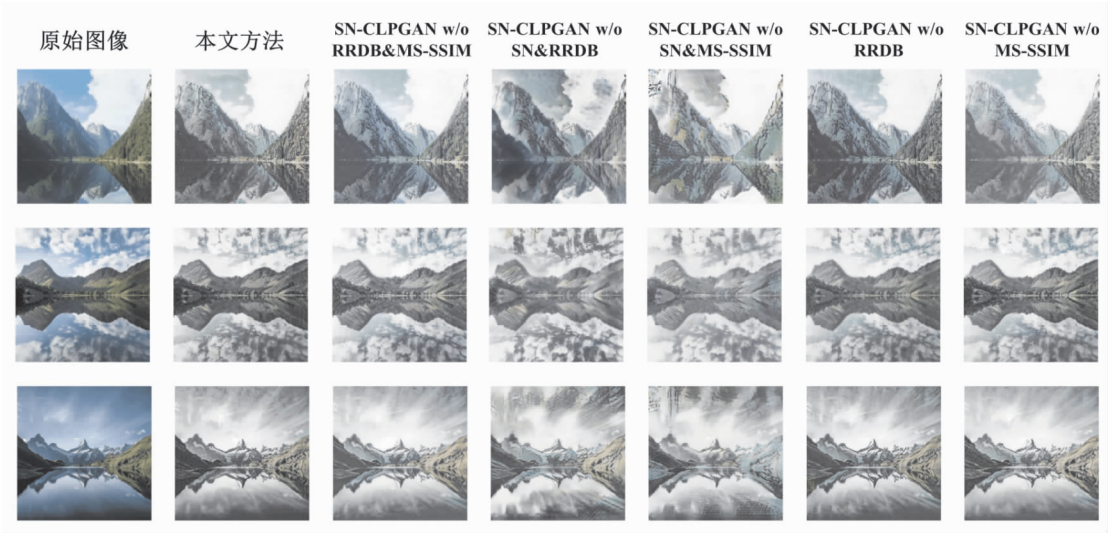


图 5 不同模型消融研究的可视化效果图

Fig. 5 Visualization of the different model ablation study

从图 5 第 1 行可以看出,本文提出的方法可以有效学习传统山水画中独特的绘画技法和细粒度纹理细节。从图 5 中的消融实验结果来看,加入 $L_{MS-SSIM}$ 损失函数可以有效减少 2 幅图之间的像素差异,更加接近传统画家所使用的丰富色彩和颜料,同时,还能为生成的图像带来更好的色彩对比度、饱和度和丰富度。引入 RRDB 模块,添加反射填充层可以进一步学习山、石、树等物体的绘画技巧和笔触。此外,还添加了反射填充层,以避免伪影和光斑,进一步提高生成图像的质量。鉴于高频纹理特征会在图像转换过程中丢失,使用带有谱归一化操作的 U-Net 判别器可以改善纹理

恢复效果。本文方法的每项改进都会影响网络的性能,2 项改进的叠加也会比 1 项改进取得更好的效果。

图 6 展示的可视化效果则进一步说明添加 $L_{MS-SSIM}$ 损失函数可以有效减少 2 张图之间的像素差异,更接近传统画家使用的丰富色彩和颜料。从第 3 列的图片对比可以看出,本文方法学习到了房屋、屋顶和树木倒影等高频纹理细节特征。第 4 列的图片显示,本文方法可以有效模仿湖水和草原的色彩效果,说明了图像生成效果更符合中国山水画的艺术风格。



图 6 本文方法与 SN-CLPGAN w/o MS-SSIM 模型对比的可视化效果图

Fig. 6 Visualization of SN-CLPGAN in comparison with SN-CLPGAN w/o MS-SSIM

3.5 其他数据集上的实验

我们采用源自谷歌艺术与文化项目的 Painting by Numbers 数据集,从该宏大的数据库中随机抽样了 4 281 幅作品,构建了一个数据子集,其中训练集包含 3 220 幅图像,测试集包含 1 061 幅图像。在确保所有训练参数固定不变的严谨条件下,我们对模型重新训练,测试结果如图 7 所示。



图 7 本文方法在 Painting by Numbers 数据集上的可视化效果图

Fig. 7 Visualization of SN-CLPGAN on Painting by Numbers

4 结语

本文创建了一个传统中国山水画风格迁移数据集 STCLP,并设计了一种创新的风格迁移模型 SN-CLPGAN。该模型通过融合反射填充层和残差稠密块 (RRDB) 构建的生成器,以及集成谱归一化技术的 U-Net 架构判别器,加之图像生成损失函数 $L_{MS-SSIM}$,在捕捉具体山水画中物象的细节和边缘处理方面取得了显著进步,并且在纹理细节的再现上展示了优秀的性能。实验结果充分验证了本文方法在进行中国山水画风格迁移任务方面的有效性与先进性。

中国山水画风格迁移的研究不仅推动生成式计算机视觉技术在中国古典文化创作领域的应用,也为文化遗产数字化保护与传承开辟了新的道路^[29]。未来计划探索去噪扩散概率模型^[30]与传统中国山水画风格迁移的结合,目的在于进一步提升图像生成能力,并实现更加令人满意的风格迁移效果。

参考文献

[1] 彭进业,余喆,屈书毅,等. 基于深度学习的图像修复方法研究综述[J]. 西北大学学报(自然科学版), 2023, 53(6): 943-963.

实验结果表明本文模型不仅成功学习到了清晰的艺术风格特征,并且在将这些风格应用于风景画的处理上达到了预期目标。通过对风格化处理的位置与强度的合理控制,最终生成的图像自然流畅、艺术效果显著,该结果有效验证了本文方法的有效性与可行性。

PENG J Y, YU Z, QU S Y, et al. A survey of image inpainting methods based on deep learning[J]. Journal of Northwest University (Natural Science Edition), 2023, 53(6): 943-963.

