引用格式：YANG Simin, ZHANG Yandong, WANG Yang, et al. Research on Structured Light Wrapped Phase Unwrapping Algorithm Based on Deep Learning[J]. Acta Photonica Sinica, 2026, 55(2):0212001
杨思敏,张衍东,王杨,等.基于深度学习的结构光包裹相位展开算法研究[J].光子学报,2026,55(2):0212001
※封面论文※

基于深度学习的结构光包裹相位展开算法研究

杨思敏[1],张衍东[1],王杨[2],尚建华[1]
(1 东华大学 信息与智能科学学院,上海 201620)
(2 中国科学院上海光学精密机械研究所 高功率激光物理联合实验室,上海 201800)

摘 要:现有的基于深度学习的绝对相位提取算法,虽然在精度和效率方面有所改进,但仍存在网络参数量过大、对不连续相位求解精度低等问题,为此提出一种基于深度学习的结构光包裹相位展开网络,以简洁高效的编码器-解码器网络框架为骨干网络,并设计自适应时钟长短期记忆模块捕获上下文和方向敏感特征;通过引入并改进多尺度注意力机制,有效提高算法的信息获取能力和特征融合的稳定性及效率。为了评估该网络模型的有效性,首先借助包含随机噪声干扰、相位不连续跳变及混叠效应的三组模拟数据集,与当前主流深度学习相位展开模型进行比较实验和消融实验,实验结果表明所提网络的性能均优于对比网络模型;其次,借助公开数据集测试,所提网络的均方根误差为0.505 2,平均绝对误差为0.180 9 rad,结构相似性指数为0.997 1,表明该网络具有良好的泛化能力。

关键词:光学测量;相位展开;深度学习;注意力机制;长短期记忆网络
中图分类号:TP391;TH741 **文献标识码**:A **doi**:10.3788/gzxb20265502.0212001

0 引言

结构光三维测量技术已广泛应用于工业设计、智能化生产、虚拟现实以及文化遗产保护等领域[1],通过投射多帧具有固定相移的余弦条纹或经过编码的条纹图像到物体表面,利用工业相机捕获变形后的条纹图案,然后借助空间几何关系和相位展开算法,最终求解得到物体表面的三维形貌。由于通过反正切函数获取的相位信息会被限制在(-π,π]区间内,呈2π周期性不连续,因此这种模糊包裹相位无法直接用于精确提取物体表面轮廓等参数;而相位展开算法则能有效消除相位跳变、恢复包裹相位的物理连续性[2],是结构光三维测量中相位信息处理的关键环节[3]。

传统相位展开算法主要分为时间相位展开和空间相位展开。空间相位展开只需单张包裹相位图,通过分析像素邻域的相位分布即可得到该像素的绝对相位,但在处理大面积不连续区域和孤立物体时效果不佳;时间相位展开则借助多组相位映射或附加编码图案为每个像素独立提供条纹级次信息,即使相位不连续也能取得较好的效果,但该方法需投射多幅条纹图案,这在一定程度上将影响测量效率。

为解决传统相位展开算法的不足,研究学者将深度学习应用于该算法的优化中,通过构建端到端的神经网络,在保证测量效率的同时提升了相位展开的精度[4]。SPOORTHI G E等[5]将相位展开视为语义分割问题,提出一种卷积编码器-解码器网络架构,在强噪声环境下能够保持较高的精度和计算效率,但对快速变化及不连续相位的处理能力有限。WANG Kaiqiang等[6]将相位展开建模为回归任务,基于U-Net(U-shaped Network,U-Net)和残差网络[7](Residual Network,ResNet)设计了DLPU模型(Deep Learning Phase Unwrapping,DLPU),实现了一步式相位展开,极大提升了相位展开效率,但在锐利边缘处,解缠后的相位表面易过度平滑。ZHANG Teng等[8]采用DeepLabV3+[9]卷积神经网络,有效消除了包裹相位图中的相位跳

基金项目:国家自然科学基金(52173219)
第一作者:杨思敏,2242240@mail.dhu.edu.cn
通讯作者:尚建华,jhshang@dhu.edu.cn
收稿日期:2025-10-30;**录用日期**:2026-01-12

跃。PERERA M V等[10]在以编码器-解码器为结构的卷积神经网络中引入空间四向长短期记忆网络[11](Long Short-Term Memory,LSTM),提升了强噪声条件下的相位展开精度,但针对相位不连续情况的研究尚显不足。HUANG Wangwang等[12]提出的高分辨率深度学习网络(High-Resolution Deep Learning Network,HRNet),通过推断每个像素的2π整数值有效消除了不连续性问题,抗噪能力强,能应对相位大范围变化的情况。ZHANG Ziheng等[13]基于Transformer提出了SFNet模型,在多噪声和不连续相位场景下,借助无需位置编码的层次化编码器与轻量MLP解码器获得了良好的鲁棒性,但语义分割在类边界处的像素模糊问题依然存在。LI Kejia等[14]针对高散斑噪声提出一种三阶段训练策略相位展开方法,将相位恢复过程分解为包裹相位去噪、包裹计数预测以及展开相位误差补偿三个顺序训练的阶段,结合空间注意力和基于DeepLabV3+的预测网络增强了模型的噪声鲁棒性,借助自动编码器进行最终相位校正,有效提升了强噪声下的相位恢复精度。LI Youxing等[15]引入可变形卷积的二维相位展开网络PUDCN,借助金字塔特征动态选择模块学习多尺度特征的表达,利用融合特征动态选择模块实现跨层特征的融合,该网络能自适应聚焦跳变相位边缘等关键区域,对复杂相位变化的适应能力强。WAN Liang等[16]提出一种两阶段相位展开方法,第一阶段对包裹相位进行去噪处理,第二阶段通过引入注意力模块实现了语义特征的重新校准,在强噪声环境下对相位跳变区域具有很强的处理能力。刘哲等[17]提出的相位展开模型(Effective Phase Unwrapping Network,EUP-Net)采用编码器-解码器架构,以包裹相位图和与之对应的单帧条纹图作为输入,经过网络的处理和运算得到绝对相位,相位展开精度高且网络轻量化,但其数据集相对单一,一定程度上限制了算法泛化能力的验证。

当前,基于深度学习的相位展开算法研究在噪声抑制和边缘保持方面已取得了较好的进展,但在强噪声和相位不连续等复杂场景下,仍需对相位求解精度、模型复杂度及模型泛化能力等方面进行协同优化。为此,本文提出一种基于深度学习的注意力导向编解码器网络(Attentive Directional Encoder-Decoder Network,ADE-Net),以编码器-解码器架构为主体框架,并融合自适应时钟长短期记忆模块(Adaptive Clockwork Long Short-Term Memory,AC-LSTM)和深度可分离高效多尺度注意力(Depthwise-separable-convolution Efficient Multi-Scale Attention,DEMA)。AC-LSTM能够突破标准卷积的局部性限制和传统LSTM的方向感知局限,可使ADE-Net更全面细致地捕捉复杂场景下相位信息的时空特征;DEMA则是在高效多尺度注意力[18](Efficient Multi-Scale Attention,EMA)的基础上改进得到的,在继承EMA关键信息聚焦能力的同时优化了计算流程和参数设置,从而有效降低了计算成本,提升了网络整体运行效率。最后,为了全面评估ADE-Net的性能,通过在含噪声、相位不连续和混叠的模拟数据集与真实数据集上进行训练和测试,定量评估了ADE-Net的相位展开性能。

1 ADE-Net网络模型的设计

相位展开的本质是从包裹相位图中恢复得出连续的绝对相位分布,因此,可将相位展开问题等效转化为回归问题进行处理。本文提出的网络模型将包裹相位图像作为输入,利用神经网络先对输入图像的空间特征进行提取,捕捉图像中的局部和全局特征信息,然后构建长程依赖关系模型,分析图像中不同位置像素之间的关联信息,最终得到高精度的解包裹相位图像,实现相位信息的准确还原。

1.1 模型框架

提出的ADE-Net网络采用编码器-解码器架构为核心主体框架,并加入AC-LSTM模块和注意力机制模块,网络的整体结构设计如图1所示,参数配置如表1所示。

编码器可划分为四个特征提取阶段,每个特征提取阶段均包含一个卷积块和一个下采样模块,逐步提取从低级到高级的特征信息。具体而言,每个卷积块由3×3卷积、批归一化和ReLU激活函数组成;下采样模块则采用哈尔小波下采样结构,相较于传统池化操作,该结构通过可分离的低通滤波器与高通滤波器实现了特征图空间维度的压缩,同时将相位跳变边缘等关键方向性高频细节结构化保留至不同子带中,进而有效避免了常规池化操作可能导致的特征信息损失问题,并为后续模块准确处理2π跳变和恢复相位连续性提供了更为完整的特征基础[19]。编码器的最终输出结果将输入到AC-LSTM模块,该模块负责多尺度与方向敏感特征的协同提取,进而为解码阶段提供丰富全面的上下文信息。

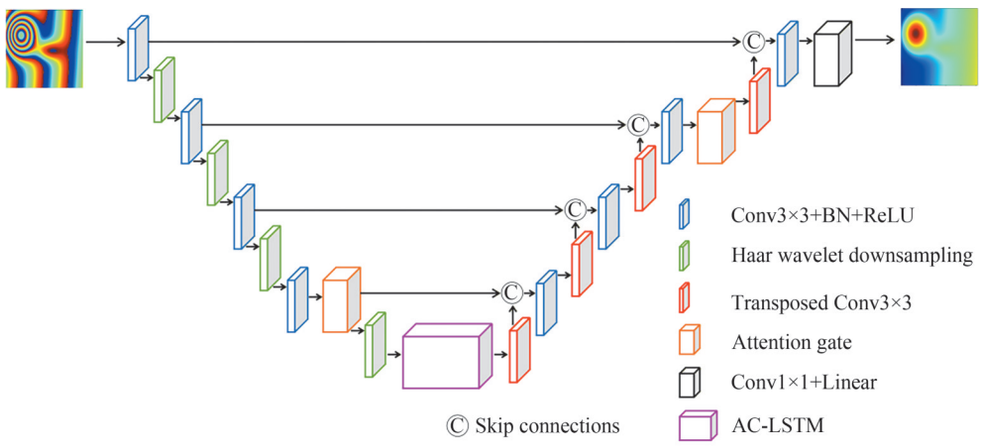


图1　ADE-Net网络模型结构

Fig.1　ADE-Net network model structure

表1　ADE-Net网络参数配置

Table 1　Parameter configurations of ADE-Net

Stage		Module type	Channels (in→out)	Module count	Output feature map size ($H\times W$)
Encoder	Stage 1	Conv block+HWD down-sampling	1→16	1	128×128
	Stage 2	Conv block+HWD down-sampling	16→32	1	64×64
	Stage 3	Conv block+HWD down-sampling	32→64	1	32×32
	Stage 4	Conv block+HWD down-sampling	64→128	1	16×16
AC-LSTM module	ASPP branch	ASPP+EMA	128→128	1	16×16
	LSTM branch	Bi-LSTM+CAA	128→128	2(horizontal/vertical)	16×16
	Feature fusion	1×1 convolution	256→128	1	16×16
Decoder	Stage 1	Transposed Conv+skip connection	128→128	1	32×32
	Stage 2	Transposed Conv+skip connection	128→64	1	64×64
	Stage 3	Transposed Conv+skip connection	64→32	1	128×128
	Stage 4	Transposed Conv+skip connection	32→16	1	256×256
	Output layer	1×1 convolution	16→1	1	256×256

