

同濟大學

TONGJI UNIVERSITY

毕业实训调研报告

课题名称 基于随机插值的高效轻量级图像修复

副标题 (如有)

学院 计算机科学与技术学院

专业 计算机科学与技术专业

学生姓名 张宇

学号 2251745

日期 2025 年 12 月 27 日

摘要

图像修复旨在从退化图像中恢复高质量的原始图像，是计算机视觉领域的基础性难题。近年来，以去噪扩散概率模型（DDPM）为代表的生成式方法在修复质量上取得了突破，但其迭代采样过程导致推理速度极慢，且庞大的参数量限制了其在移动端等资源受限场景的应用。为了解决这一效率瓶颈，基于随机插值（Stochastic Interpolation）的生成模型应运而生，它通过构建连接噪声与数据的直接概率路径，为实现“一步式”快速生成提供了理论基础。本文综述了图像修复技术的研究现状，重点分析了从传统方法、深度卷积网络到扩散模型及其加速算法的演进脉络，详细探讨了随机插值理论及轻量化网络设计的最新进展，并对未来高效图像修复技术的发展进行了展望。

关键词：图像修复，随机插值，轻量级网络，扩散模型，深度学习

ABSTRACT

Image restoration aims to recover high-quality original images from degraded observations and remains a fundamental challenge in the field of computer vision. In recent years, generative methods represented by Denoising Diffusion Probabilistic Models (DDPMs) have achieved breakthroughs in restoration quality. However, their inherent iterative sampling process results in extremely slow inference speeds, and the massive parameter count limits their deployment in resource-constrained scenarios such as mobile devices. To address this efficiency bottleneck, generative models based on Stochastic Interpolation have emerged. By constructing a direct probability path connecting noise distributions to data distributions, these models provide a theoretical foundation for achieving "one-step" rapid generation. This paper reviews the current research status of image restoration technologies, focusing on the evolutionary trajectory from traditional methods and deep convolutional neural networks to diffusion models and their acceleration algorithms. Furthermore, it discusses the latest advancements in stochastic interpolation theory and lightweight network design in detail, and concludes with an outlook on the future development of efficient image restoration technologies.

Key words: Image Restoration, Stochastic Interpolation, Lightweight Network, Diffusion Models, Deep Learning

目 录

1	引言.....	1
1.1	研究背景	1
1.1.1	图像修复的研究意义与技术演进	1
1.1.2	现有技术的瓶颈与随机插值理论的提出	1
1.2	研究难点	2
1.2.1	极少步采样下的生成质量保持	2
1.2.2	轻量化架构与特征表达的权衡	2
1.2.3	速度场学习的训练稳定性	2
2	现有方法概述.....	3
2.1	现有方法分类概览.....	3
2.2	传统图像修复方法.....	3
2.3	基于深度判别模型的修复方法.....	4
2.4	基于深度生成模型的修复方法.....	5
3	代表性方法.....	6
3.1	去噪扩散概率模型 (DDPM)	6
3.1.1	模型原理与架构.....	6
3.1.2	核心优势与局限性分析.....	6
3.2	一致性模型 (Consistency Models)	7
3.2.1	模型原理与创新点.....	7
3.2.2	与本课题的关联性分析.....	7
3.3	轻量化网络 MobileNet V2.....	7
3.3.1	模型原理与结构设计.....	7
3.3.2	优缺点及对图像修复的启示.....	8
4	结论与展望.....	9
4.1	研究现状总结.....	9
4.2	存在的问题.....	9
4.3	未来发展趋势与本课题研究展望.....	9
	参考文献.....	11

1 引言

1.1 研究背景

1.1.1 图像修复的研究意义与技术演进

图像修复旨在从受损或退化的观测数据中恢复出潜在的高质量原始图像，是计算机视觉与图像处理领域中经典的病态反问题。在实际应用中，由于成像设备本身的物理限制、拍摄环境的复杂性以及信号传输过程中的干扰，图像往往会遭受噪声污染、运动模糊、分辨率降低或局部信息缺失等退化影响。这些退化不仅降低了图像的视觉质量，更严重阻碍了医学影像分析、自动驾驶环境感知、安防监控目标检测等下游高级视觉任务的性能。因此，研究鲁棒且高效的图像修复算法，对于提升成像系统的环境适应能力和智能化水平具有重要的学术价值与现实意义。

