

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)02-0609-19

论文引用格式: Li Y, Huang J, Zou Q and Li Q Q. 2026. Multifrequency phase unwrapping and 3D imaging based on a conditional diffusion model. Journal of Image and Graphics, 31(2):0609-0627(李妍, 黄霁, 邹勤, 李清泉. 2026. 利用条件扩散模型的多频相位解包裹与三维成像. 中国图象图形学报, 31(2):0609-0627)[DOI:10.11834/jig.250133]

利用条件扩散模型的多频相位解包裹与三维成像

李妍^{1,2}, 黄霁³, 邹勤^{1,4*}, 李清泉⁴

1. 武汉大学计算机学院, 武汉 430072; 2. 武汉大学国家网络安全学院, 武汉 430079;
3. 中国科学技术大学计算机学院, 合肥 230026; 4. 人工智能与数字经济广东省实验室(深圳), 深圳 518107

摘要: 目的 相位解包裹是从受限周期范围且受噪声干扰的相位信息中恢复连续相位的技术, 是高精度结构光三维成像的关键步骤。受设备误差和环境干扰影响, 相位图易受噪声污染甚至出现大范围跳跃, 给三维成像带来困难。去噪扩散模型在图像生成方面表现突出, 但其主要面向自然图像生成, 难以保证几何模型精度, 无法直接应用于相位解包裹与三维重建。此外, 现有方法多依赖单频包裹相位, 难以兼顾全局结构与局部细节。**方法** 提出一种基于条件扩散模型的多频相位解包裹方法(DiffPhase), 结合三维成像实现精确的绝对相位重建。该方法将相位解包裹建模为条件引导生成任务, 通过构建与扩散网络对齐的多尺度特征提取模块, 并引入跨尺度交叉注意力结构, 将包裹相位特征逐步融合到扩散过程, 提升局部精度与全局一致性。训练采用两阶段策略, 先预训练特征提取模块学习结构先验, 再进行端到端优化以增强预测性能。同时设计自适应多频输入机制, 有效结合低频全局轮廓与高频局部细节, 抑制误差传播并提升鲁棒性。**结果** 在RME-multi(random matrix enlargement-multifrequency)和MoGR-multi(mixture of Gaussians with ramp-multifrequency)仿真数据集上, 本文方法的归一化均方根误差分别为0.23%、0.24%; 在NYU-phase(New York University-phase)和MS-phase(middlebury stereo-phase)真实数据集上分别为4.69%、7.50%, 优于对比的8种深度学习及传统方法。在复杂场景中, 该方法能在强噪声与遮挡下保持较高精度, 尤其在细节边缘与复杂结构区域表现更优。**结论** DiffPhase方法充分利用扩散模型的条件生成与全局建模能力, 能够在高噪声、高复杂度场景下获得准确稳健的解包裹结果, 有效提升三维重建精度与鲁棒性。

关键词: 相位解包裹; 扩散模型; 三维建模; 深度学习; 计算机视觉

Multifrequency phase unwrapping and 3D imaging based on a conditional diffusion model

Li Yan^{1,2}, Huang Ji³, Zou Qin^{1,4*}, Li Qingquan⁴

1. School of Computer Science, Wuhan University, Wuhan 430072, China; 2. School of Cyber Science and Engineering, Wuhan University, Wuhan 430079, China; 3. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China; 4. Laboratory of Artificial Intelligence and Digital Economy (Shenzhen), Shenzhen 518107, China

Abstract: Objective Phase unwrapping is a signal processing technique for recovering actual continuous phase values from phase information constrained by a periodic range and affected by noise. It represents a fundamental step in high-precision structured light 3D imaging, playing a pivotal role in ensuring the accuracy and robustness of the reconstruction

收稿日期: 2025-04-07; 修回日期: 2025-08-28; 预印本日期: 2025-09-23

* 通信作者: 邹勤 qzou@whu.edu.cn

基金项目: 国家自然科学基金项目(62171324); 人工智能与数字经济广东省实验室(深圳)开放基金项目(GML-KF-24-35)

Supported by: National Natural Science Foundation of China (62171324); Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (Shenzhen)(GML-KF-24-35)

process. This technique finds broad application in fields such as optical interferometry, synthetic aperture radar, and magnetic resonance imaging. Owing to the limitations of the measuring principle, the phases captured in real-world scenarios are “wrapped” into multiple cycles. When the variation in the measured phase exceeds 2π , the excess phase is “wrapped back” to 0, known as the wrapping phenomenon. Phase unwrapping techniques aim to restore the wrapped phase data, which span multiple cycles, to their original true phase. This restoration allows for the accurate retrieval of spatial information needed for precise analysis and computation. Because of equipment errors and environmental interference, the measured phase is often contaminated by noise in practical scenarios, sometimes even exhibiting noncontinuous phase jumps, which poses a significant challenge to accurate phase unwrapping and 3D imaging. If the unwrapping method cannot effectively model the phase globally, local jumps can lead to error propagation and result in significant unwrapping errors. Diffusion models based on denoising mechanisms have demonstrated impressive performance in high-quality image generation and shown potential in phase unwrapping. By iteratively denoising noisy images, diffusion models can effectively model large-scale phase relationships and mitigate the impact of local errors on the overall result. However, as generative models primarily designed for generating natural images, existing diffusion models focus on balancing clarity and diversity, making it challenging to ensure the precision of geometric models. Moreover, phase unwrapping is a rigorous mathematical problem governed by periodic and physical constraints, which diffusion models are not inherently equipped to handle. Therefore, the direct application of diffusion models to phase unwrapping and 3D imaging remains problematic, which requires accurate recovery of continuous phase values under strict physical constraints. Existing phase unwrapping methods also rely solely on a single-frequency wrapped phase, which struggles to simultaneously recover large-scale information and fine details of the true phase, thus affecting the high-precision reconstruction of 3D scenes. Obtaining wrapped phase at additional frequencies incurs minimal cost in many scenarios. Low-frequency wrapped phases capture large phase jumps but lose local details, while high-frequency wrapped phases preserve details but struggle with large jumps. To address the above issues, this study proposes DiffPhase, a multifrequency phase unwrapping method based on a conditional diffusion model. The proposed method can take any number of different frequency wrapped phases as input and uses multifrequency phase features as conditions. Leveraging the powerful generative capabilities and global modeling ability of the diffusion model, DiffPhase can precisely restore the phase values. In combination with 3D imaging, DiffPhase enables accurate reconstruction of the absolute phase and scene depth information. **Method** We model phase unwrapping as a conditionally guided image generation task. By incorporating multiscale feature conditioning and diffusion-based generative mechanisms, DiffPhase enhances robustness against noise and occlusions. DiffPhase consists of a feature extraction module and a generation module. In the feature extraction module, we design a multiscale feature extraction module aligned with the diffusion network architecture to extract hierarchical semantic information from the input wrapped phase. The features extracted are progressively integrated into each stage of the diffusion process by leveraging a cross-scale cross-attention mechanism based on multihead attention, thereby enhancing the local accuracy and global consistency of the generated results. The generation process is guided by the conditions provided by the feature extraction part, progressively denoising the initial Gaussian noise and generating the predicted true phase. Then, the denoising process is performed step by step on the reduced features to learn the reverse noise distribution. The interaction between feature extraction and generation parts is realized through the cross-attention layer in the diffusion model. A two-stage training strategy is adopted to improve training stability and generalization capability. The feature extraction module is first pretrained to learn structural priors, followed by end-to-end joint optimization to enhance phase prediction performance. To optimize computational efficiency, DiffPhase reduces the high-dimensional input of the diffusion model using UNet structure to obtain a compact low-dimensional representation. By introducing the stepwise denoising mechanism of diffusion models, the method can effectively handle phase unwrapping in strong noise environments and mitigate error propagation in complex scenarios. Additionally, an adaptive multifrequency input mechanism is introduced, enabling the network to flexibly process an arbitrary number of wrapped phase inputs at different frequencies. By integrating low-frequency global contours with high-frequency local details, this approach effectively suppresses error propagation and enhances the accuracy and robustness of the unwrapping results. **Result** Experiments are conducted to compare the proposed method with eight state-of-the-art deep learning and traditional approaches on two simulated datasets and two real datasets. For the two simulated methods, three

wrapped phases with different frequencies are generated for each absolute phase to simulate multifrequency data collected in real scenarios. All training data are randomly corrupted with noise ranging from 0 dB to 30 dB to train the model's denoising ability. The test sets are corrupted with noise at different levels to evaluate the model's performance under various noise conditions. The NYU Depth V2 and Middlebury Stereo datasets are also considered as real data for comparative testing. On the RME-multi and MoGR-multi simulation datasets, the normalized root-mean-square errors of phase unwrapping are 0.23% and 0.24%, respectively. On the NYU-phase and MS-phase real datasets, the normalized root-mean-square errors of phase unwrapping are 4.69% and 7.50%, respectively. The comparison experiments demonstrate that DiffPhase, using only a single-frequency phase, outperforms other methods in phase unwrapping performance. With additional frequency wrapped phases, DiffPhase achieves a stable improvement in accuracy. In real-world 3D scene reconstruction, DiffPhase demonstrates high depth reconstruction accuracy under varying levels of noise interference, exhibiting superior precision and robustness, particularly in regions with fine edges and complex structures. Furthermore, occlusion recovery experiments are conducted on the MoGR-multi dataset, which verifies the superior image restoration ability and accurate phase unwrapping capability of DiffPhase in complex occlusion scenarios. **Conclusion** In this paper, a multifrequency phase unwrapping method based on a conditional diffusion model is proposed. Experimental results show that DiffPhase outperforms several state-of-the-art methods and achieves accurate and robust unwrapping results in high-noise and complex phase scenarios, effectively improving the accuracy of 3D reconstruction.

Key words: phase unwrapping; diffusion model; 3D modeling; deep learning; computer vision

0 引言

基于相位解包裹的三维成像是一种通过分析相位信息重建物体三维形貌的技术。其核心原理是通过测量由物体表面形貌或光程差引起的相位变化,利用相位解包裹算法对因周期性(如干涉测量、结构光投影或合成孔径雷达等场景中常见的相位包裹现象)导致的相位跳变进行校正,从而恢复连续的绝对相位值。最终,通过相位与深度或高度的映射关系,精确计算出物体表面的三维坐标分布,实现高分辨率的三维形貌重构。相位解包裹是高精度结构光三维成像的关键步骤,对成像精度和鲁棒性具有决定性作用。精确的相位解包裹,对提升光学干涉测量、合成孔径雷达以及核磁共振成像等领域的重建精度与质量具有重要意义。

由于相位测量涉及反正切函数,实际场景中测量得到的相位会被“包裹”(Fang, 2023),即测量相位值的变化超过最大值(通常是 2π)时,会进入下一周期并重新回到0,使得测量采集系统无法直接区分相位值的真实变化。相位解包裹技术则尝试将多个周期包裹的相位数据还原为原本的真实相位,以获得需要测量的空间信息(Baek 和 Jung, 2024),从而进行更精确的分析和计算。图1展示了一个相位解包裹的例子:因为图1(a)中的真实相位超出了测

量范围,在测量采集后得到了图1(b)中多周期包裹的相位图。经过相位解包裹后恢复到图1(c)中的绝对相位。

在实际场景中,相位解包裹面临着诸多挑战,测量信号在采集过程中会受到环境和设备噪声的影响,使得包裹相位包含随机误差(曾庆旺等, 2023);相位数据中还存在局部的不连续区域和跳变点(Du等, 2025),会破坏相位的平滑性,增加解包裹的复杂性(Gontarz等, 2023)。如果解包裹方法不能有效地对相位进行全局建模,局部跳变会导致错误传播并造成严重的解包裹错误(Yu等, 2024)。现有的相位解包裹的研究包括传统方法和基于深度学习的方法。传统方法通常依赖于路径跟踪、最小二乘或图论等算法,利用全局优化来完成解包裹。虽然传统方法有较高的计算效率,但其抗噪声能力有限,面对复杂场景中的相位跳变,可能会累计错误并造成全局误差。基于深度学习的方法则使用深度网络从输入的包裹相位数据中自动提取特征,并通过端到端映射预测出真实的相位分布。相比传统方法,深度学习方法具备一定的抗噪能力,但现有的方法受限于网络架构,对全局结构的建模能力较弱(冯明涛等, 2025),无法利用包裹相位中隐含的大尺度信息。