[2] 胡琦瑶,杨皓文,王佳欣,等. 基于弱监督深度学习的图像检索技术研究[J]. 西北大学学报(自然科学版), 2020, 50(5): 793-801.

HU Q Y, YANG H W, WANG J X, et al. Research on image retrieval based on weakly-supervised deep learning[J]. Journal of Northwest University (Natural Science Edition), 2020, 50(5): 793-801.

[3] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [EB/OL]. (2014-06-10) [2024-04-10]. <http://arxiv.org/abs/1406.2661>.

[4] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2242-2251.

[5] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5967-5976.

[6] ZHENG Z T, LIU J Y, ZHENG N N. P²-GAN: Efficient stroke style transfer using single style image[J]. IEEE Transactions on Multimedia, 2023, 25: 6000-6012.

- [7] WANG Z Z, ZHANG Z J, ZHAO L, et al. AesUST: Towards aesthetic-enhanced universal style transfer [C] // Proceedings of the 30th ACM International Conference on Multimedia. Lisboa; ACM, 2022: 1095-1106.
- [8] WANG X T, YU K, WU S X, et al. ESRGAN: Enhanced super-resolution generative adversarial networks [C] // Computer Vision-ECCV 2018 Workshops. Cham; Springer International Publishing, 2019: 63-79.
- [9] WANG X T, XIE L B, DONG C, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data [C] // 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal; IEEE, 2021: 1905-1914.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [M] // Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham; Springer International Publishing, 2015: 234-241.
- [11] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [EB/OL]. (2018-02-16) [2024-04-10]. <https://arxiv.org/abs/1802.05957>.
- [12] ZHAO H, GALLO O, FROSIO I, et al. Loss functions for image restoration with neural networks [J]. IEEE Transactions on Computational Imaging, 2017, 3(1): 47-57.
- [13] 孟宪佳, 傅利平, 刘栋, 等. 高性能计算发展现状及其在文化遗产保护中的应用展望 [J]. 西北大学学报(自然科学版), 2021, 51(5): 807-815.
- MENG X J, FU L P, LIU D, et al. Development status of high performance computing and its application prospect in cultural heritage protection [J]. Journal of Northwest University (Natural Science Edition), 2021, 51(5): 807-815.
- [14] GATYS L A, ECKER A S, BETHGE M, et al. Texture synthesis using convolutional neural networks [C] // Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal; ACM, 2015: 262-270.
- [15] CHEN R F, HUANG W B, HUANG B H, et al. Reusing discriminators for encoding: Towards unsupervised image-to-image translation [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle; IEEE, 2020: 8165-8174.
- [16] YI Z L, ZHANG H, TAN P, et al. DualGAN: Unsupervised dual learning for image-to-image translation [C] // 2017 IEEE International Conference on Computer Vision. Venice; IEEE, 2017: 2868-2876.
- [17] KIM J, KIM M, KANG H, et al. U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation [EB/OL]. (2020-04-08) [2024-04-10]. <https://arxiv.org/abs/1907.10830v4>.
- [18] 张美玉, 刘跃辉, 秦绪佳, 等. 基于拉普拉斯算子抑制伪影的神经风格迁移方法 [J]. 计算机科学, 2020, 47(S2): 209-214.
- ZHANG M Y, LIU Y H, QIN X J, et al. Neural style transfer method based on Laplacian operator to suppress artifacts [J]. Computer Science, 2020, 47(S2): 209-214.
- [19] 茹超, 周延, 陈晓璇, 等. 一种面向文本图像的颜色迁移算法 [J]. 西北大学学报(自然科学版), 2017, 47(6): 815-821.
- RU C, ZHOU Y, CHEN X X, et al. A color transfer algorithm for text image [J]. Journal of Northwest University (Natural Science Edition), 2017, 47(6): 815-821.
- [20] PENG X L, PENG S L, HU Q Y, et al. Contour-enhanced CycleGAN framework for style transfer from scenery photos to Chinese landscape paintings [J]. Neural Computing and Applications, 2022, 34(20): 18075-18096.
- [21] ZHANG F Q, GAO H M, LAI Y P. Detail-preserving CycleGAN-AdaIN framework for image-to-ink painting translation [J]. IEEE Access, 2020, 8: 132002-132011.
- [22] IOFFE S, SZEGEDY C, PARANHOS L, et al. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02) [2024-04-10]. <https://arxiv.org/abs/1502.03167v3>.
- [23] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu; IEEE, 2017: 1132-1140.
- [24] NAH S, KIM T H, LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE, 2017: 257-265.
- [25] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity [J]. IEEE Transactions on Image Process-

ing, 2004, 13(4): 600-612.

[26] KINGMA D P, BA J, HAMMAD M M. Adam: A method for stochastic optimization[EB/OL]. (2017-01-30) [2024-04-10]. <https://arxiv.org/abs/1412.6980v9>.

[27] CHONG M J, FORSYTH D. Effectively unbiased FID and inception score and where to find them[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 6069-6078.

[28] TALEBI H, MILANFAR P. NIMA: Neural image assessment[J]. IEEE Transactions on Image Processing, 2018, 27(8): 3998-4011.

[29] 孙满利, 张景科. 文物保护学的理论探讨[J]. 西北大学学报(自然科学版), 2022, 52(2): 192-198.

SUN M L, ZHANG J K. Theoretical discussion on conservation of cultural heritages[J]. Journal of Northwest University (Natural Science Edition), 2022, 52(2): 192-198.

[30] HO J, JAIN A, ABBEEL P, et al. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2020: 6840-6851.

(编辑 李 静)