纵观图像修复技术的发展历程，大体经历了从基于物理模型和统计先验的传统方法，到基于深度卷积神经网络(CNN)的判别式模型，再到如今以扩散概率模型(Diffusion Models)为代表的深度生成式模型的演进。早期的传统方法如 BM3D、中值滤波等，主要依赖图像的局部平滑性或自相似性先验，虽然计算开销较小，但在处理复杂的非线性退化时往往力不从心，难以恢复逼真的纹理细节。随后的 CNN 方法（如 U-Net、ResNet 变体）通过大规模监督学习建立了从退化图像到清晰图像的端到端映射，显著提升了修复指标（PSNR/SSIM），但由于使用像素级回归损失函数，生成的图像往往存在过度平滑、高频细节丢失的问题，即出现“回归均值”现象。

1.1.2 现有技术的瓶颈与随机插值理论的提出

近年来，去噪扩散概率模型（DDPM）凭借其卓越的分布建模能力，成功解决了 CNN 模型生成细节不足的问题，在去噪、修复、超分辨率等任务上取得了当前的 SOTA（State-of-the-Art）效果。然而，扩散模型在带来高质量生成的同时也引入了极为严峻的效率瓶颈。传统的扩散模型依赖于模拟逆向的马尔可夫链过程，通常需要数百甚至上千步的迭代采样才能生成一张图像，这导致其推理速度极为缓慢，难以满足实时性要求。此外，为了维持高精度的生成能力，主流扩散模型往往拥有数十亿级的参数量，对 GPU 显存和计算算力要求极高，极大地限制了其在移动端、嵌入式设备等资源受限场景中的应用。

针对上述“高质量与低效率”的矛盾，基于随机插值（Stochastic Interpolation）的生成模型作为一种新兴范式应运而生。与强制依赖物理扩散过程的传统方法不同，随机插值理论允许构建连接噪声分布与数据分布的任意概率路径，并通过训练神经网络直接预测该路径上的速度场。这种灵活性在理论上使得模型能够摆脱繁琐的迭代采样，实现“一步式”或“少步式”的高质量图像生成。本课题“基于随机插值的高效轻量级图像修复算法研究”正是基于这一前沿背景，旨在结合随机插值理论的高效采样优势与轻量化网络设计的结构优势，探索一

种兼具高性能与低算力消耗的通用图像修复解决方案。

1.2 研究难点

1.2.1 极少步采样下的生成质量保持

本课题的首要挑战在于如何在将推理步数压缩至极致（如一步或数步）的同时，依然维持媲美迭代式扩散模型的生成质量。在传统扩散模型中，复杂的生成过程被分解为大量微小的去噪步骤，单步预测难度较低。而基于随机插值的一步生成模型需要网络在单次前向传播中，精准地跨越从纯噪声分布到复杂自然图像分布的巨大鸿沟。这种从高熵状态到低熵状态的剧烈跳变，极易导致模型无法准确拟合数据流形，从而在生成结果中产生边缘模糊、结构扭曲或纹理丢失等现象。因此，如何设计最优的插值路径以及速度场匹配策略，使模型在极少步数下仍能捕捉到图像的高频细节与长距离语义依赖，是理论层面必须攻克的难点。

1.2.2 轻量化架构与特征表达的权衡

架构设计的核心难点在于如何在大幅降低模型参数量与计算复杂度（FLOPs）的前提下，保证网络具备足够的特征提取与重建能力。现有的高性能修复模型通常依赖大宽度的卷积层或深层的注意力机制（Self-Attention）来捕捉全局信息，这并不适用于资源受限的移动端环境。本课题需要引入深度可分离卷积、通道混洗等轻量化算子重构骨干网络，但这些操作往往会削弱网络的特征表达能力和感受野。如何在极度压缩模型体积的同时，设计有效的特征融合与增强模块（如轻量级注意力机制），防止因模型容量下降导致的修复精度损失，是在工程实现层面面临的重大挑战。