传统的相位解包裹方法主要分为3类:路径跟踪法、最小范数法和贝叶斯正则化方法。路径跟踪算法沿着由分支切割算法等建立的路径进行线积分

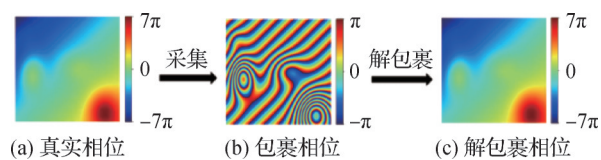


图1 相位解包裹示例

Fig. 1 An example of phase unwrapping ((a) absolute phase; (b) wrapped phase; (c) unwrapped phase)

(Goldstein 等, 1988; Su 和 Chen, 2004), 使用手动设计的度量标准量化像素的可靠性。路径跟踪算法也可以分为质量引导算法(Li 等, 2018; Arevalillo-Herráez 等, 2016)、分支切割算法(De Souza 等, 2015; Huang 等, 2015)、掩膜切割算法(Gao 和 Yin, 2012)和最小不连续性算法(Flynn, 1997)等。路径跟踪方法需要包裹相位具有良好的连续性和一致性, 当存在不连续时会影响路径的选择, 在严重噪声条件下误差会随着路径累计并产生全局误差。与关注局部积分的路径跟踪方法相反, 最小范数优化方法(Ghiglia 和 Romero, 1996; He 等, 2014)通过将解包裹转化为全局范数最小化问题, 从整体上来预测真实相位。最常用的范数优化方法为最小二乘解法(Strand 等, 1999; Guo 等, 2014), 使用快速余弦变换(Xia 等, 2016)和傅里叶变换(Haniff, 1991)高效计算解包裹相位。然而全局优化算法属于求解困难的非凸优化问题, 计算开销随着图像尺寸的增加急剧上升, 且可能优化失败。此外, 相位解包裹问题还可以通过贝叶斯框架来解决(Nico 等, 2000)。Budillon 等人(2005)考虑使用数据获取模型和相位的统计先验知识来解包裹。这一类别的代表性解决方案包括统计成本网络流算法(statistical-cost, network-flow phase-unwrapping algorithm, SNAPHU)(Chen 和 Zebker, 2002)、最大流算法(phase unwrapping max-flow/min-cut algorithm, PUMA)(Bioucas-Dias 和 Valadao, 2007)以及动态规划迭代求解算法(Ying 等, 2006)。这些方法依赖于先验信息, 在强噪声情况下可能不适用, 限制了使用场景。

近年来, 很多研究尝试使用基于深度学习的方法进行相位解包裹, 包括无监督方法和有监督方法。无监督方法中, PUDIP(phase unwrapping via deep image prior)(Yang 等, 2021)首次将深度图像先验应用于非凸相位解包裹问题, 引入自适应损失和反馈机制, 但该方法需要对每个样本重新训练, 显著增加了计算复杂性。Chen 等人(2024)提出一种基于外

部无监督深度学习的端到端相位解包裹方法 U3Net (unsupervised unrolling network for phase unwrapping), 通过迭代优化逐步提升解包裹结果, 但监督信息的缺失也限制了无监督方法在不同场景中的适用性。此外, 许多有监督方法也应用于相位解包裹。PhaseNet(Spoorthi 等, 2019)提出一种将相位解包裹公式化为语义分割的方法, 并结合聚类算法处理非连续相位; PhaseNet2.0(Spoorthi 等, 2020)在上述基础上使用改进的网络和损失函数来提高相位解包裹的准确性和计算效率。Zhang 等人(2019)将相位解包裹转化为多分类问题, 并且设计了独立的去噪网络。Dardikman 和 Shaked(2018)通过训练残差神经网络(residual network, ResNet)处理具有陡峭空间梯度的模拟相位数据。上述有监督的相位解包裹方法依赖大量高质量训练数据, 泛化能力和噪声鲁棒性有限, 且存在标签获取困难和高计算成本等问题。相比之下, 基于循环神经网络(recurrent neural network, RNN)的方法(Perera 和 De Silva, 2021)擅长捕捉长程依赖关系, 将相位像素视为序列中的一个元素, 但是这类方法由于计算成本的限制, 在实际应用中通常只能沿有限数量的解包裹路径进行建模。

去噪扩散概率模型(denoising diffusion probabilistic model, DDPM)在图像生成等领域取得了显著进展, 其独特的生成机制和优越的性能使其成为深度学习中的重要研究方向(韦炎炎 等, 2025; 黄颖 等, 2025)。扩散模型通过反复对噪声图像进行去噪得到高质量图像, 具有很好的噪声鲁棒性和全局建模能力。其对随机高斯噪声进行逐步去噪, 并在每一步去噪过程中与额外条件进行交互, 以指导图像生成(Wang 等, 2024)。

作为生成模型, 扩散模型具有较强的全局建模能力, 可以有效捕捉大范围的相位关系, 避免局部错误对整体结果的影响。与指导条件交互的交叉注意力层使得生成的真实相位在各局部区域都能与特征对齐, 从而保证了生成的一致性。图像内部的自注意力可以在每一步生成中捕捉全局信息, 对长距离关系进行建模, 有效处理大范围相位跳跃等复杂情况。此外, 在很多复杂的情况下, 单一频率的包裹相位无法兼顾大范围跳跃和局部精细结构的处理需求。低频的包裹相位由于周期更长, 可以更好地表示大范围的相位跳跃, 但会丢失一些局部的细节信

息,高频的包裹相位则刚好相反。相比于单频,在很多场景下获取多频的包裹相位并不会带来明显的额外成本,因为多频信息通常可以通过调整测量参数直接获得。例如,在3D建模或光学成像中,只需改变摄像机或激光器的发射频率即可得到不同频率下的相位数据,以帮助更精确地解包裹。

针对上述情况,本文提出一种基于条件扩散模型的多频相位解包裹方法(DiffPhase)。DiffPhase将相位解包裹建模为条件引导的图像生成问题,通过融合多尺度特征条件与扩散生成机制,提升模型对噪声与遮挡的鲁棒性。具体地,DiffPhase包括一个特征提取模块和一个生成模块。本文构建了与扩散网络结构对齐的多尺度特征提取模块,从输入的包裹相位中提取不同层级的语义信息,并通过跨尺度交叉注意力机制将其逐步融合到扩散过程的各个阶段,增强生成结果的局部精度与全局一致性。在扩散模型的每一步去噪过程中,包裹相位特征与预测的真实相位逐步交互,能够有效引导去噪过程。借助其逐步生成机制,模型能够充分利用已有信息,稳步优化预测结果,从而增强对强噪声干扰的鲁棒性,并缓解复杂场景下的误差传播问题。为了提升训练稳定性与泛化能力,采用两阶段训练策略,先对特征提取模块进行预训练以学习结构先验,再进行端到端联合优化以增强相位预测能力。此外,引入自适应多频输入机制,使网络可灵活处理任意数量的不同频率包裹相位输入,通过融合低频的全局轮廓与高频的局部细节,有效抑制误差传播并提升解包裹结果的准确性与鲁棒性。借助扩散模型的条件生成能力和全局建模能力,DiffPhase能够在高噪声、高复杂度的场景下实现稳健而准确的解包裹结果。

在RME-multi(random matrix enlargement-multifrequency)、MoGR-multi(mixture of Gaussians with ramp-multifrequency)仿真数据集和NYU-phase(New York University-phase)、MS-phase(middlebury stereo-phase)真实数据集上的实验表明,DiffPhase单频相位解包裹相较于现有方法具有更高的解包裹精度和更强的噪声鲁棒性。在加入额外频率的包裹相位后,通过有效融合不同频率的包裹相位信息,DiffPhase能够进一步稳定地提升性能。在真实三维场景重建中,DiffPhase面对不同水平的噪声干扰,仍能保持较高的重建精度,尤其在细节边缘与复杂结构区域展现出更强的准确性与鲁棒性。此外,本文还

对不同方法在被遮挡等复杂场景下的表现进行了评估,发现DiffPhase可以更有效地恢复绝对相位,更好地应对不同采集环境的影响。

总的来说,本文的贡献在于:1)提出一种基于条件扩散模型的多频相位解包裹方法,利用扩散模型的噪声鲁棒性和全局建模能力,有效提升了复杂条件下的解包裹性能。DiffPhase是第一个使用扩散模型进行相位解包裹的方法。2)设计了一种自适应的多频相位解包裹策略,使解包裹模型能够接受任意数量的不同频率的包裹相位进行解包裹,以优化模型的全局一致性和局部精度。DiffPhase是第一个可以自适应任意数量包裹相位输入的深度学习解包裹方法。3)采用两个仿真数据集和两个真实数据集验证了方法的有效性。结果表明,本文方法在解包裹精度上具有明显优势。此外,本文也评估了模型在遮挡情况下的三维成像,验证了算法的鲁棒性。

1 本文方法

1.1 问题定义

假设需要测量的真实相位是 $\phi_{\text{true}}(\mathbf{x})$,在实际测量中,相位限制在周期区间 $[-\pi, \pi]$ 内,使得测量得到的相位数据表现为包裹相位 $\phi_{\text{wrapped}}(\mathbf{x})$ 。具体为

$$\phi_{\text{wrapped}}(\mathbf{x}) = \phi_{\text{true}}(\mathbf{x}) \bmod 2\pi - \pi \quad (1)$$

为了从包裹相位中恢复真实相位,需要引入一个整数倍的 2π ,记为 $k(\mathbf{x})$,使得

$$\phi_{\text{true}}(\mathbf{x}) = \phi_{\text{wrapped}}(\mathbf{x}) + 2\pi k(\mathbf{x}) \quad (2)$$

求解 $k(\mathbf{x})$ 的过程是相位解包裹问题的核心。在理想情况下,相位分布是连续且没有噪声的,相邻像素之间的相位差满足 $|\phi_{\text{wrapped}}(\mathbf{x}+1) - \phi_{\text{wrapped}}(\mathbf{x})| < 2\pi$ 。这种情况下,可以简单地通过累积相邻像素的相位差来恢复真实相位。然而,在真实场景中的相位分布存在不连续性。例如,在具有剧烈相位梯度变化的区域,包裹相位的相邻像素相位差可能超过 2π 。此外,测量中存在不可避免的噪声 $\eta(\mathbf{x})$,会干扰相位数据,使相位差的估计偏离真实值,会进一步增加解包裹的难度。具体为

$$\phi_{\text{true}}(\mathbf{x}) = \phi_{\text{wrapped}}(\mathbf{x}) + 2\pi k(\mathbf{x}) - \eta(\mathbf{x}) \quad (3)$$

要实现精确的相位解包裹,不仅要准确求解相位跳变次数,确保还原出真实的连续相位信息,还需有效抑制噪声干扰,避免误差传播,从而提升解包裹

的稳定性与精度。

1.2 数据生成

深度学习算法需要大量数据作为实验基础,而现实采集的相位数据往往存在质量不高、噪音较多和误差较大等问题(徐映芬等,2024),所以该领域的相位数据集普遍为模拟生成的。为了模拟真实场景中的相位数据,本文利用绝对相位和包裹相位之间的物理约束,生成包含随机形状的大量模拟数据。具体地,本文使用改进的 RME (random matrix enlargement) 方法(Wang等,2019)和 MoGR (mixture of Gaussians with ramp) 方法(Perera 和 De Silva, 2021)生成了多频相位解包裹数据集。改进算法生成的 RME-multi 和 MoGR-multi 数据图例如图 2 所示,其中,上面一行是 RME-multi 数据示例,下面一行是 MoGR-multi 数据示例。每一组数据包括两类,图 2(a)为真实相位,图 2(b)为等差频率的包裹相位。仿真相位数据集地址: <https://github.com/yan35008-design/phase-data>。

