



# 南京大學

## 本科毕业论文

院 系 物理学院

专 业 物理

题 目 基于扩散模型与多源数据融合  
的显微图像超分辨率重建方法

研究

年 级 2021 学 号 211840339

学生姓名 刘梓航

指导教师 李文飞 职 称 教授

提交日期 2025 年 5 月 30 日



# 南京大学本科生毕业论文（设计、作品）中文摘要

题目：基于扩散模型与多源数据融合的显微图像超分辨率重建方法研究

院系：物理学院

专业：物理

本科生姓名：刘梓航

指导教师（姓名、职称）：李文飞 教授

摘要：

光学显微镜因受制于约 200 纳米的衍射极限，难以满足对亚细胞结构如蛋白质复合体、细胞器动态行为等的高分辨率观测需求。尽管传统超分辨成像技术（如 STED、STORM）在空间分辨率方面取得了突破，但其高光毒性、高硬件依赖性及成像速度瓶颈等问题，限制了其在活细胞成像及临床场景中的广泛应用。

近年来，深度学习在图像处理中的广泛应用为显微图像的重建带来了新机遇。以 CNN、GAN 为代表的数据驱动方法在提高图像分辨率、降低硬件依赖方面展现出良好性能，成为超分辨重建的重要技术路径。然而，现有方法在图像高频细节恢复能力、泛化性与稳定性方面仍存在显著挑战。为此，本文引入了当前在自然图像生成领域表现突出的扩散模型（Diffusion Model），并设计了一套适用于显微图像重建的评估与训练框架，系统探讨其在实际生物成像任务中的表现。

本文的研究对象包括模拟数据与真实实验数据（荧光小球、纤毛），围绕“重建质量、高频细节还原、伪影控制及模型泛化能力”展开三组实验：

首先，在预实验中，通过控制变量法对模型结构（CNN vs Diffusion）、损失函数（MSE vs JS'DIV）与归一化范围  $([0, 1] \text{ vs } [-1, 1])$  进行对比，结果表明 CNN 模型具有训练高效的优点，而 DIFFUSION 模型在细节还原和感知质量上更优；采用 JS'DIV 作为损失函数可使模型收敛速度提升近一倍，性能略优或持平；归一化为  $[-1, 1]$  可使训练速度提升约 20%，且收敛性更佳。接着，在实验一中，我们对比了 CNN 与 Diffusion 在模拟数据集上的超分效果，相比 CNN 模型，DIFFUSION 模型平均 PSNR 高 1.9dB，MS-SSIM 提升高 13%，且几乎无伪影产生，显示其在高频细节重建与抗失真方面的强大能力。最后，在实验二中，我们验证了模型对真实图像的泛化与迁移能力。在未微调的情况下，Diffusion 在荧光

小球和纤毛图像上的表现已优于 CNN，结构还原更加清晰，伪影更少；在加入 100 对真实纤毛图像微调后，模型整体性能进一步增强，Diffusion 模型在 SSIM 上提升 23%，CNN 提升约 9%。

本文的系列实验验证了 Diffusion 在图像重建质量与跨域泛化上的优势，并通过多源数据融合训练策略表明：在缺乏大规模真实标注数据的情形下，结合模拟数据训练与少量真实数据微调可实现高质量图像重建，为超分辨成像提供了一条高效高质的可行路径。

**关键词：**超分辨重建；扩散模型；卷积神经网络；训练策略；损失函数；归一化范围；图像质量评估；多源数据融合；高频细节还原

# 南京大学本科生毕业论文（设计、作品）英文摘要

THESIS: Super-Resolution Microscopy Image Reconstruction via Diffusion Models and Multi-Source Data Fusion

DEPARTMENT: Department of Physics

SPECIALIZATION: Physics

UNDERGRADUATE: Liu Zihang

MENTOR: Professor Li Wenfei

ABSTRACT:

Optical microscopy is fundamentally constrained by the diffraction limit of approximately 200 nanometers, making it challenging to observe subcellular structures such as protein complexes and dynamic organelle behaviors at high resolution. Although traditional super-resolution imaging techniques (e.g., STED, STORM) have made breakthroughs in spatial resolution, their high phototoxicity, strong hardware dependence, and imaging speed limitations restrict their broader application in live-cell imaging and clinical scenarios.

In recent years, the wide adoption of deep learning in image processing has brought new opportunities to microscopic image reconstruction. Data-driven methods, represented by CNNs and GANs, have demonstrated promising capabilities in enhancing image resolution and reducing reliance on specialized hardware, thus becoming key technical pathways in super-resolution reconstruction. However, existing methods still face significant challenges in restoring high-frequency image details, ensuring generalization, and maintaining stability. To address these issues, this study introduces the diffusion model—an approach that has recently shown remarkable performance in natural image generation—and designs a dedicated training and evaluation framework tailored for microscopic image reconstruction. The study systematically investigates the model's performance in real-world biological imaging tasks.

This work focuses on both simulated and real experimental datasets (fluorescent beads and cilia), conducting three groups of experiments centered on reconstruction quality, high-frequency detail restoration, artifact control, and model generalization ca-

pability.

We sequentially conducted pilot experiments, deep learning-based super-resolution experiments, and multi-source data fusion experiments to comprehensively evaluate the proposed model’s performance under varying conditions. First, in the pilot experiments, we employed controlled comparisons to analyze model architecture (CNN vs. Diffusion), loss function (MSE vs. JSDIV), and normalization range ( $[0, 1]$  vs.  $[-1, 1]$ ). Results show that the Diffusion model significantly improves SSIM and MS-SSIM by approximately 8.7% and 11.2%, respectively, compared to CNN, under similar MAE/NRMSE, indicating superior structural fidelity and perceptual quality. Using JSDIV as the loss function nearly doubles the convergence speed, with equal or slightly better performance; normalization to  $[-1, 1]$  enhances training speed by about 20% and results in better convergence. Next, in Experiment I, we compared the super-resolution performance of CNN and Diffusion models on simulated datasets. Diffusion showed clear advantages, with average PSNR improvement of 2.1 dB and MS-SSIM increase exceeding 10%, while effectively eliminating artifacts—demonstrating strong capabilities in high-frequency detail restoration and distortion resistance. Finally, in Experiment II, we tested the generalization and transferability of the Diffusion model on real-world images. Without fine-tuning, Diffusion already outperformed CNN in reconstructing fluorescent beads and cilia images, delivering clearer structures and fewer artifacts. After fine-tuning with just 100 paired cilia images, the model’s performance further improved: Diffusion’s SSIM increased by over 6.5%, compared to CNN’s 3.2% increase. These experiments collectively validate the Diffusion model’s advantages in training efficiency, image quality, and cross-domain generalization, and suggest that—when large-scale labeled real data is lacking—combining simulated training data with minimal real data fine-tuning offers a practical and efficient path toward high-quality image reconstruction in super-resolution microscopy.

**KEYWORDS:** Super-resolution reconstruction; Diffusion model; Convolutional neural networks; Training strategy; Loss function; Normalization range; Image quality assessment; Multi-source data fusion; High-frequency detail restoration

# 目 录

中文摘要	I
ABSTRACT	III
目 录	V
插图目录	IX
表格目录	XI
符号表	XIII
<b>第一章 绪论</b>	<b>1</b>
1.1 研究背景与意义 . . . . .	1
1.2 研究现状简介 . . . . .	2
1.2.1 超分辨显微成像技术 . . . . .	2
1.2.2 机器学习单图像超分辨技术 . . . . .	3
1.2.3 机器学习超分辨显微技术 . . . . .	5
1.3 本研究特点 . . . . .	6
1.4 本文章节安排 . . . . .	7
1.5 本章小结 . . . . .	8
<b>第二章 相关技术介绍</b>	<b>9</b>
2.1 问题设定与理论 . . . . .	9
2.1.1 傅里叶变换视角下的单透镜成像过程 . . . . .	9
2.1.2 衍射极限与成像系统缺陷引起的分辨率退化 . . . . .	10
2.1.3 不同程度干扰对成像质量的影响 . . . . .	11

2.1.4	清晰图像的还原 . . . . .	13
2.2	实验数据集介绍 . . . . .	13
2.2.1	模拟数据集 pySTED . . . . .	13
2.2.2	真实实验数据集 . . . . .	14
2.2.3	数据规范化处理 . . . . .	14
2.3	模型重建图像质量评估指标 . . . . .	16
2.3.1	峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) . . . . .	17
2.3.2	结构相似性指标 (Structural Similarity, SSIM) . . . . .	17
2.3.3	多尺度结构相似性指标 (Multi-Scale SSIM, MS-SSIM) . . . . .	18
2.3.4	像素级误差指标 (MAE / MSE / NRMSE) . . . . .	19
2.4	超分辨模型与网络框架 . . . . .	20
2.4.1	卷积神经网络 (CNN) . . . . .	20
2.4.2	生成对抗网络 (GAN) . . . . .	21
2.4.3	视觉 Transformer (ViT) . . . . .	22
2.4.4	扩散模型 (Diffusion Model) . . . . .	23
2.4.5	网络框架设计 . . . . .	23
2.4.6	工业框架与本实验框架区别: 关于 <i>scale</i> 参数的讨论 . . . . .	24
2.5	训练策略与参数 . . . . .	25
2.5.1	损失函数 (Loss Function) . . . . .	25
2.5.2	正交归一化 (Orthogonal Regularization) . . . . .	26
2.5.3	训练优化器 (Optimizer) . . . . .	26
2.6	本章小结 . . . . .	27
<b>第三章 实验</b>		<b>29</b>
3.1	预实验: 不同训练参数对训练结果的影响 . . . . .	29
3.1.1	CNN vs DIFFUSION . . . . .	29
3.1.2	MSE vs JSDEV . . . . .	30
3.1.3	归一化: $[0, 1] vs [-1, 1]$ . . . . .	31
3.2	实验一: 机器学习超分辨实验 . . . . .	32
3.3	实验二: 多源数据融合实验 . . . . .	33

3.3.1	实验 2-1：荧光小球 . . . . .	35
3.3.2	实验 2-2：纤毛 . . . . .	35
3.3.3	2-3：实验数据微调 . . . . .	35
3.3.4	本章小结 . . . . .	38
<b>第四章 总结与展望</b>		<b>41</b>
4.1	研究总结 . . . . .	41
4.2	不足与展望 . . . . .	41
<b>参考文献</b>		<b>43</b>
<b>致    谢</b>		<b>49</b>



## 插图目录

1-1 基于深度学习的单图像超分辨（SISR）方法分类 . . . . .	4
2-1 物光在空间传播并经透镜折射的过程示意图 . . . . .	9
2-2 不同程度的干扰对成像造成的影响：透镜孔径限制导致的成像模糊 (a、b)；圆环干扰因子导致的成像涟漪 (c、d)；透镜形状造成的成像模糊 (e、f) . . . . .	12
2-3 三类数据集示例图：pySTED 模拟数据集 (a, b), Beads 实验数据集 (c, d), Cilia 实验数据集 (e, f) . . . . .	16
2-4 卷积神经网络示意图 . . . . .	20
2-5 生成对抗网络示意图 . . . . .	21
2-6 视觉 Transformer 示意图 . . . . .	22
2-7 扩散模型生成过程示意图 . . . . .	23
3-1 预实验：不同训练参数对训练结果的影响。第一列表示实验所用数据集，第二列表示归一化范围与 LOSS 方案，第三列、第四列为 DIFFUSION 和 CNN 模型的实验结果。其中 loss 选取训练 100 个 epoch 后的最终 loss, time 中表示训练一个 epoch 所用的平均时间	29

- 3-2 模型 CNN(a),DIFFUSION(b) 模型在训练过程中的 loss 随 epoch 变化图。模型 CNN(c),DIFFUSION(d) 的图像质量评估结果训练效果随 epoch 变化图：第一张图为 MAE 和 NRMSE 指标；第二张图为 MS-SSIM 和 SSIM 指标；第三张图展示了 PSNR 指标。 . . . . . 30
- 3-3 不同损失函数 MSE(a),JSDIV(b) 在训练过程中的 loss 随 epoch 变化图。采用损失函数 MSE(c),JSDIV(d) 的图像质量评估结果训练效果随 epoch 变化图：第一张图为 MAE 和 NRMSE 指标；第二张图为 MS-SSIM 和 SSIM 指标；第三张图展示了 PSNR 指标。 . . . . . 31
- 3-4 采用不同归一化 [0,1](a),[-1,1](b) 在训练过程中的 loss 随 epoch 变化图。采用归一化 [0,1](c),[-1,1](d) 的图像质量评估结果训练效果随 epoch 变化图：第一张图为 MAE 和 NRMSE 指标；第二张图为 MS-SSIM 和 SSIM 指标；第三张图展示了 PSNR 指标。 . . . . . 32
- 3-5 实验 1：cnn 与 diffusion 模拟数据训练结果图表展示。其中图 a 为直观数据表，图 b、c 展示了测试集前 25 图像的具体评估参数；ipt 表示输入模糊图像的评估参数结果 . . . . . 33
- 3-6 实验 1：CNN 与 DIFFUSION 模拟数据训练结果展示。对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ；第二列展示经模型重建后的图像  $\hat{I}_0$ ；第三列展示高分辨清晰图像 (STED) $I_0$ ；第四列展示  $I$  与  $I_0$  的残差图；第五列展示  $\hat{I}_0$  与  $I_0$  的残差图；第六列展示基于前三列图像中白线划过范围内图像数值变化图 . . . . . 34
- 3-7 实验 2-1：使用 pySTED 作为训练集，CNN 与 DIFFUSION 对荧光小球测试结果。对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ；第二列展示经模型重建后的图像  $\hat{I}_0$ ；第三列展示高分辨清晰图像 (STED) $I_0$ ；第四列展示  $I$  与  $I_0$  的残差图；第五列展示  $\hat{I}_0$  与  $I_0$  的残差图；第六列展示基于前三列图像中白线划过范围内图像数值变化图 . . . . . 36
- 3-8 实验 2-2：使用 pySTED 作为训练集，CNN 与 DIFFUSION 对纤毛测试结果。对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ；第二列展示经模型重建后的图像  $\hat{I}_0$ ；第三列展示高分辨清晰图像 (STED) $I_0$ ；第四列展示  $I$  与  $I_0$  的残差图；<sup>X</sup> 第五列展示  $\hat{I}_0$  与  $I_0$  的残差图；第六列展示基于前三列图像中白线划过范围内图像数值变化图 . . . . . 37

## 表格目录

2-1 实验数据集基本参数概览 . . . . .	14
2-2 实验数据集基本参数概览 . . . . .	15
2-3 本文使用图像质量评估指标说明 . . . . .	19



## 符号表

### 各类显微镜的缩写

CM	共聚焦显微镜 (Confocal Microscope)
SXM	软 X 射线显微镜 (Soft X-ray Microscopy)
EM	电子显微镜 (Electron Microscopy)
STM	扫描隧道显微镜 (Scanning Tunneling Microscopy)
AFM	原子力显微镜 (Atomic Force Microscopy)
STED	受激发射损耗显微镜 (Stimulated Emission Depletion Microscopy)
SIM	结构光照明显微镜 (Structured Illumination Microscopy)
PALM	光激活定位显微镜 (Photoactivated Localization Microscopy)
STORM	随机光学重建显微镜 (Stochastic Optical Reconstruction Microscopy)
3D SRM	三维超分辨率显微镜 (3D Super-Resolution Microscopy)

### 机器学习模型技术的缩写

SR	超分辨 (Super-Resolution)
SISR	单图像超分辨率重建 (Single Image Super-Resolution)
CNN	卷积神经网络 (Convolutional Neural Network)
GAN	生成对抗网络 (Generative Adversarial Network)
Diffusion	去噪扩散概率模型 (Denoising Diffusion Probabilistic Model)

