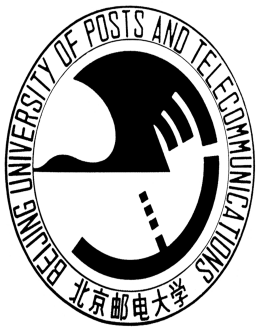


**本 科 毕 业 设 计（ 论 文 ）**



**题目: 基于能量引导扩散模型的快速图像风格迁移算法研究**

**姓 名 汤腾昊**

**学 院 人工智能学院**

**专 业 人工智能**

**班 级 2021219112**

**学 号 2021213526**

**指导教师 赵志诚**

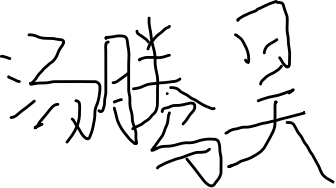
**2025年 5月**

**北 京 邮 电 大 学**

**本科毕业设计（论文）诚信声明**

本人声明所呈交的毕业设计（论文），题目《基于能量引导扩散模型的图像风格迁移算法研究》是本人在指导教师的指导下，独立进行研究工作所取得的成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。

申请学位论文与资料若有不实之处，本人承担一切相关责任。



本人签名： 日期： 2025.5.4

**关于论文使用授权的说明**

本人完全了解并同意北京邮电大学有关保留、使用学位论文的规定，即：北京邮电大学拥有以下关于学位论文的无偿使用权，具体包括：学校有权保留并向国家有关部门或机构送交学位论文，有权允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，有权允许采用影印、缩印或其它复制手段保存。汇编学位论文，将学位论文的全部或部分内容编入有关数据库进行检索。（保密的学位论文在解密后遵守此规定）

本人签名： 日期：

导师签名： 日期：

**基于能量引导扩散模型的快速图像风格迁移算法研究**

**摘要**

图像风格迁移作为计算机视觉中的一项重要任务，旨在将源图像的内容与目标风格图像的艺术风格进行融合，广泛应用于艺术创作、图像编辑、虚拟现实等领域。传统的风格迁移方法，尤其是基于卷积神经网络（CNN）和生成对抗网络（GAN）的方法，虽取得了较为显著的成果，但仍面临训练成本高，生成图像保真度低等问题。为了解决这些问题，本文设计并实现了一种基于能量函数引导扩散模型的风格迁移方法。本方法通过计算扩散中间步骤的去噪图像与风格参考图像的能量损失，并计算关于噪声特征图的梯度，将其作为每个扩散步骤的风格注入信息，从而引导扩散模型的生成过程。本文提出的方案能够让扩散生成模型无需重新训练便可应用于风格迁移任务。此外，本文还研究通过LoRA微调技术，以较低的训练计算成本进一步提高模型的风格迁移能力。实验表明，本方法在多个风格迁移任务中均取得了较为理想的效果，生成图像在风格融合、细节处理等方面均表现出优势，展示了扩散模型与LoRA技术在图像风格迁移中的潜力与应用前景。

**关键词** 扩散模型 图像风格迁移 LoRA微调

**Research on Fast Image Style Transfer Based on Energy-guided Diffusion Model**

**ABSTRACT**

Image style transfer, as a significant task in computer vision, aims to merge the content of a source image with the artistic style of a target style image. It has been widely applied in fields such as artistic creation, image editing, and virtual reality. Traditional style transfer methods, particularly those based on Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), have achieved notable results but still face challenges such as high training costs and low fidelity in generated images.

To address these issues, this thesis proposes and implements a style transfer method guided by energy functions within a diffusion model. The proposed method computes an energy loss between the denoised images from intermediate diffusion steps and the style reference image, and derives gradients with respect to the noise feature maps. These gradients are used as style injection signals at each diffusion step to guide the image generation process. This approach allows diffusion-based generative models to perform style transfer without requiring retraining.

Furthermore, this study explores the use of Low-Rank Adaptation (LoRA) fine-tuning techniques to enhance the style transfer capability of the model with minimal computational overhead. Experimental results demonstrate that the proposed method achieves desirable performance across various style transfer tasks, with superior results in style blending and fine detail preservation. These findings highlight the potential and applicability of diffusion models combined with LoRA in the domain of image style transfer.

**KEY WORDS**  Diffusion Models Image Style Transfer LoRA Fine-tuning

**目录**

【后续章节的名称根据我改的目录修改！】

[第一章 绪论 1](#_Toc197507149)

[1.1 研究背景与意义 1](#_Toc197507150)

[1.2 国内外研究现状 1](#_Toc197507151)

[1.3 本论文的主要研究内容与创新点 2](#_Toc197507152)

[1.4 论文结构安排 3](#_Toc197507155)

[第二章 图像风格迁移相关技术 4](#_Toc197507156)

[2.1 图像风格迁移技术 4](#_Toc197507157)

[2.1.1 神经风格迁移） 4](#_Toc197507158)

[2.1.2 基于生成对抗网络的风格迁移 5](#_Toc197507159)

[2.1.3 基于扩散模型的风格迁移 5](#_Toc197507160)

[2.2 扩散模型的工作原理与应用 6](#_Toc197507161)

[2.2.1 扩散模型基础 6](#_Toc197507162)

[2.2.2 Stable Diffusion模型 7](#_Toc197507163)

[2.2.3 扩散模型在图像生成中的优势 7](#_Toc197507164)

[2.3 本章小结 8](#_Toc197507165)

[第三章 基于能量函数的快速风格迁移 9](#_Toc197507166)

[3.1 基于能量函数引导的快速风格迁移框架 9](#_Toc197507167)

[3.2 预训练模型的引入 10](#_Toc197507168)

[3.2.1 Stable Diffusion 1.5模型 10](#_Toc197507169)

[3.2.2 CLIP特征编码器 10](#_Toc197507170)

[3.2.3 扩散控制结构 11](#_Toc197507171)

[3.3 能量函数的实现 11](#_Toc197507172)

[3.3.1 能量函数设计原理 11](#_Toc197507173)

[3.3.2 能量函数在扩散过程中的作用方式 11](#_Toc197507174)

[3.4 实验说明 12](#_Toc197507175)

[3.4.1 基于文本提示的风格迁移 12](#_Toc197507176)

[3.4.2 基于草图输入的风格迁移 12](#_Toc197507177)

[3.4.3 基于姿态图的风格迁移 13](#_Toc197507178)

[3.5 实验结果 13](#_Toc197507179)

[3.5.1 实验结果定量分析 13](#_Toc197507180)

[3.5.2 实验细节 15](#_Toc197507181)

[3.6 本章小结 16](#_Toc197507182)

[第四章 融合LoRA微调的风格迁移 17](#_Toc197507183)

[4.1 LoRA技术原理 17](#_Toc197507184)

[4.2 基于LoRA的模型微调方法 17](#_Toc197507185)

[4.3 微调后模型的风格迁移方法 19](#_Toc197507186)

[4.3.1 基于文本提示生成风格化图像 19](#_Toc197507187)

[4.3.2 基于已有图像的风格迁移 20](#_Toc197507188)

[4.3.3 推理过程中的关键可控参数 20](#_Toc197507189)

[4.3.4 最终输出 20](#_Toc197507190)

[4.4 实验结果 21](#_Toc197507191)

[4.4.1 实验结果的定量分析 21](#_Toc197507192)

[4.4.2 实验细节 22](#_Toc197507193)

[4.5 本章小结 23](#_Toc197507194)

[第五章 总结与展望 24](#_Toc197507195)

[5.1 工作总结 24](#_Toc197507196)

[5.2 不足与改进方向 24](#_Toc197507197)

[5.3 未来展望 24](#_Toc197507198)

[参考文献 26](#_Toc197507199)

[致谢 27](#_Toc197507200)

**第一章 绪论**

随着人工智能技术的飞速发展，图像生成与风格迁移成为计算机视觉领域的研究热点。尤其在文化创意、艺术生成、图像增强等应用场景中，图像风格迁移技术展现出广泛的应用前景。其核心目标是实现图像内容与艺术风格的融合，从而生成兼具结构保真度与艺术美感的图像作品。

近年来，深度学习推动了风格迁移方法的快速演进。从最初基于优化的神经风格迁移，到后来的生成对抗网络（GAN）方法，再到当前受到广泛关注的扩散模型（Diffusion Model），风格迁移的质量与效率均获得显著提升。然而，扩散模型在风格迁移中的实际应用仍面临建模复杂度高、推理效率低和风格控制稳定性差等问题。

针对上述挑战，本文设计并实现了一种结合能量引导机制与低秩微调（LoRA）技术的图像风格迁移方法。通过构建具有能量约束的生成路径，并引入轻量级参数调节策略，使得模型在保持生成质量的同时，大幅提升训练效率与推理速度，进而实现快速、高质量的图像风格转换。

**1.1 研究背景与意义**

图像风格迁移是计算机视觉领域的一个重要研究方向，其主要目标是将一幅图像的内容与另一幅图像的风格进行结合，从而生成具有目标风格的图像。随着深度学习的快速发展，图像风格迁移的方法也得到了显著的提升。传统的图像风格迁移方法（如基于卷积神经网络的神经风格迁移方法）在实现上往往需要较长的计算时间，并且在高分辨率图像上表现不佳。因此，如何提高风格迁移算法的效率和效果，成为了当前计算机视觉研究中的一个重要课题。