改进的 RME 算法过程主要包括绝对相位生成、多频率包裹相位生成两个部分。在绝对相位生成的过程中,随机生成矩阵后通过双线性插值放大到 256×256 像素,并设置包裹相位范围为 $[-7\pi, 7\pi]$ 。在进行包裹相位生成时,为每一个绝对相位生成了 3 个频率的包裹相位,且 3 个频率间成等差关系,以模拟现实中采集到的多频数据。改进的 MoGR 算法与改进的 RME 算法的解包裹相位原理相同,二者最主要的区别在于初始生成相位图的方式。改进的 MoGR 算法生成的相位图包含随机生成的高斯斑点,并叠加了一个带有随机梯度和偏移的斜坡。

对于两种仿真方法,分别生成了 15 000 幅绝对相位图,并按照 7:1.5:1.5 的比例划分训练、验证和测试集。然后从绝对相位图对应的 3 个频率中,随机选择一幅或两幅包裹相位来得到单频和双频的训练集。所有训练数据被随机加上 0~30 dB 的噪声,以训练模型的去噪能力。测试集则分别被加上 {0 dB, 5 dB, 10 dB, 15 dB, 20 dB, 30 dB} 的噪声,以评估模型在不同等级噪声下的性能变化。在测试时,同一个真实相位对应的不同频率包裹相位会被加上同等级的噪声。所有方法在训练集上完成训练后分别在不同噪声等级的测试集上测试。

1.3 条件去噪扩散概率模型 CDDPM

去噪扩散概率模型(denoising diffusion probabilis-

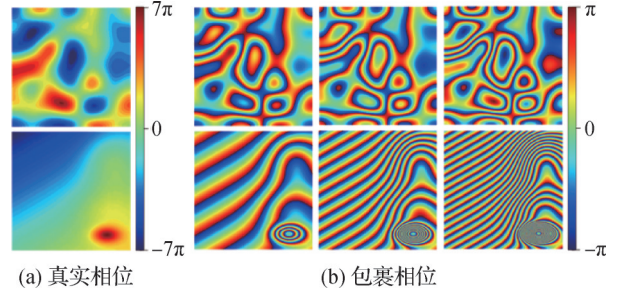


图 2 生成数据示例(上行为 RME 改进算法的生成数据,下行为 MoGR 改进算法的生成数据)

Fig. 2 Generated data examples for the improved RME and MoGR algorithms ((a) absolute phase; (b) wrapped phase)

tic model, DDPM) 是一种受非平衡热力学启发的生成模型。它通过定义一个正向扩散过程,将原始数据逐渐转换为高斯噪声,再利用参数化的逆向过程,逐步从噪声中还原出高质量样本。DDPM 通过这种逐步推理的过程,能够有效捕捉数据的复杂分布,生成的样本具有高保真度,并在图像生成等任务中表现出色。本文将相位解包裹建模为条件去噪扩散生成任务,学习条件分布 $D(\mathbf{x}|\mathbf{c})$, 其中, $\mathbf{x}$ 为绝对相位图,条件 $\mathbf{c}$ 为包裹相位图。

1.3.1 正向过程

如图 3 所示,在正向扩散过程中,噪声逐步添加至需要被重建的图像 $\mathbf{x}_0$ 中,使其逐渐退化为完全随机的高斯噪声。本文将原始图像 $\mathbf{x}_0$ 视为初始状态,定义正向加噪过程 $\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_{t-1})$, 该过程在时间步 t 处向数据 $\mathbf{x}_{t-1}$ 添加方差为 β_t 的高斯噪声,从而生成中间状态 $\mathbf{x}_t$, 具体为

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (4)$$

式中, $q(\cdot)$ 表示在正向扩散过程第 t 步从 $\mathbf{x}_{t-1}$ 生成 $\mathbf{x}_t$ 的条件分布概率, $\mathcal{N}(\cdot)$ 为正态分布, β_t 是高斯噪声方差, $\mathbf{I}$ 为单位协方差矩阵。

在整个正向扩散过程中,可以递归地表示噪声的累积效应。定义 $\bar{\alpha}_t = \prod_{i=1}^t (1 - \beta_i)$ 作为总的衰减因子,则初始状态 $\mathbf{x}_0$ 经过 t 轮加噪后的分布可以直接表示为

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (5)$$

式中, $\bar{\alpha}_t = \prod_{i=1}^t (1 - \beta_i)$ 表示信号保留的累积比例,数值越大表示信号保留越多; $\bar{\alpha}_t$ 控制了信号保留的程度,较大的 $\bar{\alpha}_t$ 说明信号保留得更多; β_t 控制每一步添

加的噪声量,较大的 β_t 使得信号衰减更快。

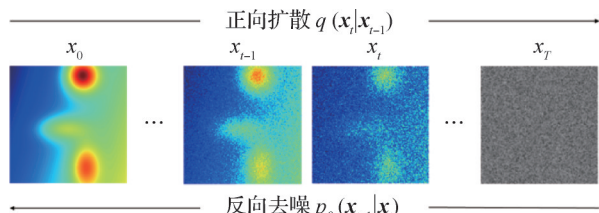


图3 去噪扩散过程

Fig. 3 Denoising diffusion process

1.3.2 逆向过程

在逆向阶段,模型学习完整的去噪过程,通过多次迭代去噪,从高斯噪声逐步重建出绝对相位。如图3所示,给定初始高斯噪声 $x_T \sim \mathcal{N}(0, I)$, I 为单位协方差矩阵,通过逐步去噪恢复出最终的原始图像 x_0 。希望学习一个参数化模型 $\epsilon_\theta(x_t, t; w)$, 用于预测在时间步 t 处的噪声 $\hat{\epsilon} \approx \epsilon_t$ 。从时间 T 开始,每一步去噪过程可表示为

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t; w) \right) + \sigma_t z \quad (6)$$

式中, $z \sim \mathcal{N}(0, I)$, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, β_t 为扩散过程的噪声方差, $\epsilon_\theta(\cdot)$ 为噪声预测函数, w 为模型权重, σ_t 控制反向采样噪声强度, z 是标准正态分布噪声向量。

为了训练模型,使其能够有效地预测噪声,需最小化噪声预测误差的期望。具体而言,优化目标为

$$L = \mathbb{E}_{x_0, \epsilon, t} \left[\left\| \epsilon - \epsilon_\theta(x_t, t; w) \right\|_2^2 \right] \quad (7)$$

式中, $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$, $\epsilon \sim \mathcal{N}(0, I)$ 是添加的高斯噪声, $t \sim U(\{1, \dots, T\})$ 表示在训练时均匀采样时间步, $\epsilon_\theta(\cdot)$ 为噪声预测函数。该损失函数的目标是使模型 $\epsilon_\theta(x_t, t; w)$ 能够准确预测不同时间步的噪声,从而在逆向扩散过程中更精准地去噪,最终恢复出原始图像 x_0 。

在实际应用中,往往希望生成的图像满足特定的条件约束,因此条件扩散模型在去噪过程中引入条件引导,通过条件信息 c 指导去噪过程,使生成的结果更加符合预期。修正后的噪声预测函数表示为 $\epsilon_\theta(x_t, t, c; w)$, 每一步去噪的更新式变为

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t, c; w) \right) + \sigma_t z \quad (8)$$

式中, $z \sim \mathcal{N}(0, I)$, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, β_t 为扩散过程的噪声方差, $\epsilon_\theta(\cdot)$ 为噪声预测函数, c 为条件信息, w 为模型权重, σ_t 是噪声调节系数, z 是标准正态分布噪声向量。

在具体实现中,条件 c 可以通过多种方式融入模型,例如直接与输入拼接、通过注意力机制增强特征,或使用额外的网络分支提取条件信息后再融合。通过引入条件引导,扩散模型能够在生成过程中调整去噪路径,使最终生成的图像更加符合指定的条件要求。

1.4 模型架构

DiffPhase 的整体框架如图4所示,包括特征提取部分和生成部分,特征提取部分从输入图像中提取多尺度的语义信息和细节特征,为生成部分提供条件指导。DiffPhase 采用与扩散模型中相同的网络架构提取特征,利用扩散模型中的交叉注意力层进行多频指导相位生成,通过对初始高斯噪声逐步去噪来精确恢复相位值。

常用的条件扩散模型需要兼顾生成的多样性,无法实现精确预测。为了精确预测包裹相位对应的真实相位,设计了如图4(a)所示的多尺度特征提取网络。具体而言,使用与扩散模型中U型网络(U-shaped network, UNet)相同的架构作为特征提取网络,将每次降维或升维后的包裹相位特征图作为提取的特征 $C = \{c'_i\}_{i=1}^n$, 指导扩散模型的预测,其中 n 为提取特征的层数。此时,每一步去噪过程变为

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t, \{c'_i\}_{i=1}^n; w) \right) + \sigma_t z \quad (9)$$

式中, $z \sim \mathcal{N}(0, I)$, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, β_t 为高斯噪声, $\epsilon_\theta(\cdot)$ 为噪声预测函数, $\{c'_i\}_{i=1}^n$ 为特征条件信息。

相比将整个图像作为一个整体特征,UNet 为图像中的每个图像块生成独立的特征表示,从而捕获更丰富的局部与全局信息,为扩散模型提供更全面的条件约束。同时,由于UNet结构本身具有降维及升维的特性,本文的特征提取网络能够更好地提取不同尺度的包裹相位信息,使模型既能关注全局结构,又能精确刻画局部细节。

此外,UNet在多尺度编码与解码路径中引入了残差连接机制,用于增强多频特征的融合效果与稳

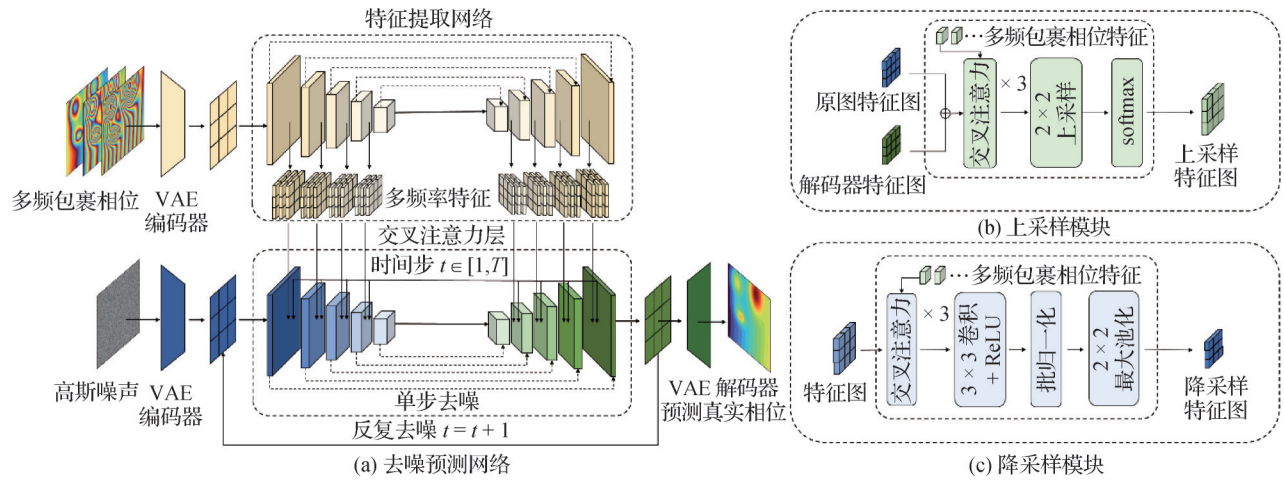


图4 DiffPhase框架

Fig. 4 DiffPhase framework ((a) denoising prediction network; (b) upsampling module; (c) downsampling module)