### 问题设定与理论的数学符号

LR	低分辨率 (Low Resolution)
HR	高分辨率 (High Resolution)
GT	Ground Truth
$I_0(x, y)/I_{GT}(x, y)$	原始物光光强分布/原始清晰图像/评价标准图像 (Ground truth image)
$I(x, y)/I_{LR}(x, y)$	实际获得的成像图像/模糊的图像/输入图像 (Observed image)
$\hat{I}_0(x, y)/I_{HR}(x, y)$	经模型重建后的图像/输出图像 (Reconstructed image)
$O(x, y)$	成像系统的干扰因子 (Optical transfer function)

$\mathcal{F}, \mathcal{F}^{-1}$	傅里叶变换与其逆变换 (Fourier Transform and Inverse Transform)
*	卷积操作 (Convolution)
$\mathcal{L}$	损失函数 (Loss function)
$\mathcal{L}_{\text{MSE}}$	均方误差损失 (Mean Squared Error Loss)
$\mathcal{L}_{\text{perceptual}}$	感知损失函数 (Perceptual Loss)
$\mathcal{L}_{\text{GAN}}$	对抗损失 (Adversarial Loss)
$\theta$	神经网络参数集合
$G_{\theta}(\cdot)$	重建模型 (Generator)
$D_{\phi}(\cdot)$	判别器模型 (Discriminator)
$\mathbb{E}$	数学期望
$x$	输入低质量图像或模糊图像
$y$	高质量图像或原始图像

# 第一章 绪论

## 1.1 研究背景与意义

光学显微镜的分辨率受限于约 200 纳米的衍射极限<sup>[1]</sup>，难以有效观察诸如蛋白质复合体和细胞器动态行为等纳米尺度的生物结构。为了突破这一物理限制，传统超分辨成像技术（如受激发射损耗显微镜 STED 和随机光学重构显微镜 STORM）依赖于复杂的物理或化学手段，虽然在空间分辨率方面取得了显著进展，但通常伴随着高光毒性、较慢的成像速度以及高昂的硬件成本，限制了其在活细胞长期成像与临床应用中的推广<sup>[2]</sup>。

近年来，随着人工智能的发展，基于深度学习的图像超分辨方法（如卷积神经网络 CNN、生成对抗网络 GAN）因其数据驱动的特性，在图像重建任务中展现出强大的能力。这类方法能够实现图像分辨率的提升，并显著降低了对高端成像设备的依赖，成为提升显微图像质量的有力工具。然而，现有深度学习方法在实际应用中仍面临诸多挑战，包括：

- **数据单一性问题** 大多数模型仅依赖模拟数据或特定实验条件下采集的数据进行训练，导致模型在真实复杂环境中的泛化能力受限；
- **细节恢复困难** 传统生成模型难以精确还原图像中的高频细节，常引入模糊区域或结构性伪影，影响图像的可用性与可信度；
- **计算效率瓶颈** 部分高性能方法（如扩散模型）由于计算资源需求高，限制了其实时处理和大规模应用的可行性。

在此背景下，探索一种能够兼顾高分辨率重建质量、广泛泛化能力及计算效率的深度学习框架，已成为显微图像重建领域的关键研究方向。该研究不仅可助力基础生命科学的深入探索，还具有推动低成本医疗设备智能化发展的潜力，因而具有重要的理论价值与应用前景。

## 1.2 研究现状简介

本节介绍研究现状。1.2.1 节介绍自 20 世纪 30 年代开始发展的超分辨显微成像技术；1.2.2 节介绍近几年来兴起的机器学习超分辨技术。

### 1.2.1 超分辨显微成像技术

超分辨显微成像技术的发展历程可追溯至 20 世纪 30 年代，随着科学家们不断探索突破光学衍射极限的方法，逐步发展出多种基于物理、化学和计算手段的成像技术。以下内容简要介绍各显微镜技术的主要原理、性能及缺点，总结如表 ?? 所示。

- **X 射线显微镜与电子显微镜** 由 Abbe 分辨率极限  $\lambda/2n$  可以知道，降低光波波长是提高分辨率最直接有效的方式。C.Jacobsen 采用大约 4nm 的相干软 X 射线作为激发光源，实现了 30nm 的分辨率<sup>[3]</sup>。但低于 400nm 的光通常会损失活细胞，因此通常 X 射线显微镜无法应用于活细胞观测<sup>[11]</sup>；电子显微镜以德布罗意波长  $10^{-3} nm$  的电子束照射样品，可以达到 0.1nm 的超高分辨率<sup>[4]</sup>，被广泛应用于细胞生物学，但由于生物样品无法存活于高真空环境，电子显微镜同样无法应用于活细胞。
- **扫描隧道显微镜 (STM) 与原子力显微镜 (AFM)** 扫描隧道显微镜利用量子力学隧穿效应，使用传感器尖端探针与样品和表面之间的隧穿电流来判断样品表面形貌，对粘附于表面的结构可以获得 0.1nm 横向和 0.01nm 纵向分辨率<sup>[5]</sup>；相似地，原子力显微镜使用探针尖端的原子与物质表面原子之间的范德华力来呈现样品的表面形貌，可以获得约 0.5nm 的横向和约 0.01nm 的纵向分辨率<sup>[6]</sup>。这两类显微镜的分辨率都可达原子尺度，但具有共同的缺点：它们仅限于样品表面成像，并且图像获取速度缓慢，实验装置（如精细探针）复杂昂贵，可能对生物组织造成破坏，无法适用于活细胞<sup>[1]</sup>。
- **受激发射耗尽显微镜 (STED)** Hell 和 Wichmann 使用一个环形的耗尽激光束抑制荧光发射，实现了高达 20 纳米的分辨率<sup>[7]</sup>。STED 显微镜在活细胞成像中具有一定优势，但存在光毒性高和设备成本昂贵等问题。
- **结构照明显微镜 (SIM)** Gustafsson 通过投射结构化光图案并利用图像重建算法，突破了传统光学显微镜的分辨率限制，实现了约 100 纳米的分辨

率<sup>[8]</sup>。SIM 具有成像速度快、光毒性低的优点，适用于活细胞成像；缺点是 SIM 系统本质上仍是线性成像系统，受到衍射极限影响，分辨率较差。

- **荧光光激活定位显微镜 (PALM)** Betzig 等人利用可光激活的荧光蛋白，通过精确定位单个分子的发射位置，重建出高分辨率图像，实现了约 20 纳米的分辨率<sup>[9]</sup>。PALM 适用于固定样品的成像，成像过程较为耗时。
- **随机光学重建显微镜 (STORM)** Rust 等人采用随机激活和定位荧光探针的方法，实现了与 PALM 相似的分辨率<sup>[10]</sup>。STORM 在多色成像和活细胞成像方面具有一定优势，但对荧光探针的选择和成像条件要求较高。
- **三维超分辨显微镜 (3D SRM)** 随着对生物结构空间分布与动态行为理解的深入需求，研究者逐步从二维超分辨成像向三维成像拓展。三维超分辨显微镜主要是以上超分辨技术的基础上，通过光学设计或图像重建算法，实现在三维空间内的纳米级分辨率重建。目前横向分辨率可达 20-50nm，纵向分辨率可达 50-100nm<sup>[12]</sup>，相比常规共聚焦显微镜（Z 轴分辨率约 500-800nm）提升显著。然而，三维超分辨成像需要更复杂的光路系统和图像处理流程，成像速度缓慢，样本制备难度更高，也有光漂白和光毒性的风险<sup>[13]</sup>。

### 1.2.2 机器学习单图像超分辨技术

近年来，机器学习单图像超分辨 (Single Image Super-Resolution, SISR) 技术在计算机视觉以及生物显微图像重建等多个领域中取得了显著进展。随着深度学习方法的持续发展，机器学习 SISR 技术已从早期的卷积神经网络 (CNN) 逐步演进至生成对抗网络 (GAN)、Transformer、近年来的扩散模型 (Diffusion Models)。图 1-1 展示了当前主流基于深度学习的 SISR 方法的技术演进路径及其代表性成果。

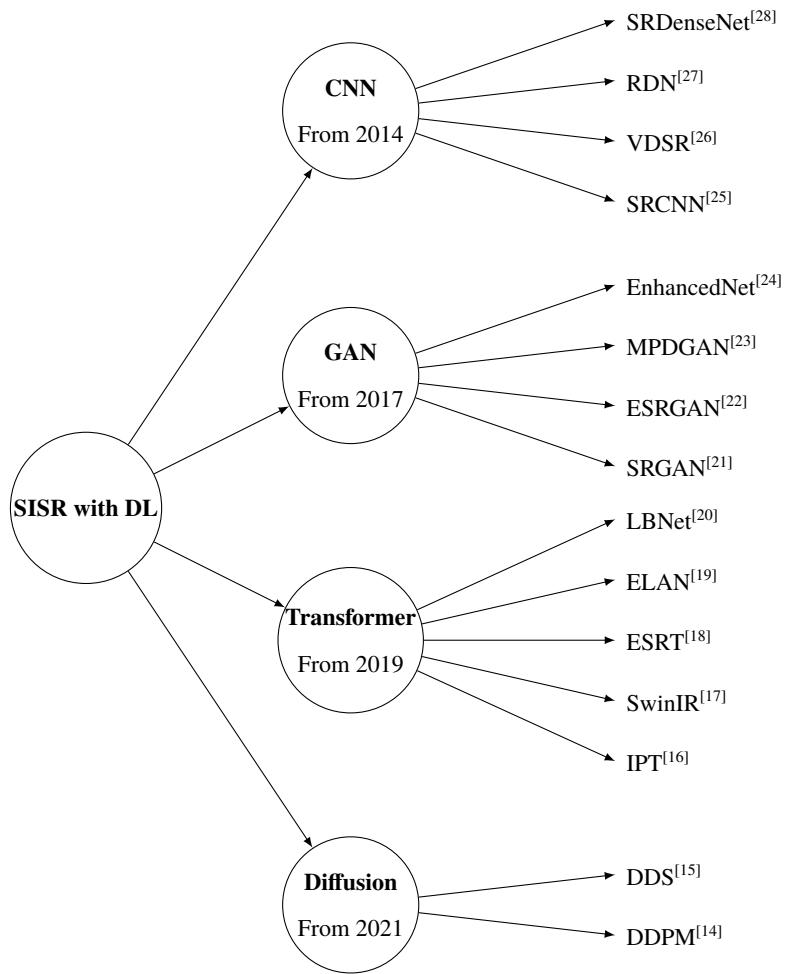


图 1-1 基于深度学习的单图像超分辨 (SISR) 方法分类

- **卷积神经网络 (CNN, 自 2014 年起)** 是最早应用于 SISR 的深度学习模型之一。Dong 等人提出的 SRCNN<sup>[25]</sup> 是首个采用端到端训练方式的深度超分辨率模型，后续方法如 VDSR<sup>[26]</sup>、SRDenseNet<sup>[28]</sup> 与 RDN<sup>[27]</sup> 通过引入残差结构与密集连接等机制，进一步提升了模型的训练深度与重建性能。CNN 类方法的主要优势是结构简洁、训练成本低、推理效率高，但在高频细节生成与全局一致性具有局限性。
- **生成对抗网络 (GAN, 自 2017 起)** 引入对抗学习机制以提升图像感知质量。SRGAN<sup>[21]</sup> 是首个应用 GAN 于超分辨率任务的模型，通过感知损失与对抗损失联合优化，显著提升了图像的细节真实感。随后 ESRGAN<sup>[22]</sup> 引入残差块改进结构，MPDGAN<sup>[23]</sup> 和 EnhancedNet<sup>[24]</sup> 则进一步探索多尺度与内容感知增强策略。GAN 类模型的优势是生成图像具有更高的视觉质量，但训练不稳定、收敛难度大，伪影问题影响严重。

- **ransformer (自 2019 起)** 通过引入自注意力机制，增强了模型对全局上下文的建模能力。Chen 等人提出的 IPT<sup>[16]</sup> 首先采用预训练 Transformer 用于多任务图像恢复，后续模型如 SwinIR<sup>[17]</sup> 利用滑动窗口结构提升计算效率，ESRT<sup>[18]</sup>、ELAN<sup>[19]</sup> 与 LBNet<sup>[20]</sup> 则在模型压缩与轻量化方面进行了改进。Transformer 系列模型在结构建模能力上优于 CNN，但训练成本与数据需求较高。
- **扩散模型 (Diffusion Model, 自 2021 起)** 是近年来兴起的一类生成模型，通过模拟图像从高斯噪声中逐步还原的过程实现图像重建。Ho 等人提出的 DDPM<sup>[14]</sup> 提出了扩散模型的原理方法，展示了该模型在生成质量上的潜力。DDS<sup>[15]</sup> 等后续工作将扩散机制与图像超分辨率任务结合，取得了优异的重建效果。扩散模型的生成结果自然逼真，在高质量图像重建方面具有显著优势；但其迭代过程计算复杂，难以满足实时成像的应用需求。

综上所述，不同的机器学习单图像超分技术具有不同的优劣势。例如 CNN 类模型，虽然图像重建质量竞争力不强，但其低成本、高效率的特性仍具有广大应用场景；Diffusion 类模型则成为高质量低伪影重建场景的首选。

### 1.2.3 机器学习超分辨显微技术

在超分辨显微技术与机器学习超分辨技术的碰撞中，诞生了更多更深入的研究。STED、STORM、PALM、SIM 等技术具有超高的分辨率，但这些方法往往依赖昂贵的实验设备，并且存在成像时间长、光毒性高等问题，难以大规模推广。深度学习模型能够通过训练数据学习低分辨率与高分辨率图像之间的映射关系，从而实现超分辨率重建，提高图像精度、加速成像速度、降低成像成本，从而扩大这些技术的应用范围。以下是几种主要的超分辨显微镜技术及其与深度学习结合的应用进展：

- **STED** 深度学习在 STED 显微镜中的应用主要体现在跨模态超分辨率成像的实现。2019 年，H. Wang 等人提出了利用深度学习方法从低分辨率荧光图像预测高分辨率图像的方案，从而在传统 STED 技术的基础上实现了更高效的分辨率提升<sup>[29]</sup>。

- **STORM** E. Nehme 等人提出的 Deep-STORM 模型，能够以更高的速度和精度处理单分子成像数据，从而生成超分辨图像<sup>[30]</sup>。
- **PALM** W. Ouyang 等人通过使用深度神经网络处理图像数据，显著缩短了成像时间，并保留了高质量的定位精度<sup>[31]</sup>。
- **SIM** C. Qiao 等人开发的深度神经网络模型，专门用于光学显微镜中的图像重建，显著提升了 SIM 技术在大视野成像与快速成像中的应用潜力<sup>[32]</sup>。
- **三维超分辨显微镜** 深度学习还被用于提升三维成像分辨率和效率。在 3D 超分辨显微镜领域，机器学习方法能够从二维切片重建高分辨率的三维结构，并用于追踪分子运动和细胞内结构。

### 1.3 本研究特点

绪论中提到，现有的机器学习超分辨显微技术在实际应用中，仍存在以下问题：

- **数据单一性问题** 大多数模型仅依赖模拟数据或特定实验条件下采集的数据进行训练，导致模型在真实复杂环境中的泛化能力受限；
- **细节恢复困难** 传统生成模型难以精确还原图像中的高频细节，常引入模糊区域或结构性伪影，影响图像的可用性与可信度；
- **计算效率瓶颈** 部分高性能方法（如扩散模型）由于计算资源需求高，限制了其实时处理和大规模应用的可行性。

本毕业论文采用的核心模型方法为扩散模型，核心训练策略为多源数据融合策略，希望回答并且解决上述问题。本文核心思路如下：

- **细节恢复困难** 为了实现最大程度发挥模型的能力，取得更好的重建效果，对训练参数进行控制变量优化（**预实验**）；采用 DIFFUSION 模型与 SRCNN 模型进行对比实验，探究扩散模型在高频细节重建的优越性（**实验一**）；采用多种评估指标，尝试回答扩散模型对图像高频细节的重建质量，尝试回答扩散模型的伪影问题以及可信度（**实验一**）。
- **数据单一性问题** 采用单一数据训练策略，用模拟数据训练，真实实验数据进行测试，尝试回答单模拟数据下模型在真实复杂环境下的超分辨能力与