扩散模型作为一种生成模型，近年来在图像生成任务中取得了显著的成功。与生成对抗网络（GAN）不同，扩散模型通过逐步去噪的方式生成图像，具有生成质量高、训练稳定的特点。然而，现有的扩散模型在图像风格迁移中的应用仍面临计算复杂度高、风格迁移效果不稳定等挑战。

为了解决这些问题，本论文设计了一种基于能量引导扩散模型的快速图像风格迁移算法，旨在通过引入能量函数来引导扩散模型的生成过程，结合 LoRA 微调技术提高风格迁移效果，并通过高效的模型训练和推理策略，提升算法的计算效率。

**1.2 国内外研究现状**

图像风格迁移的研究始于Gatys[1]等人提出的神经风格迁移（Neural Style Transfer, NST）方法，该方法通过优化图像的内容和风格损失来实现图像的风格迁移 。尽管这种方法取得了一定的成功，但由于计算开销较大，且难以处理高分辨率图像，因此需要进一步的改进。

近年来，基于生成对抗网络（GAN）的风格迁移方法得到了广泛的研究。像CycleGAN【加引用】和StyleGAN【加引用】等方法，通过对抗训练生成具有特定风格的图像，并能够在无监督的条件下进行训练。然而，生成对抗网络在训练过程中存在不稳定性的问题，并且生成的图像可能会有较强的伪影。

扩散模型作为一种新的生成模型，近年来得到了广泛关注。扩散模型通过逐步去噪的过程生成图像，具有较好的生成质量和训练稳定性。特别是，Stable Diffusion等基于扩散模型的图像生成方法，在图像生成任务中取得了突破性的进展 。然而，扩散模型在图像风格迁移中的应用仍面临效率较低、风格效果不稳定等问题。

在参数的高效微调方面，Low-Rank Adaptation（LoRA）通过引入低秩矩阵对模型进行高效的微调，避免了对大规模模型权重的全面调整，能够在较少的计算开销下提升模型在特定任务上的性能。LoRA在图像生成任务中的应用逐渐被研究人员重视，并取得了显著的效果 。例如基于LoRA微调的 StyleGAN【加引用】，基于LoRA微调的超分辨率网络【加引用】，基于LoRA 微调的神经风格迁移【加引用】等。LoRA 为图像生成任务带来了更多的灵活性和高效性，成为了深度生成模型中不可忽视的一个技术工具。

**1.3 本论文的主要研究内容与创新点**

本论文的研究内容主要集中在基于扩散模型的图像风格迁移算法上，结合能量引导扩散模型和LoRA微调技术，设计并实现了一种兼顾效率与生成质量的图像风格迁移方法。

1.基于能量引导的扩散模型：设计了一种基于能量引导的扩散模型框架，利用能量函数引导扩散模型的图像生成过程，确保生成的图像符合目标风格。

2.LoRA微调技术的应用：将LoRA微调技术应用于扩散模型，通过低秩矩阵对模型进行微调，提高了风格迁移效果，同时避免了大规模参数调整带来的计算开销。

3.风格迁移算法优化：结合LoRA微调后的扩散模型和风格迁移算法，通过文本和图像的融合，提升了风格迁移的生成质量和计算效率。

4.算法实验与性能评估：通过对比实验，验证了本算法在生成效果、计算效率等方面的优势，并展示了其在多种风格迁移任务中的应用。

**1.4 论文结构安排**

本文各章节叙述内容概述如下：

第一章，绪论。介绍了图像风格迁移在计算机视觉中的重要性，特别是在艺术创作和视觉效果增强中的应用，指出传统方法存在计算效率低和内容保真性不足的问题。回顾了基于CNN和GAN的相关研究，指出现有方法的局限性。本文提出利用扩散模型结合低秩适配（LoRA）方法，通过能量函数引导和多条件协同控制，优化风格迁移效果与效率，精准控制图像内容和风格。最后，简要介绍了论文的结构安排。

第二章，图像风格迁移相关技术。回顾了图像风格迁移的技术，包括神经风格迁移、基于生成对抗网络的风格迁移、以及基于扩散模型的风格迁移，重点介绍了扩散模型的原理及其在图像生成中的应用。

第三章基于能量函数的快速风格迁移。提出了一种基于能量函数引导的快速风格迁移方法，结合Stable Diffusion模型、CLIP特征编码器，改进了风格迁移的效率和质量。

第四章，融合LoRA微调的风格迁移。详细介绍了LoRA技术在扩散模型中的引入，通过低秩适配微调扩散模型，提升风格迁移效果并显著减少计算开销。

第五章．总结与展望。总结了论文的主要贡献，并对未来研究方向进行展望，提出了进一步优化模型和拓展应用的可能性。

**第二章 图像风格迁移相关技术**

本章扼要地介绍图像风格迁移与扩散模型相关的核心技术，为后续模型设计与算法实现奠定理论基础。首先回顾图像风格迁移的发展历程，包括神经风格迁移方法、生成对抗网络在风格转换中的应用，以及近年来迅速兴起的基于扩散模型的风格迁移方法。其次，重点介绍扩散模型的工作原理、核心架构与训练机制，并进一步分析当前主流模型如Stable Diffusion在图像生成任务中的实际应用。

**2.1 图像风格迁移技术**

**2.1.1 神经风格迁移**

图像风格迁移最早由Gatys[1]等人于2015年提出，其核心思想是通过卷积神经网络（CNN）对图像的内容特征和风格特征进行分离和重组。具体来说，神经风格迁移（Neural Style Transfer, NST）方法通过预训练的卷积神经网络（如VGG-19），提取图像在不同层次的特征表示。内容图像的高层特征主要保留了物体的结构信息，而风格图像的低层特征则体现了纹理、颜色等局部信息。在训练过程中，算法通过定义两类损失函数：内容损失（Content Loss）和风格损失（Style Loss）。内容损失用于保持生成图像的结构与原始内容图一致，风格损失则通过最小化生成图像和风格图像的特征Gram矩阵之间的差异，使生成图像呈现目标风格。最终，优化目标是同时最小化这两种损失，从而实现图像风格迁移。尽管神经风格迁移开创了这一研究方向，但其存在显著缺点，例如生成速度慢、无法实时处理高分辨率图像，且在迁移复杂风格时容易失真。

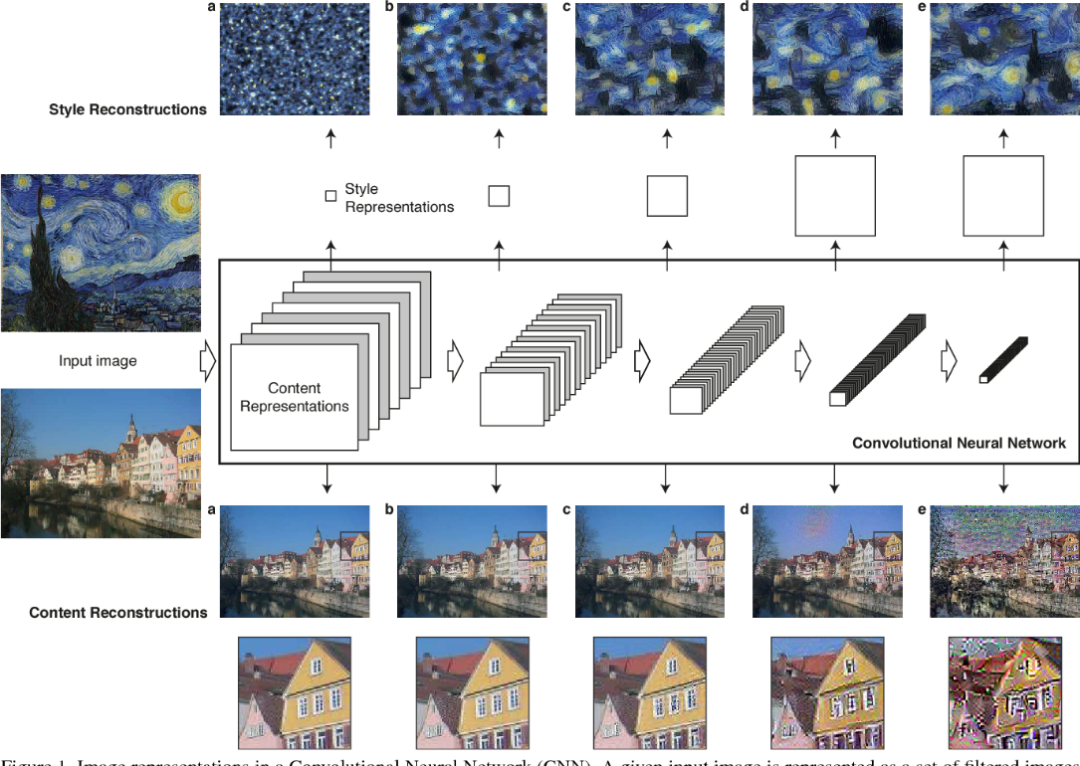


图2-1 NST模型结构[1]

(b)

(a)

**2.1.2 基于生成对抗网络的风格迁移**

为了解决神经风格迁移方法生成速度慢的问题，后续研究引入了生成对抗网络（GANs）以加速图像风格迁移。代表性方法包括：

1. CycleGAN[2]：提出了循环一致性损失（Cycle Consistency Loss），实现了无需成对数据的无监督图像到图像翻译，极大地拓展了风格迁移的应用范围。

2．Pix2Pix[3]：基于条件生成对抗网络（Conditional GANs），通过有监督学习方式，实现了高质量的图像转换，如图像着色、语义分割图到真实图像的转换。