1.2.3 速度场学习的训练稳定性

相较于传统的像素级回归任务，基于随机插值的速度场学习具有更高的优化难度。该类模型的训练本质上是求解高维空间中的常微分方程（ODE）或随机微分方程（SDE）的近似解。轻量级网络由于参数空间较小，其拟合复杂非线性速度场的能力相对较弱，在训练过程中极易出现梯度消失、梯度爆炸或损失函数难以收敛的震荡现象。此外，如何设计合理的混合损失函数，平衡像素级一致性（L1/L2 Loss）、感知质量（Perceptual Loss）以及分布匹配（Matching Loss）之间的关系，以引导轻量级模型在复杂的优化景观中稳定收敛，也是本研究需要解决的关键问题。

2 现有方法概述

2.1 现有方法分类概览

图像修复作为计算机视觉领域的经典问题，其研究历史跨越了数十年，积累了丰富的理论成果与技术看案。纵观该领域的发展脉络，现有的图像修复方法依据其核心原理、数学模型及优化目标的不同，大体上可归纳为三大类：传统图像修复方法、基于深度判别模型的修复方法以及基于深度生成模型的修复方法。

第一类是传统图像修复方法。这一阶段的研究主要建立在信号处理、凸优化理论以及图像统计学先验的基础之上。其核心思想是利用图像自身的物理特性（如局部平滑性、稀疏性、自相似性）来构建从退化图像到清晰图像的逆问题求解模型。这类方法无需大规模数据训练，具有良好的数学解释性，但在处理复杂的非线性退化时表现有限。

第二类是基于深度判别模型的修复方法。随着深度学习技术的兴起，以卷积神经网络（CNN）为代表的驱动方法逐渐占据主导地位。这类方法通常将图像修复视为一个端到端的监督回归问题，通过学习从退化域到清晰域的非线性映射函数来实现修复。为了解决回归模型导致的图像模糊问题，生成对抗网络（GAN）随后被引入，通过对抗训练引入感知层面的约束。这类方法推理速度较快，但在生成多样性和极端退化下的鲁棒性方面存在瓶颈。

第三类是基于深度生成模型的修复方法。这是近年来发展最为迅猛的前沿方向，主要以去噪扩散概率模型（Diffusion Models）及其高效采样变体为代表。该方法不再局限于学习点对点的映射，而是致力于显式地建模高质量图像的概率分布。凭借强大的分布拟合能力，生成式模型在纹理还原和细节生成上取得了突破性进展。然而，高昂的计算成本与漫长的推理时间是该类方法面临的主要挑战，这也催生了基于随机插值等高效采样理论的最新研究。本章后续部分将对这三类方法的基本原理、代表性工作及优缺点进行详细阐述。

2.2 传统图像修复方法

在深度学习技术兴起之前，图像修复领域的研究主要建立在信号处理理论、数学变分法以及图像统计学先验的基础之上。这一阶段的方法虽然无法像现代神经网络那样从海量数据中学习复杂的语义特征，但其严谨的数学推导和无需训练的特性，使其在特定场景下依然具有参考价值，并为后续基于学习的方法提供了理论基石。

早期的图像修复主要采用基于滤波和扩散的算法。这类方法的直观思想是利用图像的局部平滑性，假设受损像素与其邻域像素具有相似的灰度或色彩值。例如，中值滤波和维纳滤波通过统计局部窗口内的像素信息来抑制高斯噪声或椒盐噪声，计算简单且速度极快，但在处理高强度噪声时容易导致图像边缘模糊。基于偏微分方程（PDE）的方法则将图像视为连续的曲面，利用热扩散方程或各向异性扩散方程（如 Perona-Malik 方程）将周围的信息平滑地传播到受损区域。虽然这类方法能较好地保持明显的边缘结构，但在面对大面积纹理缺失时，往往只能产生平滑的插值结果，无法恢复出具有意义的纹理细节。