泛化能力（实验二）；采用多源数据融合策略对比试验，用模拟数据融合部分实验数据，测试模型超分辨能力与泛化能力的提升（实验二）。

- **计算效率瓶颈** 采用多源数据融合策略，使用大量模拟数据进行预训练，少量实验数据进行复训练微调，解决实际应用场景下对训练数据的依赖性与重复训练计算资源需求量大的问题（实验二）。

对以上问题的回答与解决结果，请查看本文第三章实验与第五章总结与展望。

## 1.4 本文章节安排

本文各章节的内容安排如下：

第一章：绪论。本章介绍本研究的背景与意义，综述当前超分辨显微成像、基于机器学习的单图像超分辨技术及其在显微图像中的应用现状，指出传统方法在高频细节还原、模型泛化能力等方面不足，进而引出本文的研究目标与技术路线。

第二章：相关技术介绍。本章系统梳理了本研究中涉及的理论基础与关键技术，包括成像过程中的物理建模与分辨率退化分析、模拟与真实实验数据的来源与处理方法、图像质量评估指标的选用依据、各类超分模型的技术特点，以及训练过程中的优化策略与参数设置。

第三章：实验设计与结果分析。本章基于第二章中提出的模型架构与评估体系，开展三组实验以验证本文方法的有效性与实用性。首先通过预实验探讨 Diffusion 模型在结构还原、训练收敛性等方面相较于 CNN 模型的优势；随后在模拟数据上评估模型在高频细节还原与伪影控制上的性能表现；最后在真实实验数据上进一步分析其跨域泛化与少量微调下的迁移能力，全面验证所提方法的鲁棒性与应用潜力。

第四章：总结与展望。本章对全文的研究内容与主要贡献进行总结，并对当前研究中仍存在的不足与未来可能的研究方向进行展望，指出 Diffusion 模型在显微图像超分重建领域的拓展潜力与进一步优化路径。

## 1.5 本章小结

本章作为绪论部分，系统性地阐述了超分辨显微成像技术的研究背景、发展现状及本论文的研究目标。首先分析了光学显微镜在生物医学研究中的分辨率限制问题，指出突破衍射极限对纳米级生物结构观测的重要意义；其次详细梳理了传统超分辨显微技术（如 STED、STORM、PALM 等）和基于深度学习的单图像超分辨方法（CNN、GAN、Transformer、扩散模型等）的技术原理、演进历程及各自的局限性；进而针对当前机器学习超分辨显微技术存在的三大核心挑战——数据单一性、细节恢复困难和计算效率瓶颈，提出了本研究的创新性解决方案与技术路线。最后概述了全文的章节安排，为后续研究内容的展开奠定理论基础并明确研究方向。

## 第二章 相关技术介绍

### 2.1 问题设定与理论

从数学角度来看，来自物体的光在空间中的传播以及通过透镜的折射过程，可视为一个二维傅里叶变换的过程。因此，其衍射极限成像过程可通过傅里叶变换中的干扰因子进行建模。本文的理论分析与模型实现均建立在这一物理-数学视角之上。

#### 2.1.1 傅里叶变换视角下的单透镜成像过程

物体发出的光在空间中传播并经过透镜后，其波前变化可等效描述为一次二维傅里叶变换，如图 2-1 所示。

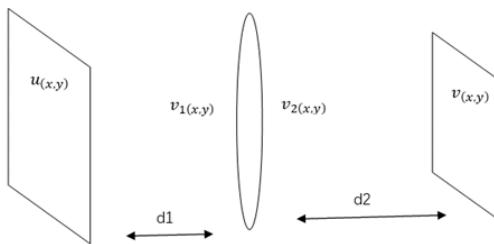


图 2-1 物光在空间传播并经透镜折射的过程示意图

在远场近似（即传播距离远大于横向变量变化）下，物光场  $u(\xi, \eta)$  经过自由传播与透镜相位调制后，其传播至透镜前的光场表达为：

$$v_1(x, y) \approx e^{i\phi_1(x, y)} \cdot \mathcal{F}[u(\xi, \eta)] \quad (2-1)$$

其中， $\phi_1(x, y)$  为与空间坐标相关的相位项， $\mathcal{F}[\cdot]$  表示傅里叶变换。

透镜的作用等价于对该波前叠加一个额外的相移项，得到：

$$v_2(x, y) = v_1(x, y) \cdot e^{iP(x, y)} \quad (2-2)$$

若相移函数  $P(x, y)$  恰当设计，其整体传播过程等价于对原始光场的调制与傅里叶变换的组合。

进一步传播至像平面后，最终光场  $v(x, y)$  可表示为：

$$v(x, y) \approx e^{i\phi_2(x, y)} \cdot (u(x, y) \cdot e^{ip(x, y)}) \quad (2-3)$$

在合适的相位补偿  $p(x, y)$  下，最终像场仅与原始光场相差一个相位因子，表示为：

$$v(x, y) = u(x, y) \cdot e^{ip(x, y)} \quad (2-4)$$

对于光强分布而言，由于光强为振幅平方，其相位因子不影响观测结果，因此可简化表示成像过程为：

$$I(x, y) = \mathcal{F}^{-1} (\mathcal{F}[I_0(x, y)]) \quad (2-5)$$

其中， $I_0(x, y)$  表示物光的光强分布， $I(x, y)$  为成像平面上的光强分布。由此可见，单透镜成像过程可视为傅里叶变换与其逆变换的组合过程。

### 2.1.2 衍射极限与成像系统缺陷引起的分辨率退化

在公式 2-2 中，我们假设来自物体的所有光均可被透镜调制，且所引入的相位因子完全符合理想模型  $P(x, y)$ ，从而实现无损成像。然而，在实际成像系统中，这一理想情况难以满足。首先，透镜的尺寸有限，无法覆盖所有入射光；其次，高于波数的空间频率成分由于形成倏逝波（evanescent wave）而无法传播至远场；此外，透镜所引入的相位调制也可能与理想模型存在偏差。这些因素均导致最终像面上的光场与物面光场不再一致。

这种不一致可建模为傅里叶面上的光场不再是理想物光场傅里叶变换  $\mathcal{F}(I_0(x, y))$ ，而是与成像系统响应函数（干扰因子） $O(x, y)$  的乘积。因此，最终像场可表示为：

$$I(x, y) = \mathcal{F}^{-1} [\mathcal{F}(I_0(x, y)) \cdot O(x, y)] \quad (2-6)$$

进一步推导可得：

$$I(x, y) = I_0(x, y) * \mathcal{F}(O(x, y)) \quad (2-7)$$

即最终像场等效于原始物光场与干扰因子频谱的卷积。

### 2.1.3 不同程度干扰对成像质量的影响

(1) 无干扰情况：若干扰因子  $O(x, y) = 1$ ，则其傅里叶变换为  $\delta(x, y)$ ，此时成像过程为：

$$I(x, y) = I_0(x, y) * \delta(x, y) = I_0(x, y) \quad (2-8)$$

即系统实现了理想的、无失真的成像。

(2) 有限孔径干扰：若透镜孔径有限，部分光线无法通过成像，此时干扰因子可表示为：

$$O(x, y) = u(r^2 - x^2 - y^2) \quad (2-9)$$

其傅里叶变换具有中心亮斑及环状副瓣（如图 3-4c、3-4b）。该干扰频谱与物光强度分布卷积后，会导致图像条纹拓宽、边缘模糊，并在周围形成明显涟漪（图 3-4d、2-3d）。

(3) 透镜形状误差干扰：若透镜存在制造误差，则其调制函数不再是二值函数，而变为一个在  $[0, 1]$  范围内变化的函数，可设置为：

$$O(x, y) = u(r^2 - x^2 - y^2) \cdot P(x, y) \quad (2-10)$$

其中  $P(x, y)$  为形状误差引起的相位扰动函数。则成像表达为：

$$I(x, y) = I_0(x, y) * \mathcal{F}(O(x, y)) = I_0(x, y) * [\mathcal{F}(u(\cdot)) * \mathcal{F}(P(x, y))] \quad (2-11)$$

这表示干扰项为有限孔径带来的衍射图样与相位误差频谱之间的卷积。

若设定

$$P(x, y) = 0.1 \sin(x + y) + 0.9 \quad (2-12)$$

其结果如图 2-3e、2-3f 所示。在相位扰动较小的情况下，频谱仍呈现出以中心亮斑为主的环状结构，但图像质量明显下降，表现为边缘模糊与周期性涟漪增强。

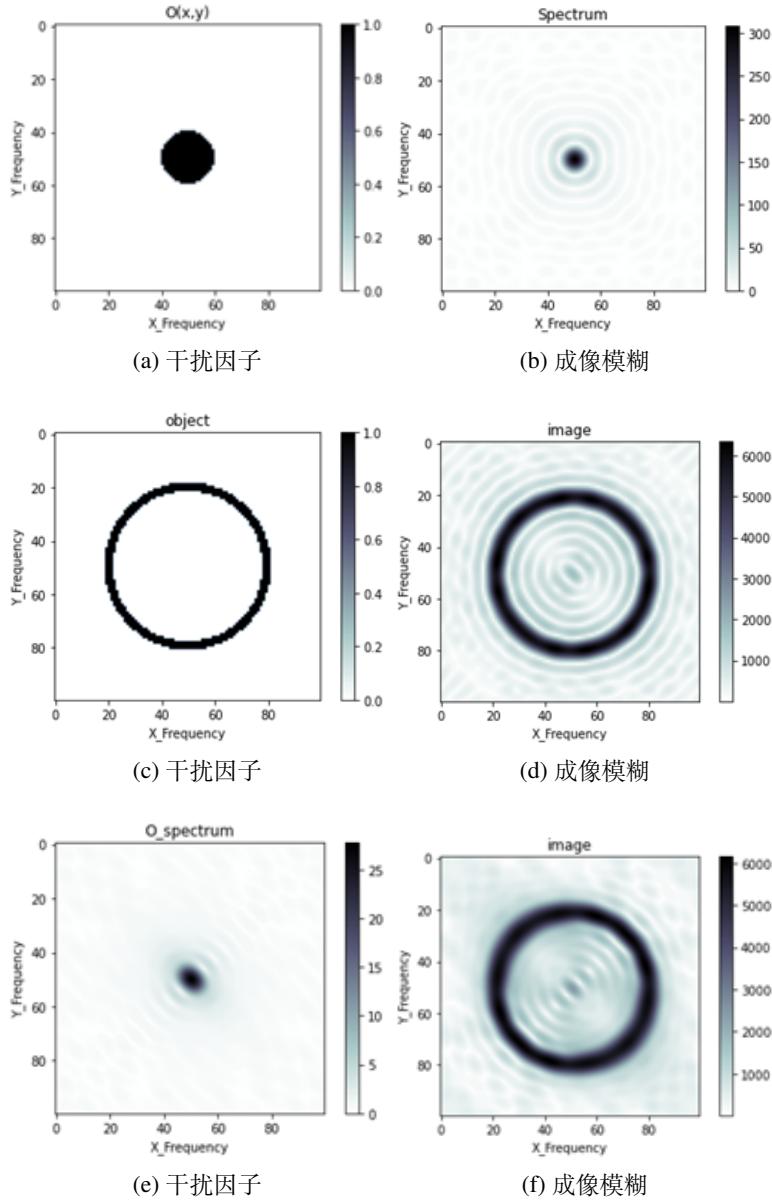


图 2-2 不同程度的干扰对成像造成的影响：透镜孔径限制导致的成像模糊 (a、b); 圆环干扰因子导致的成像涟漪 (c、d); 透镜形状造成的成像模糊 (e、f)

## 2.1.4 清晰图像的还原

如前所述，原始物光的强度分布  $I_0(x, y)$  与成像系统干扰频谱  $F(O(x, y))$  发生卷积，导致最终成像图像  $I(x, y)$  出现模糊、失真现象。对于给定的光学系统，干扰因子  $O(x, y)$  是已知的。因此，关键问题在于如何从如下形式的卷积关系中有效地还原出原始图像：

$$I(x, y) = I_0(x, y) * F(O(x, y)) \quad (2-13)$$

传统方法如反卷积在处理强干扰情况下常常面临稳定性与精度的挑战。为此，本研究拟采用人工智能方法，学习图像与系统干扰之间的非线性映射关系，从而实现对原始图像  $I_0(x, y)$  的高质量重建与恢复。

## 2.2 实验数据集介绍

在本研究中，为充分验证扩散模型的重建效果以及多源数据融合的有效性，我们构建并使用了两个类型的数据集：其一为基于仿真平台工具生成的模拟超分辨数据集，另一类为来自真实显微成像系统的实拍图像数据集。

### 2.2.1 模拟数据集 pySTED

为获取高质量且具有真实光学退化特性的训练样本，我们基于高精度仿真框架 pySTED<sup>[33]</sup> 构建了大规模的超分辨显微模拟数据集。该框架可通过基于光学物理建模的模拟流程，精确重现 STED 显微镜与共聚焦显微镜在不同成像条件下的图像对。

我们采用 pySTED 对简单荧光量子点仿真了 1500 对成像图像，每对包含一幅 Confocal 模拟图像（分辨率 200nm）与对应的 STED 超分辨图像（分辨率 50nm），图像大小统一为  $256 \times 256$  像素。所有仿真图像均基于荧光点扩散函数（PSF）模拟，同时引入了真实成像过程中可能出现的物理干扰因素，如信号衰减、光漂白（photobleaching）、噪声扰动及点扩散失真等，以提升模拟数据的真实度与训练泛化能力。

## 2.2.2 真实实验数据集

在实际实验条件下，我们在生物医学光学实验室中采集了两个真实样本的数据集，分别针对标准荧光微球（Fluorescent Beads）与生物纤毛结构（Cilia Structure），在相同样本视野下，使用 Confocal 显微镜与 STED 显微镜分别采集图像，并进行配准与切片处理，构建真实图像对。

具体地：

- **荧光小球数据集（Beads Dataset）**：我们使用标准 100nm 荧光聚苯乙烯微球进行标定，分别采集 50 对图像。每对图像包含一幅共聚焦图像与对应视野的 STED 图像，图像尺寸为  $1024 \times 1024$  像素。
- **纤毛数据集（Cilia Dataset）**：使用固定组织样本并标记纤毛结构的荧光染料，采集 200 对共聚焦 STED 图像对。图像大小统一为  $1024 \times 1024$  像素。

三组数据集详细参数可见表 2-1。

表 2-1 实验数据集基本参数概览

数据集	图像对数	分辨率对比	图像尺寸（像素）	样本类型
pySTED	1500	200/50nm	$256 \times 256$	模拟量子点结构（物理建模）
Beads	50	200/50nm	$1024 \times 1024$	荧光微球（CM/STED 成像对）
Cilia	200	200/50nm	$1024 \times 1024$	生物纤毛（CM/STED 成像对）

## 2.2.3 数据规范化处理

仿真平台获得的模拟数据以及拍摄得到的实验数据集一般需要经过预处理才能被送入神经网络训练。我们对原始数据进行了标准化预处理与多种数据增强操作，主要包括：

- **图像尺寸统一化** 为了方便训练以及后续评估，我们将所有数据集统一裁剪为  $64 \times 64$  的尺寸大小；
- **数据集灰度化** 为了方便标准归一化，我们将所有获取到的图像做灰度化处理，化为单通道的灰度图片映射在 0-255 灰度上；
- **归一化处理（Normalization）** 所有图像像素强度均被线性缩放至  $[0, 1]$  或  $[-1, 1]$  区间，这一步是为了训练能够快速拟合，属于机器学习常规数据操作；

- **数据集扩张** 我们采用了经典的数据预处理技术来进行数据集扩张，包括随机裁剪：训练时从原图中随机裁剪  $128 \times 128$  的子图块，模拟不同视野变化；随机翻转与旋转：包括水平翻转、垂直翻转、 $90^\circ$ 、 $180^\circ$  和  $270^\circ$  旋转，以增强空间不变性；
- **数据集清理** 经过切割的图像可能对应原图像中信息量极低的“空白”部分，这对于训练神经网络是没用的。我们通过计算图像平均灰度，删去了一些低质量的数据。
- **数据集增强** 为了提高训练效果，我们对高分辨的图像应用 Gamma 函数，以增强对比度。我们认为，图像的低灰度部分对应的基本是噪音，我们利用 Gamma 函数放大图像在高亮度部分的细节，以方便评估还原效果。