GANs[4]在风格迁移中的应用显著提高了生成速度，并能在一次前向传播中完成风格迁移。此外，GAN能够更好地学习复杂的风格模式，生成高质量且风格一致的图像。然而，GANs在训练过程中常常面临模式崩溃（Mode Collapse）和训练不稳定等问题，这些问题导致生成结果在一定程度上缺乏多样性和稳定性，尤其在处理极复杂或细粒度风格时效果仍有待提升。

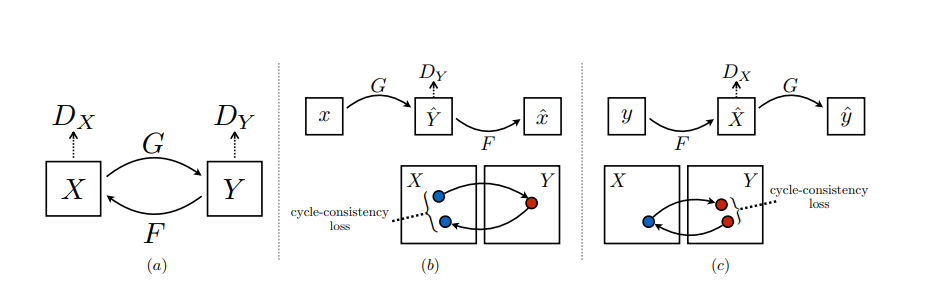


图2-2 CycleGAN模型原理[2]

**2.1.3 基于扩散模型的风格迁移**

近年来，扩散模型（Diffusion Models）在图像生成领域迅速崛起，代表方法如Stable Diffusion、Imagen、DALL-E 2[5][6]等。这类方法通过逐步地去除噪声的方式生成图像，与GANs不同，扩散模型具有更高的训练稳定性和生成质量。

在风格迁移任务中，扩散模型通过将输入条件（如文本描述、草图、轮廓图等）与噪声注入过程结合，实现了对生成图像风格的灵活控制。例如，Stable Diffusion在基本架构中引入了条件扩散过程，使得用户可以通过文字描述控制生成内容，同时应用特定风格。尽管基于扩散模型的风格迁移在生成质量上超越了传统方法，但由于其推理步骤繁多（通常需要数百步去噪），导致推理速度较慢，难以满足实时应用需求。此外，现有扩散模型在复杂风格的一致性保持方面仍存在一定挑战。

**2.2 扩散模型的工作原理与应用**

**2.2.1 扩散模型基础**

扩散模型（Diffusion Models, DMs）是一类基于概率图模型构建的生成模型，近年来在图像生成、图像编辑、风格迁移等领域取得了显著成果。其核心思想源于逐步扰动与逐步恢复的思想：首先将真实图像数据通过一系列加噪操作扰动至高斯噪声，然后通过训练一个反向神经网络模型逐步去噪，从而恢复出原始图像。

扩散模型包含两个阶段：前向扩散过程（Forward Process）和反向去噪过程（Reverse Process）。

前向过程将真实样本 在个步骤中逐渐加入噪声，最终得到几乎纯噪声的样本。这一过程定义为马尔可夫链，其每一步满足以下条件概率分布：

式（2-1）

【左对齐！】其中，∈(0,1) 是预设的噪声调度系数，控制每一步加入的噪声强度。经过 步之后，，接近各向同性高斯分布。

通过迭代展开，可以直接从原始图像构造任意时刻的加噪样本 ：

式（2-2）

其中为累计保持因子。

理想状态下，反向过程应为：

式（2-3）

但由于真实数据分布不可得，无法直接建模该过程，因此引入参数化神经网络（通常是UNet结构），近似学习一个反向采样分布：

式（2-4）

现代扩散模型往往简化为直接预测噪声 ，并以此反推出样本，形成所谓的“噪声预测模型”。整个采样过程通过多次迭代逐步还原噪声样本 至，恢复出目标图像。

扩散模型采用变分推断（Variational Inference）框架训练，其简化的训练目标为预测噪声误差的均方差：

式（2-5）

其中：ϵ∼N(0,I)为标准高斯噪声,为模型预测的噪声残差,为累计保持因子。

该损失函数本质上等价于最大化生成样本的对数似然，因此训练过程稳定、理论上具有收敛保证。

与GAN模型的对比，该模型具有更高的稳定性，无需对抗训练机制，从而避免了模式崩溃（Mode Collapse）的问题；在生成质量上，它在高分辨率细节生成方面优于传统的GAN模型；此外，它的灵活性更强，可以接入多模态条件（如文本、草图、姿态图等）进行控制生成，提供了更广泛的应用场景。

**2.2.2 Stable Diffusion模型**

Stable Diffusion[7]是当前最具影响力的扩散模型之一，由Stability AI发布。它在传统扩散模型基础上，引入了潜空间扩散（Latent Diffusion Models, LDM）策略，以显著降低计算资源消耗并加速推理。

Stable Diffusion模型架构主要包括以下几个部分。首先是编码器（VAE）[8]：使用变分自编码器（Variational Autoencoder）将高维图像压缩到潜在表示。这样可以在低维潜空间中进行扩散建模，大幅减少计算量。潜在扩散过程（Latent Diffusion）：在潜空间上进行前向加噪和反向去噪。具体而言，扩散过程在VAE编码后的紧凑特征空间中进行，不直接作用于高维像素空间，从而加快了训练和采样速度。条件机制（Conditioning）：为了实现可控生成，Stable Diffusion在扩散过程中引入了条件机制。最常见的是文本条件，通过使用CLIP文本编码器将自然语言提示嵌入潜空间，引导图像生成过程。此外，也支持姿态图、草图、深度图等视觉条件，扩展了模型应用范围。推理过程中，Stable Diffusion通过调节扩散步数、采样策略（如DDIM、PLMS等）实现生成质量与速度之间的灵活平衡。Stable Diffusion首次在开源社区广泛传播，极大推动了扩散模型在艺术创作、图像编辑、风格迁移等领域的应用。

**2.2.3 扩散模型在图像生成中的优势**

扩散模型在图像生成领域展现出多方面的优势，主要体现在以下几个方面。

高质量生成效果：扩散模型生成的图像通常细节丰富、纹理自然，能够逼真地还原复杂的场景结构，超越了GAN等传统生成方法，尤其在高分辨率图像生成方面效果显著。训练过程稳定：相较于GANs需要对抗优化（Generator与Discriminator相互博弈），扩散模型的训练基于最大化数据似然，无需引入对抗机制，因此避免了训练不稳定、模式崩溃（mode collapse）等常见问题。多模态条件控制能力强：扩散模型天然支持多种条件输入，如文本描述、姿态图、边缘图、深度图等，能够实现灵活、可控的图像生成。这使得扩散模型在风格迁移、图像编辑、图像到图像转换（Image-to-Image Translation）等应用场景中具有极大潜力。

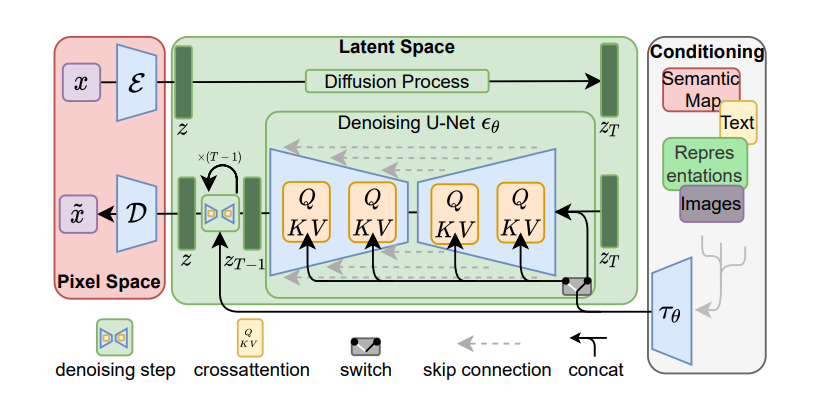


图2-3 Stable Diffusion模型原理[7]

**2.3 本章小结**

本章围绕图像风格迁移与扩散模型两大核心领域，系统性地回顾了其主要技术路径与关键演化过程。在图像风格迁移方面，从早期的神经风格迁移到基于生成对抗网络的方法，再到当前以扩散模型为基础的迁移策略，展现出技术从“优化驱动”向“数据驱动”的演进趋势。与此同时，扩散模型凭借其高生成质量、稳定训练机制以及强大的多模态条件控制能力，已成为图像生成领域的主流方法之一。

然而，扩散模型在风格迁移应用中仍存在推理效率低、风格控制稳定性差等问题，有待进一步优化。为此，后续将围绕LoRA微调与能量函数设计，介绍本论文所提出的风格迁移算法的模型架构与实现方法。

**第三章 基于能量函数的快速风格迁移**

尽管当前扩散模型在图像生成任务中表现出强大的能力，但其在风格迁移中的应用仍存在诸如采样效率低、风格一致性不足、条件控制粒度有限等问题。为提升风格迁移质量并缩短推理时间，亟需一种更高效、稳定且具备可控性的风格引导机制。