定性。具体而言,每个频率的包裹相位图经过编码后提取不同层级的特征,并通过残差连接将低级信息与高层语义特征显式结合,保留结构细节,减缓梯度消失与信息损失。在解码阶段,残差特征与跳跃连接协同作用,进一步提升特征还原能力。该机制有效实现了多频信息的深度整合,相较简单拼接方式,在特征表达的一致性与鲁棒性方面表现更优。

由于该特征提取网络与扩散模型采用相同的UNet架构,在扩散模型的每个集成多头注意力机制的交叉注意力层中,直接注入提取的多尺度特征,以精准引导扩散分支的预测。这些交叉注意力层通过将Query(查询,来自扩散过程)与Key(键,来自条件特征)和Value(值,来自条件特征)进行多子空间并行匹配,有效建模当前状态与条件信息之间的复杂对应关系,从而实现更加细致且可控的特征引导。采用相同架构提取的特征具有更强的控制力和一致性,使得模型在各个层次上都能保持稳定、精准的相位预测能力,从而在保证扩散模型生成能力的同时,提高相位预测的可控性和精确度。

生成部分则基于特征提取部分提供的指导条件,对初始高斯噪声进行逐步去噪,生成预测的真实相位。为优化计算效率,DiffPhase首先通过变分自编码器(variational autoencoder, VAE)对输入的图像进行降维,得到紧凑的低维表示。然后在降维后的特征上进行逐步去噪,学习反向生成的噪声分布。这种设计通过较小的计算量完成扩散模型的去噪过程,从而提升解包裹的生成效率。

在去噪引导过程中,通过建立包裹相位特征与

预测真实相位之间的残差约束,保证了引导过程的正向性。具体而言,在扩散模型的每一步迭代中,计算预测相位与输入包裹相位之间的残差分布,并利用残差均值和梯度方向作为约束信号。由于真实相位在数值域上相较包裹相位具有连续性和物理一致性,该残差信号能够为去噪过程提供稳定的收敛方向。同时,在训练阶段引入基于残差最小化的监督损失,使模型学习到从包裹相位逐步逼近真实相位的单调映射关系,避免了负向更新的可能。换言之,去噪过程的每一步更新都受到残差约束与物理一致性的双重引导,从而确保了预测逐步向真实相位收敛的正向性。

特征提取部分和生成部分的交互是通过扩散模型中的交叉注意力层实现的。具体而言,在扩散模型的每个去噪时间步,将特征提取网络对应层提取的条件特征注入到在生成部分的交叉注意力层。待生成的相位特征 b_i 经过自注意力和前馈网络进行特征提取后,在交叉注意力层引入特征提取网络的条件特征 c_i ,使其作为额外的键和值,而当前生成特征 b_i 作为查询,具体计算为

$$f_{\text{Attention}}(Q_b, K_c, V_c) = \text{softmax}\left(\frac{Q_b K_c^T}{\sqrt{d}}\right) V_c \quad (10)$$

式中, $Q_b = W_Q b_i$ 是待生成相位特征的查询表示, $K_c = W_K c_i$ 和 $V_c = W_V c_i$ 分别是条件特征 c_i 生成的键和值, d 是特征维度的缩放因子,以稳定梯度计算。这一过程确保了扩散模型在生成相位的每个时间步,都能在特征空间上适应性结合输入图像的条件信息,从而增强模型在生成过程中的可控性。

1.5 多频指导相位生成

对于多频包裹相位同时输入的情况, DiffPhase 单独提取每幅包裹相位图的特征 C_j , 然后在每个交叉注意力层中, 使生成图像的特征与所有条件特征 $\{C_j\}_{j=1}^m$ 同时交互, 其中 m 是输入包裹相位的数量。一方面, 同时与多频特征交互使得生成模型能够充分考虑多个包裹相位的不同特征, 利用它们提供的全局结构和局部细节共同指导生成过程, 提升了网络对复杂输入的处理能力; 另一方面, 这种交互方式允许任意多个频率的包裹相位输入, 提高模型对不同应用场景的适应性。使用同样的网络提取特征要求特征提取模型能同时提取不同尺度的信息, 从而确保能够生成模型的指导条件中包含不同尺度的细节与全局语义。

1.6 两阶段训练

采用 UNet 作为特征提取网络, 虽然可以实现精确控制生成, 但也带来了额外的挑战。首先, 本文方法需要在每个尺度上提取并注入特征, 使得网络的计算量大幅增加, 导致训练速度减慢。此外, 由于额外的特征引导, 模型的优化过程变得更加复杂, 容易引入额外的梯度传播路径, 增加训练的不稳定性, 从而导致更难收敛。尤其是在早期训练阶段, 模型需要同时学习如何从提取的特征中获取有效信息, 并结合扩散过程进行合理预测, 增加了优化难度。

假设特征提取网络为 F_{Feature} , 生成网络部分为 F_{Generate} 。为了加快模型的训练过程, 本文设计了如下的两阶段训练策略。

1.6.1 第1阶段: 相位解包裹预训练

在该阶段, 将特征提取网络 F_{Feature} 视为一个独立的相位解包裹网络, 以包裹相位作为输入, 并训练其直接预测真实相位 w 。

训练目标是 minimized 预测相位 $\hat{w} = F_{\text{Feature}}(x)$ 与真实相位 w 之间的误差, 使用均方误差 (mean squared error, MSE) 作为损失函数。具体为

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (\hat{w}_i - w_i)^2 \quad (11)$$

为了进一步抑制过拟合、提升模型的泛化能力, 本阶段同时加入了 L2 正则化, 对网络权重 W 施加惩罚项, 使得第 1 阶段的损失函数整体定义为

$$L_{\text{first_stage}} = L_{\text{MSE}} + \lambda \|W\|_2^2 \quad (12)$$

式中, λ 为正则化系数, 权衡预测误差与权重惩罚的比重, W 表示网络所有可训练参数。

该阶段的主要目的是使特征提取网络学会从包裹相位中提取有效信息, 为下一阶段的真实相位预测奠定基础。

1.6.2 第2阶段: 端到端去噪生成训练

在第 1 阶段训练好的 F_{Feature} 的基础上, 继续训练整个去噪生成网络 F_{Feature} 和 F_{Generate} , 使其能在提取特征的基础上生成去噪后的真实相位 $\hat{w}$ 。

该阶段的训练通过交叉熵损失 (cross-entropy loss, CE) 和 MSE, 提高网络的整体性能。具体为

$$L_{\text{CE}} = -\frac{1}{N} \sum_{x,y} Y(x,y) \log(\hat{Y}(x,y)) \quad (13)$$

式中, N 为参与损失计算的图像像素点数量, $Y(x,y)$ 和 $\hat{Y}(x,y)$ 分别是像素位置 (x,y) 处的真实值和预测的像素值。

$$L_{\text{MSE}} = \frac{1}{N} \sum_{x,y} (\phi(x,y) - \hat{\phi}(x,y))^2 \quad (14)$$

式中, $\phi(x,y)$ 是真实展开相位, $\hat{\phi}(x,y)$ 是预测生成相位。第 2 阶段同样加入了 L2 正则化, 损失函数整体定义为

$$L_{\text{second_stage}} = \mu_1 L_{\text{CE}} + \mu_2 L_{\text{MSE}} + \lambda \|W\|_2^2 \quad (15)$$

式中, λ 为 L2 正则化系数, 用于抑制权重过大, 防止过拟合。

在本文实验中, 采用 $\mu_1 = \mu_2 = 0.5$, 以保证模型逐像素和整体的预测精确度。交叉熵损失用于优化生成网络的分类能力, 使其能够更准确地预测相位分布, 而均方误差损失则保证生成结果在数值上尽可能接近真实相位。此外, 第 2 阶段 F_{Generate} 的特征融合和解码过程中引入了 dropout 正则化, 以缓解过拟合风险并提升模型的泛化性能。

这种两阶段训练策略先让特征提取网络单独学习相位特征, 提高初始阶段的学习效率, 从而减少端到端训练的难度。经过预训练的特征提取网络能够提供更好的特征表示, 从而加快模型的收敛速度。

2 实验与分析

2.1 实施细节

2.1.1 影响因素分析与实验设置

为了分析不同因素对相位解包裹性能的影响, 并且更贴近实际应用场景, 从两个关键方面进行探讨: 噪声水平和相位数据来源。首先, 通过不同的信噪比 (signal-to-noise ratio, SNR) 模拟实际场景中的

噪声变化,以评估模型在不同噪声条件下的性能。其次,研究了不同类型输入数据、输入包裹相位的数量以及输入单频、双频和三频等多种情况对解包裹效果的影响。由于现有方法主要针对单频数据进行研究,重点对比了单频的效果,并进一步分析了双频和三频输入对模型性能的提升。此外,探讨了输入数据不完整的情况,即在部分相位数据缺失的情况下,模型是否仍能实现有效重建,验证 DiffPhase 对于数据缺失的鲁棒性。

超参数的设定对实验结果有显著影响。为了保证实验的可复现性,列出了主要超参数及其设定范围,如表 1 所示。其中,扩散步数影响生成结果的精细程度,步数越多,预测结果越准确,但计算开销也随之增加;输入频率数量的增加则可以提供不同尺度的信息,提升解包裹精度,尤其是在高噪声条件下,额外的频率信息可提供更强的约束,从而降低噪声干扰;条件引导权重决定了模型对先验信息的依赖程度,较大的权重可以增强重建精度;输入分辨率影响细节重建能力,较高的分辨率有助于恢复更精细的相位结构,但计算成本相应增加;包裹层数决定模型的多尺度特性,适当增加包裹层数可以提升模型的鲁棒性,但过多层数可能导致包裹相位过于复杂。

在本文实验中,用 DiffPhase- n 表示不同输入频率的模型,其中 n 代表输入频率的数量。例如,DiffPhase-3 代表以 3 个不同频率的相位图作为输入进行解包裹。

表 1 超参数设置
Table 1 Hyperparameter settings

超参数	取值范围
扩散步数 n	3
频率数 f	1 ~ 3(单频、双频、三频)
条件引导权重 μ	1.0
分辨率/像素	256 × 256
包裹层数 m	0 ~ 7

2.1.2 数据集与评估指标

为了评估方法在不同场景下的性能,本文在两个模拟数据集和两个真实数据集上评估了模型性能。根据 1.2 节中的数据仿真方法,本文分别生成了 RME-multi 数据集和 MoGR-multi 数据集。此外,

本文还考虑了 NYU depth V2 数据集(Silberman 等, 2012)和 Middlebury Stereo 数据集(Scharstein 等, 2014),作为真实数据进行测试对比。NYU depth V2 和 Middlebury Stereo 数据集中包含使用结构光相机拍摄的真实场景深度图像,遵循先前工作中的处理方法(Chen 等, 2024; Sica 等, 2022),本文将其转为绝对相位,然后采用与模拟数据集相同的方法获取不同频率的包裹相位,并加上不同等级的噪声,从而得到 NYU-phase 和 MS-phase 数据集。

由于相对相位值才是相位解包裹的主要关注点,在计算误差之前,测试的解包裹相位主值需要进行均值对齐,所以本文使用归一化均方根误差(normalized root mean square error, NRMSE)作为性能指标。具体为

$$f_{\text{NRMSE}} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\phi_{\text{unwrapped}, i} - \phi_{\text{GT}, i})^2}}{\phi_{\text{max}} - \phi_{\text{min}}} \quad (16)$$

式中, $\phi_{\text{unwrapped}, i}$ 为第 i 个像素的解包裹相位值, $\phi_{\text{GT}, i}$ 为第 i 个像素的真实相位值(ground truth, GT), N 为图像中的像素总数, ϕ_{max} 、 ϕ_{min} 分别为真实相位值的最大值和最小值。