解码器则采用基于3×3转置卷积的对称上采样结构，用以有效恢复特征图的空间分辨率。为了确保细节信息的有效传递，每个上采样阶段均引入跳跃连接，将编码器中对应尺度的特征与上采样结果进行融合。此外，在编码器末端卷积块和解码器的最后一个上采样操作之前加入注意力机制，以在不破坏浅层特征细节信息的前提下，以最小的计算成本增强高级特征的表达能力。最后，网络借助线性激活函数的1×1卷积层调整通道数，进而将解码器输出的高维特征映射为单通道相位图。

1.2　AC-LSTM模块

针对标准卷积操作固有的局部性限制及传统LSTM缺乏方向感知能力的问题，提出AC-LSTM模块，该模块能同步捕获多尺度上下文信息和方向敏感的局部细节，增强模型的特征表达能力，为后续任务提供更丰富的可学习特征。

AC-LSTM模块的结构如图2(a)所示，由两个并行主干分支构成，其输入为形状为$C\times H\times W$的输入特征X，其中，C为特征通道数，H为特征图高度，W为特征图宽度。第一个分支包含一个空洞空间金字塔池化模块[20](Atrous Spatial Pyramid Pooling，ASPP)，网络结构如图2(d)，通过多个不同膨胀率的空洞卷积和全局平均池化，协同捕获多尺度上下文信息以及特征图中的微小细节，其计算表达式为

$$F_{\text{aspp}}=\text{Conv}_{1\times1}\left(\text{Concat}\left(\left\{\text{Con}_{3\times3}^{(d_i)}(X)\right\}_{i=1}^{N},\text{u}\left(\text{Conv}_{1\times1}\left(P_{\text{avg}}(X)\right)\right)\right)\right) \tag{1}$$

式中，F_{aspp}为ASPP模块的输出特征图，d_i为第i个空洞卷积的扩张率，N为空洞卷积分支的数量，P_{avg}代表全局平均池化操作，u代表上采样操作，$Conv_{k\times k}$代表$k\times k$卷积运算，Concat代表通道维度的特征拼接运算。

第二个分支采用两个独立的双向长短期记忆网络（Bidirectional Long Short-Term Memory，Bi-LSTM）处理输入特征，该结构使模型能够同时学习“过去到现在”和“未来到现在”的长程双向依赖关系。如图2(b)所示，水平Bi-LSTM将特征图的每一行特征视作一个一维序列，通过正向LSTM（从左到右）和反向LSTM（从右到左）捕捉水平方向的上下文关系；垂直Bi-LSTM则将每一列的特征视为一个一维序列，并沿正向LSTM（从上到下）和反向LSTM（从下到上）捕捉垂直方向的上下文关系。上述两个方向的Bi-LSTM协同工作，其输出共同构成该分支网络的最终输出特征，表达式为

$$F_{hor}=\mathrm{BiLSTM}_{horizontal}(X),F_{ver}=\mathrm{BiLSTM}_{vertical}(X) \tag{2}$$

尽管Bi-LSTM能够有效构建序列之间的依赖关系，但其顺序处理的方式使它无法判断空间维度特征的重要性，在抑制冗余背景信息方面存在着明显不足。为了突破这一限制，受CAI Xinhao等[21]提出的上下文锚点注意力模块（Context Anchor Attention，CAA）的启发，将水平与垂直方向的Bi-LSTM分别与对应方向的CAA模块进行串联，从而分别得到F_{hor}^{caa}和F_{ver}^{caa}。如图2(c)所示，CAA模块通过水平与垂直方向的分离卷积操作，准确捕捉到方向敏感的局部特征；同时，上下文锚点计算注意力权重，恰好能弥补Bi-LSTM在空间感知能力上的不足，进而实现序列上下文特征的精细化调控；最后，将两个方向感知特征进行拼接，并使用3×3卷积层进行融合，具体实现过程数学表达式为

$$F_{lstm}=\mathrm{Conv}_{3\times 3}\left(\mathrm{Concat}\left(F_{hor}^{caa},F_{ver}^{caa}\right)\right) \tag{3}$$

随后，将第二分支得到的融合特征与第一分支中ASPP模块输出的特征沿通道维度进行拼接，再采用1×1卷积层对拼接后的特征进行融合降维处理，最终得到一个形状为$C_{out}\times H\times W$的输出特征图Y，即

$$Y=\mathrm{Conv}_{1\times 1}\left(\mathrm{Concat}\left(F_{aspp},F_{lstm}\right)\right) \tag{4}$$

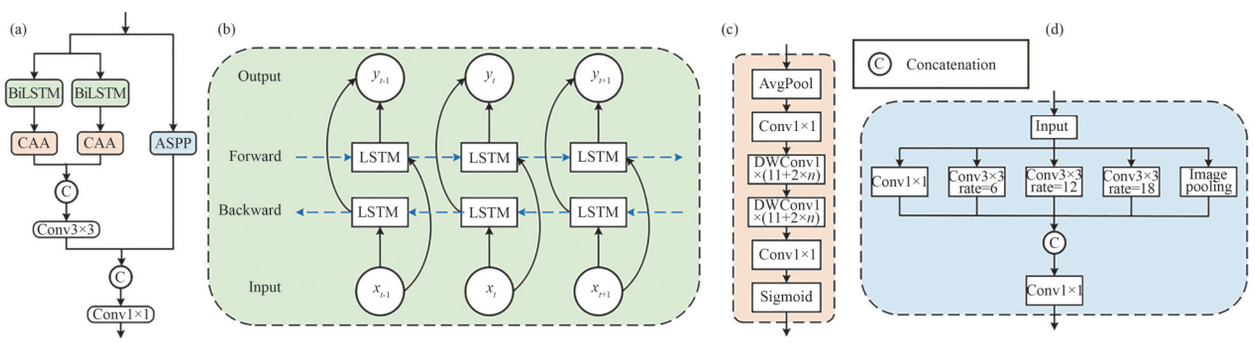


图2 AC-LSTM结构示意图。(a)AC-LSTM模块；(b)Bi-LSTM模块；(c)CAA模块；(d)ASPP模块

Fig. 2 AC-LSTM overview. (a) AC-LSTM module; (b) Bi-LSTM module; (c) CAA module; (d) ASPP module

1.3 注意力机制模块DEMA

在深度学习模型中，注意力机制已成为增强网络表征能力的关键技术，高效多尺度注意力模块凭借特征分组策略、多尺度并行分支架构和跨空间学习能力，能够准确捕捉并强化关键信息。然而，现有的注意力机制模块在特征融合时，存在维度对齐方面的缺陷，且计算效率有待进一步提升。鉴于此，提出一种改进型高效多尺度注意力模块DEMA。

本文在原始EMA模块架构的基础上进行了两处改进，得到DEMA模块，其结构示意图如图3(a)所示。首先，引入深度可分离卷积代替传统的标准3×3卷积操作，可以在保持感受野范围不变的前提下，显著降低计算复杂度，提升模块的运行效率；然后，使用更加稳定可靠的批量矩阵乘法（Batch Matrix Multiply，BMM）取代通用矩阵乘法Matmul，提升批量数据并行处理的效率。如图3所示，对于输入的特征图F，其形状为$C\times H\times W$，DEMA模块首先通过分组操作，将通道划分为G组，即每组通道数C'等于$C//G$；其次，对分组后的特征分别进行自适应高度池化和自适应宽度池化操作，得到形状分别为$C'\times 1\times W$和$C'\times H\times 1$的特征图；将上述两个特征图沿通道维度进行拼接，并用1×1融合卷积对其进行压缩处理，再经Sigmoid激活函数

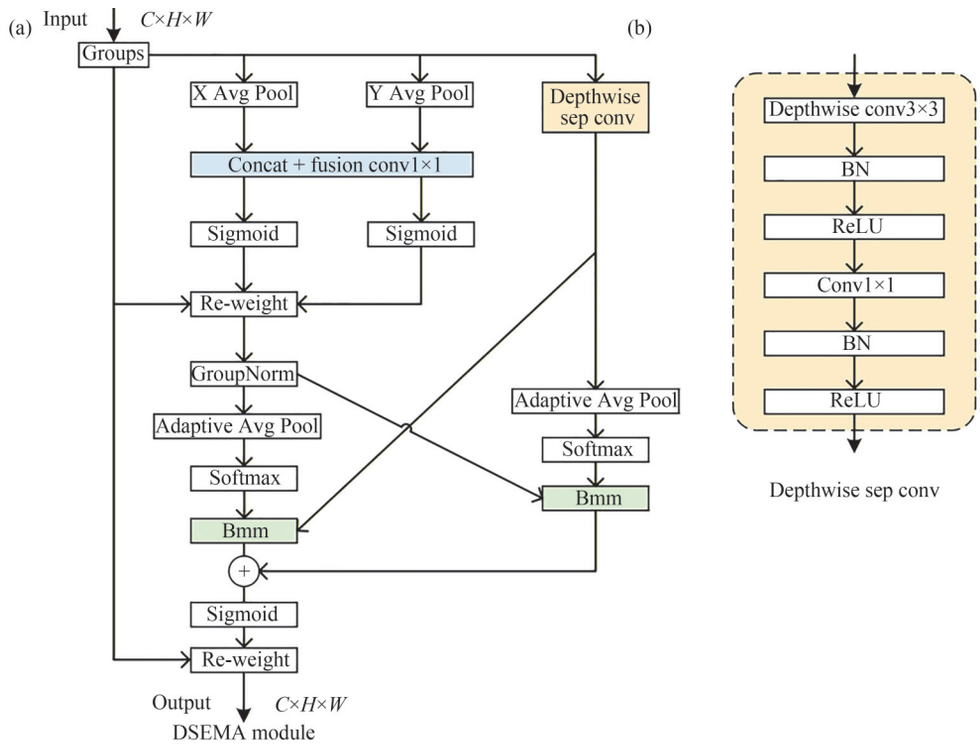


图3　DEMA结构示意图。(a)DSEMA模块；(b)深度可分离卷积

Fig. 3　DEMA overview. (a) DSEMA module; (b) Depthwise-separable-convolution

作用，生成形状为$C'\times H\times W$的注意力权重。与此同时，DEMA模块还设置了另一个并行分支，如图3(b)所示，该分支采用深度可分离卷积对分组特征进行处理，以轻量化的方式捕获局部空间特征并完成通道间的信息融合，最终得到形状为$C'\times H\times W$的特征图。