本章设计了一种结合能量函数引导与LoRA参数微调的快速图像风格迁移方法。该方法以Stable Diffusion模型为基础，融合CLIP编码器特征提取能力与能量引导机制，在采样过程中通过特征一致性损失动态调整生成轨迹，从而实现风格表达与内容保持的双重优化。系统通过引入预训练模型（如Stable Diffusion、CLIP）降低整体训练成本，并借助控制模块（ControlNet机制）增强输入引导能力，使得风格迁移过程具备更强的灵活性与生成质量。

此外，为进一步提高方法的适应性，本章还设计并实现了文本引导、草图引导与姿态图引导三类风格化图像生成流程，展示了该方法在不同输入控制条件下的表现能力，为后续实验部分的效果评估提供技术基础。

**3.1 基于能量函数引导的快速风格迁移框架**

本章设计的基于能量函数引导的快速图像风格迁移框架如图3-1所示。



图3-1 能量引导风格迁移结构图

具体来说，通过引入预训练的扩散模型及扩散控制结构在扩散采样过程中结合能量函数（Energy Function）对中间生成样本施加引导，促使图像在每一步采样时朝向目标风格特征收敛。

这种方法兼顾了生成质量与推理效率，能够在较少采样步数下实现风格高度一致的图像生成，同时显著降低了计算资源消耗，满足实际应用中对速度与质量的双重要求。

本方法的主要优势包括：高效性，通过减少扩散步数显著加速风格迁移过程；高质量生成，保持原始扩散模型高保真度的细节还原能力；风格一致性增强，利用能量引导动态优化生成方向，提高与目标风格的一致性；灵活性，能够适应多种输入控制条件，如姿态图、涂鸦图和文本描述等。

**3.2 预训练模型的引入**

为了保证风格迁移过程中生成图像的内容准确性、细节丰富性与风格特征保持能力，本章在方法设计中充分利用了现有强大的预训练模型资源，具体包括以下几个方面：

**3.2.1 Stable Diffusion 1.5模型**

本小节以Stable Diffusion 1.5版本为基础生成框架。Stable Diffusion是一种基于潜空间扩散（Latent Diffusion Models, LDM）的方法，其核心思想是：高分辨率图像首先通过变分自编码器（VAE）编码为潜空间表示，随后在这一潜空间中进行扩散建模，即通过加噪与去噪过程逐步生成潜空间样本，最后再由VAE解码器将采样完成的潜空间结果还原为最终图像。

由于潜空间维度远小于原始像素空间，潜空间扩散大幅降低了模型计算负担，并加快了推理速度，使高分辨率图像生成成为可能。引入经过大规模数据集预训练的Stable Diffusion模型，使本章无需从零开始训练扩散模型，避免了高昂的计算资源和时间开销，同时直接获得了优异的图像生成能力。这一设计大幅缩短了本章开发周期，为快速实现风格迁移系统提供了基础保障。

**3.2.2 CLIP特征编码器**

为了在扩散采样过程中定义和计算能量函数，本小节引入了OpenAI发布的CLIP（Contrastive Language-Image Pretraining）模型[9]。CLIP模型能够将图像和文本编码到统一的特征空间，提取高层次语义表示，适用于风格特征比较与相似性度量。在本节中，CLIP编码器的作用主要包括：对目标风格图像提取风格特征向量；对当前生成样本在采样过程中的中间图像提取特征向量；计算两者之间的特征距离，作为能量函数的基础，用于引导采样方向。这种基于深层特征的一致性度量，相较于传统像素空间差异（如L2距离），更加符合人类感知，能够更有效地实现风格迁移。通过直接利用预训练的CLIP模型，避免了重新训练特征提取器的过程，不仅保证了风格特征提取的一致性与可靠性，还进一步提升了整体系统的开发效率与运行速度。

**3.2.3 扩散控制结构**

在模型架构设计上，参考了ControlNet[10]提出的条件控制扩展机制：在原扩散UNet网络的基础上，增加了轻量级的条件支路（如姿态引导支路、草图引导支路）；将控制输入（如姿态图、scribble图）经过小卷积网络编码后，以残差方式注入到UNet主干的各层中；实现了在保持主干生成能力的同时，通过外部控制条件调整生成特征，使生成图更符合控制信号和风格要求。这种设计确保了在风格迁移过程中，不仅可以保持内容结构的准确性，还能根据不同的风格图引导生成多样化风格效果。

**3.3 能量函数的实现**

**3.3.1 能量函数设计原理**

为了在扩散采样过程中实现高效的风格迁移控制，本节在基础潜空间扩散模型（Stable Diffusion）上引入了基于CLIP编码器特征的一致性能量函数（Feature Consistency Energy Function）。具体而言，设定参考风格图像，当前扩散过程中某时刻生成的样本为，通过预训练的CLIP模型提取两者在语义特征空间中的表示，并定义能量函数衡量二者特征的一致性：

式（3-1）

此外，为了更细粒度地捕获风格结构信息，本节在特征提取阶段，进一步计算了CLIP中间层输出的Gram矩阵残差，并以此定义更加稳定的能量约束：

式（3-2）

其中，表示Gram矩阵操作，为Frobenius范数。Gram矩阵特征能够有效捕获风格中的纹理、局部相关性与整体排布信息。

最终，扩散过程中的总能量约束项可表示为：

式（3-3）

其中，与超参数，用于调节不同能量分量的权重。

**3.3.2 能量函数在扩散过程中的作用方式**

在标准的扩散去噪推理流程中，给定当前噪声图，模型预测其对应的噪声残差，并依此进行一步还原更新。

为了引入风格特征一致性约束，本节在每次扩散迭代步，在预测的过程中，进行以下改进。首先是计算能量梯度在采样步骤中（p\_sample\_ddim\_conditional函数​），以为变量，对能量函数进行反向传播，得到能量梯度。接着是修正采样结果，将标准采样得到的基于能量梯度进行修正，更新公式为：

式（3-4）

其中，为自适应步长因子，按无条件引导强度动态调整。最后是局部步长控制，能量引导仅在扩散中后期（如第70步至30步之间）启用，以确保早期主要建模图像内容，中后期逐渐迁移风格特征，避免生成初期过早引入风格扭曲。

通过上述机制，能量函数在扩散采样的每一步中持续起作用，引导生成结果在内容还原的基础上，逐步趋向于参考风格特征，实现自然且细腻的风格迁移效果。

**3.4 实验说明**

本节基于前一节方法分别实现了基于文本、草图以及姿态图的多条件图像生成。各生成流程在引入预训练模块的基础上，进一步结合特定的损失约束，有效提升了生成图像的内容准确性、风格一致性与细节丰富性。以下分别对不同生成流程进行详细阐述。

**3.4.1 基于文本提示的风格迁移**

在基于文本提示的图像生成任务中，首先通过加载预训练的Stable Diffusion 1.5模型，建立潜空间扩散建模体系。输入的文本提示（Prompt）经由文本编码器处理，映射至潜空间条件向量，作为引导扩散过程的语义指引。

为实现风格迁移效果，系统进一步引入了风格参考图像，并利用CLIP编码器提取其高层语义特征。采样过程中，定义特征能量损失项，计算当前生成样本与参考图特征之间的差异：

式（3-5）

其中， 表示CLIP编码器提取的特征向量，为当前采样样本，为参考风格图。

通过在每步扩散去噪过程中引入作为附加优化信号，系统在保持文本语义一致性的同时，实现了风格特征的渐进式迁移。最终，经VAE解码器还原潜空间样本至像素空间，输出符合语义与风格要求的高质量图像。在本流程中，Stable Diffusion承担了潜空间采样与高分辨率生成主导任务，CLIP编码器负责风格特征提取与引导，控制结构则通过隐式风格注入机制辅助扩散轨迹调整。

**3.4.2 基于草图输入的风格迁移**

在草图引导生成任务中，系统进一步引入显式的内容控制信号，以保障生成图像在布局结构上的准确性。

输入草图首先经过预处理，包括图像归一化、边缘提取与尺寸调整，形成标准化控制图（Scribble Map）。随后，控制图经轻量卷积网络编码后，作为残差特征注入至扩散模型各层特征流中，实现对生成过程的显式约束。

在潜空间扩散过程中，系统同时引入风格参考图特征，定义综合损失函数：

式（3-6）

其中，为原扩散过程的噪声预测损失；为风格特征一致性损失；衡量生成特征与控制图特征之间的一致性。控制约束 强化了生成图在局部结构上的符合性，使得最终输出图像能够严格遵循输入草图的轮廓布局，并在此基础上实现风格化渲染。

在本流程中，Stable Diffusion作为生成基础，CLIP编码器提供风格一致性约束，控制结构通过控制支路注入草图特征，确保内容结构保持。

**3.4.3 基于姿态图的风格迁移**

在姿态引导生成任务中，系统进一步拓展了控制结构的应用，以实现对人物动作姿态的精准约束。

输入的人像图首先通过OpenPose检测模块提取骨架姿态图，并作为控制条件输入。姿态控制图经过卷积编码后，以残差形式注入至Stable Diffusion模型各尺度特征层中，有效引导扩散过程中姿态信息的保持。

与前述流程一致，系统在采样过程中同时引入风格参考图特征，并定义能量损失进行风格一致性约束，同时引入控制一致性损失以衡量生成样本与目标姿态图的一致性。最终，潜空间采样完成后通过VAE解码器重建成像，得到符合输入动作姿态、文本提示语义与目标风格要求的图像结果。在本流程中，Stable Diffusion完成了扩散与生成主体任务，CLIP编码器用于实现风格迁移一致性优化，控制结构则确保了动作姿态信息的高保真度保持。