2.1.3 实验环境

DiffPhase 在单块 NVIDIA GeForce GTX 1080 Ti GPU 上运行,训练 20 个周期大约耗时 10 h。在测试阶段,相位解包裹的速度非常快,对于 256 × 256 像素的图像,处理时间在 0.2 s 左右。为了减少计算时间,本文取去噪步数为 3,并通过 Kaiming Uniform 初始化模型权重。训练过程中采用 Adam 优化器,批量大小为 16。学习率初始值为 1×10^{-6} ,并在每个 epoch 中以 0.99 的因子进行指数衰减。实验表明,网络在 20 个 epoch 内可以收敛。2.2—2.5 节实验结果表明,相比于其他方法,DiffPhase 以较快的推理速度带来显著的性能提升。DiffPhase 可以接受单幅或任意多幅图像作为输入,并综合输入来预测真实相位。因此,在测试阶段,可以根据需求输入不同数量或尺寸的图像以适应各种应用场景。

2.1.4 对比方法

本文选用了 8 种最新的相位解包裹方法与 DiffPhase 进行对比,包括 4 种有监督深度学习方法 Uformer、Restormer、SQD-LSTM (spatial quad-directional long short-term memory)、PhaseNet2.0, 2 种无监督深度学习方法 PUDIP (phase unwrapping

via deep image prior)、U3Net 以及 2 种传统方法 LSPU (least-squares phase unwrapping) 和 QGPU (quality-guided phase unwrapping)。

1)Uformer(Wang 等, 2022)。Uformer 利用局部增强窗口 Transformer 模块和跳跃连接机制构建了一种分层编解码网络,提升了图像复原过程中捕获有用依赖关系的能力。

2)Restormer(Zamir 等, 2022)。Restormer 通过优化多头自注意力和前馈网络的设计,能够有效地捕捉高分辨率图像中的长程像素交互,同时保持较低的计算复杂度。

3)SQD-LSTM (Perera 和 De Silva, 2021)。SQD-LSTM 将相位解包裹问题转化为回归问题,并通过结合空间四向长短期记忆网络增强模型在全局空间依赖学习中的能力。

4)PhaseNet2. 0(Spoorthi 等, 2020)。PhaseNet2. 0 将相位解包裹问题转化为密集分类问题,通过预测每个像素的包裹计数来恢复真实相位。

5)PUDIP(Yang 等, 2021)。PUDIP 提出一种基于未训练卷积神经网络和深度图像先验的二维相位解包裹框架,引入了自适应损失和反馈机制。

6)U3Net(Chen 等, 2024)。U3Net 是一种以包裹相位梯度作为噪声标签,通过迭代过程逐步优化解

包裹结果的无监督方法。

7)LSPU (Ghiglia 和 Romero, 1994)。LSPU 是一种基于快速余弦变换的加权最小二乘相位解包裹方法,可在存在路径不一致的情况下有效处理区域性相位剪切问题。

8)QGPU(Herráez 等, 2002)。QGPU 是一种基于离散路径的相位解包裹算法,通过优先解包裹高可靠性像素来克服传统算法在噪声处理和路径依赖性上的局限性。

2. 2 仿真数据集对比

噪声普遍存在于相位解包裹的各种应用领域中,这对真实相位的恢复造成了困难。为了评估不同方法在不同噪声水平下的表现,本文在 RME-multi 和 MoGR-multi 数据集上进行对比实验。为了与其他方法对齐,本文在仅使用单频包裹相位的场景下对 DiffPhase 和其他方法的解包裹精度进行对比,不同数据集和噪声等级下的实验结果如表 2 所示。评价指标为 NRMSE,数值越小表示性能越佳。

从表 2 可以看出,DiffPhase 在不同仿真数据集和噪声等级下表现出领先或与最优方法相当的解包裹精度。在噪声水平较高的场景中,包裹相位的连续性被破坏,LSPU 和 QGPU 等传统方法无法处理噪声的干扰,出现了明显的性能下降。而 DiffPhase 在

表 2 仿真数据集上基于 NRMSE 的性能对比
Table 2 Performance comparison on simulated datasets in terms of NRMSE

		RME-multi 数据集					MoGR-multi 数据集					/%
方法		SNR = 0	SNR = 5	SNR = 10	SNR = 20	SNR = 30	SNR = 0	SNR = 5	SNR = 10	SNR = 20	SNR = 30	
传统方法	LSPU	9.52	2.33	1.66	0.51	0.20	8.42	2.81	1.66	0.51	0.17	
	QGPU	20.11	2.58	1.83	0.49	0.16	15.26	2.76	1.77	0.48	0.14	
深度学习 方法	PUDIP	15.50	9.34	3.11	0.56	0.18	23.11	20.13	8.91	0.74	0.14	
	U3Net	1.23	0.52	0.34	0.27	0.16	0.81	0.47	0.41	0.35	0.18	
	Uformer	1.04	0.90	0.75	0.67	0.57	0.70	0.67	0.64	0.59	0.55	
	Restormer	0.83	0.80	0.61	0.57	0.52	0.52	0.50	0.46	0.44	0.42	
	SQD-LSTM	1.41	1.30	1.12	0.90	0.69	1.81	1.69	1.17	0.92	0.76	
	PhaseNet2.0	10.19	9.68	9.04	8.29	7.91	11.99	11.46	10.90	9.88	9.47	
	DiffPhase-1	0.38	0.33	0.28	0.26	0.25	0.48	0.40	0.36	0.33	0.28	
	DiffPhase-2	0.33	0.32	0.27	0.25	0.24	0.47	0.39	0.34	0.30	0.26	
	DiffPhase-3	0.31	0.27	0.26	0.25	0.23	0.46	0.37	0.32	0.29	0.24	

注:加粗字体表示各列最优结果(限于输入单频相位)。

高噪声下仍然保持很高的解包裹精度,表明其去噪生成架构对噪声具有较高的鲁棒性。在低噪声下,DiffPhase也展现出与基于路径或优化方法类似的精度,说明本文方法可以应用到不同场景。

表2的最后两行给出了网络在输入双频和三频包裹相位时,DiffPhase的预测精度。随着额外频率包裹相位的加入,DiffPhase可以带来稳定的精度提升。这得益于DiffPhase能够有效提取并融合不同频率包裹相位中的信息,并以此指导真实相位预测。在很多应用场景中,获取额外频率的包裹相位不会有太多额外的成本,例如在三维重建中,只需要调整投影参数即可获得不同频率的包裹相位,选择性地保存大尺度信息或局部细节。DiffPhase可以融合任意数量的多频率包裹相位信息,这种灵活性允许其在各种场景中取得更高的解包裹精度,满足不同的需求。在信号强度较弱或噪声较大的情况下,输入多频相位图进行解包裹可以显著提升效果。多频率包裹相位的额外加入为解包裹过程提供了互补信息,有助于网络识别和纠正由噪声或其他因素引起

的相位跳变。通过引入多频率的信息,DiffPhase在面对复杂的噪声环境时表现出更强的鲁棒性。

图5展示了在0 dB信噪比条件下,使用不同方法进行相位解包裹的示例可视化。图5(d)是不同方法的解包裹相位图(上行)以及相比于GT的误差映射图(下行),颜色越浅则误差越大。实验表明,在高噪声场景中,DiffPhase可以最精确地还原出真实相位。LSPU和QGPU等传统方法由于优化条件的限制,在高噪声的影响下产生大量局部误差,并逐步累积形成全局误差,造成预测结果的偏离。可以观察到,基于深度学习的方法主要在高梯度区域产生误差。而DiffPhase凭借扩散模型强大的全局建模能力,能够更好地捕捉数据的全局结构,从而提高困难区域的预测精度。使用多频包裹相位则进一步减少了DiffPhase在高梯度区域的误差,不同频率的包裹相位信息提升了其在复杂场景中的表现。

2.3 真实数据集对比

除了仿真多频相位数据集,结合2.1.1节内容,在使用真实数据构建的NYU-phase和MS-phase

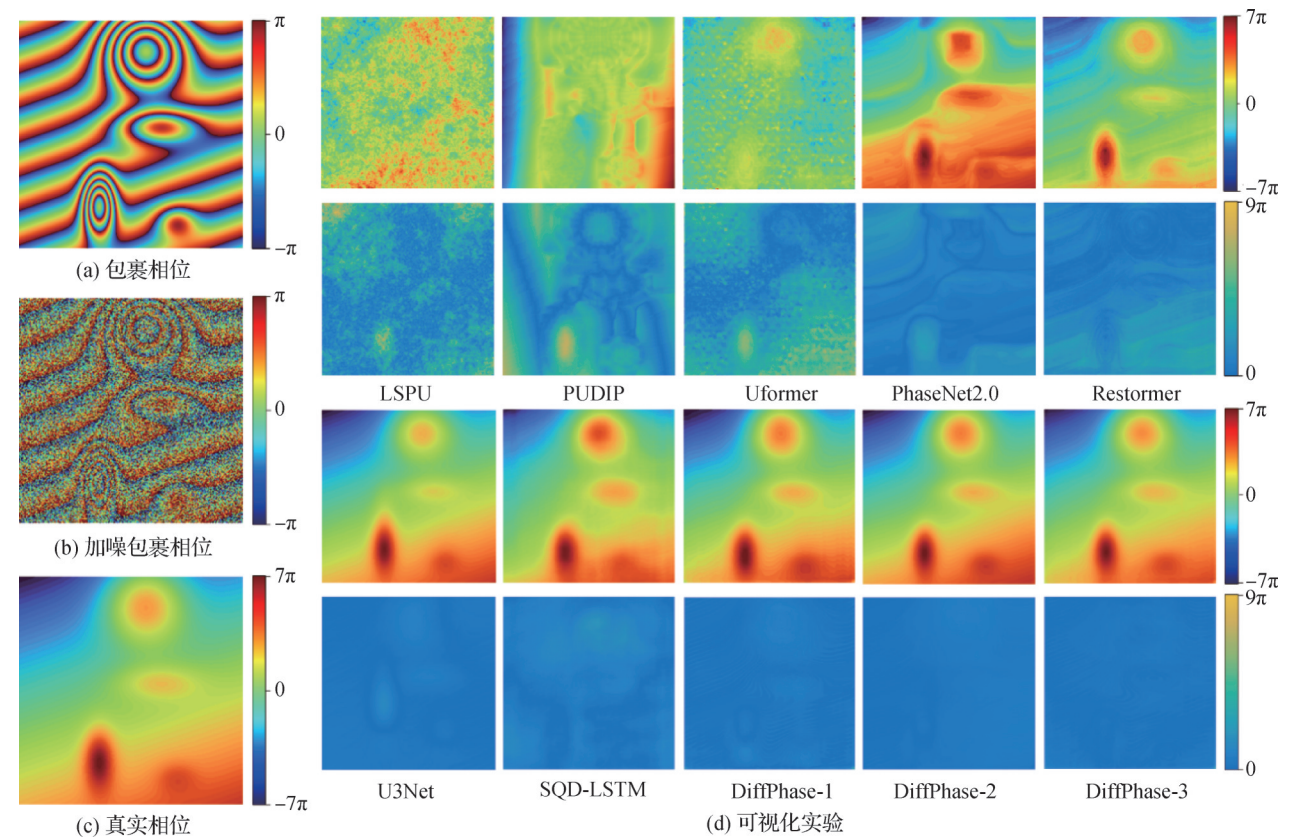


图5 仿真数据实验结果

Fig. 5 Experimental results of simulation data

((a) wrapped phase; (b) noisy wrapped phase; (c) absolute phase; (d) experiment visualization)