对于DEMA模块两个并行分支的输出，首先，它们各自进行全局自适应池化处理，得到相应的全局特征，经Softmax函数作用后被分别赋予权重；然后，使用批量矩阵乘法实现跨分支信息的融合；最后，将原始的分组特征与处理后得到的全局注意力权重进行逐元素相乘，并让G组相乘后的结果沿通道维度拼接，最终得到形状为$C\times H\times W$的输出特征图。

因此，DEMA注意力机制可在保持多尺度注意力机制捕捉不同尺度特征信息这一优势的同时，引入深度可分离卷积和批量矩阵乘法，通过两者的结构互补与计算优化，大幅降低计算复杂度，提升特征融合的稳定性与效率，最终实现模型整体性能的提升。

2　数据集构建与实验设计

使用模拟数据集和公开的真实数据集，对构建的ADE-Net进行训练与验证，即借助模拟数据集检验网络的优势和整体性能，借助真实数据测试网络在实际相位展开任务中的实用性、鲁棒性和泛化能力。

2.1　数据集

使用的模拟数据集包含由随机形状合成的绝对相位图像及其对应的包裹相位图，其中，采用多组不同参数的高斯函数的组合运算生成随机形状，即对多个具有不同形状和空间坐标的高斯函数做叠加处理，得到无固定模式的连续不规则形状；为进一步丰富相位模式，模拟数据集还融入了由随机斜率和偏移量共同决定的斜坡相位。合成的绝对相位$\phi(x,y)$和相应的包裹相位$\varphi(x,y)$如式(5)所示，式中，(x,y)为像素空间坐标，j为虚数单位。

$$\varphi(x,y)=\angle\exp\left(\mathrm{j}\phi(x,y)\right) \tag{5}$$

为了模拟实际场景中相位图像存在含噪声、不连续区域以及混叠区域等情况，基于上述方法分别生成三组数据集，每个数据集均包含2 000张尺寸为256×256的包裹相位图以及对应的绝对相位图，且相位值在−44至44之间随机设定，按照8∶2的比例将数据集划分为训练集(1 600组场景数据)和测试集(400组场

景数据)。具体而言,第一组数据集模拟实际场景下包裹相位图像存在噪声干扰的情况,即在对合成相位图像进行包裹处理前,随机添加信噪比为0、5 dB、20 dB、40 dB和80 dB的高斯噪声;第二组数据集模拟实际场景中包裹相位图像存在不连续区域的情况,即在对合成相位图像进行包裹处理前,在无噪声的条件下随机生成1至2个位置、大小和相位值都随机变化的方块,进而得到含遮挡的包裹相位图像;第三组数据集模拟前两种情况同时出现的场景,即生成既有噪声又有遮挡的包裹相位图像。

在真实数据选取方面,使用公开数据集Single-input dual-output 3D shape reconstruction[22]。该数据集基于实际搭建的条纹投影轮廓术结构光三维测量系统构建,以数十种不同的小型石膏雕塑和实验器具为核心样本,采用随机旋转角度调整、放置位置变化及单物体/多物体组合排列三种维度的操作,并投射多帧具有固定相移的余弦条纹图案,然后借助高精度相机捕捉物体表面形变导致的变形条纹图像,并利用相移法对该变形条纹图像进行处理,进而精确解算出包裹相位数据,最终形成1 500组场景数据,每组数据均含复杂三维形状的精细化表征及对应的多通道彩色目标图像。在该数据集划分时,按照8:1:1的比例将数据分为训练集、验证集和测试集。模拟数据集和真实数据集的样式如图4所示。

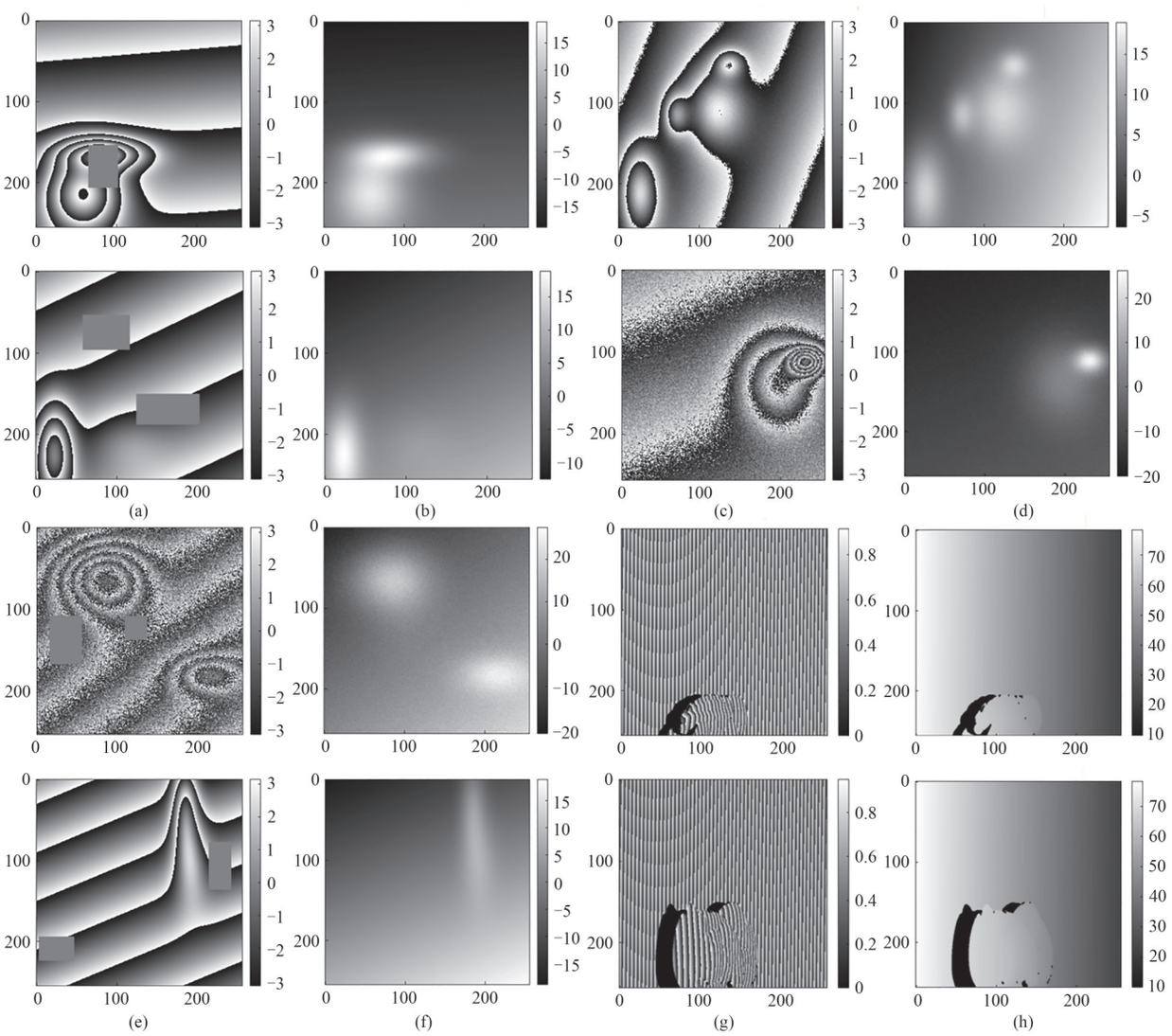


图4 模拟数据集和真实数据集的样式。(a)包裹相位图(模拟数据集1);(b)绝对相位图(模拟数据集1);(c)包裹相位图(模拟数据集2);(d)绝对相位图(模拟数据集2);(e)包裹相位图(模拟数据集3);(f)绝对相位图(模拟数据集3);(g)包裹相位图(真实数据集);(h)绝对相位图(真实数据集)

Fig. 4 The simulated datasets and real datasets. (a) Wrapped phase (simulated data set 1); (b) Absolute phase (simulated data set 1); (c) Wrapped phase (simulated data set 2); (d) Absolute phase (simulated data set 2); (e) Wrapped phase (simulated data set 3); (f) Absolute phase (simulated data set 3); (g) Wrapped phase (real data set); (h) Absolute phase (real data set)

2.2 实验环境及参数

本实验在Windows 10操作系统下，借助Pycharm平台进行网络训练，深度学习框架为Pytorch 1.10.0，编程语言为Python 3.8。在模型训练中，将模型迭代次数设定为200轮，批量大小设为4。为平衡训练的收敛速度和稳定性，选用学习率为0.001的Adam优化器，同时采用均方误差损失（MSELoss）作为训练神经网络的损失函数，其计算公式为

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 \tag{6}$$

式中，n为样本数量，y_i为真实结果，$\hat{y}_i$为网络模型的预测结果。

在相位展开回归任务中，MSELoss通过计算模型预测值与真实值差值的平方均值以量化整体偏差程度，该损失函数是连续可微的凸函数且其梯度计算简洁，能有效避免局部最优陷阱，并能直观反映模型的收敛状态与拟合效果。网络训练时，MSELoss与Adam等梯度下降类优化器能够高效配合，Adam优化器能结合MSELoss提供的梯度方向快速更新模型参数，加速模型的收敛过程。