**3.5 实验结果**

所有实验在配备两张 NVIDIA GeForce RTX 3090 GPU 的服务器上进行，单卡显存为 24GB。系统安装的驱动版本为 550.127.05，CUDA 版本为 12.4

**3.5.1 实验结果定量分析**

实验安排：在3个prompt，利用3个风格图像通过Frobenius 范数，评估风格图像和生成图像的风格一致性。

式（3-7）

其中，是生成图特征和是参考图特征。

通过三个提示词“A cat wearing glass”，“A colorful photo of a tower”，“A fantasy photo of volcanoes”以及三个不同的风格参考图在Freedom下的Stable Diffusion模型与在UGD (Universal Guidance for Diffusion Models)【加引用】下的Stable Diffusion模型，得到如下图3-6，3-7，3-8的结果。图片上方是具体的提示词，图片下方的数值为计算 Gram 矩阵残差得到的 Frobenius 范数。

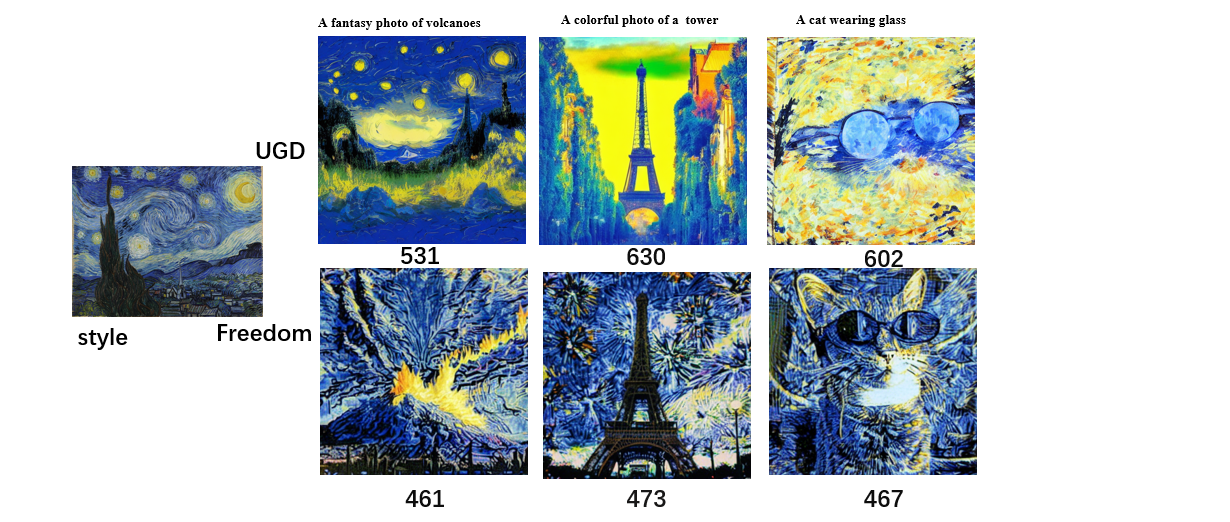


图3-2 效果对比图1

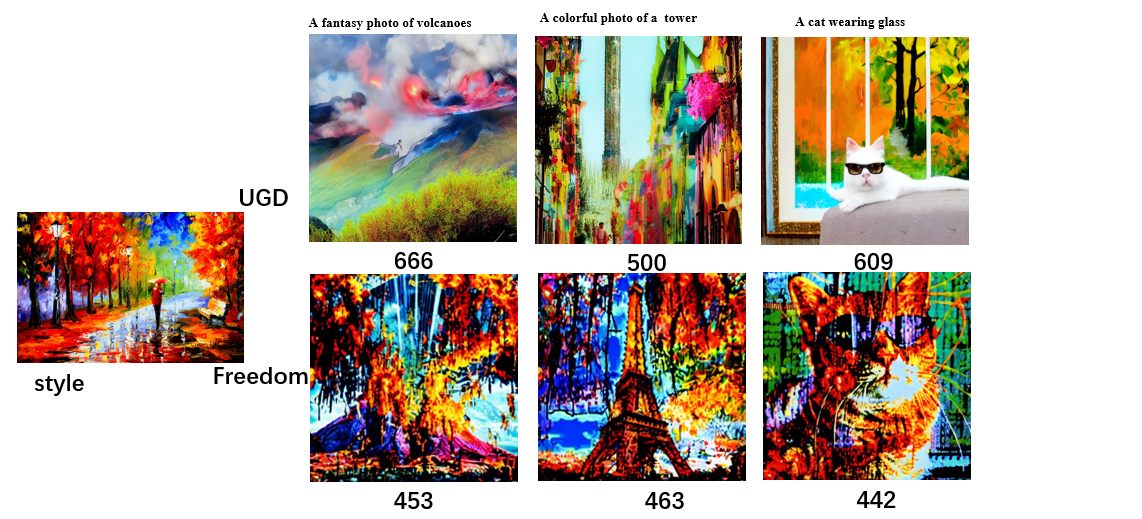


图3-3 效果对比图2

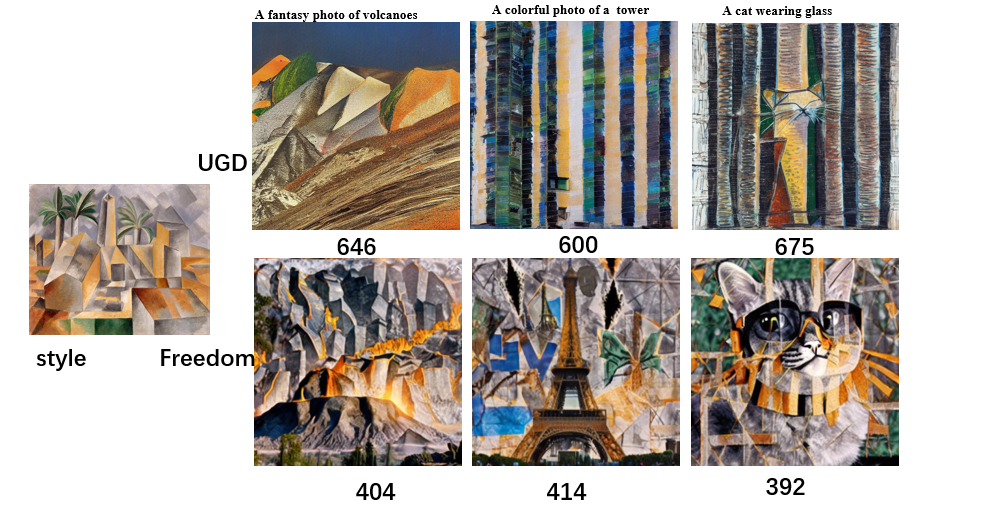


图3-4 效果对比图3

从给出的风格迁移结果图可以得出，与UGD方法相比，Freedom 方法在风格迁移的一致性和精度上表现较好，生成的图像能够较好地保持参考风格的细节和笔触特征。较低的 Frobenius 范数上表明该方法在风格迁移中的控制能力要强于UGD方法。

**3.5.2 实验细节**

首先本节以“戴眼镜的猫”作为目标图像，通过引入不同艺术风格的参考图，对目标图进行风格迁移。所得到的结果如下图3-5所示。

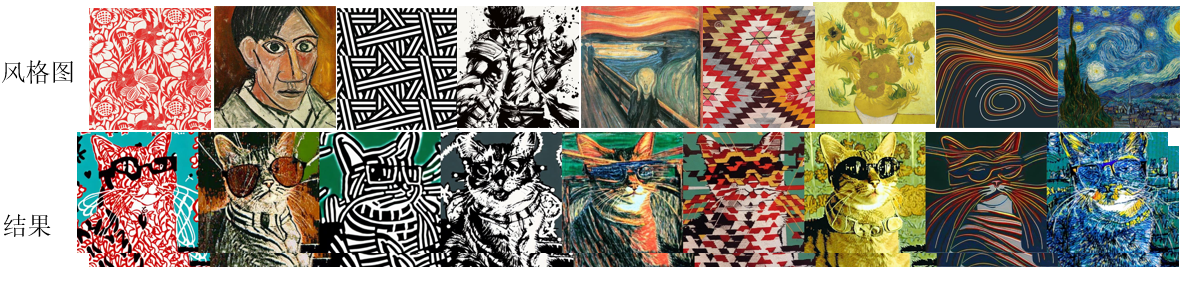


图3-5 各个风格下戴眼睛的猫

从图3-5可以看出，本方法较好地实现了风格迁移任务。生成图像在保留“戴眼镜的猫”语义结构的同时，成功融合了参考风格图的艺术特征，整体视觉风格统一。几何图形重构、黑白漫画和《星夜》风格的迁移效果尤为显著，而立体派肖像与《咆哮》在结构和笔触还原上相对较弱。总体而言，该方法在实现风格表达的同时有效保持了图像内容的一致性，展现出良好的适应性与生成质量。

引入CLIP多模态模型，提取图像的语义特征，将图像特征与文本特征对齐到同一嵌入空间，实现跨模态的特征比较和融合。并且通过引入可调节的融合系数α实现加权融合策略，平衡风格特征与文本特征的影响。其中α∈[0,1]控制风格特征的强度，当α=0时系统退化为原始文本到图像生成，当α=1时则完全由风格图像主导生成过程。各个系数下相同提示词下的风格迁移结果如下图3-6（α系数分别为0 0.1 0.3 0.5 0.7 0.9 1）。