经过数据集扩张以及数据规范化处理后，三组数据集详细参数可见表 2-2

表 2-2 实验数据集基本参数概览

数据集	图像对数	分辨率对比	图像尺寸（像素）	样本类型
pySTED	10408	200/50nm	$64 \times 64$	模拟量子点结构（物理建模）
Beads	1504	200/50nm	$64 \times 64$	荧光微球（CM/STED 成像对）
Cilia	6512	200/50nm	$64 \times 64$	生物纤毛（CM/STED 成像对）

经过数据规范化处理后的三组数据集样本示例可见图 2-3

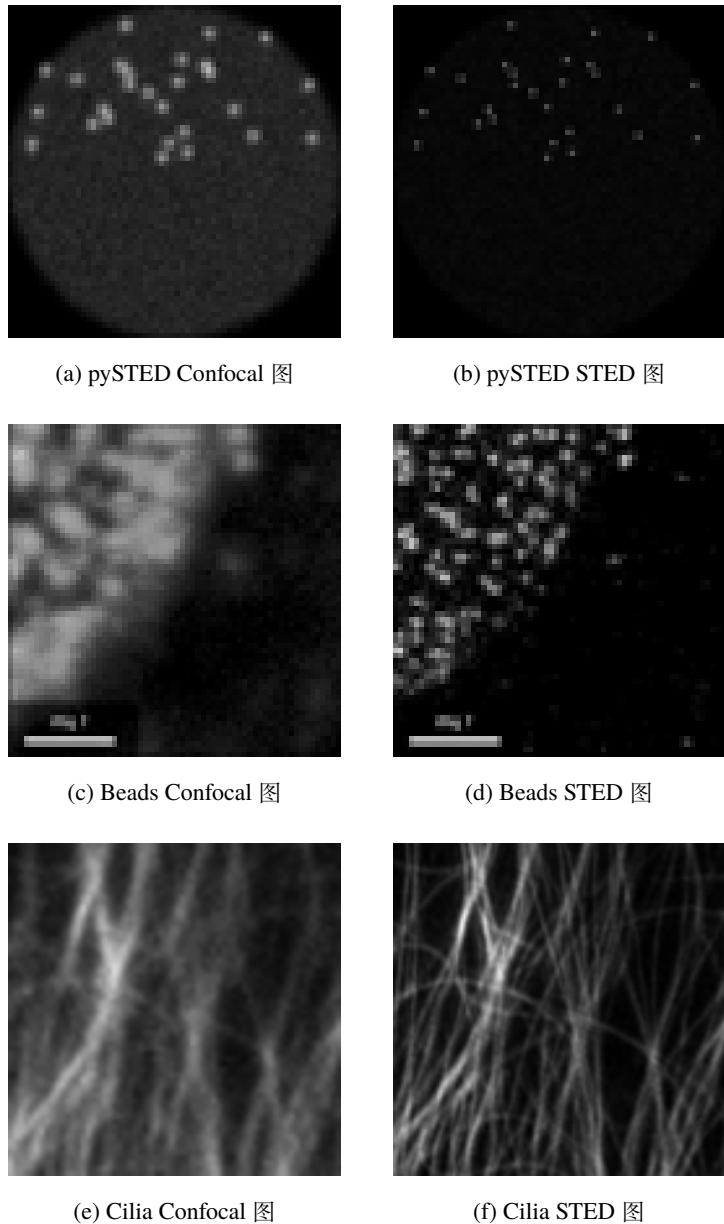


图 2-3 三类数据集示例图：pySTED 模拟数据集 (a, b), Beads 实验数据集 (c, d), Cilia 实验数据集 (e, f)

### 2.3 模型重建图像质量评估指标

图像质量 (Image quality) 指图像的视觉属性，并关注观看者的感知评价。一般而言，图像质量评估 (IQA) 方法包括基于人类感知的主观方法（即图像看起来有多真实）和客观计算方法。前者更符合我们的需求，但通常耗时且成本高昂，因此后者是目前的主流方法。然而，这些方法之间并不一定一致，因为客观方法往往无法精准捕捉人类视觉感知，可能导致 IQA 结果存在较大差异<sup>[21]</sup>。本

小节将介绍几种最常用的 IQA 方法，并在之后的图像重建质量评估中运用。

### 2.3.1 峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)

峰值信噪比 (PSNR) 是最常用的有损变换（如图像压缩、图像修复）重建质量评估指标之一。在图像超分辨率任务中，PSNR 通过最大像素值（记为  $L$ ）和图像间的均方误差（Mean Squared Error, MSE）计算得出。给定真实图像  $I_0$ （像素数为  $N$ ）及其重建结果  $\hat{I}_0$ ，两者之间的 PSNR 定义如下：

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{L^2}{\frac{1}{N} \sum_{i=1}^N (I_0(i) - \hat{I}_0(i))^2} \right), \quad (2-14)$$

其中，对于常规的 8 位图像表示， $L$  通常取值为 255。由于 PSNR 仅与像素级均方误差 (MSE) 相关，仅关注对应像素的差异而忽略视觉感知，因此在真实场景中评估重建质量时表现往往欠佳——而实际应用中我们通常更关注人类主观感知。然而，出于与现有文献对比的必要性，以及目前缺乏完全准确的感知评价指标，PSNR 至今仍是超分辨率模型最广泛使用的评估标准<sup>[34]</sup>。

### 2.3.2 结构相似性指标 (Structural Similarity, SSIM)

考虑到人类视觉系统 (HVS) 高度适应于图像结构的提取，结构相似性指数 (Structural Similarity Index, SSIM<sup>[35]</sup>) 被提出用于评估图像之间的结构相似度。SSIM 通过亮度、对比度和结构三个方面的独立比较，对图像质量进行感知驱动的评估。

对于一个包含  $N$  像素的图像  $I_0$ ，其亮度  $\mu_{I_0}$  与对比度  $\sigma_{I_0}$  分别可由像素均值与标准差计算得出：

$$\mu_{I_0} = \frac{1}{N} \sum_{i=1}^N I_0(i), \quad \sigma_{I_0} = \left( \frac{1}{N-1} \sum_{i=1}^N (I_0(i) - \mu_{I_0})^2 \right)^{1/2} \quad (2-15)$$

对亮度与对比度的比较函数分别定义为：

$$C_l(I_0, \hat{I}_0) = \frac{2\mu_{I_0}\mu_{\hat{I}_0} + C_1}{\mu_{I_0}^2 + \mu_{\hat{I}_0}^2 + C_1}, \quad C_c(I_0, \hat{I}_0) = \frac{2\sigma_{I_0}\sigma_{\hat{I}_0} + C_2}{\sigma_{I_0}^2 + \sigma_{\hat{I}_0}^2 + C_2} \quad (2-16)$$

其中  $C_1 = (k_1 L)^2$ ,  $C_2 = (k_2 L)^2$  是用于避免分母为零的稳定性常数, 通常  $k_1, k_2 \ll 1$ ,  $L$  表示像素最大动态范围 (如  $L = 255$ )。

图像结构部分以标准化图像 (即减去均值并除以标准差) 之间的内积为度量, 即协方差形式。结构比较函数定义如下:

$$\sigma_{I_0 \hat{I}_0} = \frac{1}{N-1} \sum_{i=1}^N (I_0(i) - \mu_{I_0})(\hat{I}_0(i) - \mu_{\hat{I}_0}) \quad (2-17)$$

$$C_s(I_0, \hat{I}_0) = \frac{\sigma_{I_0 \hat{I}_0} + C_3}{\sigma_{I_0 \hat{I}_0} + C_3} \quad (2-18)$$

最终, 结构相似性指数 (SSIM) 定义为三者加权乘积的形式:

$$\text{SSIM}(I_0, \hat{I}_0) = [C_l(I_0, \hat{I}_0)]^\alpha \cdot [C_c(I_0, \hat{I}_0)]^\beta \cdot [C_s(I_0, \hat{I}_0)]^\gamma \quad (2-19)$$

其中  $\alpha$ 、 $\beta$ 、 $\gamma$  为控制每一项权重的重要性参数。由于 SSIM 能更好地从感知角度评估图像重建质量, 因此被广泛用于超分辨率和图像增强等图像复原任务中<sup>[36]</sup>。

### 2.3.3 多尺度结构相似性指标 (Multi-Scale SSIM, MS-SSIM)

为了增强 SSIM 对不同尺度图像结构的建模能力, Wang 等人进一步提出了多尺度结构相似性指标 (Multi-Scale SSIM, MS-SSIM)<sup>[37]</sup>。该方法将图像在多个分辨率下进行评估, 更加符合人类视觉系统的多尺度感知特性。

MS-SSIM 首先构造图像金字塔, 对图像逐层进行下采样, 并分别在每一尺度上计算亮度、对比度与结构相似度。设共有  $M$  个尺度, 第  $j$  层尺度的亮度、对比度、结构比较函数分别记为  $C_l^{(j)}$ 、 $C_c^{(j)}$ 、 $C_s^{(j)}$ , 则 MS-SSIM 定义为:

$$\text{MS-SSIM}(I_0, \hat{I}_0) = [C_l^{(M)}(I_0, \hat{I}_0)]^{\alpha_M} \cdot \prod_{j=1}^M [C_c^{(j)}(I_0, \hat{I}_0)]^{\beta_j} \cdot [C_s^{(j)}(I_0, \hat{I}_0)]^{\gamma_j} \quad (2-20)$$

其中,  $\alpha_M$ 、 $\beta_j$  和  $\gamma_j$  为各尺度的加权系数, 通常通过经验或学习方式设定。MS-SSIM 融合了不同尺度的结构信息, 因此相比单尺度 SSIM, 更加稳健地反映

了图像整体感知质量，尤其适用于高分辨率重建或存在模糊、局部失真的场景。

### 2.3.4 像素级误差指标 (MAE / MSE / NRMSE)

除了感知驱动的评价指标外，本文仍将传统的像素差异指标作为参考，主要包括平均绝对误差 (Mean Absolute Error, MAE)、均方误差 (Mean Squared Error, MSE) 与归一化均方根误差 (Normalized Root Mean Squared Error, NRMSE)。

(1) MAE 衡量重建图像与原图之间像素差值的平均绝对值，定义为：

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |I_0(i) - \hat{I}_0(i)| \quad (2-21)$$

(2) MSE 是平方误差的平均值，与 PSNR 密切相关，定义为：

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (I_0(i) - \hat{I}_0(i))^2 \quad (2-22)$$

(3) NRMSE 是 MSE 的平方根，并进一步除以像素值范围（例如  $L = 255$ ），从而具有归一化尺度：

$$\text{NRMSE} = \frac{\sqrt{\text{MSE}}}{L} \quad (2-23)$$

其中， $N$  表示图像像素总数， $L$  为最大像素值 (8-bit 图像中为 255)。这些指标直观易算，符合直觉，但并不能有效刻画图像结构和感知质量，仅适用于评估重构图像的整体误差水平。因此，在实际应用中，更广泛地应用于 LOSS 函数中，在 IQA 中仅辅助 PSNR, MS-SSIM 和 SSIM 等指标进行综合评估。

综上，将本文使用的 5 类评价指标及其特点作入表 2-3：

表 2-3 本文使用图像质量评估指标说明

参数	数值	范围	意义
MAE	越小越好	$[0, 1]$	像素级误差指标，直观易算
NRMSE	越小越好	$[-1, 1]$	可以一定程度反应表示模型泛化能力强弱
SSIM	越大越好	$[-1, 1]$	从亮度、对比度、结构三方面评估图像相似性，关注整体结构
MS-SSIM	越大越好	$[-1, 1]$	更稳定地反映图像整体感知质量
PSNR	越大越好	$[0, \infty)$	传统图像质量评估指标，值越高表示失真越小

## 2.4 超分辨模型与网络框架

### 2.4.1 卷积神经网络 (CNN)

卷积神经网络 (CNN) 典型架构如图 2-4 所示，可分为卷积层、池化层与全连接层三部分。其中，卷积层提取多通道特征：

$$F_k^{(l)}(m, n) = \sigma \left( \sum_{i,u,v} W_{k,i}^{(l)}(u, v) F_i^{(l-1)}(m+u, n+v) + b_k^{(l)} \right) \quad (2-24)$$

池化层进行下采样以增强平移不变性：

$$P_k(m, n) = \max_{0 \leq i, j < p} F_k(pm+i, pn+j) \quad (2-25)$$

全连接层最后将展平后的向量映射到输出空间：

$$\mathbf{h}^{(l)} = \phi(W^{(l)}\mathbf{h}^{(l-1)} + b^{(l)}) \quad (2-26)$$

整体网络可视为端到端函数  $\hat{I}_0 = f_\theta(I_{LR})$ ，通过最小化均方误差损失  $\mathcal{L}_{MSE} = \frac{1}{N} \sum_i (I_0(i) - \hat{I}_0(i))^2$  完成超分辨率重建。

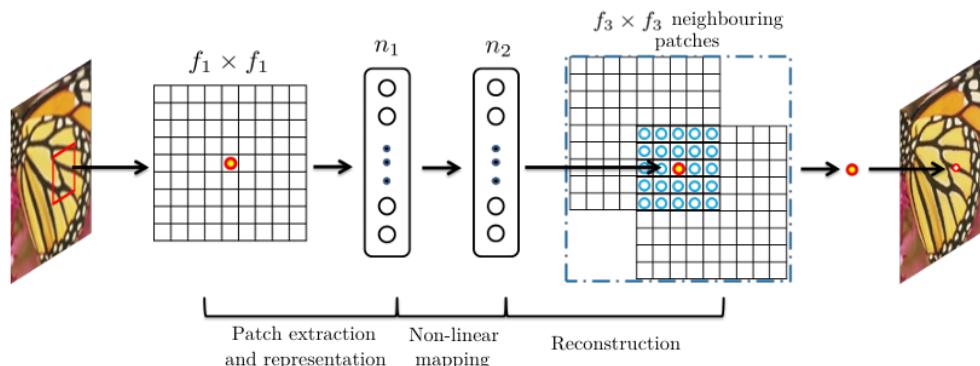


图 2-4 卷积神经网络示意图

## 2.4.2 生成对抗网络 (GAN)

生成对抗网络 (GAN) 由生成器  $G$  和判别器  $D$  两部分组成, 如图 2-5 所示。生成器尝试将低分辨率图像  $I_{\text{LR}}$  映射为高分辨率重建  $\hat{I}_0 = G_\theta(I_{\text{LR}})$ , 而判别器则区分重建图像与真实高分辨率图像  $I_0$ 。二者通过对抗损失共同训练, 生成器优化目标为

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{I_0}[\log D(I_0)] + \mathbb{E}_{I_{\text{LR}}}[\log(1 - D(G(I_{\text{LR}})))] \quad (2-27)$$

为了兼顾像素精度和感知质量, 通常在此基础上添加内容损失 (如 MSE 或感知损失):

$$\mathcal{L}_G = \mathcal{L}_{\text{GAN}}(G, D) + \lambda \mathcal{L}_{\text{MSE}}(G), \quad (2-28)$$

$$\mathcal{L}_{\text{MSE}}(G) = \frac{1}{N} \sum_i (I_0(i) - G(I_{\text{LR}})(i))^2. \quad (2-29)$$