2.3 模型性能评价指标

为全面客观地评估相位展开模型ADE-Net的性能，选取归一化均方根误差（Normalized Root Mean Square Error，NRMSE）、平均绝对误差（Mean Absolute Error，MAE）和结构相似性指数（Structural Similarity Index Measure，SSIM）三种评价指标，分别从整体偏差、平均误差和结构相似性三个层面对模型预测结果进行度量，进而反映模型在不同场景下的相位展开效果，为后续实验结果的对比分析提供量化依据。

NRMSE用于衡量预测相位与真实相位之间的整体偏差程度，通过引入真实相位的数值范围对均方根误差做归一化处理，能够更客观地反映模型的相对误差情况，其数值以百分比表示，数值越小，表明模型预测的整体精度越高。公式为

$$\mathrm{NRMSE}=\frac{1}{N}\left(\frac{\sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(y_{i,j}-\hat{y}_{i,j}\right)^2}}{r_i}\right) \tag{7}$$

式中，N为测试样本数量，M为单样本像素的总数，$y_{i,j}$为第i个样本的第j个像素的真实相位值，$\hat{y}_{i,j}$为网络对应的预测值，$r_i=\max\left(y_i\right)-\min\left(y_i\right)$为第$i$个样本的真实相位数值范围。

MAE用于计算预测值与真实值之间绝对差值的平均值，能直观反映预测误差的平均幅度。由于该指标对异常值不敏感，因而可作为衡量模型平均性的一个稳健可靠的度量标准，其值越小，说明模型在像素级别的预测准确性越高。公式为

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{8}$$

式中，n为有效像素点的个数，y_i为真实相位值，$\hat{y}_i$为对应的网络预测值。

SSIM用于评估预测相位与真实相位在结构层面的相似程度，该指标从亮度、对比度和结构三个维度对图像的相似性进行量化分析，其值越接近1，表明预测图像与真实图像的结构相似性越高、模型的预测性能越好，其表达式为

$$\mathrm{SSIM}\left(x,y\right)=\frac{\left(2\mu_x\mu_y+C_1\right)\left(2\sigma_{xy}+C_2\right)}{\left(\mu_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)} \tag{9}$$

式中，x、y为预测和真实相位图像；μ_x、μ_y为预测图像x和预测图像y的均值；σ_x、σ_y为图像x和y的标准差，σ_{xy}为协方差；C_1、C_2为常数，防止分母为零。

3 实验结果与分析

3.1 消融实验

为验证本文提出的ADE-Net中的AC-LSTM模块和DEMA模块对相位展开精度的提升效果，设计六个模型进行实验，进而评估两个模块的性能。六个模型依次为：1）未添加任何模块的原始编码器-解码器模型，记作模型A；2）在模型A的基础上，加入AC-LSTM模块，构建模型B；3）将EMA模块嵌入模型A，得到

模型C；4）把DEMA模块加入模型A，形成模型D；5）在模型A的基础上仅加入基础四向LSTM模块，记作模型E；6）在模型A上同时加入AC-LSTM和DEMA模块，得到模型ADE-Net。

六组实验均在包含随机噪声的数据集上开展，迭代次数统一设定为200次。消融实验所得的定量结果指标如表2所示，可以看出，未加入任何模块的模型A在NRMSE、MAE和SSIM三项指标上表现欠佳，表明该模型预测的解包裹相位与真实值之间存在较大偏差。当单独引入AC-LSTM模块后，NRMSE下降到5.15%，MAE降至1.329 6，同时SSIM也有明显提升，说明AC-LSTM模块对模型特征表达能力的增强作用，并且仅加入AC-LSTM模块的模型B在各项性能指标上均优于仅加入基础四向LSTM模块的模型E，更加证明了AC-LSTM模块对精度提升的有效性。对于仅加入DEMA模块的模型D，较模型A而言有一定程度的性能改善，且其性能优于仅加入EMA模块的模型C，这表明改进后的注意力机制DEMA能够更有效地学习数据中的关键信息，并对噪声有更强的抑制能力。而同时引入AC-LSTM模块与DEMA模块的模型ADE-Net取得了最优的综合性能，其NRMSE、MAE和SSIM分别达到4.58%、1.216 5 rad和0.838 9，有效证明了通过同时集成这两个模块，模型的预测精度得到了显著提升。

表2 消融实验结果

Table 2 Ablation experiment results

Model	Baseline	Four-way space LSTM	AC-LSTM	EMA	DEMA	NRMSE	MAE/rad	SSIM
A	√					13.69%	3.677 1	0.503 2
B	√		√			5.15%	1.329 6	0.793 3
C	√			√		9.33%	2.355 4	0.643 6
D	√				√	7.49%	1.972 8	0.718 8
E	√	√				6.09%	1.618 0	0.783 6
ADE-Net	**√**		**√**		**√**	**4.58%**	**1.216 5**	**0.838 9**

3.2 对比实验

为评估本文的ADE-Net网络模型的综合性能，将该模型与U-Net[23]、Res-UNet[24]、TNUNet[13]及Perera[10]等提出的网络模型进行对比实验，所有实验均在配备RTX 4090显卡的计算环境下进行。在数据集的选择上，采用2.1节提到的存在不连续区域、含有随机噪声和存在混叠的三组模拟数据集，对模型进行训练和测试，全面检验各模型在不同数据特性下的性能表现。其次，在评价各模型条纹边界区域的相位展开精度时，借助相对误差图进行量化显示，且相对误差图采用百分比的表达形式以有效放大局部误差的视觉对比度，从而更为直观地展现误差在空间维度上的分布趋势，精准定位各模型在相位跳变、噪声干扰等关键区域的误差集中情况。

3.2.1 基于不连续数据集的对比实验

图5展示了不同模型在不连续数据集上的相位展开结果，可以看出U-Net和Res-UNet生成的相位展开结果和实际相位之间存在较明显的偏差，这种偏差不仅体现在整体相位值的差异上，还反映在对于数据集中不连续特征的捕捉能力方面。而Perera′s Net与TNUNet虽然在相位展开结果上更接近真实相位，但在处理图像数据中的细微结构时，这两个模型未能精准还原真实相位的特征，导致部分细节信息丢失或扭曲。对于第一组图像数据，U-Net、Res-UNet、Perera′s Net、TNUNet和ADE-Net模型的绝对误差均值分别是2.1 rad、2.932 rad、1.567 rad、1.253 rad和0.565 rad，ADE-Net的绝对误差均值占原始相位动态范围的2.24%，绝对误差最大值分别是7.636 rad、11.742 rad、6.129 rad、4.760 rad和3.242 rad；对于第二组图像数据，上述网络的绝对误差均值分别为1.873 rad、1.109 rad、1.251 rad、0.855 rad和0.679 rad，ADE-Net的绝对误差均值占原始相位动态范围的2.70%，绝对误差最大值分别是7.093 rad、12.129 rad、3.293 rad、7.433 rad和3.593 rad。

表3通过NRMSE、MAE和SSIM三个指标定量对比分析了各模型的相位展开性能，结果显示，ADE-Net在三项指标上均有明显优势：NRMSE低至4.35%，表明该模型对于不连续区域造成的相位跳变具有很好的纠正能力，整体相位恢复最为准确；MAE为1.094 5 rad，证明了其在像素级别的预测误差最小，能更好地保留相位的细节信息；SSIM最接近1，体现了其输出相位图在整体结构和轮廓边缘方面与真实相位分布情况的一致性。相较于其它网络模型，ADE-Net在不连续场景下具有更高精度的相位展开能力，其预测结果与真实相位最为一致，对于不连续特征和细节信息的捕捉均具有明显优势。

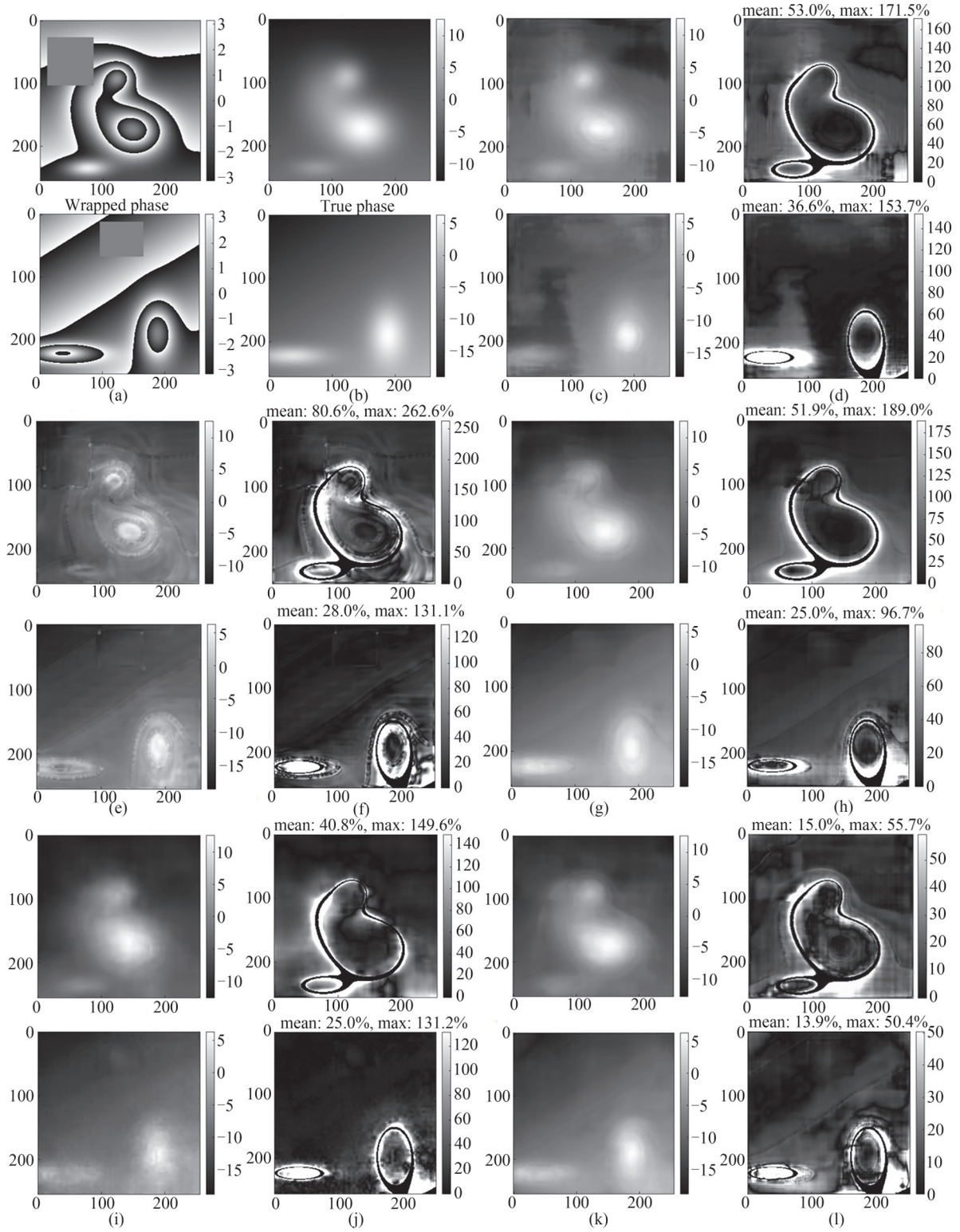


图5 不连续数据集各模型相位展开结果。(a)包裹相位图;(b)绝对相位图;(c)U-Net预测的绝对相位图;(d)U-Net预测结果的相对误差图;(e)Res-UNet预测的绝对相位图;(f)Res-UNet预测结果的相对误差图;(g)Perera′s Net预测的绝对相位图;(h)Perera′s Net预测结果的相对误差图;(i)TNUNet预测的绝对相位图;(j)TNUNet预测结果的相对误差图;(k)ADE-Net预测的绝对相位图;(l)ADE-Net预测结果的相对误差图