为了克服局部方法的局限性，研究者们提出了基于非局部自相似性（Non-local Self-similarity）的算法，其中最具代表性的是 BM3D（Block-Matching and 3D Filtering）算法。BM3D 的核心原理是在整幅图像范围内寻找相似的二维图像块，将这些相似块堆叠成三维矩阵，并在变换域（如小波变换或离散余弦变换域）进行协同滤波。这种方法充分利用了自然图像中广泛存在的重复结构（如草地、建筑纹理），在去噪性能上达到了传统算法的巅峰。此外，基于稀疏表示（Sparse Representation）的方法也是这一时期的主流方向，如 K-SVD 算法。该类方法假设自然图像块可以由一个过完备字典中的少量原子线性组合而成，通过求解稀疏编码系数来重建清晰图像。尽管传统方法在数学上具有良好的解释性，但其由于缺乏对高层语义的理解，难以处理复杂的非线性退化（如严重的运动模糊或人脸遮挡），且迭代优化过程通常计算耗时较长。

2.3 基于深度判别模型的修复方法

随着卷积神经网络（CNN）在计算机视觉任务中的成功应用，基于深度学习的图像修复方法逐渐取代传统算法成为主流。这类方法通常被建模为判别式任务，即直接学习从退化图像域到清晰图像域的非线性映射函数。根据网络结构和优化目标的不同，该类方法主要可以分为基于 CNN 的回归模型和基于生成对抗网络（GAN）的模型。

基于 CNN 的回归模型是最早期的深度修复尝试，其基本范式是通过构建深层网络，以最小化预测图像与真实图像之间的像素级误差（如 L1 或 L2 损失）为优化目标。经典的 SRCNN 模型首次将三层卷积网络引入超分辨率任务，证明了端到端学习相比传统插值方法的优越性。随后，DnCNN 模型引入了残差学习（Residual Learning）策略，通过预测残差图像（即噪声分布）而非直接预测清晰图像，显著降低了训练难度并提升了收敛速度。为了进一步扩大感受野并整合多尺度特征，U-Net 架构被广泛应用于图像修复。U-Net 采用编码器-解码器结构，通过下采样提取抽象语义特征，再通过上采样逐步恢复空间分辨率，同时利用跳跃连接（Skip Connection）将底层的高频细节直接传递到解码器，有效解决了深层网络中的信息丢失问题。尽管基于 CNN 的方法在 PSNR 等客观指标上表现优异，且推理速度较快（通常只需一次前向传播），但由于其使用的像素级损失函数倾向于对所有可能的解求平均，导致生成的图像往往过于平滑，缺乏逼真的高频纹理，这种现象被称为“回归均值”问题。

为了解决图像模糊和纹理缺失的问题，生成对抗网络（GAN）被引入图像修复领域。基于 GAN 的方法（如 Pix2Pix、SRGAN）通过引入一个判别器网络与生成器网络进行博弈训练。判别器的目标是区分生成图像与真实图像，而生成器的目标是欺骗判别器。这种对抗性损失实际上是对生成图像分布与真实图像分布之间的距离进行了约束，迫使生成器输出更符合自然图像统计规律的结果，从而产生锐利的边缘和丰富的纹理。然而，GAN 模型的训练极不稳定，容易出现模式崩塌（Mode Collapse）现象，即生成器只能生成有限种类的样本。此外，GAN 生成的图像虽然视觉效果逼真，但在像素级准确性（PSNR）上往往不如纯 CNN 方法，且容易产生与原图不符的虚假伪影（Hallucination）。

2.4 基于深度生成模型的修复方法

近年来，深度生成模型，特别是去噪扩散概率模型（DDPM），在图像合成与修复领域引发了范式转移。与判别式模型直接学习映射关系不同，生成式模型旨在显式地建模数据的概率分布，从而具备极强的样本生成多样性和高质量细节还原能力。目前主流的研究主要集中在扩散模型及其快速采样变体上。