数据集上将 DiffPhase 与其他方法进行对比,用于定量评估 DiffPhase 在现实场景中三维重建任务中的表现。由于室内物品摆放的前后关系,这两个数据集中存在大量的相位跳变,在跳变部分梯度信息丢失无法恢复。所以相比于模拟数据,NYU-phase 和 MS-phase 更需要解包裹方法结合包裹相位中的全局信息(例如不同物品接触部分的梯度)来辅助恢复。

表 3 和表 4 分别展示了 DiffPhase 及其他方法在 NYU-phase 和 MS-phase 数据集上的解包裹精度,数值越小表示性能越佳。实验结果表明,DiffPhase 始终给出最好的结果,并且多频包裹相位输入能进一步提升相位解包裹精度。

图 6 展示了在 NYU-phase 真实数据集中对示例数据进行解包裹的可视化。图 6(d)为不同方法的解包裹相位图(上行)和相比于 GT 的误差映射图(下行),颜色越浅则误差越大。从图 6 可以看出,由于场景中物品的前后关系,包裹相位的梯度仅在物品相互接触的部分连续。传统方法难以处理物体边界处的梯度突变,造成相位预测明显错误;基于深度学习的方法受限于感受野大小,在复杂的场景中极易出现局部误差。而 DiffPhase 通过全局建模去噪,可以相对准确地预测出实例图的绝对相位。

表 3 NYU-phase 数据集实验性能对比

Table 3 Performance comparison on NYU-phase dataset

%

方法	NRMSE		
	0 dB	15 dB	30 dB
LSPU	29.16	26.76	25.26
QGPU	37.43	29.71	19.57
PUDIP	19.19	18.1	17.01
U3Net	13.96	12.24	11.8
PhaseNet2.0	13.27	12.92	10.63
Uformer	12.87	9.90	8.03
Restormer	14.36	10.87	7.72
SQD-LSTM	17.51	11.76	6.50
DiffPhase-1	11.34	9.08	5.82
DiffPhase-2	10.84	8.53	5.26
DiffPhase-3	9.51	7.54	4.69

注:加粗字体表示各列最优结果(限于输入单频相位)。

表 4 MS-phase 数据集实验性能对比

Table 4 Performance comparison on MS-phase dataset

%

方法	NRMSE		
	0 dB	15 dB	30 dB
LSPU	25.36	23.93	22.70
QGPU	25.21	24.61	23.15
PUDIP	19.07	16.82	14.49
U3Net	18.21	16.77	14.36
PhaseNet2.0	17.28	15.39	13.98
Uformer	14.29	12.05	9.16
Restormer	15.40	13.82	11.87
SQD-LSTM	14.63	12.15	10.23
DiffPhase-1	13.28	10.72	8.11
DiffPhase-2	12.02	10.46	7.85
DiffPhase-3	11.09	9.28	7.50

注:加粗字体表示各列最优结果(限于输入单频相位)。

图 7 展示了在 MS-phase 真实数据集中对示例数据进行解包裹的可视化。图 7(d)为不同方法的解包裹相位图(上行)和相比于 GT 的误差映射图(下行),颜色越浅则误差越大。与图 6 相比,图 7 的真实场景更加复杂,画面中梯子的包裹相位与背景间存在大量不连续相位,恢复真实相位需要准确识别梯子的细节信息和梯子与地面连接处的信息。传统方法中只有 QGPU 能将梯子与背景区分开来,但仍然错误地判断了梯子中台阶与扶手的位置关系。在基于深度学习的方法中,Restormer 正确识别了梯子的轮廓,但地面与背景部分的信息丢失;Uformer 正确恢复了部分背景和地面的真实相位,但无法处理梯子部分复杂的不连续包裹相位。SQD-LSTM 利用包裹相位中上下左右 4 个方向的信息,更适用于连续的包裹相位,而无法处理图 7 中梯子这种不连续的情况。相比于其他方法,本文提出的 DiffPhase 在使用单频的情况下最准确地识别了场景中 3 个物体的大致轮廓和与背景的关系。在使用多频率包裹相位输入的情况下,DiffPhase 可以进一步分离梯子中间的台阶与背景,得到更为准确的结果。

2.4 恢复能力对比

本文实验也评估了所提方法在一些具有挑战性场景下的表现。在测量过程中,真实相位可能会受

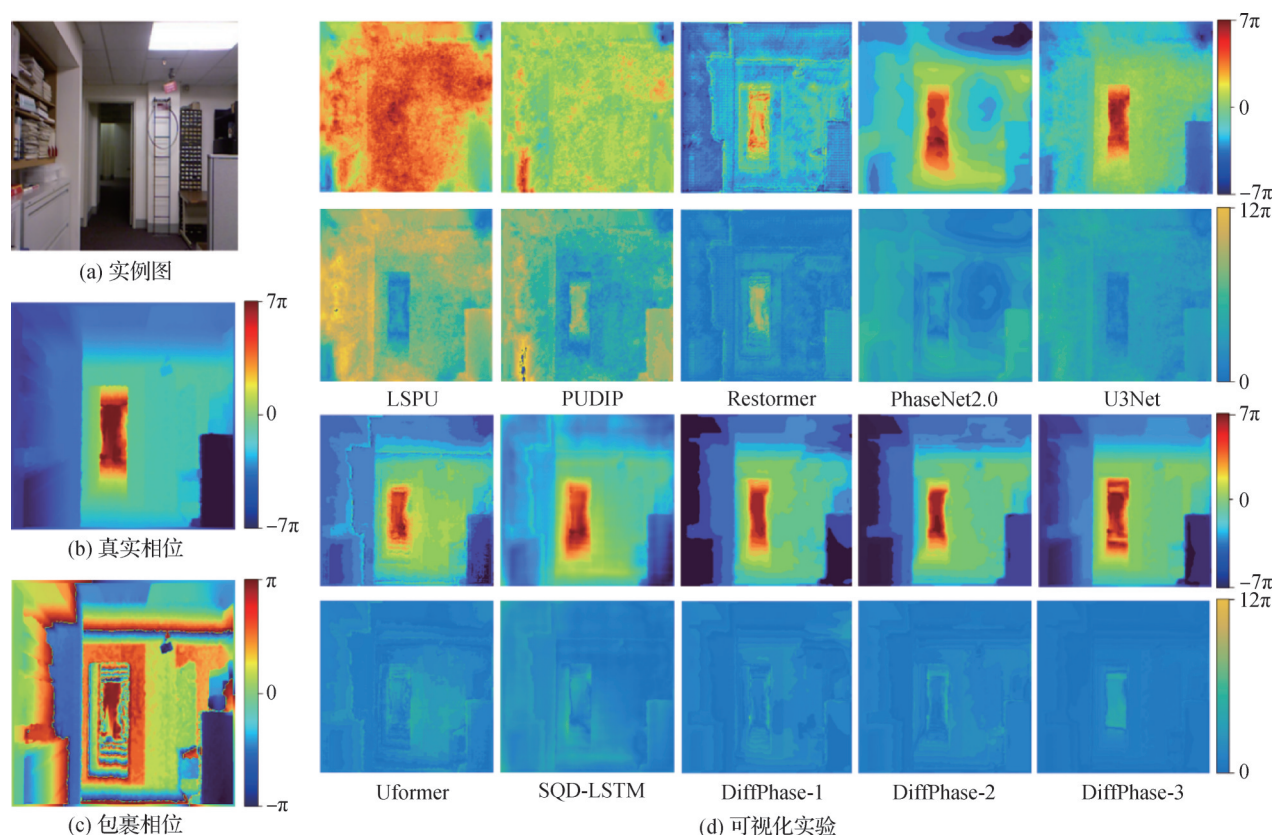


图6 NYU-phase数据集实验结果

Fig. 6 NYU-phase dataset experimental results

((a) instance image; (b) absolute phase; (c) wrapped phase; (d) experiment visualization)

到干扰或遮挡,导致局部信息丢失。在这种情况下,相位解包裹方法需要有效提取包裹相位其他部分的特征,并以此恢复丢失的信息。为进一步评估 DiffPhase 在复杂相位解包裹任务中的表现,在被遮挡的数据上对比了不同方法的解包裹精度。

从 MoGR-multi 模拟数据集中随机选取图像,如图 8 所示,将其左上 1/4 像素用中间相位值替代。图 8(b) 为被遮挡的包裹相位,图像左上部分的信息被相位中值覆盖,解包裹方法需要根据图像的其他部分恢复这一部分。图 8(d) 为不同方法的解包裹相位图(上行)和相比于 GT 的误差映射图(下行),颜色越浅则误差越大。通过误差图可以发现, LSPU 方法旨在最小化不连续性,导致相位过于平滑。无监督方法依赖于连续性假设,难以处理遮挡存在的情况。

其他基于深度学习的方法普遍在梯度相对平滑的包裹相位区域能够正确解包裹,但是在遮挡区域则出现较大误差。而 DiffPhase 继承了扩散模型的优势,具备较强的局部恢复能力:即使包裹相位中部分信息缺失, DiffPhase 仍能通过提取图像其他部分

的特征,有效恢复缺失的内容。

2.5 计算速度对比

为了全面评估模型的计算效率,在 MoGR-multi 数据集上进行了计算速度对比实验,并以每幅图像的处理时间(ms)作为衡量标准,噪声水平为 30 dB,实验结果如表 5 所示。

从实验结果可知,在推理速度方面,本文方法 DiffPhase 表现适中,比部分有监督方法要快。传统方法如 LSPU 和 QGPU 不需要训练或依赖于其他先验知识,因此推理速度非常快。无监督方法中, U3Net 通过端到端的预训练模型实现了快速推理,而 PUDIP 由于每幅图像都需要单独计算推理,导致速度非常缓慢。尽管 DiffPhase 的模型体量较大,但得益于其高效的架构设计与优化策略,推理速度仍具备一定的竞争力。由于本文方法采用了更复杂的网络结构和参数量来提升性能,不可避免地增加了计算成本,这些计算开销是为了在极端噪声和高复杂度的场景下实现更强的鲁棒性和更高的精度。

结合应用场景的 2.3 节和 2.4 节的实验结果可

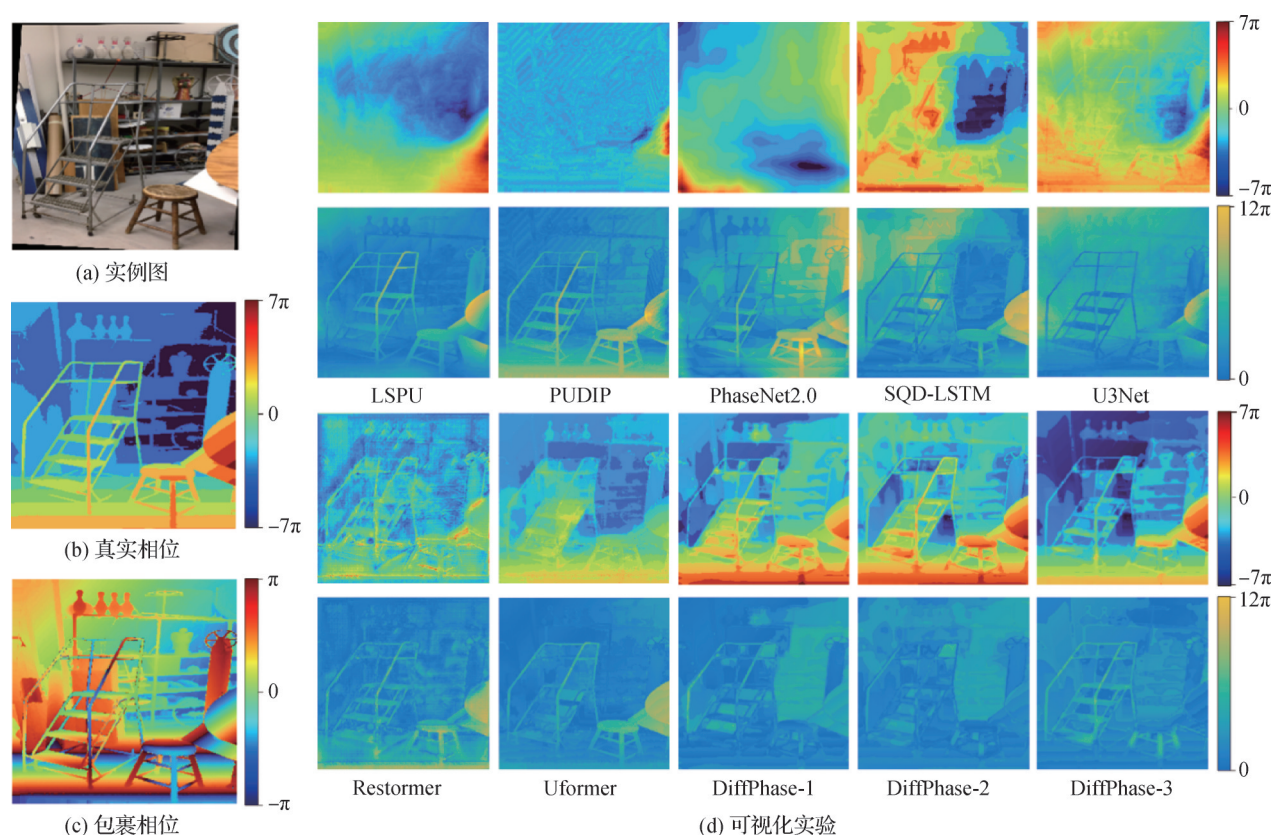


图7 MS-phase数据集实验结果

Fig. 7 MS-phase dataset experimental results

((a) instance image; (b) absolute phase; (c) wrapped phase; (d) experiment visualization)