Fig. 5 Phase unwrapping results of various models on discontinuous datasets. (a) Wrapped phase; (b) Absolute phase; (c) Absolute phase predicted by U-Net; (d) Relative error map of U-Net; (e) Absolute phase predicted by Res-UNet; (f) Relative error map of Res-UNet; (g) Absolute phase predicted by Perera′s Net; (h) Relative error map of Perera′s Net; (i) Absolute phase predicted by TNUNet; (j) Relative error map of TNUNet; (k) Absolute phase predicted by ADE-Net; (l) Relative error map of ADE-Net

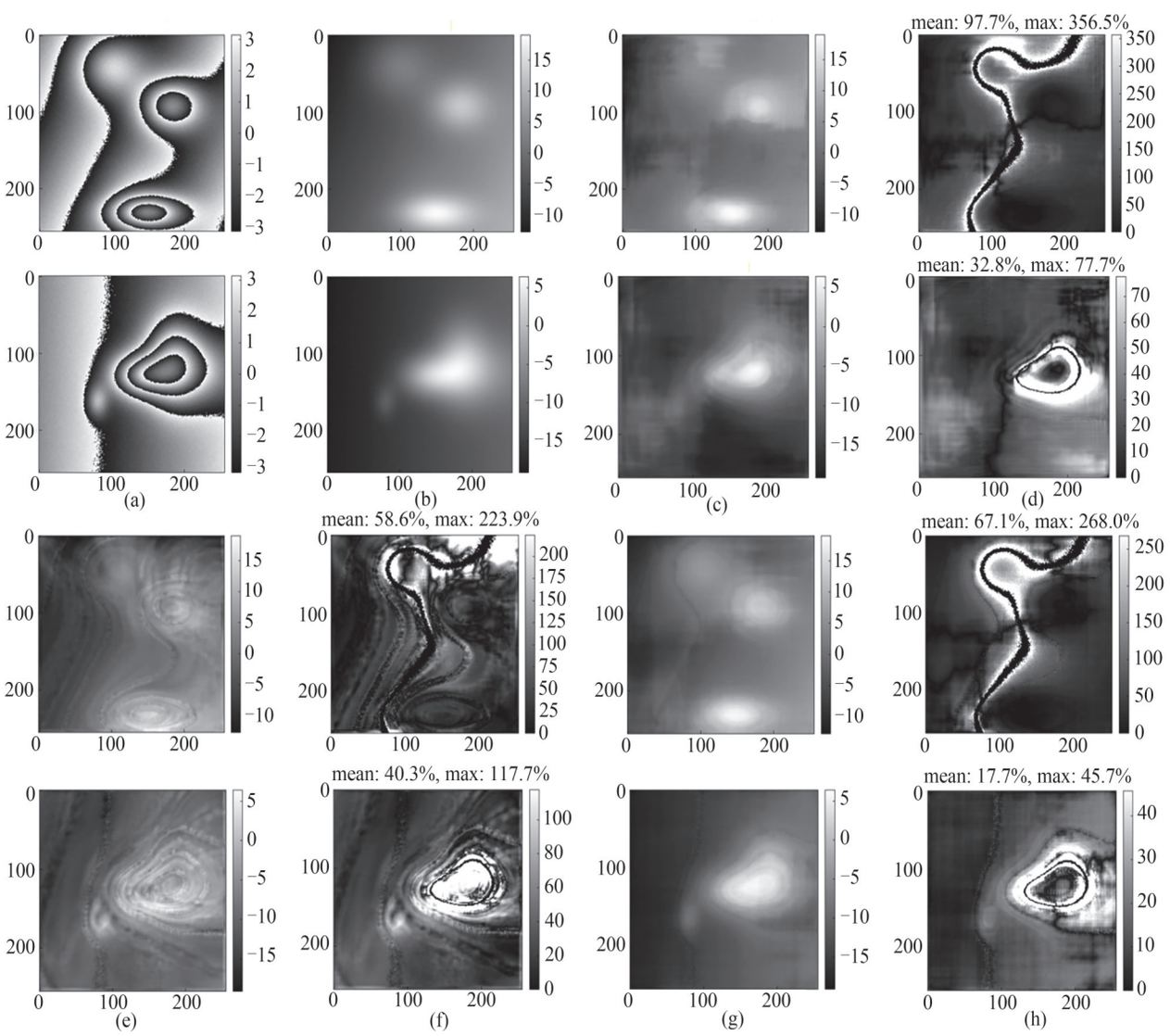
表3 不连续数据集上的定量评估结果

Table 3 Quantitative evaluation results on discontinuous datasets

Method	NRMSE	MAE/rad	SSIM
U-Net	10.53%	2.519 9	0.667 2
Res-UNet	9.10%	2.212 3	0.673 2
Perera's Net	5.60%	1.342 1	0.824 8
TNUNet	5.28%	1.265 3	0.831 0
ADE-Net	**4.35%**	**1.094 5**	**0.871 4**

3.2.2 基于随机噪声数据集的对比实验

图6呈现了各模型在随机噪声数据集上的相位展开结果，如图所示，含噪声的包裹相位经U-Net和Res-UNet处理后，其展开结果与真实相位相比存在明显偏差，且噪声干扰突出；Perera's Net与TNUNet虽在性能上有所提升，但展开结果中仍有残留噪声；同时，U-Net、Res-UNet、Perera's Net、TNUNet和ADE-Net在第一组图像数据上的绝对误差均值分别为2.847 rad、1.941 rad、1.907 rad、0.648 rad和0.843 rad，ADE-Net的绝对误差均值占原始相位动态范围的2.65%，绝对误差最大值分别是9.402 rad、13.606 rad、5.216 rad、4.564 rad和4.326 rad；在第二组图像数据上的绝对误差均值分别是3.005 rad、3.862 rad、1.458 rad、1.242 rad和0.861 rad，ADE-Net的绝对误差均值占原始相位动态范围的3.36%，绝对误差最大值分别是9.058 rad、14.526 rad、4.600 rad、5.748 rad和3.282 rad。



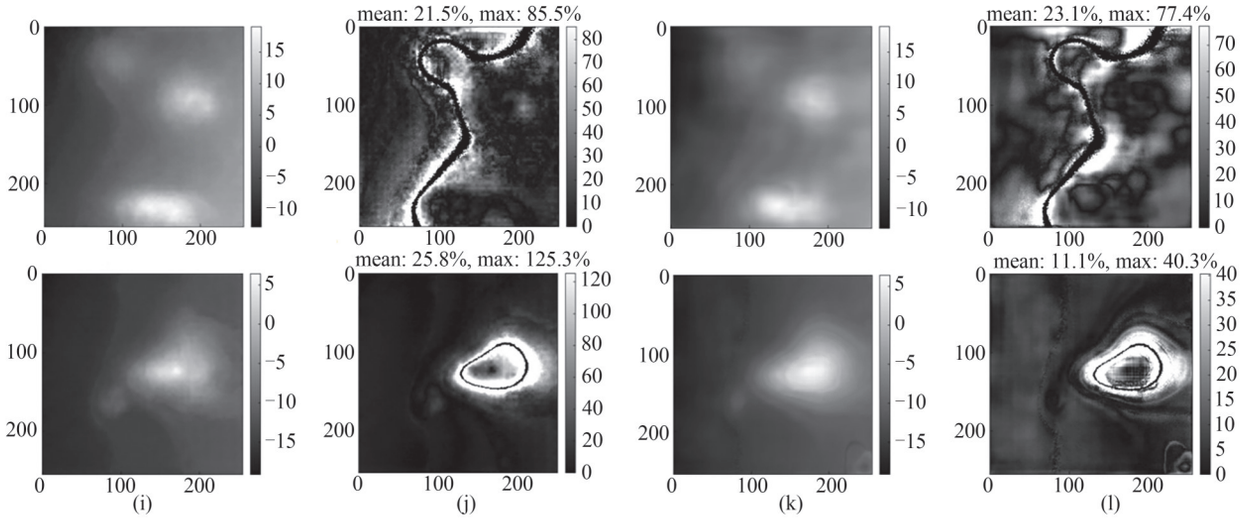


图6　随机噪声数据集各模型相位展开结果。(a)包裹相位图；(b)绝对相位图；(c)U-Net预测的绝对相位图；(d)U-Net预测结果的相对误差图；(e)Res-UNet预测的绝对相位图；(f)Res-UNet预测结果的相对误差图；(g)Perera′s Net预测的绝对相位图；(h)Perera′s Net预测结果的相对误差图；(i)TNUNet预测的绝对相位图；(j)TNUNet预测结果的相对误差图；(k)ADE-Net预测的绝对相位图；(l)ADE-Net预测结果的相对误差图