通过交替更新  $G$  和  $D$ , GAN 能在重建细节和整体真实感之间取得平衡, 从而显著提升超分图像的视觉质量。

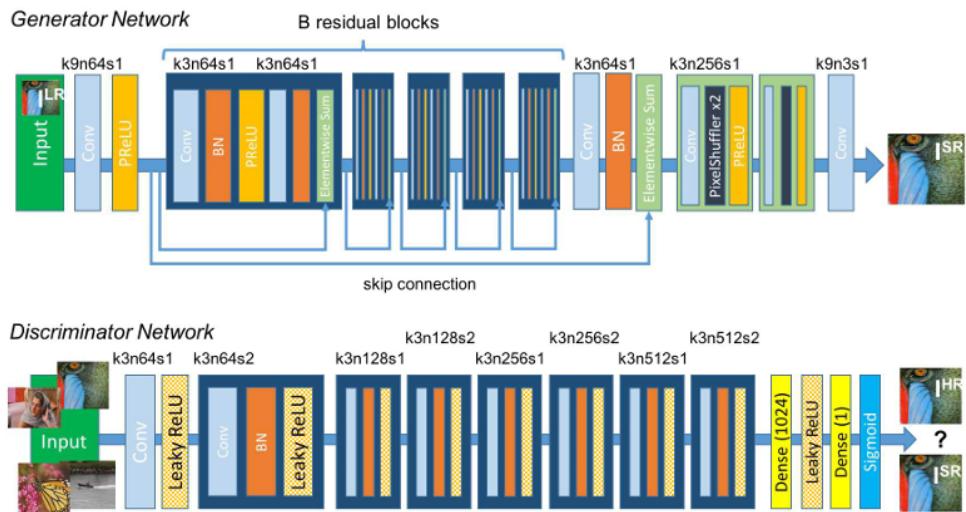


图 2-5 生成对抗网络示意图

### 2.4.3 视觉 Transformer (ViT)

Transformer<sup>[38]</sup>最初用于自然语言处理，近年来被引入图像建模任务。如图 2-6 所示，其核心思想是将图像划分为  $N$  个不重叠 patch（每个大小为  $P \times P$ ），展平后形成序列输入，每个 patch 通过线性映射成维度为  $D$  的 token，形成嵌入矩阵  $\mathbf{X} \in \mathbb{R}^{N \times D}$ ，再与位置编码  $\mathbf{E}_{\text{pos}}$  相加以保持空间结构： $\mathbf{Z}_0 = \mathbf{X} + \mathbf{E}_{\text{pos}}$ 。

Transformer 的每一层均由多头自注意力（Multi-Head Self-Attention, MHSA）和前馈网络（Feed Forward Network, FFN）构成。MHSA 模块计算每个 token 之间的全局相关性，核心注意力机制定义为：

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (2-30)$$

其中  $\mathbf{Q} = \mathbf{Z}\mathbf{W}^Q, \mathbf{K} = \mathbf{Z}\mathbf{W}^K, \mathbf{V} = \mathbf{Z}\mathbf{W}^V$ ， $d_k$  为 key 维度。最终输出图像由最后一层 Transformer 表示经过重建头（如反卷积或 MLP）还原为高分辨率图像  $\hat{I}_0$ 。

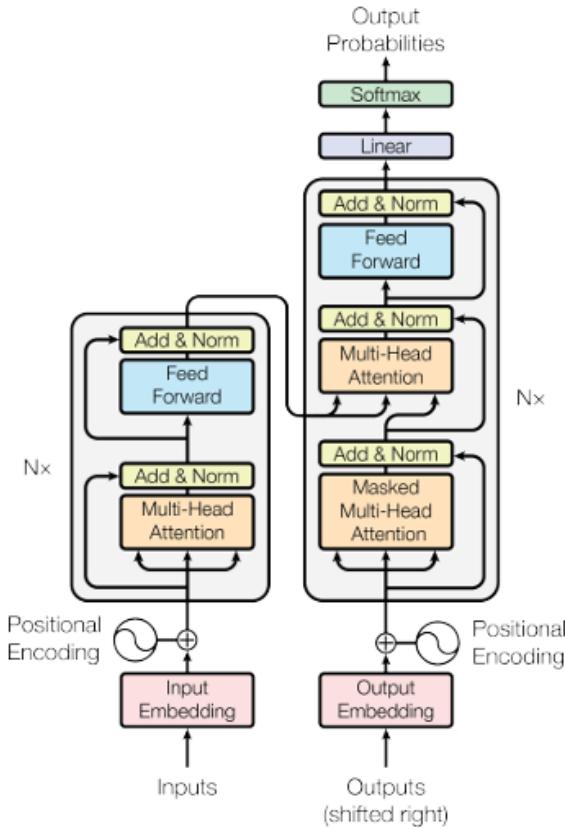


图 2-6 视觉 Transformer 示意图

#### 2.4.4 扩散模型 (Diffusion Model)

扩散模型是一类基于逐步添加噪声和反向去噪的生成模型，近年来在图像超分辨率等任务中取得显著效果。其基本思想如图 2-7 所示：前向过程逐步将高质量图像  $I_0$  添加噪声转变为近似高斯分布，反向过程则学习一个去噪网络逐步还原图像。

前向扩散过程将图像  $I_0$  添加噪声生成序列  $\{x_t\}_{t=0}^T$ ，定义如下：

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t \mathbf{I}) \quad (2-31)$$

其中  $\beta_t$  为时间步  $t$  的噪声强度，通常缓慢增长。该过程整体表示为：

$$q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}) \quad (2-32)$$

反向去噪过程由参数化模型  $p_\theta$  学习估计，形式为：

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (2-33)$$

训练目标是最小化变分下界 (ELBO)，通常简化为噪声预测损失：

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, \epsilon \sim \mathcal{N}(0, \mathbf{I})} \left[ \|\epsilon - \epsilon_\theta(x_t, t)\|^2 \right] \quad (2-34)$$

在超分辨率任务中，给定低分辨图像  $I_{LR}$ ，可作为条件输入对每步去噪过程进行引导，从而恢复高分辨图像  $\hat{I}_0$ 。

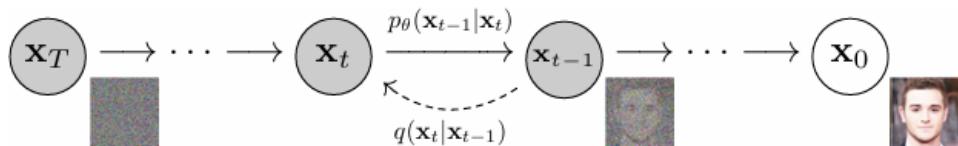


图 2-7 扩散模型生成过程示意图

#### 2.4.5 网络框架设计

综合上述模型架构，为实现更高效的超分任务，现代网络设计趋向于结合不同模块以取长补短。例如 SwinIR 结合了 Transformer 与 CNN 特征提取优势，

Real-ESRGAN 则引入残差结构与对抗训练协同优化。典型的端到端超分网络可分为以下几个模块：

- **特征提取模块**：提取 LR 图像的底层特征；
- **特征增强模块**：使用注意力机制、残差连接或变换器结构增强特征表达；
- **上采样模块**：将特征图重建为目标分辨率（如亚像素卷积、最近邻插值 + 卷积）；
- **重建模块**：输出最终的超分图像。

此外，为进一步提高泛化能力与恢复质量，损失函数设计也趋于多样化，例如联合使用像素损失 (MSE)、感知损失 (VGG)、对抗损失 (GAN) 与边缘损失等，使模型在还原细节的同时更贴近人类感知标准。

#### 2.4.6 工业框架与本实验框架区别：关于 $scale$ 参数的讨论

在多数图像超分辨率任务中，训练数据集通常由原始高分辨率图像 (HR) 通过下采样操作构造低分辨率图像 (LR)，形成 (LR, HR) 对。其中，常用的工业数据集如 DIV2K，即采用 bicubic 插值等方法以固定缩放因子  $scale = 2, 3, 4$  进行降采样。这类数据的生成方式可表述为：

$$I_{\text{LR}} = \mathcal{D}_{\text{scale}}(I_{\text{HR}}) \quad (2-35)$$

其中  $\mathcal{D}_{\text{scale}}(\cdot)$  表示基于缩放因子  $scale$  的下采样算子。训练过程中，模型需学习从  $I_{\text{LR}}$  恢复至原始  $I_{\text{HR}}$  的映射关系：

$$\hat{I}_{\text{HR}} = f_{\theta}(I_{\text{LR}}) \quad (2-36)$$

相比之下，本文所采用的数据集来源于真实显微成像过程：分别使用低分辨率与高分辨率显微镜采集同一样本，从而直接获得 LR 和 HR 图像对。此处的 LR 并非通过插值合成，而是物理采集所得。因此，该任务更贴近实际应用场景，其输入输出映射不可再依赖统一的  $scale$  比例建模。

本文所采用的模型对现有超分辨框架中的数据预处理、模型结构、损失函数等关键部分进行调整，实现对实际  $\text{LR} \rightarrow \text{HR}$  显微映射的端到端建模。

## 2.5 训练策略与参数

### 2.5.1 损失函数 (Loss Function)

在图像超分任务中，模型训练的目标是学习一个从低分辨率图像  $I$  映射至高分辨率图像  $\hat{I}_0$  的函数  $\hat{I}_0 = f_\theta(I)$ 。最常见的损失函数为像素级回归损失，即均方误差（Mean Squared Error, MSE）损失：

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (I_0(i) - \hat{I}_0(i))^2 \quad (2-37)$$

$I_0$  是本征的清晰图片，也就是评价标准图像。MSE 损失在图像重建中具有优化目标明确、收敛稳定的优点，应用广泛；然而由于其以像素差为唯一指标，往往忽略图像的结构信息与感知质量，容易导致图像重建过于平滑、细节丢失。

为进一步提升生成图像的感知质量，本文引入另一种基于概率分布差异度量的损失函数：Jensen-Shannon 散度损失 (JSDiv Loss)，这是衡量两个概率分布相似度的一种对称且有界的指标，相较于另一个常用的损失函数 Kullback-Leibler (KL) 散度，其具有更好的数值稳定性和解释性<sup>[39]</sup>。其数学定义如下：

给定两个概率分布  $P$  和  $Q$ ，其 Jensen-Shannon 散度为：

$$\text{JSD}(P \parallel Q) = \frac{1}{2}\text{KL}(P \parallel M) + \frac{1}{2}\text{KL}(Q \parallel M), \quad (2-38)$$

其中  $M = \frac{1}{2}(P + Q)$  为中间分布， $\text{KL}(P \parallel Q)$  表示 Kullback-Leibler 散度<sup>[40]</sup>。在图像超分任务中，我们可将原图  $I_0$  和重建图  $\hat{I}_0$  通过特征提取函数  $\phi(\cdot)$  映射至高维特征空间，转化为特征分布  $P = \phi(I_0)$  与  $Q = \phi(\hat{I}_0)$ ，从而定义 JSDiv 损失函数为：

$$\mathcal{L}_{\text{JSDiv}} = \text{JSD}(\phi(I_0) \parallel \phi(\hat{I}_0)) = \frac{1}{2}\text{KL}(\phi(I_0) \parallel M') + \frac{1}{2}\text{KL}(\phi(\hat{I}_0) \parallel M') \quad (2-39)$$

与 MSE 不同，JSDiv 更关注生成图像与真实图像在高维感知空间的分布一致性，尤其对纹理细节、结构形态等具有较强的感知能力。因此，本文在不同模型中对 MSE Loss 与 JSDiv Loss 进行对比实验，结果将在实验章节中详细分析。

## 2.5.2 正交归一化 (Orthogonal Regularization)

在图像处理任务中，输入图像像素值的数值范围通常不固定。为了提升实验的可重复性与模型的收敛速度，需对图像进行归一化处理，将其缩放至标准数值区间。本文采用以下两种常见归一化方式进行对比分析：

### (1) [0, 1] 区间归一化

该方法将输入图像的每个像素值线性缩放至 [0, 1] 区间，归一化公式为：

$$I' = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (2-40)$$

其中  $I$  表示原始像素值， $I_{\min}$  和  $I_{\max}$  分别表示图像中的最小值与最大值。

### (2) [-1, 1] 区间归一化

该方法在 [0, 1] 归一化的基础上进一步将数据中心化，使其数值分布更均衡，有助于加快网络收敛速度。其归一化公式为：

$$I'' = 2 \cdot \frac{I - I_{\min}}{I_{\max} - I_{\min}} - 1 \quad (2-41)$$

本文在后续实验中将对比这两种归一化策略在不同模型结构与损失设置下的性能表现，探讨归一化范围对训练收敛速度和重建质量的影响。

## 2.5.3 训练优化器 (Optimizer)

优化器指的是模型在训练过程中通过改变学习率参数变化的大小。本文在训练过程中均采用 Adam 优化器<sup>[41]</sup>，通过自适应一阶与二阶矩估计动态调整学习率，具有良好收敛速度。其更新规则为：

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla \mathcal{L}_t \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla \mathcal{L}_t)^2 \quad \hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (2-42)$$

$$\hat{\theta}_t = \frac{v_t}{1 - \beta_2^t} \quad \theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (2-43)$$

其中  $\eta$  为初始学习率， $\beta_1, \beta_2$  是一阶与二阶动量参数。

## 2.6 本章小结

本章系统性地介绍了本文所涉及的关键技术与理论基础。首先从傅里叶光学角度建立了显微成像的数学模型，分析了衍射极限与系统缺陷导致的分辨率退化机制；其次详细阐述了实验数据集的构建与预处理流程，包括模拟数据集 pySTED 和真实实验数据集（荧光微球与生物纤毛）的采集规范与增强策略；进而详细阐述了本文运用的图像质量评估指标体系，包括常用的 PSNR、SSIM 以及像素级误差指标；随后重点分析了 CNN、GAN、Transformer 和扩散模型等主流超分辨网络架构的技术特点与适用场景，并针对显微图像特性做出了网络框架的适配方案；最后详细说明了训练过程中的优化策略与参数设置，包括损失函数设计、正则化方法和优化器选择。



## 第三章 实验

### 3.1 预实验：不同训练参数对训练结果的影响

本节旨在探索不同训练方法与参数配置对模型性能的影响，并对 CNN 与 Diffusion 模型进行了对比分析。

我们采用控制变量法，分别考察以下三个因素的影响：1) 模型结构 (CNN vs Diffusion)；2) 损失函数 (MSE vs JS)；3) 归一化范围 ( $[0, 1]$  vs  $[-1, 1]$ )。实验评估指标包括：1) 训练所需的 epoch 数以衡量训练速度；2) 图像质量评估指标（详见第 2.3 节）衡量重建效果；3) Loss-Epoch 曲线的斜率衡量模型收敛速度。

实验数据集为 10408 对的 pySTED 数据集以及 1504 对的 Beads 数据集，采用 Adam 优化器，实验结果如图 3-1 所示。

EXP	Exp	DIFFUSION		CNN	
		loss	time	loss	time
pySTED	$[-1,1]$ +MSE	0.0151	1'45	0.0187	8"
	$[0,1]$ +MSE	0.0086	2'08		
Beads	$[0,1]$ +JS	0.0025	1'25		
	$[-1,1]$ +MSE	0.0064	45"		None
	$[0,1]$ +MSE	0.0029	1'6"		
	$[0,1]$ +JS	0.0010	34"		

图 3-1 预实验：不同训练参数对训练结果的影响。第一列表示实验所用数据集，第二列表示归一化范围与 LOSS 方案，第三列、第四列为 DIFFUSION 和 CNN 模型的实验结果。其中 loss 选取训练 100 个 epoch 后的最终 loss，time 中表示训练一个 epoch 所用的平均时间

#### 3.1.1 CNN vs DIFFUSION

对比模型结构：CNN 与 Diffusion，并统一训练参数：损失函数为 MSE，归一化范围为  $[-1, 1]$ 。

结果显示(图 3-1 及 3-2): CNN 的训练速度快约 10 倍，收敛速度略快；从图 3-

2可见，Diffusion 模型在 MAE/NRMSE 上略优，SSIM/MS-SSIM 显著更高，PSNR 也优于 CNN。由不同评估参数的物理意义来看，CNN 训练更高效，Diffusion 拥有更强的模型能力和泛化能力；DIFFUSION 模型的结构还原、人类视觉感知、细节表现和抗失真能力均优于 CNN。