扩散模型的核心思想受到非平衡热力学的启发，包含前向扩散过程和逆向去噪过程。在前向过程中，模型通过逐步向真实图像添加高斯噪声，将其转化为纯噪声分布；在逆向过程中，通过训练一个神经网络（通常是 U-Net 结构）来预测每一步添加的噪声，从而实现从纯噪声逐步还原回清晰图像。DDPM 通过变分推断导出了简单的均方误差损失函数，使得训练过程极其稳定，避免了 GAN 中的对抗博弈难题。由于扩散模型能够精确拟合复杂的数据分布，其在图像去噪、超分辨率和填补等任务上生成的图像细节极其丰富，视觉真实感达到了前所未有的高度。然而，DDPM 最大的缺陷在于其推理效率极其低下。标准的 DDPM 采样过程需要模拟完整的马尔可夫链，通常包含 1000 个离散的时间步，这意味着生成一张图像需要进行 1000 次网络前向推理，计算成本高昂，完全无法满足实时应用需求。

为了缓解扩散模型的速度瓶颈，研究者们提出了多种加速策略。DDIM（Denoising Diffusion Implicit Models）通过将扩散过程重新表述为非马尔可夫过程，允许在采样时跳步，从而将采样步数从 1000 步压缩至 50-100 步，但仍难以达到实时水平。最近，基于“一致性模型”（Consistency Models）和“随机插值”的研究开始受到关注。这类方法试图打破迭代去噪的限制，通过学习从任意噪声状态直接映射到数据起点的函数，或者构建连接噪声与数据的直线传输路径，在理论上支持单步或极少步采样。本课题所采用的“随机插值”理论正是这一方向的最前沿探索，它为解决扩散模型的高延迟问题提供了坚实的数学基础。

3 代表性方法

阐述代表性方法的基本原理本章将选取与本课题密切相关的三个代表性方法进行深入剖析。首先介绍去噪扩散概率模型（DDPM），它是当前生成式图像修复技术的核心基石；其次详细阐述一致性模型（Consistency Models），它是基于随机插值思想实现高效采样的前沿代表；最后分析轻量化网络 MobileNet V2，探讨其如何通过结构创新降低计算复杂度。通过对这三个方法的原理、架构及优缺点的解构，为本课题的算法设计提供理论依据。

3.1 去噪扩散概率模型（DDPM）

3.1.1 模型原理与架构

去噪扩散概率模型（Denoising Diffusion Probabilistic Models, DDPM）是一种基于非平衡热力学原理的生成模型，其核心思想是通过模拟一个逐渐破坏数据分布的前向扩散过程和一个逐渐恢复数据分布的逆向去噪过程来生成高质量图像。DDPM 的数学框架主要包含两个离散时间的马尔可夫链。

在前向扩散过程中，模型定义了一个固定的噪声添加策略。给定真实图像数据分布 x_0 ，模型在每个时间步 t 向图像中添加微量的高斯噪声，直到时刻 T ，图像 x_T 近似为各向同性的标准高斯分布。这一过程可以由状态转移方程描述，且得益于高斯分布的特性，任意时刻 t 的噪声图像 x_t 可以直接从 x_0 采样得到，无需逐步迭代，极大地简化了训练数据的构建过程。

逆向去噪过程是 DDPM 的核心，旨在训练一个参数化的深度神经网络（通常采用基于 U-Net 的架构）来逆转上述扩散过程。由于精确的逆向后验分布在数学上是难以处理的，DDPM 利用变分推断将其近似为高斯分布，并训练神经网络来预测每一步加入的噪声分量。在生成阶段，模型从纯高斯噪声 x_T 出发，通过神经网络预测并减去噪声，逐步迭代 T 次（通常为 1000 次），最终还原出清晰图像 x_0 。这种迭代式的生成方式使得模型能够逐步细化图像的粗糙结构和精细纹理，从而生成具有极高保真度的结果。