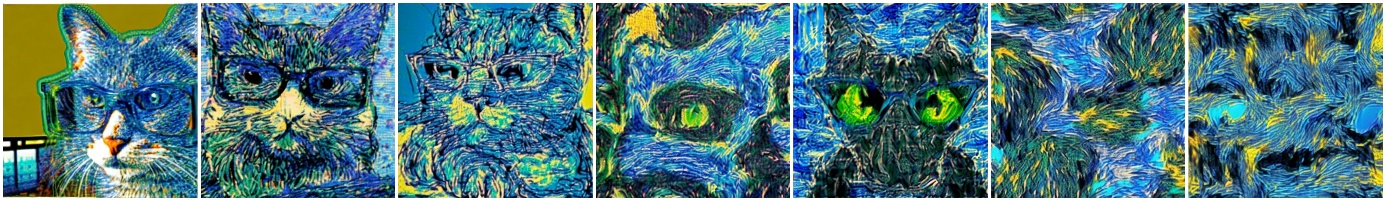


图3-6 不同融合系数α下梵高风格的戴眼睛的猫

基于草图输入的风格迁移和基于姿态图的风格迁移结果的展示如下图3-7，3-8所示。

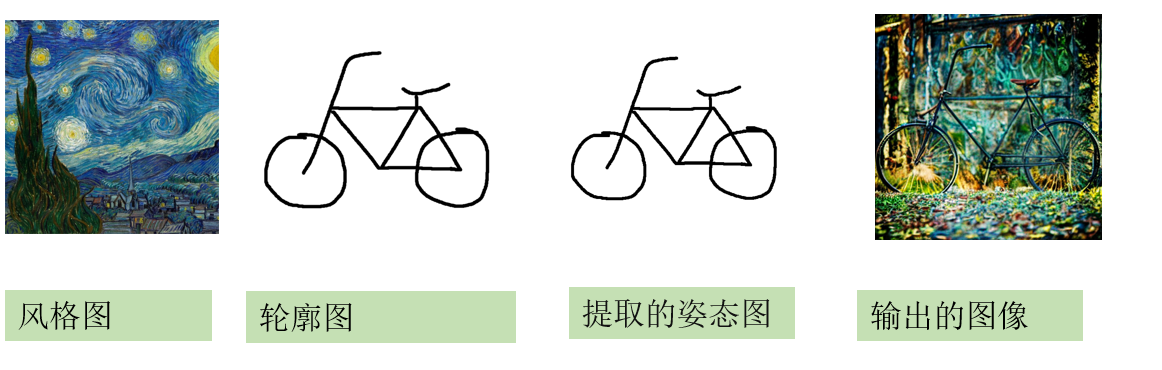


图3-7 基于草图输入风格迁移结果图

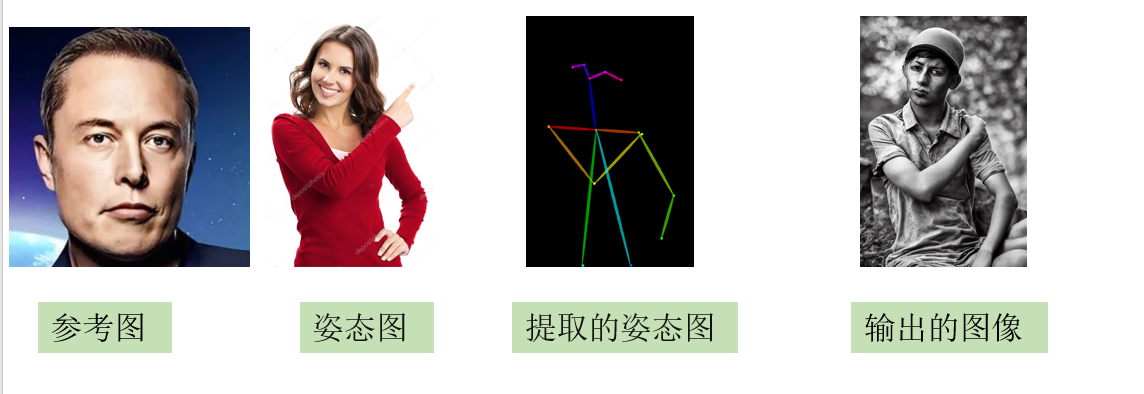


图3-8 基于姿态图的风格迁移结果

在“涂鸦图 + 风格图”任务中，模型以“自行车”为文本提示词，结合涂鸦图和《星夜》风格图生成图像。借助 ControlNet 引入的结构控制，生成图准确保留了自行车的几何轮廓，同时有效融合了《星夜》的典型笔触与色彩特征，整体风格统一、语义清晰。在“姿态图 + 身份图”任务中，模型根据姿态图生成准确的动作姿势，并通过身份图引导面部特征生成，图像在外观细节与真实感上表现良好。结果表明，ControlNet 能够显著提升结构控制效果，确保控制信息在生成图中的有效表达，支持复杂条件下的高质量图像合成。

**3.6 本章小结**

本章介绍了基于能量函数的快速图像风格迁移方法，详细阐述了核心技术路线与模块协同机制。通过引入Stable Diffusion作为基础扩散模型，并融合CLIP特征空间中的一致性损失设计，构建了能够在扩散过程中持续引导采样方向的能量函数。该机制有效增强了风格特征的传达力，同时避免了传统方法中常见的风格失真与细节丢失问题。此外，为进一步提高方法的适应性，本章还设计并实现了文本引导、草图引导与姿态图引导三类风格化图像生成流程，展示了该方法在不同输入控制条件下的表现能力，为后续实验奠定了基础。

**第四章 融合LoRA微调的风格迁移**

为解决扩散模型在图像风格迁移任务中存在的训练成本高、微调效率低和风格控制力不足等问题，本章重点介绍低秩适配（Low-Rank Adaptation, LoRA）技术在扩散模型中的引入与应用。LoRA作为一种轻量级微调方法，通过引入可训练的低秩增量矩阵，实现在冻结主干模型参数的前提下高效调整特定网络层的权重，从而显著降低微调计算资源开销并增强模型的任务适应性。

本章首先从LoRA方法的技术原理出发，介绍其在深度神经网络权重矩阵上的结构设计与更新机制。随后，结合本研究所采用的扩散模型架构，详细阐述了如何基于LoRA技术进行模型微调，包括损失函数设计、特征引导机制与参数优化策略。最后，本章还分析了微调后模型在图像生成任务中的具体推理流程，说明LoRA技术如何有效提升风格一致性与内容保真度，进而实现高质量的风格迁移效果。

**4.1 LoRA技术原理**

随着深度学习模型规模的不断扩张，传统的微调（Fine-tuning）方法逐渐暴露出高计算成本、高内存占用和易过拟合等问题。针对这一挑战，Hu等人提出了低秩适配（Low-Rank Adaptation，LoRA）方法[13]，旨在通过引入低秩矩阵的形式，对预训练权重进行高效调整，以降低微调开销并提升适应性。

在经典微调过程中，通常需要对模型的大部分甚至全部参数进行更新，假设一个权重矩阵为，则标准微调需要更新全部个参数。而LoRA的核心思想是：冻结原始权重，只引入一个可学习的低秩增量矩阵，并假设可以被两个小矩阵的乘积所近似表示，即：

式（4-1）

其中，、，远小于和（通常）。微调时仅需更新和两个小矩阵，从而大幅度减少了参数规模。更新后的权重为：

【缺公式】

其中，为缩放因子，用于平衡微调引入的扰动大小。

这种方法不仅有效减小了存储和计算负担，还因参数更新空间受到限制而在一定程度上抑制了过拟合风险，因而在大规模预训练模型的下游任务中展现出良好的迁移能力。

在扩散模型（Diffusion Models）场景下，LoRA可以应用于模型的UNet主干网络，通过低秩调整卷积层权重，快速适配特定风格或领域，无需大规模完整微调整个模型。

**4.2 基于LoRA的模型微调方法**

为了实现图像风格迁移任务，本研究以扩散生成模型为基础，结合LoRA机制对模型进行高效微调。微调后的风格注入模型的结构图如下图4-1。

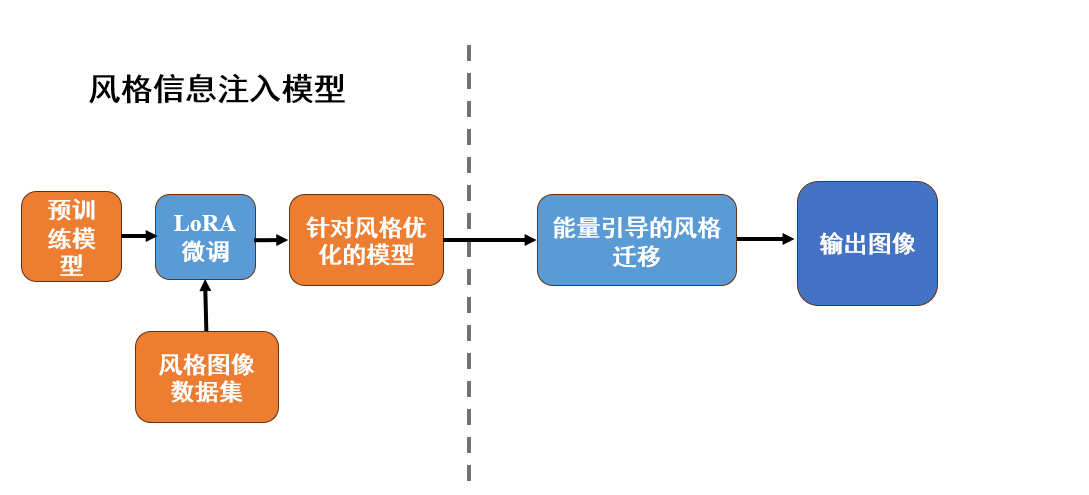


图4-1 风格信息注入模型结构

微调过程包括以下关键步骤：

首先，选取一批具有代表性风格特征的图像作为训练数据。每一张风格图像被送入编码器进行特征提取，并与扩散模型中间步骤的噪声预测结果进行对齐。训练的基本目标是最小化模型对噪声的预测误差，同时引导模型在特征空间内保持对目标风格的契合度。具体地，训练过程中使用了两个主要的损失函数。一是扩散重建损失，设扩散过程的噪声预测为，真实加噪后的样本为，则基本的扩散重建损失为：