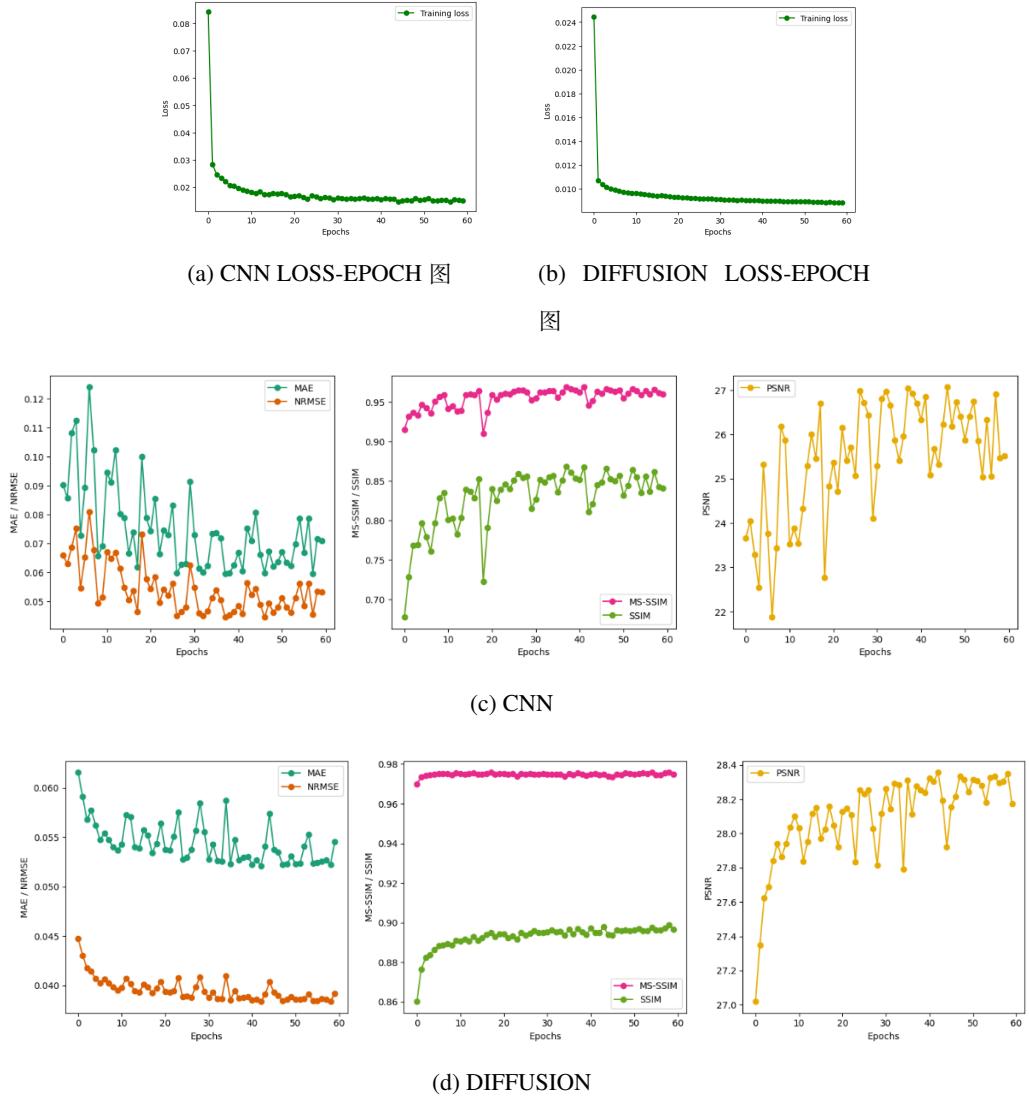


图 3-2 模型 CNN(a),DIFFUSION(b) 模型在训练过程中的 loss 随 epoch 变化图。模型 CNN(c),DIFFUSION(d) 的图像质量评估结果训练效果随 epoch 变化图：第一张图为 MAE 和 NRMSE 指标；第二张图为 MS-SSIM 和 SSIM 指标；第三张图展示了 PSNR 指标。

### 3.1.2 MSE vs JSDEV

对比损失函数：MSE 与 JSDEV，并统一模型结构为 Diffusion，归一化范围为  $[-1, 1]$ 。

结果显示（图 3-1 及 3-3）：JSDEV 的训练速度接近 MSE 的两倍，收敛速度也

更快；从图 3-3 可见，MAE/NRMSE 两者相近，其他 JSDIV 均略优。表明 JSDIV 在训练效率与收敛速度方面具有优势，且在性能上并不逊色。

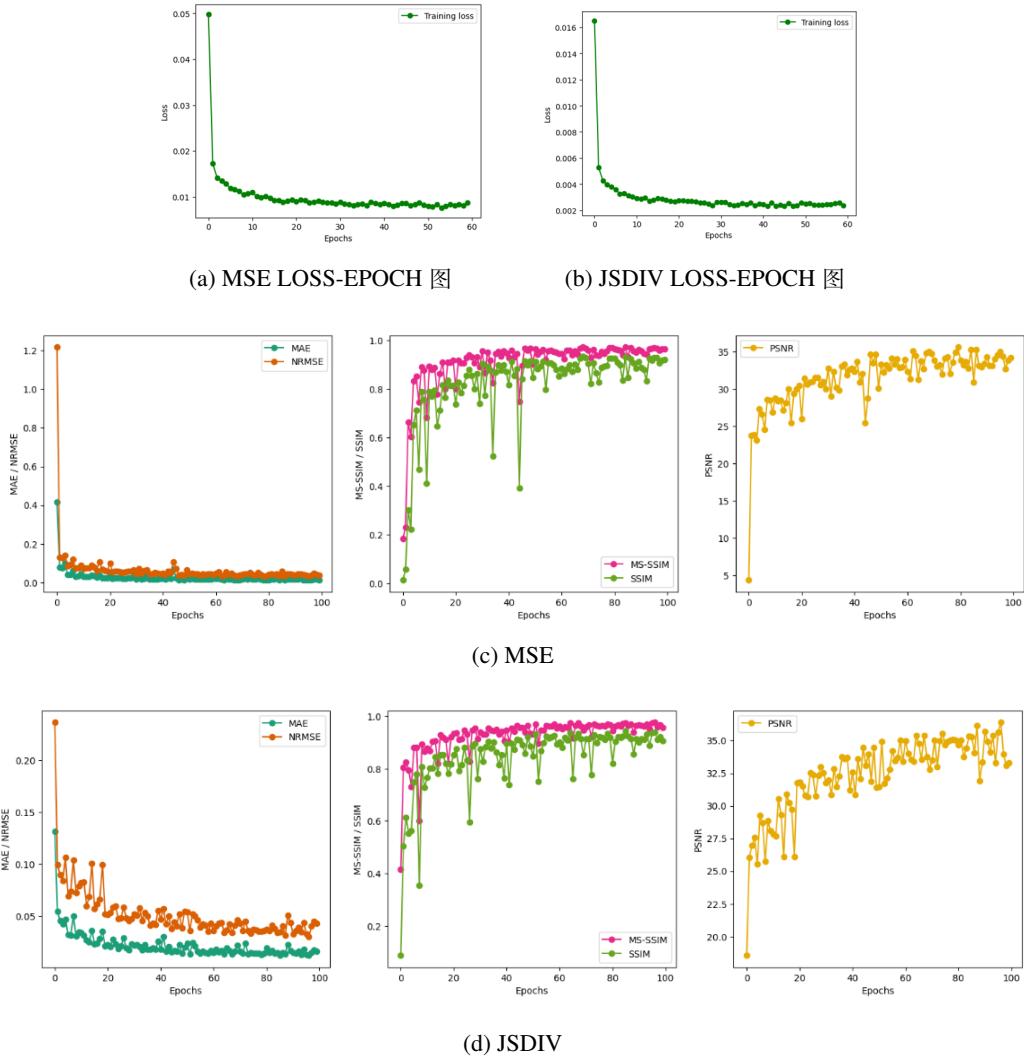


图 3-3 不同损失函数 MSE(a),JSDIV(b) 在训练过程中的 loss 随 epoch 变化图。采用损失函数 MSE(c),JSDIV(d) 的图像质量评估结果训练效果随 epoch 变化图：第一张图为 MAE 和 NRMSE 指标；第二张图为 MS-SSIM 和 SSIM 指标；第三张图展示了 PSNR 指标。

### 3.1.3 归一化： $[0, 1] vs [-1, 1]$

对比归一化范围： $[0, 1]$  与  $[-1, 1]$ ，并统一模型结构为 Diffusion，损失函数为 MSE。

如图 3-1 及 3-4 所示：采用  $[-1, 1]$  归一化训练速度提升约 20%，且收敛速度更快，这是因为  $[-1, 1]$  归一化具有对称性。训练效果如图 3-4 所示，由于归一化范围不同导致的数值尺度差异，各项指标有所差异，但整体效果相近。

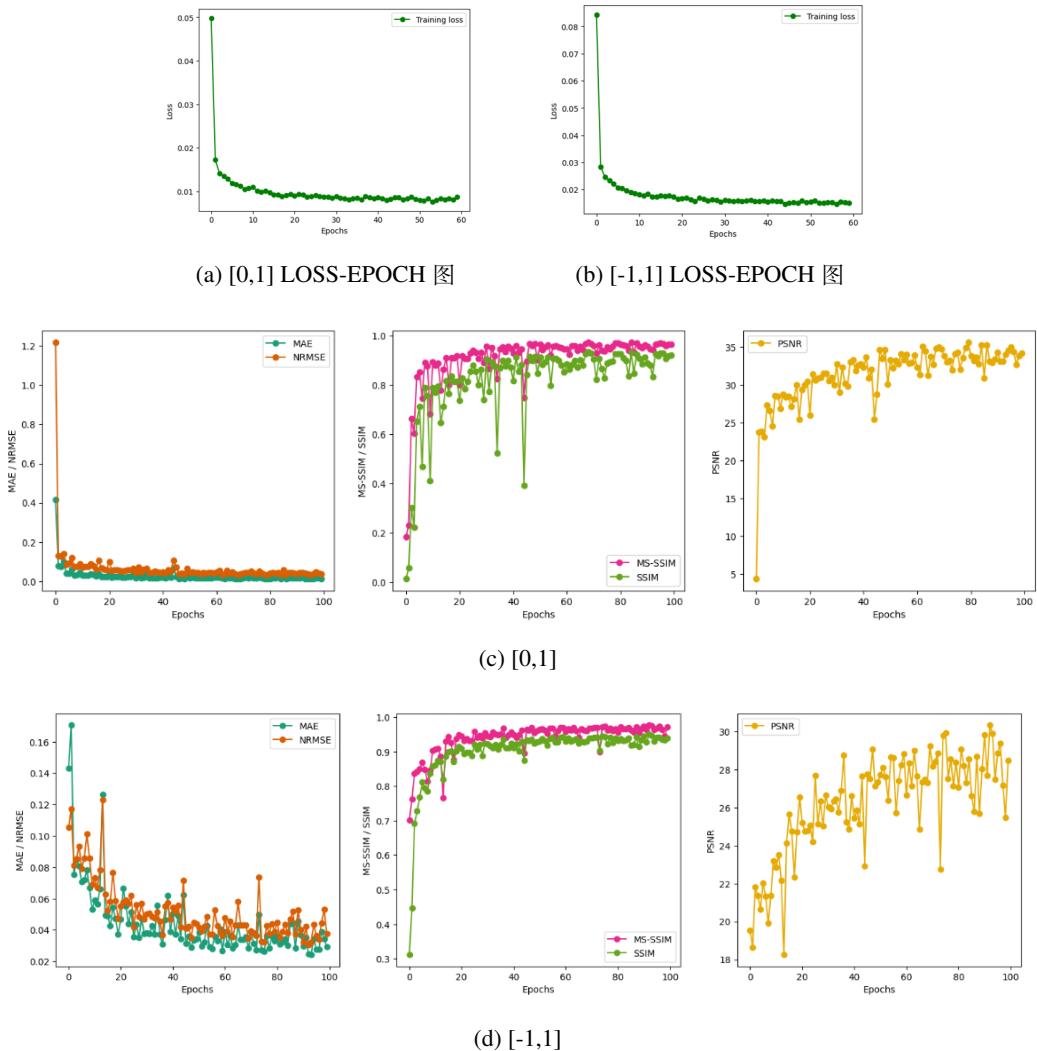


图 3-4 采用不同归一化 [0,1](a),[-1,1](b) 在训练过程中的 loss 随 epoch 变化图。采用归一化 [0,1](c),[-1,1](d) 的图像质量评估结果训练效果随 epoch 变化图：第一张图为 MAE 和 NRMSE 指标；第二张图为 MS-SSIM 和 SSIM 指标；第三张图展示了 PSNR 指标。

### 3.2 实验一：机器学习超分辨率实验

本实验将 DIFFUSION 模型与 SRCNN 模型进行对比实验，采用多种评估指标，评估扩散模型的图像高频细节重建质量、伪影现象以及重建图像可信度。

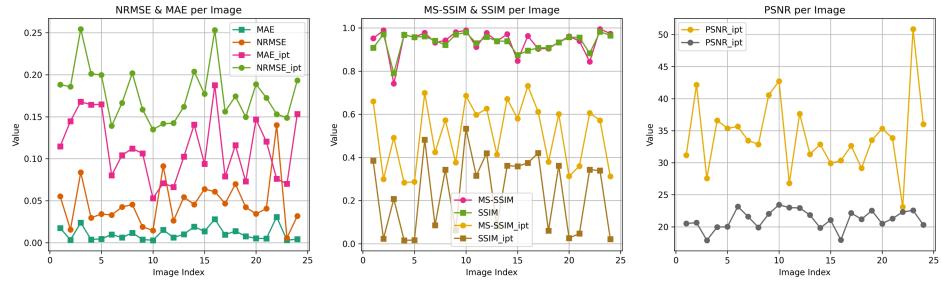
按照预实验的分析结果，本实验训练参数：采用 JSDEV 为损失函数、[-1,1] 为归一化范围、Adam 优化器。

图 3-5a 展示了输入的模糊图像、CNN 重建结果和 Diffusion 重建结果的整体评估情况；图 3-5b 与图 3-5c 展示了测试集中前 25 张图像的详细评估指标。可以看出，两种模型均提升了图像质量，且 Diffusion 的提升幅度更为显著；相比

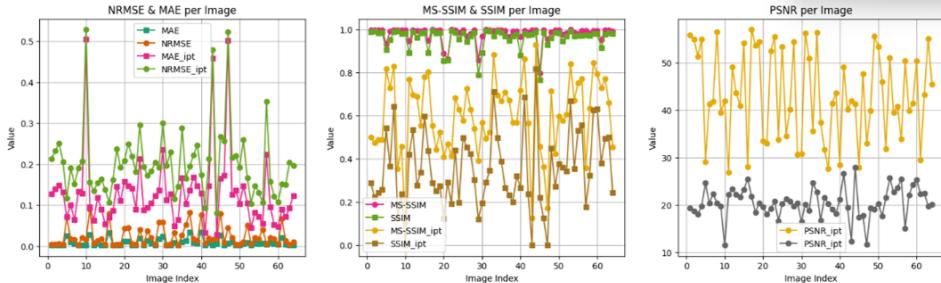
CNN 模型, DIFFUSION 模型平均 PSNR 高 1.9dB, SSIM 提升多 13%, MS-SSIM 提升多 4%。

Datatype	NRMSE	MAE	MS-SSIM	SSIM	PSNR
Input images	0.176968	0.095377	0.913752	0.730675	24.459363
CNN	0.069020	0.047819	0.942188	0.816428 (10%↑)	30.024266
DIFFUSION	0.048614	0.035058	0.972249	0.896244 (23%↑)	31.906473