以看出,在高噪声且对精度要求较高的应用中,Diff-Phase的表现优于其他方法,展现出很高的应用价值。这种计算成本与性能提升的权衡表明,本文方法在特定条件下具有明显的优势和实用性。

2.6 不同特征融合方式对比

为了探索不同特征融合方式对模型性能的具体影响,本文设置双频相位条件引导、30 dB噪声的条件,在MoGR-multi数据集上进行对比实验,实验结果如表6所示,评价指标为NRMSE,数值越小表示性能越佳。在实验中尝试了多种不同的特征融合方法,包括直接相加、拼接和多头注意力机制。实验结果表明,采用注意力机制的方案效果最佳,显著提升了模型的特征融合能力及整体性能。

相比于相加和拼接,使用多头注意力机制进行特征融合和交互有一个显著的优点,即交叉注意力层可以接受任意数量的包裹相位特征输入,使得本文方法可以自适应提取不同频率包裹相位中的信息,提高模型对不同应用场景的适应性。交叉注意力层能够同时综合多个包裹相位的不同特征,利用

它们提供的全局结构和局部细节共同指导生成过程,提升了网络对复杂输入的处理能力。这一过程在网络的多个扩散阶段反复进行,实现了分层、分频和动态的融合策略,能够有效缓解简单拼接方式中可能存在的特征冗余或冲突问题。

2.7 消融实验

为了探索方法中各组成模块对模型性能的具体影响,本文设置30 dB噪声的条件,在MoGR-multi数据集上进行系统的消融实验,结果如表7所示,评价指标为NRMSE,数值越小表示性能越佳。

表7的结果表明,所设计的基于UNet的特征提取模块在性能上优于传统卷积神经网络(convolutional neural network, CNN)和视觉Transformer(vision Transformer, ViT)。与传统CNN相比,UNet能更有效地捕捉包裹相位中的多尺度特征;与ViT相比,UNet在处理局部细节与空间连续性方面更具优势,特别是在有限样本下表现出更高的稳定性。其编码—解码架构在降维与升维过程中引入的多层特征图显著增强了对不同尺度信息的表达能力。同

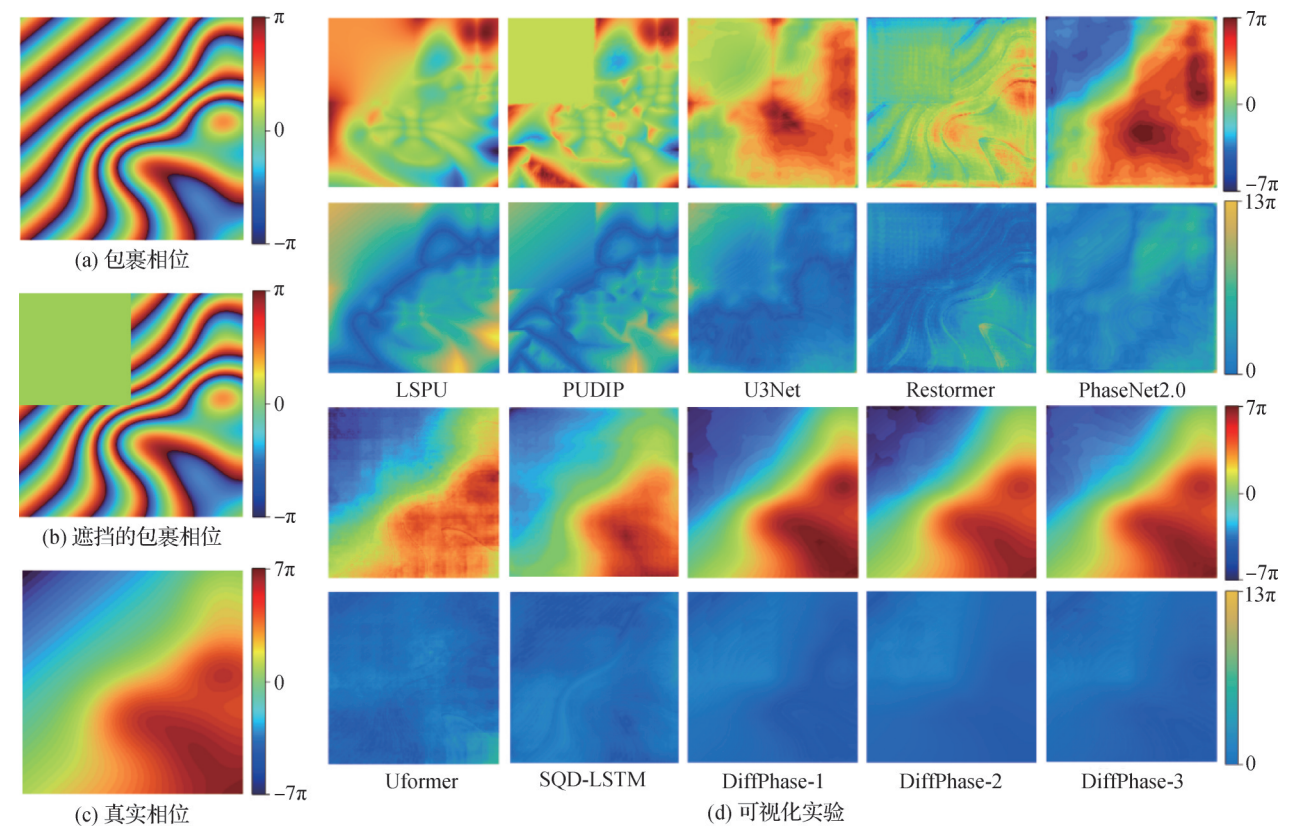


图8 恢复实验结果

Fig. 8 Restoration experimental results

((a) wrapped phase; (b) occluded wrapped phase; (c) absolute phase; (d) experiment visualization)

表5 模型计算速度对比

Table 5 Computation speed comparison of models	
方法	时间/ms
LSPU	0.018
QGPU	0.014
PUDIP	99.65
U3Net	0.071
SQD-LSTM	0.078
PhaseNet2.0	0.14
Uformer	0.25
Restormer	0.27
DiffPhase-1	0.15
DiffPhase-2	0.27
DiffPhase-3	0.36

注:加粗字体表示最优结果。

时,由于该结构与扩散模型保持一致,使得在扩散过程中的每一交叉注意力层中能够准确注入对应尺度的特征,引导生成过程更加精准与一致,从而进一步

表6 特征融合方式对比实验

Table 6 Feature fusion comparison experiments			
相加	拼接	注意力机制	NRMSE/%
√	—	—	0.47
—	√	—	0.51
—	—	√	0.26

注:加粗字体表示最优结果。“—”表示不使用,“√”表示使用。

表7 消融实验

Table 7 Ablation experiments					
CNN	ViT	UNet	三步去噪	三频相位条件引导	NRMSE/%
√	—	—	—	—	0.81
—	√	—	—	—	0.67
—	—	√	—	—	0.40
—	—	√	√	—	0.33
—	—	√	√	√	0.26

注:加粗字体表示最优结果。“—”表示不使用,“√”表示使用。

提升了解包裹精度和三维重建质量。

在扩散过程方面,三步扩散去噪机制通过分阶段建模和逐步优化,显著提升了模型的抗噪性与对复杂结构的恢复能力。与一步或两步去噪方式相比,虽然计算开销略有增加,但在性能提升和生成过程的可解释性方面展现出明显优势。一般而言,随着噪声水平的升高,适当增加去噪步数可以逐步消除更多高斯噪声成分,从而在高噪声环境下获得更精细的相位预测结果。然而,实验结果表明,在缺少真实图像、仅有噪声图像情况下,即使保持较小的去噪步数,模型也能凭借其强大的全局建模能力与多尺度特征约束,在高噪声场景中保持较高的预测精度。这说明模型本身在噪声干扰下的鲁棒性已经足够强,去噪步数的调整仅能带来边际上的性能提升。

在条件引导方面,多频特征交互机制(即三频包裹相位条件引导)使模型能够充分融合不同频率包裹相位图所携带的互补信息,有效结合低频提供的全局结构和高频捕捉的局部细节,从而提升了模型对复杂输入的适应性和解包裹精度。值得说明的是,表中未使用三频条件引导的实验,默认仅使用单幅包裹相位图作为条件输入,该设置类似于现有方法中常见的单图输入策略。对比结果表明,多频条件引导在整体性能上具有显著优势,也反映出当前诸多仅使用单幅图像的方法在表达能力上的不足。

3 结 论

本文提出一种基于条件扩散模型的多频相位解包裹方法,结合三维成像充分发挥扩散模型在噪声鲁棒性和全局建模能力方面的优势,有效提升了复杂条件下的解包裹性能与三维成像精度。本文设计了一种自适应多频相位输入的多频相位解包裹方法,构建了与扩散网络结构对齐的多尺度特征提取模块,采用两阶段训练策略提升训练稳定性与泛化能力。模型有效融合多频率信息的全局结构与局部特征,增强了特征传递性能,优化了基于相位解包裹的场景深度恢复效果。

将 DiffPhase 与最新的相位解包裹方法在强噪声和遮挡条件下进行了性能比较。实验结果表明,所提方法在相位解包裹精度上优于现有方法,并能在真实三维场景重建中仍保持较高的重建精度,尤其在细节边缘与复杂结构区域展现出更优的准确性

与鲁棒性。此外,本文评估了 DiffPhase 在一些具有挑战性问题上的表现,例如在遮挡情况下的恢复能力,实验表明 DiffPhase 能够有效恢复绝对相位,实现稳健的真实场景恢复。

参考文献 (References)