式（4-2）

二是风格一致性损失，为了保证生成图像在风格特征上的接近性，引入基于CLIP编码器的特征损失。设表示输入图像在CLIP特征空间的嵌入，则风格损失包括两个部分，内容风格差异损失如式（4-3）和Gram矩阵差异损失如式（4-4）

式（4-3）

式（4-3）

其中，为风格参考图像，Gram矩阵可以刻画特征之间的相关性， 表示Frobenius范数。

最终综合损失函数为：

式（4-5）

其中，和为平衡各损失项的重要性权重。

微调策略上，冻结了原有扩散模型的全部主干参数，仅优化插入到UNet编码器各关键卷积模块中的LoRA分支参数。这种方式显著降低了训练时的计算负荷，并且由于参数更新量小，训练可以在单张或极少量风格参考图像下完成。

训练过程中还采用了逐步采样（progressive sampling）与噪声调制机制，提升模型对不同风格细节的感知能力。

**4.3 微调后模型的风格迁移方法**

在完成微调后，风格化图像的生成流程围绕微调后的扩散模型展开。生成过程主要包括模型加载、条件初始化、扩散推理、以及风格引导机制四个关键阶段。

首先，在推理阶段加载经过微调的模型权重。此时模型包含两部分参数：一是原始扩散模型的基础权重，用以保持内容信息及基本结构建模能力；二是微调过程中学习到的LoRA参数，作为低秩增量，动态注入至扩散模型内部指定层位，调整特定卷积运算的特征映射，从而在生成过程中注入目标风格特征。LoRA参数的注入遵循以下公式构建有效权重：

式（4-6）

其中，为冻结的原始权重，与分别为训练得到的低秩矩阵，为注入缩放因子，用以控制LoRA分支对输出特征的影响力度。

**4.3.1 基于文本提示生成风格化图像**

在基于文本生成任务中，首先输入一段自然语言文本描述[14][15]。该文本通过预训练文本编码器（CLIP文本编码器）进行特征提取，得到一组高维语义向量。该向量作为条件信息，与扩散过程中的潜变量进行融合，引导模型在去噪过程中朝向符合文本语义方向进行调整。

扩散过程自随机高斯噪声初始化，设初始噪声样本为。通过逐步反向去噪迭代，在每个时间步，模型预测噪声残差，其中为编码后的文本条件。采用采样器（DDIM采样器）根据预测残差进行逆扩散更新，在每一去噪步中，注入的LoRA参数动态影响特征变换，使得生成结果在整体风格上逐步向训练时学习到的目标风格靠拢。得到的结果如图4-2所示。



图4-2 莫奈风格的风景图

**4.3.2 基于已有图像的风格迁移**

在已有图像风格迁移任务中，输入为一张原始图像。首先通过图像编码器（如VAE编码器）将其编码为潜空间表征。这一潜表示包含了输入图像的结构与内容信息，同时具备紧凑的连续性，适合扩散建模。

在扩散推理阶段，以作为初始潜变量，通过条件去噪过程逐步生成最终图像。不同于完全随机噪声初始化的文本生成，这种以潜空间为起点的推理过程能够更好地保留原始图像的局部细节和内容结构。

为了实现风格转移，在每次去噪更新中引入风格特征引导。具体地，当前生成图像的潜变量通过解码器重建为图像，再输入CLIP图像编码器，提取中间特征。该特征与参考风格图像在特征空间中进行比较，并根据两者差异调整去噪方向，确保生成过程在保持原有内容基础上，逐步引入目标风格特性。在梵高风格以及写实派风格的画作上进行迁移的结果如下图4-3。



图4-3 莫奈风格迁移图像示例

**4.3.3 推理过程中的关键可控参数**

推理过程中可调控多个关键超参数，以进一步细化生成效果：

去噪步数（Inference Steps）：指扩散过程中的迭代次数，步数越多生成结果越细致，但推理时间增加。

LoRA注入强度（Injection Scale）：调整系数大小，控制LoRA分支对主干网络的影响程度。数值增大则风格特征更明显，但可能牺牲部分内容保真度。

**4.3.4 最终输出**

经过若干轮逐步去噪和条件引导后，最终生成的输出图像不仅在整体语义上符合输入文本或原图内容，还能够充分体现目标风格特征，如色调变化、笔触质感或局部结构变形，形成具有高审美一致性和内容保真的风格化图像。整个推理过程可概括为以下数学框架：

初始化潜变量（噪声或编码图），并进行条件编码（文本或图像特征）。通过多步条件去噪引导和CLIP特征空间风格约束，结合LoRA动态权重调整，最终将潜变量解码为输出图像。整体流程体现了潜空间建模、风格引导优化与动态权重适配的协同作用，从而实现了高效且高质量的图像风格迁移与生成。

**4.4 实验结果**

所有实验在配备两张 NVIDIA GeForce RTX 3090 GPU 的服务器上进行，单卡显存为 24GB。系统安装的驱动版本为 550.127.05，CUDA 版本为 12.4

**4.4.1 实验结果的定量分析**

实验安排:同一个提示词和3个风格图像在三个模型上的表现，通过Frobenius 范数，评估风格图像和生成图像的风格一致性。

式（4-8）

其中，是生成图特征和是参考图特征。

在相同的提示词下：“A  photo of a eiffel tower”，在Freedom下的Stable Diffusion模型与Lora微调后的模型以及在UGD (Universal Guidance for Diffusion Models)下的Stable Diffusion得到的结果如下图4-4所示。图片下方的数值为计算 Gram 矩阵残差得到的 Frobenius 范数。在相同的提示词的情况下及多各种风格下，微调后的模型无论是生成图片的质量，还是Frobenius 范数的数值上，都要明显优于其他两个模型或者方法。这说明通过特定的具有代表性风格特征的图像作为训练数据，将这些图像的特征信息通过Lora能够很好的注入到微调后的模型，对于特定风格的图片进行迁移的效果较好。

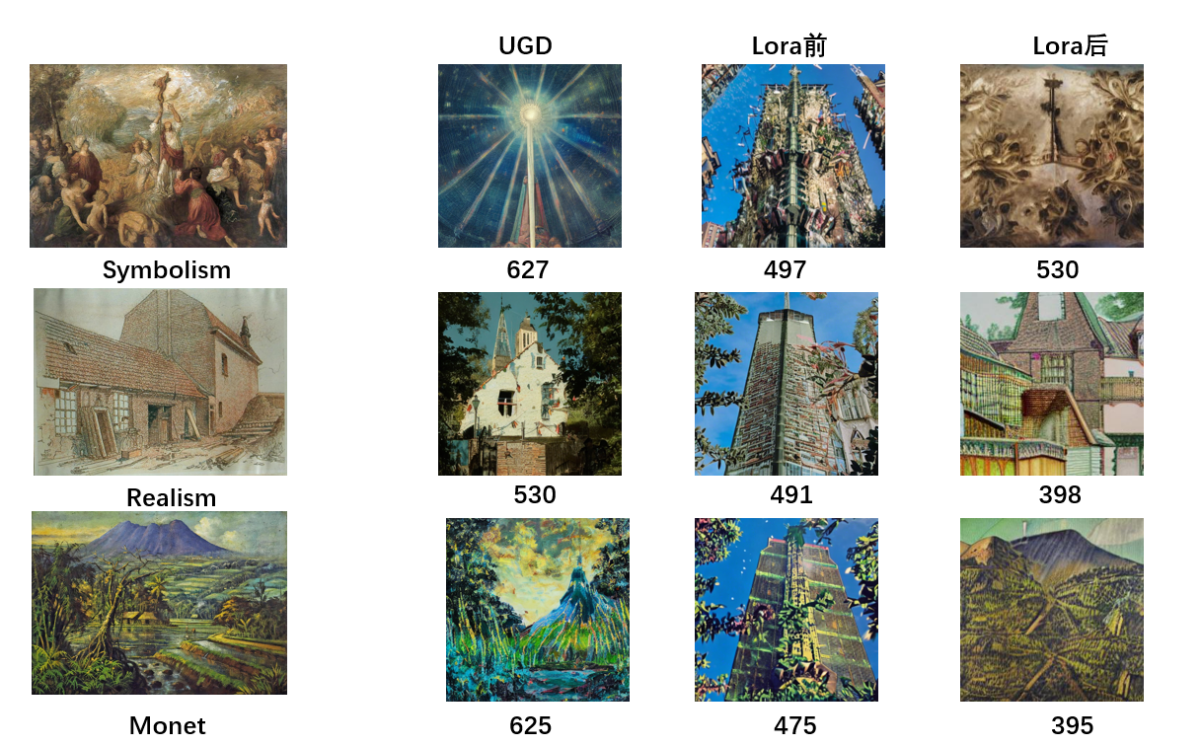


图4-4 微调前后模型效果迁移图像对比

**4.4.2 实验细节**

首先在提示词相同的情况下：A beautiful Monet style painting of a peaceful landscape，进行试验，使用不同的缩放因子系数微调得到模型或者在同一系数微调后的模型下，在不同的去噪步数呈现出不同的效果。针对不同的去噪步数和缩放因子系数进行实验，得到的结果图如下图4-5和4-6。