(a) 输入模糊图像、cnn 重建图像、diffusion 重建图像的评估结果



(b) 测试集前 25 张图像 CNN 重建具体评估参数



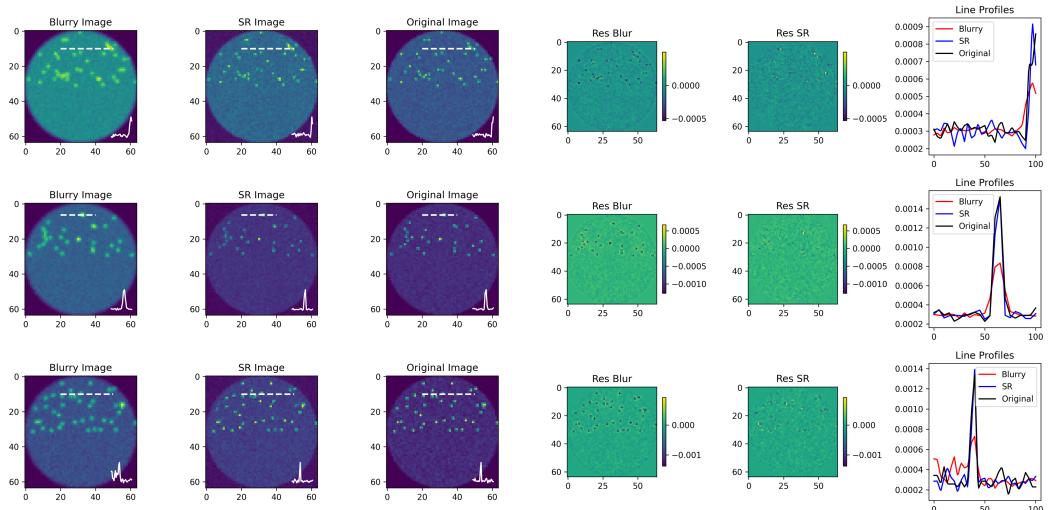
(c) 测试集前 25 张图像 DIFFUSION 重建的具体评估参数

图 3-5 实验 1: cnn 与 diffusion 模拟数据训练结果图表展示。其中图 a 为直观数据表, 图 b、c 展示了测试集前 25 图像的具体评估参数; ipt 表示输入模糊图像的评估参数结果

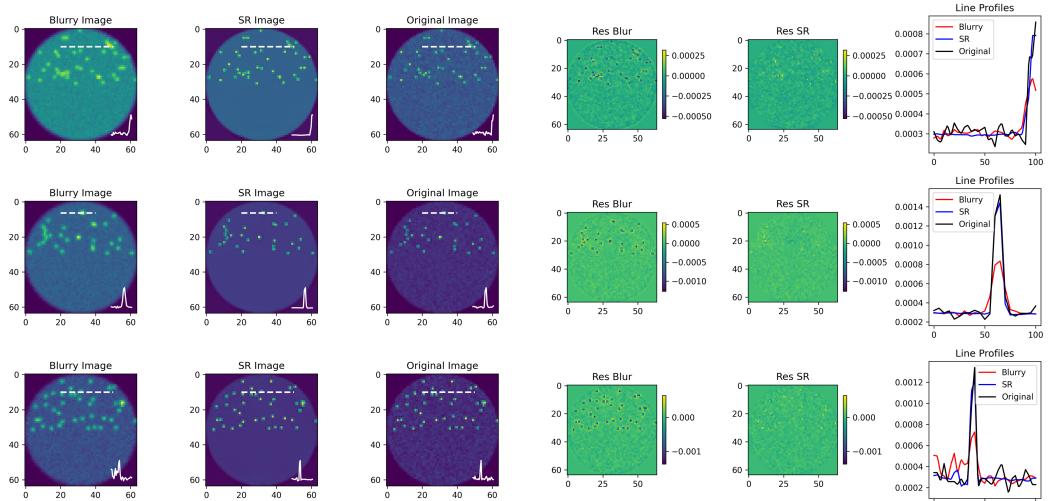
图 3-6a 与图 3-6b 展示了具体的图像重建效果。结果表明, CNN 与 Diffusion 模型均成功完成了模拟数据的超分辨重建任务, 且评估结果优良; 其中, Diffusion 的视觉和量化指标表现更为优异。此外, 未观察到明显的伪影现象, 说明两种方法在生成结果的真实性与稳定性方面均较为可靠。

### 3.3 实验二：多源数据融合实验

本实验从数据集入手, 基于多源数据融合的思想, 探讨使用模拟数据训练的模型对真实实验数据的重建能力, 并评估引入少量真实数据进行微调后模型性能的变化。具体实验包括: 实验 2-1 (荧光小球)、实验 2-2 (纤毛) 以及实验 2-3 (少量真实数据微调)。



(a) CNN 模型重建效果图像



(b) DIFFUSION 模型重建效果图像

**图 3-6 实验 1: CNN 与 DIFFUSION 模拟数据训练结果展示。**对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ; 第二列展示经模型重建后的图像  $\hat{I}_0$ ; 第三列展示高分辨清晰图像 (STED)  $I_0$ ; 第四列展示  $I$  与  $I_0$  的残差图; 第五列展示  $\hat{I}_0$  与  $I_0$  的残差图; 第六列展示基于前三列图像中白线划过范围内图像数值变化图

按照预实验的分析结果，本实验训练参数：采用 JSDEV 为损失函数、[-1,1] 为归一化范围、Adam 优化器。

### 3.3.1 实验 2-1：荧光小球

本实验使用 pySTED 模拟训练集进行训练，使用 Beads 真实实验数据集进行测试，图 3-7a 展示了输入模糊图像、CNN 重建图像与 Diffusion 重建图像的整体评估结果。对于 SSIM 指标，使用 CNN 模型后 SSIM 提升了 11%，使用 DIFFUSION 模型后提升了 34%。可见，两种模型均对图像质量有所提升，Diffusion 的提升幅度更为显著。

图 3-7b 与图 3-7c 展示了重建效果的图像对比。结果表明，Diffusion 模型能够完成超出训练数据分布范围的图像重建任务；相比之下，CNN 重建结果中较易出现伪影。综上，Diffusion 模型在图像结构复原中表现出较强的反卷积能力，能够完成对超出训练范围图像的重建任务。

### 3.3.2 实验 2-2：纤毛

本实验使用 pySTED 模拟训练集进行训练，使用 Cilia 真实实验数据集进行测试。图 3-8a 展示了输入模糊图像、CNN 重建图像和 Diffusion 重建图像的评估对比。对于 SSIM 指标，使用 CNN 模型重建后 SSIM 下降了 21%，而使用 DIFFUSION 模型重建后提升了 10%。图 3-8b 与图 3-8c 展示了具体的重建结果。Diffusion 模型再次展现出其对非训练集图像的泛化能力，而使用 CNN 模型的重建图像不仅产生了较多的伪影，大幅度的 SSIM 下降展示其无法完成对纤毛的重建任务。

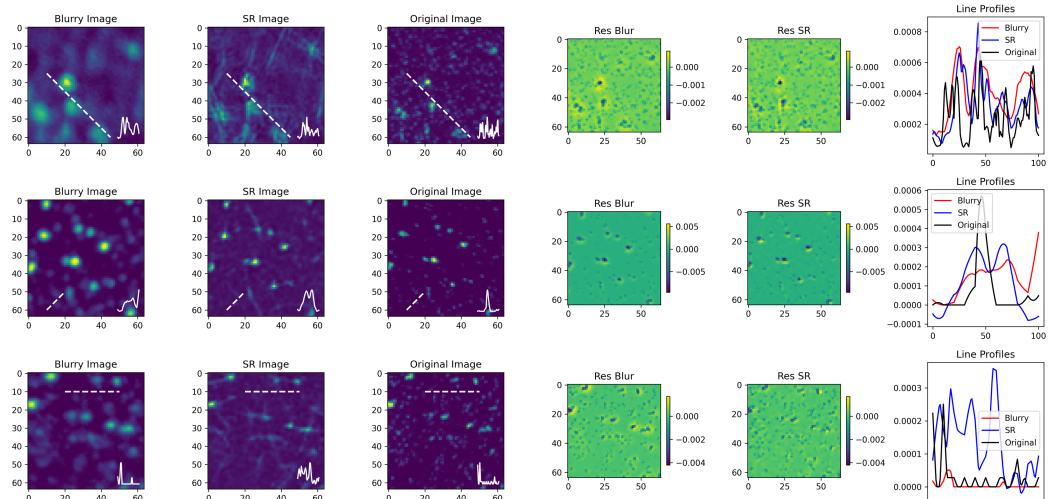
综上，Diffusion 仍然能够完成对超出训练范围图像的重建任务，展现出比较强的泛化能力。

### 3.3.3 2-3：实验数据微调

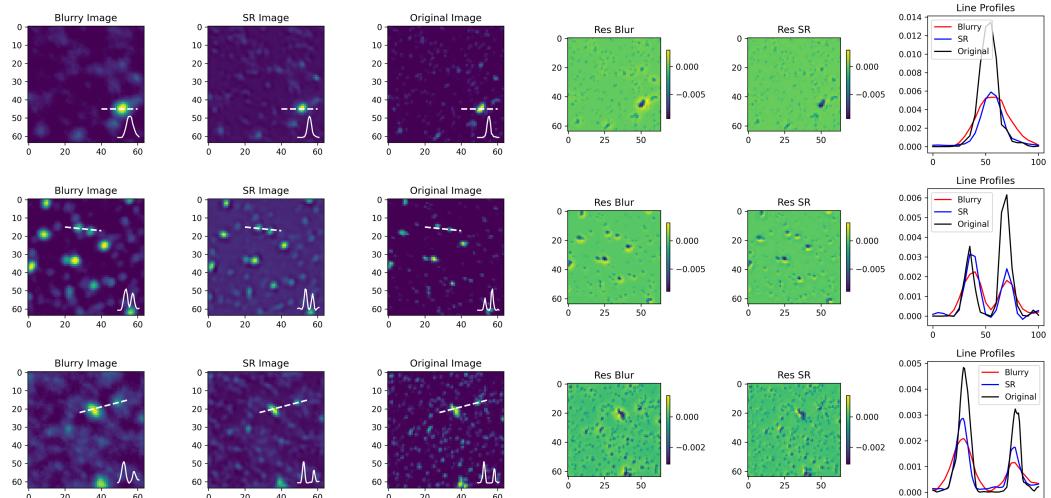
本实验在实验 2-1 基础上，进一步使用 100 对纤毛实验数据进行微调再训练，并与实验 1 的结果进行对比，以验证多源数据融合是否能够增强模型对真实图像的重建性能。

Datatype	NRMSE	MAE	MS-SSIM	SSIM	PSNR
Input images	0.123724	0.088622	0.761519	0.512868	24.912680
CNN	0.091702	0.068716	0.808208	0.568635 (11%↑)	31.906473
DIFFUSION	0.062927	0.036879	0.903208	0.684188 (34%↑)	31.906473

(a) 输入模糊图像、cnn 重建图像、diffusion 重建图像的评估结果



(b) CNN 模型重建效果图像

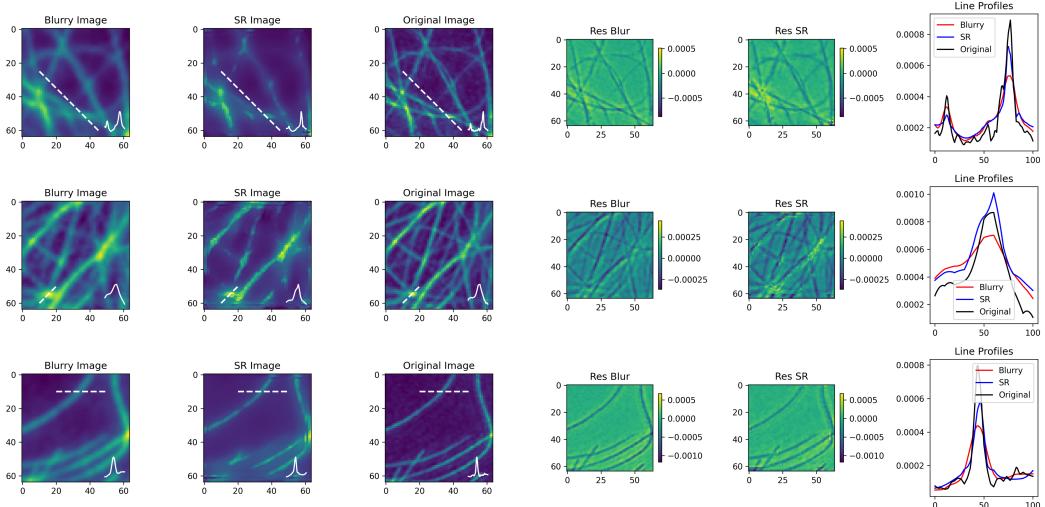


(c) DIFFUSION 模型重建效果图像

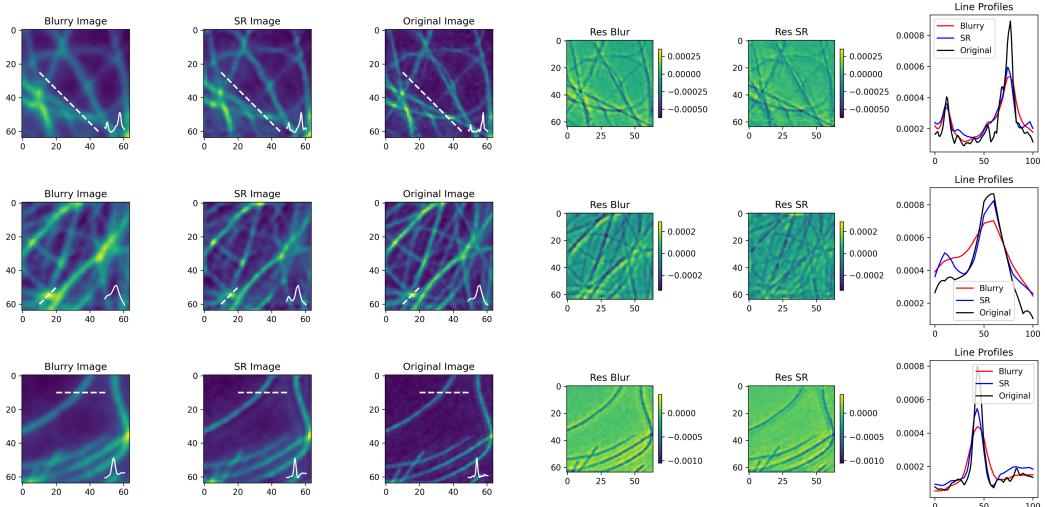
图 3-7 实验 2-1：使用 pySTED 作为训练集，CNN 与 DIFFUSION 对荧光小球测试结果。对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ；第二列展示经模型重建后的图像  $\hat{I}$ ；第三列展示高分辨清晰图像 (STED)  $I_0$ ；第四列展示  $I$  与  $I_0$  的残差图；第五列展示  $\hat{I}$  与  $I_0$  的残差图；第六列展示基于前三列图像中白线划过范围内图像数值变化图

Datatype	NRMSE	MAE	MS-SSIM	SSIM	PSNR
Input images	0.115086	0.089774	0.915862	0.737833	25.082701
CNN	0.137258	0.095564	0.752670	0.576953 (21%↓)	23.516902
DIFFUSION	0.069973	0.049244	0.938330	0.807728 (10%↑)	29.801025

(a) 输入模糊图像、cnn 重建图像、diffusion 重建图像的评估结果



(b) CNN



(c) DIFFUSION

图 3-8 实验 2-2：使用 pySTED 作为训练集，CNN 与 DIFFUSION 对纤毛测试结果。对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ；第二列展示经模型重建后的图像  $\hat{I}_0$ ；第三列展示高分辨清晰图像 (STED)  $I_0$ ；第四列展示  $I$  与  $I_0$  的残差图；第五列展示  $\hat{I}_0$  与  $I_0$  的残差图；第六列展示基于前三列图像中白线划过范围内图像数值变化图