3.1.2 核心优势与局限性分析

DDPM 最大的优势在于其卓越的生成质量和训练稳定性。与生成对抗网络（GAN）不同，DDPM 不需要训练不稳定的判别器，其优化目标简化为预测噪声的均方误差（MSE），使得训练过程平滑且不易出现模式崩塌（Mode Collapse）。这意味着 DDPM 能够更好地覆盖整个数据分布，生成样本的多样性极高，特别适合处理具有不确定性的图像修复任务（如大面积遮挡补全）。

然而，DDPM 的局限性同样显著，主要体现在极低的推理效率上。由于其生成过程必须严格遵循逆向马尔可夫链的时间步序，生成一张图像通常需要进行上千次神经网络的前向推理。这导致其推理时间比 GAN 慢两到三个数量级，在消费级 GPU 上生成单张高分辨率图

像往往需要数秒甚至更长时间，完全无法满足实时交互或视频流处理的需求。这也正是本课题试图通过随机插值理论解决的核心痛点。

3.2 一致性模型 (Consistency Models)

3.2.1 模型原理与创新点

一致性模型 (Consistency Models) 是近年来针对扩散模型采样速度慢这一问题提出的颠覆性解决方案，其理论基础与本课题关注的随机插值思想紧密相关。该模型的核心创新在于提出了一种新的生成范式，旨在打破传统扩散模型必须逐步去噪的限制，实现单步或极少步的高质量生成。

一致性模型的设计灵感来源于概率流常微分方程 (PF-ODE) 解的轨迹特性。在扩散模型的理论中，数据到噪声的变换过程可以看作是高维空间中沿特定轨迹的平滑移动。一致性模型的核心思想是学习一个“一致性函数”，该函数能够将 PF-ODE 轨迹上的任意一点（无论是纯噪声还是中间状态）直接映射回轨迹的起点（即清晰图像）。根据自一致性属性，对于同一条轨迹上的任意两个时刻点，该函数的输出应当是相同的。

基于这一思想，一致性模型引入了“一致性蒸馏” (Consistency Distillation) 和“一致性训练” (Consistency Training) 两种策略。前者利用预训练的扩散模型作为教师网络来指导学习，后者则直接从数据中单独训练。通过优化一致性损失函数，模型被强制要求在相邻时间步的预测结果保持一致。这种设计使得模型在推理阶段不再需要逐步求解微分方程，而是可以在任意时刻直接“跳跃”到终点，从而在理论上支持一步生成 (One-step Generation)。

3.2.2 与本课题的关联性分析

一致性模型是随机插值理论在高效生成领域的成功实践。它证明了通过构建特定的映射关系，可以在不牺牲太多生成质量的前提下，将推理步数从上千步压缩至一步或几步。这一特性与本课题“高效”的目标高度契合。然而，现有的一致性模型大多基于庞大的模型架构（如大型 Transformer 或深层 U-Net），虽然步数减少了，但单次推理的计算量依然巨大，难以在移动端运行。本课题将借鉴一致性模型的路径映射思想，并将其与轻量化网络设计相结合，探索更加极致的效率优化方案。

3.3 轻量化网络 MobileNet V2

3.2.1 模型原理与结构设计

MobileNet V2 是 Google 团队提出的轻量化卷积神经网络，专为移动端和嵌入式视觉应用设计，其核心理念对于本课题构建轻量级修复网络具有重要的参考价值。MobileNet V2 在第一代的基础上，引入了“倒残差结构” (Inverted Residuals) 和“线性瓶颈” (Linear Bottlenecks) 两个关键创新，显著提升了模型在低计算资源下的特征表达能力。

传统的残差块（Residual Block）通常采用“降维-卷积-升维”的沙漏型结构，即先通过 1×1 卷积压缩通道数，再进行 3×3 卷积，最后恢复通道数。这种设计虽然减少了计算量，但在通道数较少时容易丢失信息。MobileNet V2 反其道而行之，提出了倒残差结构，即“升维-卷积-降维”。首先利用 1×1 卷积将低维特征扩展到高维空间（通常扩展倍率为 6），然后在高维空间进行深度可分离卷积（Depthwise Convolution）提取特征，最后再通过 1×1 卷积投影回低维空间。这种设计基于一个直觉：高维空间能够编码更丰富的流形信息，而在通过激活函数（如 ReLU）时，低维特征容易丢失信息，高维特征则能保留更多细节。