- Arealillo-Herráez M, Villatoro F R and Gdeisat M A. 2016. A robust and simple measure for quality-guided 2D phase unwrapping algorithms. *IEEE Transactions on Image Processing*, 25 (6) : 2601-2609 [DOI: 10.1109/TIP.2016.2551370]
- Baek W K and Jung H S. 2024. Phase unwrapping of SAR interferogram from modified U-net via training data simulation and network structure optimization. *Remote Sensing of Environment*, 314: #114392 [DOI: 10.1016/j.rse.2024.114392]
- Bioucas-Dias J M and Valadao G. 2007. Phase unwrapping via graph cuts. *IEEE Transactions on Image Processing*, 16 (3) : 698-709 [DOI: 10.1109/TIP.2006.888351]
- Budillon A, Ferraiuolo G, Pascasio V and Schirrinzi G. 2005. Multichannel SAR interferometry via classical and Bayesian estimation techniques. *EURASIP Journal on Advances in Signal Processing*, 2005(20): #791816 [DOI: 10.1155/ASP.2005.3180]
- Chen C W and Zebker H A. 2002. Phase unwrapping for large SAR interferograms: statistical segmentation and generalized network models. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8): 1709-1719 [DOI: 10.1109/TGRS.2002.802453]
- Chen Z L, Quan Y H and Ji H. 2024. Unsupervised deep unrolling networks for phase unwrapping//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 25182-25192 [DOI: 10.1109/CVPR52733.2024.02379]
- Dardikman G and Shaked N T. 2018. Phase unwrapping using residual neural networks//*Proceedings of the Computational Optical Sensing and Imaging 2018*. Orlando, USA: Optica Publishing Group [DOI: 10.1364/COSI.2018.CW3B.5]
- De Souza J C, Oliveira M E and Dos Santos P A M. 2015. Branch-cut algorithm for optical phase unwrapping. *Optics Letters*, 40 (15) : 3456-3459 [DOI: 10.1364/OL.40.003456]
- Du S, Zou Q, Chen C and Yang B S. 2025. Dual-frequency spatio-temporal phase unwrapping//*Proceedings of 2025 ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing*. Hyderabad, India: IEEE, 2025: 1-5 [DOI: 10.1109/ICASSP49660.2025.10890711]
- Fang M Q. 2023. Wrap-type phase unwrapping utilizing arctangent operation as a differential amplifier through shifted phase maps. *Optics Express*, 31 (19) : 31622-31633 [DOI: 10.1364/OE.500826]
- Feng M T, Shen J H, Wu Z J, Peng W X, Zhong H, Guo Y L, et al.

2025. Advancements in 3D vision understanding using multimodal large language models. *Journal of Image and Graphics*, 30(6): 1744-1791 (冯明涛, 沈军豪, 武子杰, 彭伟星, 钟杭, 郭裕兰, 等. 2025. 多模态大模型驱动的三维视觉理解技术前沿进展. *中国图象图形学报*, 30(6): 1744-1791) [DOI: 10.11834/jig.240588]
- Flynn T J. 1997. Two-dimensional phase unwrapping with minimum weighted discontinuity. *Journal of the Optical Society of America A*, 14(10): 2692-2701 [DOI: 10.1364/JOSAA.14.002692]
- Gao D P and Yin F L. 2012. Mask cut optimization in two-dimensional phase unwrapping. *IEEE Geoscience and Remote Sensing Letters*, 9(3): 338-342 [DOI: 10.1109/LGRS.2011.2168940]
- Ghiglia D C and Romero L A. 1994. Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *Journal of the Optical Society of America A*, 11(1): 107-117 [DOI: 10.1364/JOSAA.11.000107]
- Ghiglia D C and Romero L A. 1996. Minimum L^p -norm two-dimensional phase unwrapping. *Journal of the Optical Society of America A*, 13(10): 1999-2013 [DOI: 10.1364/JOSAA.13.001999]
- Goldstein R M, Zebker H A and Werner C L. 1988. Satellite radar interferometry: two-dimensional phase unwrapping. *Radio Science*, 23(4): 713-720 [DOI: 10.1029/RS023i004p00713]
- Gontarz M, Dutta V, Kujawińska M and Krauze W. 2023. Phase unwrapping using deep learning in holographic tomography. *Optics Express*, 31(12): 18964-18992 [DOI: 10.1364/OE.486984]
- Guo Y, Chen X T and Zhang T. 2014. Robust phase unwrapping algorithm based on least squares. *Optics and Lasers in Engineering*, 63: 25-29 [DOI: 10.1016/j.optlaseng.2014.06.007]
- Haniff C A. 1991. Least-squares Fourier phase estimation from the modulo 2π bispectrum phase. *Journal of the Optical Society of America A*, 8(1): 134-140 [DOI: 10.1364/JOSAA.8.000134]
- He W, Xia L and Liu F. 2014. Sparse-representation-based direct minimum L^p -norm algorithm for MRI phase unwrapping. *Computational and mathematical methods in medicine*, 2014(1): #134058 [DOI: 10.1155/2014/134058]
- Herráez M A, Burton D R, Lalor M J and Gdeisat M A. 2002. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied Optics*, 41(35): 7437-7444 [DOI: 10.1364/AO.41.007437]
- Huang Q, Zhou H Q, Dong S C and Xu S J. 2015. Parallel branch-cut algorithm based on simulated annealing for large-scale phase unwrapping. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7): 3833-3846 [DOI: 10.1109/TGRS.2014.2385482]
- Huang Y, Cheng B, Fang S J and Liu X. 2025. Shadow removal with wavelet-based non-uniform diffusion model. *Journal of Image and Graphics*, 30(1): 66-82 (黄颖, 程彬, 房少杰, 刘歆. 2025. 用于阴影去除的小波非均匀扩散模型. *中国图象图形学报*, 30(1): 66-82) [DOI: 10.11834/jig.230904]
- Li Q L, Bao C Y, Zhao J S and Jiang Z G. 2018. A new fast quality-guided flood-fill phase unwrapping algorithm. *Journal of Physics: Conference Series*, 1069(1): #012182 [DOI: 10.1088/1742-6596/1069/1/012182]
- Nico G, Palubinskas G and Datcu M. 2000. Bayesian approaches to phase unwrapping: theoretical study. *IEEE Transactions on Signal Processing*, 48(9): 2545-2556 [DOI: 10.1109/78.863057]
- Perera M V and De Silva A. 2021. A joint convolutional and spatial quad-directional LSTM network for phase unwrapping//*Proceedings of 2021 ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing*. Toronto, Canada: IEEE: 4055-4059 [DOI: 10.1109/ICASSP39728.2021.9414748]
- Scharstein D, Hirschmüller H, Kitajima Y, Krathwohl G, Nešić N, Wang X, et al. 2014. High-resolution stereo datasets with subpixel-accurate ground truth//*Proceedings of the 36th German Conference on Pattern Recognition*. Münster, Germany: Springer: 31-42 [DOI: 10.1007/978-3-319-11752-2_3]
- Sica F, Calvanese F, Scarpa G and Rizzoli P. 2022. A CNN-based coherence-driven approach for InSAR phase unwrapping. *IEEE Geoscience and Remote Sensing Letters*, 19: #4003705 [DOI: 10.1109/LGRS.2020.3029565]
- Silberman N, Hoiem D, Kohli P and Fergus R. 2012. Indoor segmentation and support inference from RGBD images//*Proceedings of the 12th European Conference on Computer Vision*. Florence, Italy: Springer: 746-760 [DOI: 10.1007/978-3-642-33715-4_54]
- Spoorthi G E, Gorthi R K S S and Gorthi S. 2020. PhaseNet 2.0: phase unwrapping of noisy data based on deep learning approach. *IEEE Transactions on Image Processing*, 29: 4862-4872 [DOI: 10.1109/TIP.2020.2977213]
- Spoorthi G E, Gorthi S and Gorthi R K S S. 2019. PhaseNet: a deep convolutional neural network for two-dimensional phase unwrapping. *IEEE Signal Processing Letters*, 26(1): 54-58 [DOI: 10.1109/LSP.2018.2879184]
- Strand J, Tact T and Jain A K. 1999. Two-dimensional phase unwrapping using a block least-squares method. *IEEE Transactions on Image Processing*, 8(3): 375-386 [DOI: 10.1109/83.748892]
- Su X Y and Chen W J. 2004. Reliability-guided phase unwrapping algorithm: a review. *Optics and Lasers in Engineering*, 42(3): 245-261 [DOI: 10.1016/j.optlaseng.2003.11.002]
- Wang L, Yi L N, Zhang Y T, Wang X F, Wang W, Wang X J, et al. 2024. 3D reconstruction method based on N -step phase unwrapping. *The Visual Computer*, 40(5): 3601-3613 [DOI: 10.1007/s00371-023-03054-y]
- Wang K Q, Li Y, Kemao Q, Di J L and Zhao J L. 2019. One-step robust deep learning phase unwrapping. *Optics Express*, 27(10): 15100-15115 [DOI: 10.1364/OE.27.015100]
- Wang Z D, Cun X D, Bao J M, Zhou W G, Liu J Z and Li H Q. 2022. Uformer: a general U-shaped transformer for image restoration//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 17662-17672

- [DOI: 10.1109/CVPR52688.2022.01716]
- Wei Y Y, Mao T Y, Li B A, Wang F, Li F, Zhang Z, et al. 2025. Visual and large multimodal models promote image restoration and enhancement: research progress. *Journal of Image and Graphics*, 30(5): 1197-1219 (韦炎炎, 毛天一, 李柏昂, 王飞, 李锋, 张召, 等. 2025. 视觉模型及多模态大模型推进图像复原增强: 研究进展. *中国图象图形学报*, 30(5): 1197-1219) [DOI: 10.11834/jig.240436]
- Xia H T, Montresor S, Guo R X, Li J C, Yan F, Cheng H M, et al. 2016. Phase calibration unwrapping algorithm for phase data corrupted by strong decorrelation speckle noise. *Optics Express*, 24(25): 28713-28730 [DOI: 10.1364/OE.24.028713]
- Xu Y F, Hu X M, Huang T Y, Li S and Chen L. 2024. Precise image translation based on conditional diffusion model for driving scenarios. *Journal of Image and Graphics*, 29(11): 3305-3318 (徐映芬, 胡学敏, 黄婷玉, 李桑, 陈龙. 2024. 面向驾驶场景精准图像翻译的条件扩散模型. *中国图象图形学报*, 29(11): 3305-3318) [DOI: 10.11834/jig.230785]
- Yang F S, Pham T A, Brandenburg N, Lütolf M P, Ma J W and Unser M. 2021. Robust phase unwrapping via deep image prior for quantitative phase imaging. *IEEE Transactions on Image Processing*, 30: 7025-7037 [DOI: 10.1109/TIP.2021.3099956]
- Ying L, Liang Z P, Munson D C, Koetter R and Frey B J. 2006. Unwrapping of MR phase images using a Markov random field model. *IEEE Transactions on Medical Imaging*, 25(1): 128-136 [DOI: 10.1109/TMI.2005.861021]
- Yu R M, Yu H S, Sun W and Akhtar N. 2024. Color phase order coding and interleaved phase unwrapping for three-dimensional shape measurement with few projected pattern. *Optics & Laser Technology*, 168: #109842 [DOI: 10.1016/j.optlastec.2023.109842]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: efficient transformer for high-resolution image restoration//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]
- Zeng Q W, Dong Z Y, Yang X Z and Chong F T. 2023. Interferometric phase denoising combining global context and fused attention. *Journal of Image and Graphics*, 28(9): 2956-2968 (曾庆旺, 董张玉, 杨学志, 种法亭. 2023. 结合全局上下文与融合注意力的干涉相位去噪. *中国图象图形学报*, 28(9): 2956-2968) [DOI: 10.11834/jig.220562]
- Zhang J C, Tian X B, Shao J B, Luo H B and Liang R G. 2019. Phase unwrapping in optical metrology via denoised and convolutional segmentation networks. *Optics Express*, 27(10): 14903-14912 [DOI: 10.1364/OE.27.014903]

作者简介

李妍,女,硕士研究生,主要研究方向为机器视觉。

E-mail: yan_li@whu.edu.cn

邹勤,通信作者,男,教授,主要研究方向为机器视觉与机器人。E-mail: qzou@whu.edu.cn

黄霁,男,硕士研究生,主要研究方向为深度学习和图像处理。E-mail: huangjustc@mail.ustc.edu.cn

李清泉,男,教授,主要研究方向为动态精密工程测量、时空数据智能处理与分析。E-mail: liqq@szu.edu.cn