Fig. 6　Phase unwrapping results of various models on random noise datasets. (a) Wrapped phase; (b) Absolute phase; (c) Absolute phase predicted by U-Net; (d) Relative error map of U-Net; (e) Absolute phase predicted by Res-UNet; (f) Relative error map of Res-UNet; (g) Absolute phase predicted by Perera′s Net; (h) Relative error map of Perera′s Net; (i) Absolute phase predicted by TNUNet; (j) Relative error map of TNUNet; (k) Absolute phase predicted by ADE-Net; (l) Relative error map of ADE-Net

借助NRMSE、MAE和SSIM三个指标对多模型在噪声场景下的相位展开性能进行定量对比，如表4所示，ADE-Net的NRMSE仅为4.58%、MAE为1.216 5 rad，两项误差指标均显著低于其它对比模型；同时，结构相似度也最高，SSIM指标达到0.838 9。上述量化结果表明ADE-Net在复杂噪声环境下的鲁棒性优势，噪声抑制能力强，相位还原精度高。

表4　随机噪声数据集上的定量评估结果

Table 4　Quantitative evaluation results on random noise dataset

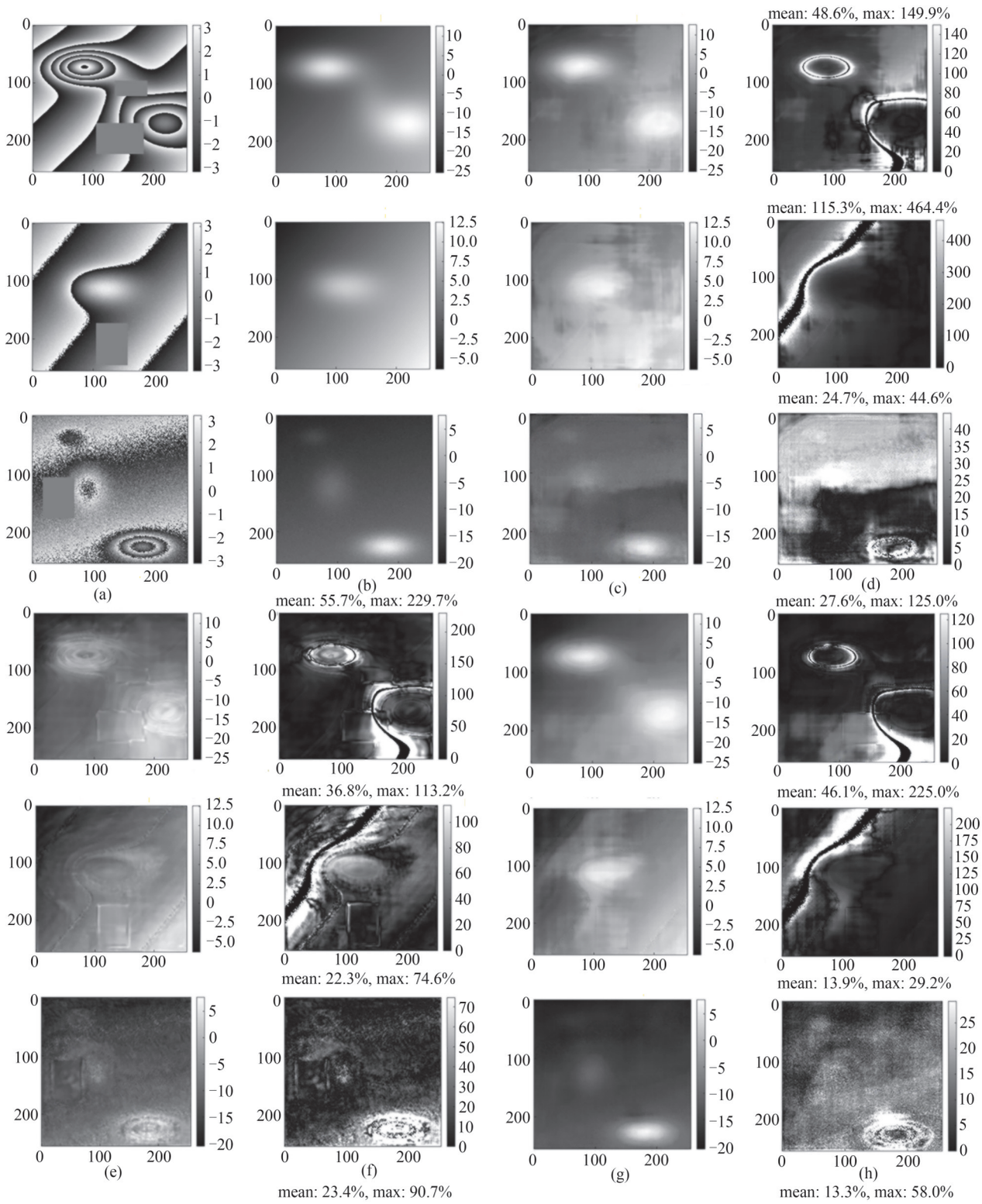
Method	NRMSE	MAE/rad	SSIM
U-Net	10.11%	2.620 4	0.666 9
Res-UNet	11.41%	3.034 0	0.548 4
Perera′s Net	5.63%	1.465 8	0.800 6
TNUNet	5.79%	1.520 8	0.768 5
ADE-Net	**4.58%**	**1.216 5**	**0.838 9**

3.2.3　基于混叠数据集的对比实验

图7展示了各模型在含有随机噪声且存在不连续区域的复杂数据集上的相位展开结果，实验结果表明U-Net和Res-UNet的展开结果与真实相位存在显著偏差，尤其在噪声干扰强的区域更为明显；Perera′s Net与TNUNet虽然性能有所改进，但在相位突变等复杂混叠区域仍存在相位还原差距；同时，U-Net、Res-UNet、Perera′s Net、TNUNet和ADE-Net网络在第一组图像数据上的绝对误差均值分别是3.149 rad、2.385 rad、1.276 rad、1.118 rad和0.588 rad，ADE-Net的绝对误差均值占原始相位动态范围的1.55%，绝对误差最大值分别是9.647 rad、18.295 rad、5.800 rad、5.782 rad和4.747 rad；在第二组图像数据上的绝对误差均值分别是3.021 rad、1.338 rad、1.225 rad、0.584 rad和0.528 rad，ADE-Net的绝对误差均值占原始相位动态范围的2.80%，绝对误差最大值分别是6.612 rad、7.248 rad、4.770 rad、3.254 rad和3.177 rad；在第三组图像数据上的绝对误差均值分别是3.236 rad、1.949 rad、1.448 rad、0.986 rad和0.583 rad，ADE-Net的绝对误差均值占原始相位动态范围的2.06%，绝对误差最大值分别

是 8.649 rad、11.938 rad、6.104 rad、6.917 rad 和 4.091 rad。相比之下，ADE-Net 的预测结果与真实相位之间的匹配度最优。

表 5 通过 NRMSE、MAE 和 SSIM 三个指标，对各模型在复杂噪声场景下的相位展开综合性能进行了评估，结果表明 ADE-Net 模型表现最佳，NRMSE 低至 4.34%，MAE 仅为 1.137 7 rad，并且 SSIM 达到最高，结构相似度最好，充分证明了本文的 ADE-Net 在更复杂场景下的综合性能优势，其在相位展开任务中具有更优的噪声鲁棒性、相位还原精度及细节保持能力。



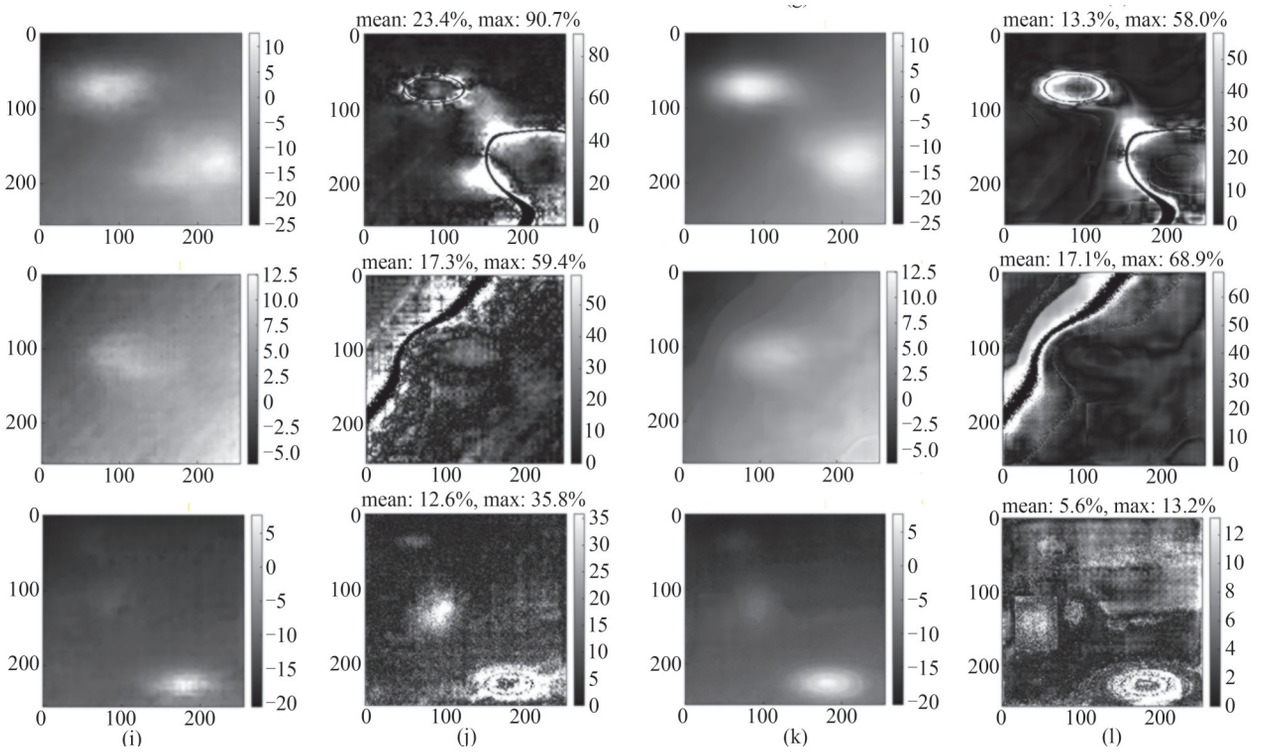


图7　混叠数据集各模型相位展开结果。(a)包裹相位图；(b)绝对相位图；(c)U-Net预测的绝对相位图；(d)U-Net预测结果的相对误差图；(e)Res-UNet预测的绝对相位图；(f)Res-UNet预测结果的相对误差图；(g)Perera′s Net预测的绝对相位图；(h)Perera′s Net预测结果的相对误差图；(i)TNUNet预测的绝对相位图；(j)TNUNet预测结果的相对误差图；(k)ADE-Net预测的绝对相位图；(l)ADE-Net预测结果的相对误差图