此外，MobileNet V2 去除了输出层的非线性激活函数，直接输出线性结果，即所谓的“线性瓶颈”。研究表明，在低维流形上使用 ReLU 等非线性激活函数会造成不可逆的信息损耗，而线性层则能更好地保留特征的完整性。

3.2.2 优缺点及对图像修复的启示

MobileNet V2 的优点在于其极高的参数效率和推理速度。深度可分离卷积将标准卷积的计算量降低了约一个数量级，而倒残差结构保证了在参数受限的情况下依然具备强大的特征提取能力。这使得它非常适合作为图像修复任务中的特征提取骨干（Backbone）。

然而，直接将 MobileNet V2 用于图像修复也存在不足。原版网络是为图像分类设计的，经过多次下采样后，空间分辨率大幅降低，导致位置信息和纹理细节丢失。而图像修复任务对像素级的精度要求极高。因此，在本课题中，我们需要对 MobileNet V2 进行适应性改造，例如减少下采样倍率、引入跳跃连接或增加针对图像重建任务的解码器结构，使其在保持轻量化的同时，能够满足图像修复对高频细节恢复的需求。

4 结论与展望

4.1 研究现状总结

本文围绕图像修复技术的研究现状进行了系统的调研与梳理。纵观该领域的发展历程，可以清晰地看到技术范式从基于手工先验的传统方法向基于数据驱动的深度学习方法演进的脉络。早期的滤波与块匹配算法虽然奠定了理论基础，但在处理复杂非线性退化时显得力不从心。随着深度卷积神经网络（CNN）的兴起，判别式模型通过端到端的监督学习显著提升了修复精度，但其固有的回归均值特性导致了生成图像的纹理细节缺失。近年来，以去噪扩散概率模型（DDPM）为代表的生成式模型凭借强大的分布拟合能力，成功解决了纹理平滑问题，将图像修复的视觉质量推向了新的高度，成为当前学术界的研究热点。

然而，现有的 SOTA 技术在迈向实际应用的过程中，依然面临着“高质量与低效率”之间难以调和的矛盾。扩散模型虽然生成效果惊人，但其依赖数千步迭代采样的推理机制导致了极高的时间延迟，且其庞大的参数量对硬件算力提出了苛刻要求。尽管已有研究提出了 DDIM、一致性模型等加速策略，以及 MobileNet 等轻量化设计思想，但如何将两者有机结合，构建一个既能实现“一步式”快速生成，又能在移动端低功耗运行的高性能修复框架，仍是当前领域亟待解决的空白与挑战。

4.2 存在的问题

基于上述调研，目前图像修复领域主要存在以下两个亟待突破的技术瓶颈：

首先是生成式模型的推理延迟问题。尽管一致性模型等前沿工作尝试将采样步数压缩至单步，但这些方法大多仍基于参数量巨大的重型网络架构（如 Large Vision Transformer）。在实际部署中，即便步数减少了，单次推理的浮点运算量（FLOPs）依然巨大，难以在算力受限的边缘设备上实现真正的实时响应。

其次是轻量化网络在生成任务中的性能衰减问题。现有的轻量化设计（如深度可分离卷积）主要针对分类或检测等判别式任务优化，其核心逻辑是压缩冗余信息。然而，图像修复是一个信息重建过程，对高频纹理和微小细节极其敏感。直接将现有的轻量化骨干应用于生成式模型，往往会导致严重的特征丢失，使得修复后的图像出现模糊、伪影或结构扭曲。

4.3 未来发展趋势与本课题研究展望

展望未来，图像修复技术的发展趋势必然是朝着“高效化、轻量化、通用化”的方向演进。一方面，基于随机插值（Stochastic Interpolation）等新型生成理论的探索将进一步打破物理扩散过程的束缚，为设计任意速度场和概率路径提供数学工具，从而实现更极致的采样效率；另一方面，结合神经架构搜索（NAS）与硬件感知的模型压缩技术，将推动生成式模型真正落地于手机、VR/AR 设备等终端。