图4-5 不同去噪步数下（为16）的迁移图像



图4-6 不同系数下（步数为150）迁移图像

对实验结果有如下表4-1和4-2的分析

表4-1 不同步数下的效果分析

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **指标项** | **50步** | **100步** | **150步** | **200步** |
| **图像清晰度** | 图像较为模糊，细节缺失明显 | 图像较清晰，结构轮廓明确 | 图像细节表现良好，纹理丰富 | 图像非常清晰，纹理精细 |
| **风格表现力** | 风格特征初步显现，但不充分 | 风格特征较为明显，已有代表性笔触与色调 | 风格特征饱满，艺术感增强 | 风格特征表现最强，具有浓厚艺术感 |
| **内容保真度** | 内容结构简略，部分区域失真 | 内容表达基本完整，语义清晰 | 内容还原度较高，结构清晰自然 | 内容表达充分但略受风格压制 |
| **风格与内容融合性** | 风格和内容较为分离，边界不协调 | 风格与内容基本协调，融合度适中 | 风格与内容高度融合，过渡自然 | 风格略微主导，可能影响局部内容准确性 |
| **主观评价总结** | 图像整体质量较低，生成结果不稳定，风格迁移效果有限 | 风格与内容之间达到一定平衡，适合作为默认采样步数 | 整体生成质量较高，风格与语义协调，适合高质量生成任务 | 整体效果最佳，风格表现力最强，但存在一定过拟合趋势 |

表4-2 不同系数下的效果分析

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评估维度​​ | ​​图像1：蜿蜒河流 (α=4)​​ | ​​图像2：茂密植被 (α=8)​​ | ​​图像3：水域 倒影 (α=16)​​ | ​​图像4：树林草地 (α=32)​​ |
| 风格特征强度​​ | 中等（笔触初步破碎） | 强（色彩交融明显） | 极强（典型印象派光影） | 过强（细节丢失，抽象化） |
| 内容保真度​​ | 高（河流结构清晰） | 中（植被部分风格化） | 高（倒影物理合理） | 低（草地透视失真） |
| 色彩分布​​ | 淡蓝/浅紫主导，色彩较保守 | 绿色层次丰富，天空灰黄渐变 | 蓝白水面+多层级绿树 | 高饱和绿+云朵色块化 |
| 关键风格缺陷​​ | 天空光斑未完全风格化 | 前景叶簇笔触一致性不足 | 无 | 树木形态抽象过度 |

从这两个表格的定性分析可以得出在步数为150，系数为16时的既能兼顾时间效率，生成的图片质量也较好，能达到比较理想的风格迁移效果。

**4.5 本章小结**

本章围绕LoRA技术在图像风格迁移扩散模型中的引入与应用展开，从微调方法的技术背景入手，深入解析了其低秩参数化策略对模型可调性的提升作用。通过构建风格一致性与内容还原并重的损失体系，并在UNet主干中灵活注入LoRA分支，本节实现了一种训练高效、调节灵活的风格迁移模型。

在推理阶段，LoRA微调后的模型在保持原始扩散模型结构优势的基础上，结合CLIP特征引导机制和多模态控制信号，完成了多种输入条件下的风格化图像生成。最终实验验证显示，LoRA技术不仅显著压缩了训练时间和参数开销，还提升了生成图像在风格表达和内容结构保持上的整体质量。

**第五章 总结与展望**

**5.1 工作总结**

本论文针对当前图像风格迁移任务中存在的风格表达力不足、生成速度较慢以及模型训练成本较高等问题，提出了一种融合能量引导机制与LoRA微调策略的快速图像风格迁移方法。该方法以Stable Diffusion为生成核心，引入CLIP编码器构建风格特征一致性损失，设计能量函数引导扩散采样轨迹，同时结合LoRA技术实现低开销微调，显著提升了模型的风格迁移能力与推理效率。

论文的主要研究内容与成果如下：

构建了基于能量函数的图像风格引导机制，通过CLIP特征空间中的风格相似度度量，在扩散过程引入动态能量约束，有效增强了图像风格的一致性与表达力；

将LoRA技术引入扩散模型微调流程，在保持主干网络结构不变的前提下，仅通过低秩矩阵实现高效参数调节，显著降低了模型训练与部署的计算负担；

实现了三种条件下的风格迁移流程，包括文本提示、草图控制与姿态图控制，展现了所提方法在多模态输入条件下的通用性与扩展性；

通过实验验证了方法的有效性，在多个输入控制场景中均能生成具有较高风格一致性与内容保真度的图像结果，优于传统方法与基础扩散模型。

整体而言，本研究在保证生成质量的同时，显著提升了风格迁移的灵活性与高效性，为实际应用中的图像风格生成任务提供了一种可行且易扩展的解决方案。

**5.2 不足与改进方向**

尽管本论文所提出的方法在风格表达与计算效率方面取得了良好效果，但仍存在一些限制与改进空间，主要包括以下几点：

风格表征的丰富性仍有限：当前风格表示主要依赖CLIP提取的特征，在某些细粒度风格表达（如笔触、纹理动态）方面仍存在不足，后续可考虑融合多层特征或引入专门风格识别模型；

推理过程仍较为依赖经验性参数调整：如能量函数权重、LoRA注入强度等需要人工调节，未来可考虑引入自适应调节机制或学习型参数优化策略；

实验覆盖范围有限：本研究主要针对静态图像的风格迁移，未来可进一步扩展至视频风格迁移、三维风格重建等更具挑战性的任务场景中。

**5.3 未来展望**

针对当前工作中存在的不足与更广泛的应用需求，未来研究可从以下几个方向展开：

多模态风格融合机制：研究如何结合多源风格输入（如多幅风格图、文字+图像联合描述等），提升模型对复杂风格组合的适应能力；

风格迁移与编辑的统一框架构建：在风格迁移基础上加入风格编辑接口，实现用户可控的风格调节，如调整颜色、纹理强度、笔触方向等；

轻量化部署与移动端适配：结合模型剪枝、蒸馏等技术，探索在边缘设备或移动平台上的高效部署策略，使得风格迁移模型具备更强的实用性与普适性。

综上所述，本文所提出的方法为图像风格迁移任务提供了一个性能优异、结构灵活的研究范式，为后续多模态生成、可控生成以及艺术创作系统等方向的研究奠定了坚实的技术基础。

# 参考文献 【偏少，补一些到25篇左右】

1. Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]∥Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 2414–2423.
2. Zhu J Y, Park T, Isola P, Efros A A. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]∥IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 2223–2232.
3. Isola P, Zhu J Y, Zhou T, Efros A A. Image-to-image translation with conditional adversarial networks[C]∥IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 1125–1134.
4. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models[C]∥Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 10684–10695.
5. Ramesh A, Dhariwal P, Nichol A, et al. Hierarchical text-conditional image generation with CLIP latents[J/OL]. arXiv:2204.06125, 2022.
6. Zhang L, Agrawala M. Adding conditional control to text-to-image diffusion models[J/OL]. arXiv:2302.05543, 2023.
7. Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]∥Proceedings of the International Conference on Machine Learning (ICML). 2021: 8748–8763.
8. Hu E J, Shen Y, Wallis P, et al. LoRA: Low-rank adaptation of large language models[J/OL]. arXiv:2106.09685, 2021.
9. Song J, Meng C, Ermon S. Denoising diffusion implicit models[J/OL]. arXiv:2010.02502, 2020.
10. Kingma D P, Welling M. Auto-encoding variational Bayes[J/OL]. arXiv:1312.6114, 2013.
11. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]∥Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 4401–4410.
12. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]∥International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015: 234–241.
13. Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding[J/OL]. arXiv:2205.11487, 2022.
14. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train prompt tune: Towards efficient foundation model adaptation[J/OL]. arXiv:2110.07602, 2021.
15. Couairon G, Thoret G, Leclerc Y, et al. DiffEdit: Diffusion-based semantic image editing with mask guidance[J/OL]. arXiv:2210.11427, 2022.

# 致谢

在本次毕业论文的撰写过程中，我深切体会到科研道路上的点滴进步，离不开师长的悉心指导与身边人的无私帮助。首先，衷心感谢我的指导老师赵志诚老师，在选题、研究方法与论文撰写等各个阶段都给予了我耐心指导与宝贵建议，使我能够顺利完成本课题的研究工作。同时，也特别感谢学长蔡哲明在实验搭建与技术实现方面给予的耐心帮助与经验分享，让我在实际操作中少走了许多弯路。  
 再次向所有在我本科学习过程中给予帮助的老师与同学表示诚挚的感谢！