Fig. 7　Phase unwrapping results of various models on the aliased dataset. (a) Wrapped phase; (b) Absolute phase; (c) Absolute phase predicted by U-Net; (d) Relative error map of U-Net; (e) Absolute phase predicted by Res-UNet; (f) Relative error map of Res-UNet; (g) Absolute phase predicted by Perera′s Net; (h) Relative error map of Perera′s Net; (i) Absolute phase predicted by TNUNet; (j) Relative error map of TNUNet; (k) Absolute phase predicted by ADE-Net; (l) Relative error map of ADE-Net

表5　混叠数据集上的定量评估结果

Table 5　Quantitative evaluation results on aliased datasets

Method	NRMSE	MAE/rad	SSIM
U-Net	11.71%	3.121 3	0.616 0
Res-UNet	9.51%	2.489 8	0.619 3
Perera′s Net	5.75%	1.515 7	0.798 3
TNUNet	5.53%	1.417 4	0.765 4
ADE-Net	**4.34%**	**1.137 7**	**0.834 0**

3.2.4　参数量与计算效率分析

为评估ADE-Net的计算效率，将其与U-Net、Res-UNet、Perera′s Net和TNUNet四种主流深度学习模型在模型复杂度与运行效率方面进行对比，结果如表6所示。在模型复杂度方面，ADE-Net的参数量仅为2.012 M，相较于U-Net参数量(17.266 M)、Res-UNet参数量(15.940 M)和TNUNet参数量(8.434 M)分别降低了88.35%、87.38%、76.14%；ADE-Net的GFLOPs为2.655，较U-Net(40.111)与Res-UNet(87.014)而言，大幅降低了计算负担，有效实现了轻量化设计。

在训练阶段，ADE-Net平均每轮训练时间为16.97 s，训练显存占用451.21 MB，远低于U-Net、Res-UNet及TNUNet，体现了更低的硬件资源门槛与更优的训练友好性。在预测阶段，经典U-Net虽因结构简洁具有最快的推理速度，但其预测精度无法满足相位展开任务对高精度的严苛要求，而ADE-Net对单张图像的推理时间仅需12.71 ms，能够较好满足实时相位处理场景的应用需求；较TNUNet(单张推理24.74 ms)而言，ADE-Net的推理速度实现了近一倍的提升，高效性优势显著；尽管与轻量级Perera′s Net相比，ADE-Net在参数量和计算效率上略有增加，但其在模型复杂度和计算效率之间达到了更具工程实用价

表6 模型复杂度与运行效率对比

Table 6 Comparison of model complexity and runtime efficiency

Method	Parameters/M	GFLOPs	Average training time per epoch/s	Training GPU memory usage/MB	FPS	Inference time per image/ms
U-Net	17.266	40.111	14.49	1 650.69	322.30	3.10
Res-UNet	15.940	87.014	35.96	3 251.53	101.15	9.89
Perera's Net	0.760	2.279	13.13	304.82	97.13	10.30
TNUNet	8.434	8.270 9	34.07	2 272.63	40.41	24.74
ADE-Net	**2.012**	**2.655**	**16.97**	**451.21**	**78.67**	**12.71**

值的动态平衡，既可保障相位展开任务中高精度相位重建的需求，又能满足实时处理场景下的效率要求。

3.3 测试实验

上述仿真数据集对比实验已验证了所提的ADE-Net在各类复杂困难场景下的相位展开性能优势，为进一步检验该方法在真实数据场景中的泛化能力，在2.1节提及的公开真实数据集Single-input dual-output 3D shape reconstruction上开展训练与测试。如图8所示，真实场景下的相位展开结果包括包裹相位图(图8(a))、实际的解包裹相位图(图8(b))、ADE-Net预测的解包裹相位图(图8(c))和展开相位误差图(图8(d))。从视觉对比可观察到，预测相位图与实际相位图在相位分布趋势、细节纹理特征及噪声抑制效果等方面均展现出高度的一致性，尤其在相位突变区域仍能保持精准的相位还原能力。

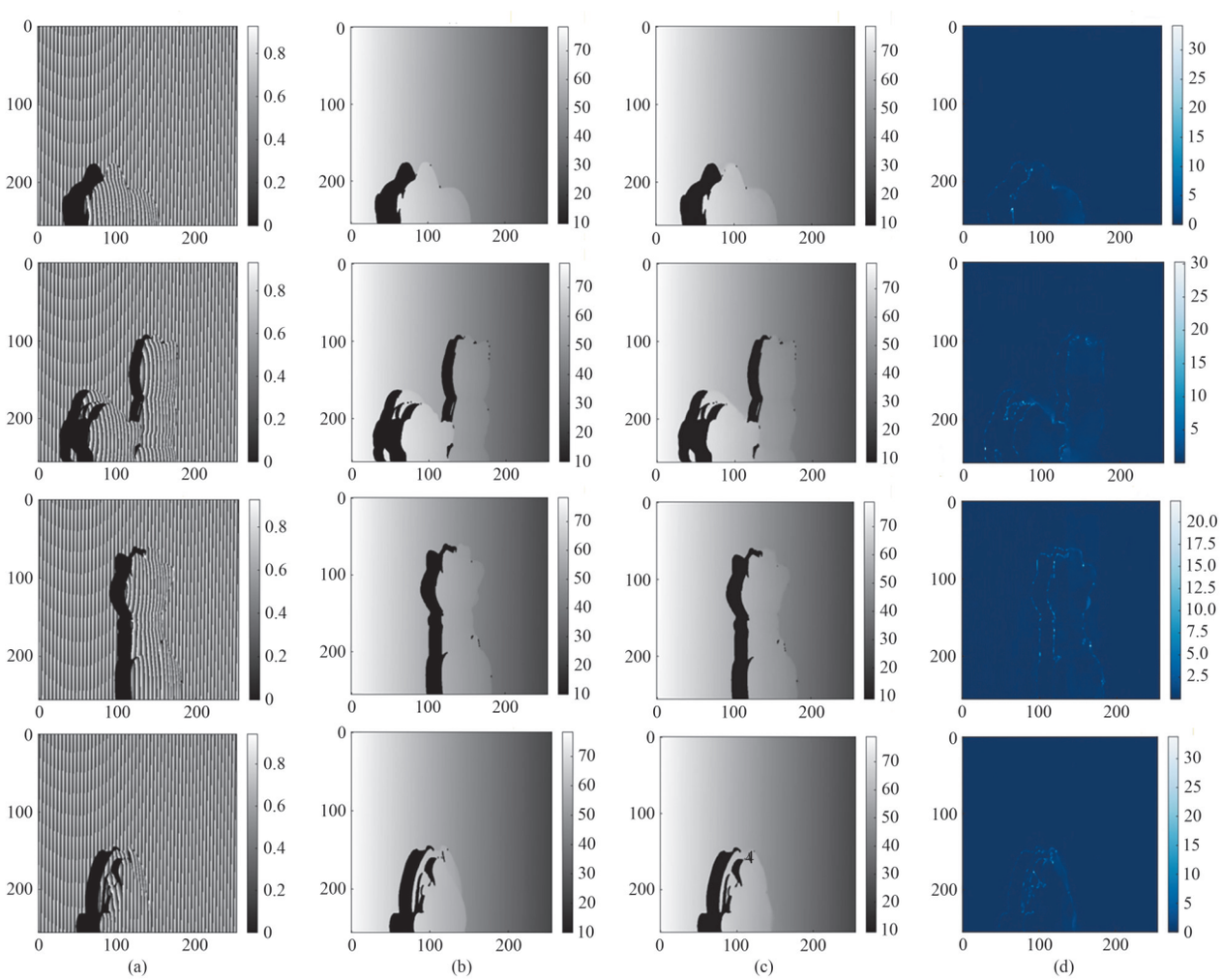


图8 公开真实数据集下ADE-Net模型的相位展开结果。(a)包裹相位图；(b)绝对相位图；(c)ADE-Net预测的绝对相位图；(d)展开相位误差图

Fig. 8 Phase unwrapping results of the ADE-Net on the public real dataset. (a) Wrapped phase; (b) Absolute phase; (c) Absolute phase predicted by ADE-Net; (d) Phase unwrapping error map

利用全部真实数据集进行定量测试，结果表明本文设计的ADE-Net同样具备优异的相位展开性能，NRMSE低至0.74%，表明网络的相位还原误差极小；RMSE为0.505 2且MAE为0.180 9 rad，这两项绝对误差指标均达到了亚像素精度；同时，结构相似性SSIM高达0.997 1，证明ADE-Net在真实场景下的实际相位展开任务中具有很好的结构保持能力，可实现精度与鲁棒性的双重突破。

4 结论

为提升网络模型在随机噪声干扰、相位不连续跳变及混叠等多类复杂场景下的相位展开精度与鲁棒性，本文提出一种轻量级且高效的相位展开深度学习网络模型ADE-Net，设计并引入AC-LSTM模块与DEMA注意力机制。通过多场景的对比实验验证，ADE-Net在NRMSE、MAE与SSIM等关键量化指标上均优于U-Net、Res-UNet、Perera′s Net和TNUNet四种主流深度学习模型，在真实场景下能够稳定高效地实现高精度的相位展开，其噪声抑制能力、相位连续性、保持精度及细节还原方面均优势明显，并具有较强的泛化能力，能以更轻量化的网络架构实现精度与计算效率的优化和平衡，可为工程实践中的高精度相位展开任务提供可靠的技术手段。

参考文献

[1] LI Yifeng, ZHOU Shun, LI Pangyue, et al. Three-dimensional reconstruction method integrating active and passive polarization based on structured light [J]. Acta Photonica Sinica, 2024, 53(12): 1210002.
李宜峰, 周顺, 李庞跃, 等. 基于结构光的偏振主被动融合三维重建方法[J]. 光子学报, 2024, 53(12): 1210002.

[2] HOU Lei, XI Dejun, LUO Jun, et al. Deep learning-based correction of defocused fringe patterns for high-speed 3D measurement [J]. Advanced Engineering Informatics, 2023, 58: 102221.