基于此，本课题“基于随机插值的高效轻量级图像修复算法研究”拟从以下三个维度展开深入探索：

第一，构建基于随机插值的高效生成框架。利用随机插值理论构建连接噪声与数据的直接概率路径，设计非迭代式的速度场预测网络，旨在实现“一步式”图像修复，彻底解决传统扩散模型的耗时痛点。

第二，设计适配生成任务的轻量级骨干网络。借鉴 MobileNet V2 的倒残差结构与 ShuffleNet 的通道混洗思想，重构修复网络的特征提取模块。我们将探索如何在大幅降低参数量的同时，通过引入轻量级注意力机制或多尺度特征融合模块，增强网络对图像纹理细节的捕捉能力，平衡模型体积与修复质量。

第三，探索多任务通用的修复能力。我们将利用 Places2、CelebA-HQ 等大规模数据集，研究如何通过单一模型处理去噪、去模糊、图像补全等多种退化类型，提升模型的泛化能力与实际应用价值。

通过上述研究，本课题期望能够开发出一款兼具学术创新性与工程实用价值的图像修复算法，为解决生成式模型在资源受限场景下的落地应用提供切实可行的技术方案。

装

订

线

参考文献

- [1] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.
- [2] Song Y, Dhariwal P, Chen M, et al. Consistency models[J]. 2023.
- [3] Albergo M S, Vanden-Eijnden E. Building normalizing flows with stochastic interpolants[J]. arXiv preprint arXiv:2209.15571, 2022.
- [4] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684-10695.
- [5] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [6] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [7] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.
- [8] 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. Cham: Springer international publishing, 2015: 234-241.
- [9] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(2): 295-307.
- [10] Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising[J]. IEEE transactions on image processing, 2017, 26(7): 3142-3155.
- [11] Zamir S W, Arora A, Khan S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5728-5739.
- [12] Liang J, Cao J, Sun G, et al. Swinir: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1833-1844.
- [13] Song J, Meng C, Ermon S. Denoising diffusion implicit models[J]. arXiv preprint arXiv:2010.02502, 2020.
- [14] Luo S, Tan Y, Huang L, et al. Latent consistency models: Synthesizing high-resolution images with few-step inference[J]. arXiv preprint arXiv:2310.04378, 2023.
- [15] Liu X, Zhang X, Ma J, et al. InstafLOW: One step is enough for high-quality diffusion-based text-to-image generation[J]. arXiv preprint arXiv:2309.06380, 2023.
- [16] Lipman Y, Chen R T Q, Ben-Hamu H, et al. Flow matching for generative modeling[J]. arXiv

preprint arXiv:2210.02747, 2022.

- [17] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [18] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.
- [19] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on image processing, 2007, 16(8): 2080-2095.
- [20] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries[J]. IEEE Transactions on Image processing, 2006, 15(12): 3736-3745.
- [21] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2536-2544.
- [22] Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5505-5514.
- [23] Kong D, Li F, Wang Z, et al. Dual prompting image restoration with diffusion transformers[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 12809-12819.
- [24] Hu J K, Jin L, Yao Z, et al. Universal image restoration pre-training via degradation classification[J]. arXiv preprint arXiv:2501.15510, 2025.
- [25] Serrano-Lozano D, Herranz L, Su S, et al. Adaptive Blind All-in-One Image Restoration[J]. arXiv preprint arXiv:2411.18412, 2024.
- [26] Zhao H, Tian L, Xiao X, et al. AverNet: All-in-one video restoration for time-varying unknown degradations[J]. Advances in Neural Information Processing Systems, 2024, 37: 127296-127316.
- [27] Li X, Wang Z, Chen C, et al. SemID: blind image inpainting with semantic inconsistency detection[J]. Tsinghua Science and Technology, 2024, 29(4): 1053-1068.
- [28] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.
- [29] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.
- [30] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.