图 3-9a 显示了微调后模型对输入模糊图像的重建评估结果。对于 SSIM 指标, CNN 多源重建后提升 9%, DIFFUSION 多源重建后提升 23%, 这与实验一中使用对 pySTED 测试集进行测试的 10% 与 23% 几乎一致。说明对于使用 10000 对模拟数据集训练的通用模型, 在应用时只需 100 对图片进行针对性的微调, 就能使得微调后的模型对其它形态的图片具有很强的超分辨能力。

再讨论伪影问题。图 3-9b 与图 3-9c 展示了具体图像重建效果。微调后的两个模型均高质量完成了新的超分辨任务, Diffusion 模型优势更为显著, 且未观察到明显伪影, 显示出良好的泛化能力与重建稳定性。

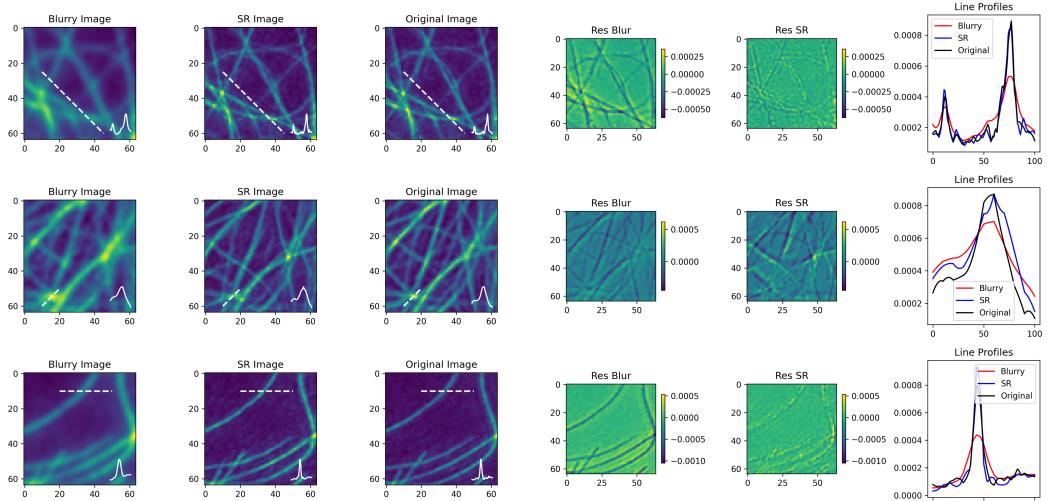
### 3.3.4 本章小结

本章通过系统性的实验设计, 验证了扩散模型在显微图像超分辨重建任务中的有效性。在预实验阶段, 我们通过控制变量法深入分析了不同训练参数对模型性能的影响, 确立了最优的训练策略; 实验一重点对比了 CNN 与扩散模型在模拟数据上的表现, 证实了扩散模型在细节还原和感知质量方面的显著优势; 实验二进一步探讨了模型在真实实验数据上的泛化能力, 创新性地提出并验证了“模拟数据预训练 + 少量真实数据微调”的多源数据融合策略。

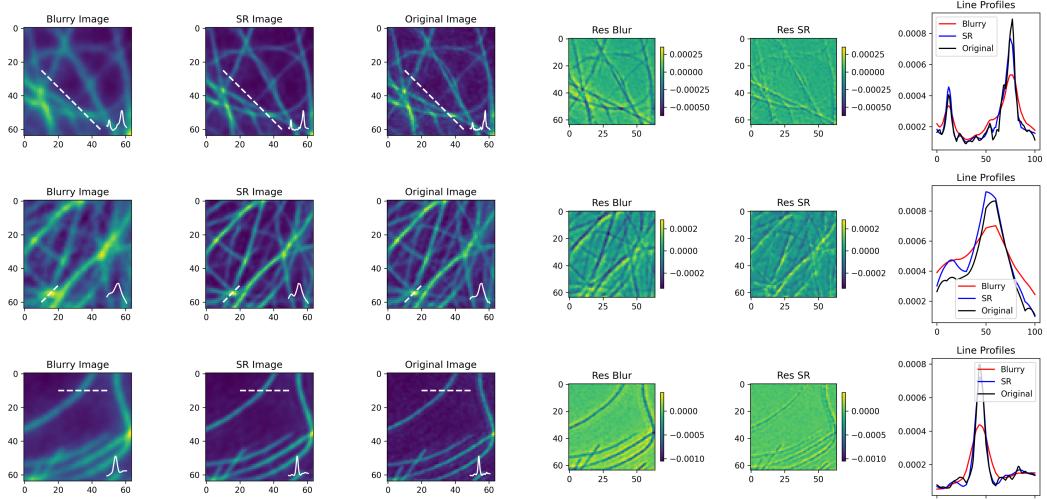
实验结果表明: 1) 扩散模型在各项评估指标上均优于传统 CNN 方法, 尤其在结构相似性和高频细节保留方面表现突出; 2) 基于物理建模的模拟数据能够有效支持模型学习超分辨映射关系, 即使面对真实样本也展现出良好的泛化能力; 3) 引入少量真实数据进行微调可显著提升模型在目标域的表现, 验证了多源数据融合策略的实用价值。这些发现为在实验成本受限条件下实现高质量显微图像重建提供了可行的技术路径, 同时也为深度学习在计算光学成像领域的应用提供了新的思路和方法参考。

Datatype	NRMSE	MAE	MS-SSIM	SSIM	PSNR
Input images	0.176968	0.095377	0.913752	0.730675	24.459363
CNN	0.069973	0.049244	0.938330	0.807728 (9%↑)	29.801025
DIFFUSION	0.048614	0.035058	0.972249	0.896244 (23%↑)	31.906473

(a) 输入模糊图像、cnn 重建图像、diffusion 重建图像的评估结果



(b) CNN



(c) DIFFUSION

图 3-9 实验 2-3：使用 pySTED 作为训练集，极小占比的 Cilia 数据集进行微调再训练后，CNN 与 DIFFUSION 对荧光小球测试结果。对于每一个模型，展示了相同的三张测试集图像的重建效果。第一列展示低分辨输入图像 (Confocal)  $I$ ；第二列展示经模型重建后的图像  $\hat{I}_0$ ；第三列展示高分辨清晰图像 (STED)  $I_0$ ；第四列展示  $I$  与  $I_0$  的残差图；第五列展示  $\hat{I}_0$  与  $I_0$  的残差图；第六列展示基于前三列图像中白线划过范围内图像数值变化图



## 第四章 总结与展望

### 4.1 研究总结

本研究面向超分辨显微成像任务，针对传统光学分辨率受限与高精度成像成本过高的问题，结合当前深度学习技术的发展，设计并实现了一套以机器学习为核心的超分辨图像重建实验流程，初步达成了提升低分辨图像空间解析度的研究目标。本课题主要完成了以下几方面的工作：

1. 系统梳理并分析了超分辨显微成像的研究背景、基本原理及发展现状，介绍了不同类型的超分模型在显微成像任务中的适用性与局限性，明确了本研究聚焦的问题方向。
2. 结合成像系统的物理建模，从傅里叶变换视角出发，对分辨率退化的主要影响因素进行理论剖析，明确干扰因素与图像模糊之间的逻辑关联，为后续模型设计提供了物理支撑。
3. 构建了涵盖模拟数据（pySTED）与真实显微图像的双数据集评估框架，并对图像进行规范化处理，为实验训练提供了统一的数据基础。
4. 在实验设计方面，实施多组实验探讨了不同模型结构（CNN与扩散模型）、损失函数类型（MSE与JS散度）、归一化方式（[0,1]与[-1,1]）等超参数选择对训练效果的影响。在多源数据融合实验中，验证了扩散模型在高频结构还原、跨域泛化与细节保持等方面的优越性，形成了一套具备实践价值的超分建模策略。
5. 在评估方面，选用了PSNR、SSIM、MS-SSIM、MAE等多种图像重建指标，对模型性能进行了全面评价，确保模型效果既有全局一致性，也有局部细节保持能力。

### 4.2 不足与展望

尽管本研究在显微图像超分辨重建方面取得了阶段性成果，但在模型设计、数据获取与评价体系等方面仍存在一定不足，未来有待进一步优化与拓展。具

体展望如下：

1. 数据获取维度有限：目前实验主要基于外部模拟数据和有限的真实样本，在数据质量与场景多样性方面尚有不足。未来可考虑引入更多高质量、高动态范围的显微图像数据，特别是来自不同成像技术（如 STORM、PALM、SIM）的实拍样本，以增强模型的泛化能力。
2. 模型结构仍可优化：虽然本文尝试了 Diffusion 模型等先进架构，但在重建速度、显存效率与推理稳定性上仍存在优化空间。未来可探索轻量化 Transformer 结构、多尺度注意力融合机制、以及物理引导学习策略等新型方法，以兼顾重建质量与资源效率。
3. 物理一致性建模不足：目前模型主要通过数据驱动方式实现超分重建，尚未充分融合成像系统的物理过程信息。未来可进一步引入成像点扩散函数 (PSF) 建模、光学衍射模型等作为网络先验，提升模型对复杂成像条件的适应性。
4. 评价指标不够充分：尽管采用了常见的图像质量指标，但仍偏重数值层面，缺乏面向实际应用的感知质量评估。未来可结合人类视觉系统模型、微观结构识别准确率等下游指标，构建更符合显微分析需求的多维度评估体系。

## 参考文献

- [1] 毛峥乐, 王琛, 程亚, 等. 超分辨远场生物荧光成像——突破光学衍射极限[J]. Chinese Journal of Lasers, 2008, 35(9): 1283-1307.
- [2] 乌拉, 郑玉祥. 突破衍射极限的成像方法综述[J]. 光学仪器, 2017, 39(1): 81-87.
- [3] JACOBSEN C. Soft x-ray microscopy[J]. Trends in cell biology, 1999, 9(2): 44-47.
- [4] KNOLL M, RUSKA E. Das elektronenmikroskop[J]. Zeitschrift für physik, 1932, 78: 318-339.
- [5] BINNIG G, ROHRER H, GERBER C, et al. Surface studies by scanning tunneling microscopy[J]. Physical review letters, 1982, 49(1): 57.
- [6] BINNIG G, QUATE C F, GERBER C. Atomic force microscope[J]. Physical review letters, 1986, 56(9): 930.
- [7] HELL S W, WICHHMANN J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy[J]. Optics letters, 1994, 19(11): 780-782.
- [8] GUSTAFSSON M G. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy[J]. Journal of microscopy, 2000, 198(2): 82-87.
- [9] BETZIG E, PATTERSON G H, SOUGRAT R, et al. Imaging intracellular fluorescent proteins at nanometer resolution[J]. science, 2006, 313(5793): 1642-1645.
- [10] RUST M J, BATES M, ZHUANG X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)[J]. Nature methods, 2006, 3(10):

793-796.

- [11] KOENIG K, KRASIEVA T B, BAUER E, et al. Cell damage by UVA radiation of a mercury microscopy lamp probed by autofluorescence modifications, cloning assay, and comet assay[J]. *Journal of biomedical optics*, 1996, 1(2): 217-222.
- [12] HUANG B, WANG W, BATES M, et al. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy[J]. *Science*, 2008, 319(5864): 810-813.
- [13] DANI A, HUANG B, BERGAN J, et al. Superresolution imaging of chemical synapses in the brain[J]. *Neuron*, 2010, 68(5): 843-856.
- [14] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. *Advances in neural information processing systems*, 2020, 33: 6840-6851.
- [15] CHUNG H, LEE S, YE J C. Decomposed diffusion sampler for accelerating large-scale inverse problems[J]. *arXiv preprint arXiv:2303.05754*, 2023.
- [16] CHEN H, WANG Y, GUO T, et al. Pre-trained image processing transformer [C] / / Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 12299-12310.
- [17] 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.
- [18] LU Z, LI J, LIU H, et al. Transformer for single image super-resolution[C] / / Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 457-466.
- [19] ZHANG X, ZENG H, GUO S, et al. Efficient long-range attention network for image super-resolution[C] / / European conference on computer vision. 2022: 649-667.
- [20] GAO G, WANG Z, LI J, et al. Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive transformer[J]. *arXiv preprint arXiv:2204.13286*, 2022.

- [21] 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.
- [22] WANG X, YU K, WU S, et al. Esrgan: Enhanced super-resolution generative adversarial networks[C] / / Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0–0.
- [23] LEE O Y, SHIN Y H, KIM J O. Multi-perspective discriminators-based generative adversarial network for image super resolution[J]. IEEE Access, 2019, 7: 136496-136510.
- [24] SAJJADI M S, SCHOLKOPF B, HIRSCH M. Enhancenet: Single image super-resolution through automated texture synthesis[C] / / Proceedings of the IEEE international conference on computer vision. 2017: 4491-4500.
- [25] 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.
- [26] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C] / / Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.
- [27] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[C] / / Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2472-2481.
- [28] TONG T, LI G, LIU X, et al. Image super-resolution using dense skip connections[C] / / Proceedings of the IEEE international conference on computer vision. 2017: 4799-4807.
- [29] WANG H, RIVENSON Y, JIN Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy[J]. Nature methods, 2019, 16(1): 103-110.
- [30] NEHME E, WEISS L E, MICHAELI T, et al. Deep-STORM: super-resolution

- single-molecule microscopy by deep learning[J]. Optica, 2018, 5(4): 458-464.
- [31] OUYANG W, ARISTOV A, LELEK M, et al. Deep learning massively accelerates super-resolution localization microscopy[J]. Nature biotechnology, 2018, 36(5): 460-468.
- [32] QIAO C, LI D, GUO Y, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy[J]. Nature methods, 2021, 18(2): 194-202.
- [33] BILODEAU A, MICHAUD-GAGNON A, CHABBERT J, et al. Development of AI-assisted microscopy frameworks through realistic simulation with pySTED [J]. Nature Machine Intelligence, 2024, 6(10): 1197-1215.
- [34] WANG Z, CHEN J, HOI S C. Deep learning for image super-resolution: A survey [J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 43(10): 3365-3387.
- [35] 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.
- [36] SHEIKH H R, SABIR M F, BOVIK A C. A statistical evaluation of recent full reference image quality assessment algorithms[J]. IEEE Transactions on image processing, 2006, 15(11): 3440-3451.
- [37] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C] / /The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003: vol. 2. 2003: 1398-1402.
- [38] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [39] MENÉNDEZ M L, PARDO J A, PARDO L, et al. The jensen-shannon divergence [J]. Journal of the Franklin Institute, 1997, 334(2): 307-318.
- [40] VAN ERVEN T, HARREMOS P. Rényi divergence and Kullback-Leibler divergence[J]. IEEE Transactions on Information Theory, 2014, 60(7): 3797-3820.

- [41] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.



## 致 谢

本科四年的时光过的很快，在此论文完成之际，我想对所有给予过我关怀与帮助的人表示衷心的感谢。

感谢我的导师李文飞老师。大二上李老师的《计算物理》课，我第一次接触机器学习，我以课程小组作业的契机跟随李老师探索并学习新的技术。李老师人很平和，学术思维敏锐，总能敏锐地发现关键问题。他不仅是老师，更像一位朋友，这些年给了我很多的关心和鼓励，带着我这个科研小白，一点一点学会了怎么搞科研、怎么端正态度。在此前的研究方向以及本次毕业论文的撰写过程中，李老师都为我提供了许多建设性意见，让我能够明确方向、理清思路。

感谢和我一同探索机器学习超分辨技术的组员：林杨，陈亦安，王梓烨。我们一同踏入科研的道路，应该都忘不了共同探讨未知问题的激动，忘不了提出新的办法使结果取得进展时的兴奋，还有闲暇之余在烧烤摊放松、探讨时的惬意。朋友让科研，或者是探索的道路不再无聊。

还有很多想感谢的同学和老师，但是实在不想写致谢了，打字力竭。

最后，谨以本文献给养育我多年的父母与家人，感谢他们多年来对我的理解与支持，他们最无私的爱和付出，一直激励我能够积极乐观地面对学习和生活。

另外，感谢 LUG@NJU 的工作，这是一个好用的模板，让我能如此快速地完成一篇精美的论文。