[3] LI Leyang, WU Zhoujie, ZHANG Qican. Phase error compensation technique based on phase-shifting fringe analysis: a review [J]. Laser & Optoelectronics Progress, 2024, 61(2): 144-163.
李乐阳, 吴周杰, 张启灿. 基于相移条纹分析的相位误差补偿技术发展综述[J]. 激光与光电子学进展, 2024, 61(2): 144-163.

[4] ZHANG Zonghua, LI Yanling, GAO Feng, et al. Phase unwrapping technology for structured light three-dimensional measurement: a review [J]. Infrared and Laser Engineering, 2023, 52(8): 23-45.
张宗华, 李雁玲, 高峰, 等. 面向结构光三维测量的相位展开技术综述[J]. 红外与激光工程, 2023, 52(8): 23-45.

[5] SPOORTHI G E, GORTHI S, GORTHI R K S. A deep learning-based model for phase unwrapping [C]. Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing, 2018: 1-8.

[6] WANG Kaiqiang, LI ying, QIAN Kemao, et al. One-step robust deep learning phase unwrapping [J]. Optics Express, 2019, 27(10): 15100-15115.

[7] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[8] ZHANG Teng, JIANG Shaowei, ZHAO Zixin, et al. Rapid and robust two-dimensional phase unwrapping via deep learning [J]. Optics Express, 2019, 27(16): 23173-23185.

[9] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. Proceedings of the European Conference on Computer Vision (ECCV), Springer, 2018: 801-818.

[10] PERERA M V, DE SILVA A. A joint convolutional and spatial quad-directional LSTM network for phase unwrapping [C]. ICASSP 2021IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021: 4055-4059.

[11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[12] HUANG Wangwang, MEI Xuesong, WANG Yage, et al. Two-dimensional phase unwrapping by a high-resolution deep learning network [J]. Measurement, 2022, 200: 111566.

[13] ZHANG Ziheng, WANG Xiaoxu, LIU Chengxiu, et al. Efficient and robust phase unwrapping method based on SFNet [J]. Optics Express, 2024, 32(9): 15410-15432.

[14] LI Kejia, ZHAO Zixin, ZHAO Hong, et al. Three-stage training strategy phase unwrapping method for high speckle noises [J]. Optics Express, 2024, 32(27): 48895-48914.

[15] LI Youxing, MENG Lingzhi, ZHANG Kai, et al. PUDCN: two-dimensional phase unwrapping with a deformable convolutional network [J]. Optics Express, 2024, 32(16): 27206-27220.

[16] WAN Liang, LIU Jiayang, QIAN Yi′nan, et al. A two-stage phase unwrapping method by deep learning on noisy wrapped phases [J]. Optics and Laser Technology, 2025, 192(PD): 113787.

[17] LIU Zhe, WANG Linlin, LI Xiaodan, et al. A fast and efficient phase unwrapping method based on deep learning [J]. Laser & Optoelectronics Progress, 2025, 62(20): 2039001.

刘哲，王琳霖，李晓丹，等. 一种基于深度学习的快速且高效的相位展开方法[J]. 激光与光电子学进展，2025，62(20)：2039001.

[18] OUYANG Daliang, HE Su, ZHANG Guozhong, et al. Efficient multi-scale attention module with cross-spatial learning [C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023: 1-5.

[19] SHANG Jianhua, WANG Gang, LIU Yang, et al. Multi-scale phase extraction network for structured light based on discrete wavelet transform and attention mechanism [J]. Infrared and Laser Engineering, 2025, 54(6): 334-346.
尚建华，王刚，刘洋，等. 基于离散小波注意力机制的结构光多尺度相位提取网络[J]. 红外与激光工程，2025，54(6)：334-346.

[20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[21] CAI Xinhao, LAI Qiuxia, WANG Yuwei, et al. Poly Kernel inception network for remote sensing detection [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 27706-27716.

[22] NGUYEN A H, REES O, WANG Z Y. Learning-based 3D imaging from single structured-light image [J]. Graphical Models, 2023, 126: 101171.

[23] ZHANG Junchao, TIAN Xiaobo, SHAO Jianbo, et al. Phase unwrapping in optical metrology via denoised and convolutional segmentation networks [J]. Optics Express, 2019, 27(10): 14903-14912.

[24] ZHANG Zhengxin, LIU Qingjie, WANG Yunhong. Road extraction by deep residual U-net [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.

Research on Structured Light Wrapped Phase Unwrapping Algorithm Based on Deep Learning

YANG Simin[1], ZHANG Yandong[1], WANG Yang[2], SHANG Jianhua[1]

(1 *School of Information and Intelligent Science, Donghua University, Shanghai* 201620, *China*)

(2 *Joint Laboratory of High Power Laser and Physics, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai* 201800, *China*)

Abstract: Phase unwrapping serves as a critical step in high-precision optical measurement, aiming to recover the absolute phase from its wrapped counterpart. While current deep learning-based approaches for absolute phase extraction have achieved improvements in both accuracy and computational efficiency, critical challenges persist, including excessively large network parameter counts and low accuracy in resolving discontinuous phases. To overcome these limitations, this paper introduces the Attentive Directional Encoder-Decoder Network (ADE-Net) for structured light phase unwrapping applications. The primary objective is to develop a lightweight, efficient, and robust deep learning framework that achieves consistently high unwrapping accuracy under complex conditions such as noise corruption, phase discontinuities, and aliasing artefacts, thereby providing a practical, engineering-ready solution capable for reliable development in real-world optical measurement systems.

The proposed ADE-Net is built upon a concise and efficient encoder-decoder backbone architecture. The encoder consists of four feature extraction stages, each comprising a convolutional block and a Haar Wavelet Down-sampling (HWD) module. The HWD structure preserves high-frequency directional details, such as phase jump edges, during down-sampling, providing a more complete feature foundation for subsequent processing. The decoder employs a symmetric up-sampling structure based on transposed convolutions, with skip connections to fuse multi-scale features from the encoder, ensuring effective detail recovery. To enhance feature representation, two innovative modules are integrated: the Adaptive Clockwork Long Short-Term Memory (AC-LSTM) module and the Depthwise-separable-convolution Efficient Multi-Scale Attention (DEMA) module. The AC-LSTM module is designed to capture multi-scale contextual and direction-sensitive features simultaneously. It consists of two parallel branches: an Atrous Spatial Pyramid Pooling (ASPP) branch for multi-scale context aggregation, and a bidirectional LSTM (Bi-LSTM) branch enhanced with Context Anchor Attention (CAA) to model long-range dependencies along horizontal and vertical directions with spatial awareness. The outputs of both branches are fused to produce rich, direction-aware features. The DEMA module is an improved version of the Efficient Multi-Scale Attention (EMA) mechanism. It introduces depthwise separable convolution to replace standard 3×3 convolutions, significantly reducing computational complexity while maintaining

receptive field coverage. Additionally, Batch Matrix Multiply (BMM) is adopted for more efficient parallel processing of batch data. DEMA operates by grouping input features, performing adaptive height and width pooling, and applying depthwise separable convolution in parallel branches, followed by cross-branch fusion via BMM and element-wise multiplication with the original features. This design enhances the model's ability to focus on critical information across scales while optimizing computational efficiency.

To comprehensively evaluate the performance of ADE-Net, systematic experiments were conducted using both simulated and real-world datasets. For simulation, three distinct datasets were constructed, each comprising 2 000 pairs of wrapped and absolute phase images with a size of 256×256. Dataset 1 simulated noise interference by introducing Gaussian noise with signal-to-noise ratios of 0, 5, 20, 40, and 80 dB. Dataset 2 simulated phase discontinuities by randomly generating 1 to 2 rectangular jump regions with random positions, sizes, and phase values. Dataset 3 combined both noise and discontinuities to emulate aliasing effects. Additionally, the public real dataset 'Single-input dual-output 3D shape reconstruction' was adopted to assess generalization capability, which contains 1 500 real-scene samples acquired by a structured-light 3D measurement system. The evaluation was performed using three metrics: Normalized Root Mean Square Error (NRMSE), Mean Absolute Error (MAE), and Structural Similarity Index Measure (SSIM). Benchmark comparisons were carried out against mainstream models such as U-Net, Res-UNet, Perera's Net, and TNUNet. In the comparative experiments, ADE-Net achieved the best performance across all three simulated datasets. On the discontinuous dataset, it attained an NRMSE of 4.35%, an MAE of 1.094 5 rad, and an SSIM of 0.871 4. On the random-noise dataset, the results were 4.58% NRMSE, 1.216 5 rad MAE, and 0.838 9 SSIM. On the aliased dataset, ADE-Net reached 4.34% NRMSE, 1.137 7 rad MAE, and 0.834 0 SSIM. Cross-validation on the public dataset demonstrated excellent generalization ability, with a Root Mean Square Error (RMSE) of 0.505 2, an MAE of 0.180 9 rad, and an SSIM of 0.997 1. Ablation studies further confirmed the individual performance contributions of both the AC-LSTM and DEMA modules. Moreover, while maintaining the aforementioned high accuracy, ADE-Net also exhibited remarkable computational efficiency advantages. With only 2.012 M parameters and 2.655 GFLOPs, it is significantly lighter than the compared models. The inference time per single image is only 12.71 ms, meeting real-time processing requirements without compromising accuracy, thereby achieving an effective balance between model complexity and computational performance.

In conclusion, ADE-Net presents a lightweight, efficient, and highly accurate deep learning framework for phase unwrapping in structured light 3D measurement. By integrating the AC-LSTM module for capturing direction-sensitive multiscale features and the DEMA module for computationally streamlined attention-driven feature refinement, the models achieve an ideal balance between performance and computational complexity. It demonstrates superior resilience in challenging conditions involving noise corruption, phase discontinuities, and aliasing artefacts, outperforming leading benchmark models in both accuracy and processing efficiency. Rigorous validation on real-world datasets confirms its robust generalization capability, while its fast inference speed and low resource footprint make it well-suited for high-precision phase reconstruction tasks. By addressing practical constrains in deployment, this work promotes the feasibility of structured light 3D measurement system for real-world, resource-aware applications.

Key words: Optical measurement; Phase unwrapping; Deep learning; Attention mechanism; Long short-term memory network

OCIS Codes: 120.5050; 110.5086; 100.5070; 100.4996

CSTR: 32255.14.gzxb20265502.0212001

Foundation item: National Natural Science Foundation of China (52173219)