

基于多尺度注意力机制相位展开的 三维人脸建模

朱江平^{1,2}, 王睿珂¹, 段智涓^{1,2}, 黄怡洁¹, 何国欢¹, 周佩^{1,2*}

¹四川大学计算机学院, 四川 成都 610065;

²四川大学视觉合成图形图像技术国防重点学科实验室, 四川 成都 610065

摘要 相位展开作为三维(3D)测量技术中的关键环节,其解析精度直接影响 3D 建模的精度。由于存在欠采样和相位不连续等问题,故传统空间相位展开难以得到正确的相位信息,而时间相位展开又需要额外的信息辅助。针对复杂场景中的 3D 人脸建模,提出了基于多尺度注意力机制的相位展开网络。在所提网络中,利用编码器-解码器结构融合多尺度特征,并在解码网络中嵌入注意力子网络以获取上下文信息。构建一个包含 5000 组样本的 FACE 数据集和一个包含 100 组样本的 MASK 数据集,每组样本均包含截断相位和连续相位的真值,这些真值可用于相位展开的训练及测试。所提网络在 FACE 数据集和 MASK 数据集上的均方根误差分别为 0.0387 rad 和 0.0273 rad,结构相似性分别为 0.9850 和 0.9793。在欠采样、相位不连续等区域中,所提网络可快速准确地提取相位特征,进而保证了相位展开的正确性。最后,通过对比实验证实了所提网络的有效性和可行性。

关键词 测量; 三维人脸建模; 相位展开; 多尺度注意力机制融合; 上下文特征信息; 编码器-解码器结构

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/AOS202242.0112005

Three-Dimensional Face Modeling Based on Multi-Scale Attention Phase Unwrapping

Zhu Jiangping^{1,2}, Wang Ruike¹, Duan Zhijuan^{1,2}, Huang Yijie¹, He Guohuan¹, Zhou Pei^{1,2*}

¹College of Computer Science, Sichuan University, Chengdu, Sichuan 610065, China;

²National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu, Sichuan 610065, China

Abstract Phase unwrapping plays an important role in three-dimensional (3D) measurement technologies, and its analytical accuracy directly affects the accuracy of 3D modeling. Due to undersampling and discontinuity of the wrapped phase, it is difficult to obtain correct phase information for traditional spatial phase unwrapping, while temporal phase unwrapping requires additional auxiliary information. For 3D face modeling in complex scenarios, a phase unwrapping network based on multi-scale attention is proposed in this paper. In this network, the encoder-decoder structure is used to fuse multi-scale features, and an attention sub-network is embedded into the decoding network for contextual information collection. A FACE dataset of 5000 samples and a MASK dataset of 100 samples are constructed, and each sample contains the truth values of wrapped phases and continuous phases for training and testing of phase unwrapping. The root-mean-square errors of the proposed network are 0.0387 rad and 0.0273 rad on the FACE dataset and the MASK dataset. The structural similarities are 0.9850 and 0.9793 respectively. The phase features can be extracted quickly and accurately in areas such as undersampled and phase discontinuous ones to ensure the correctness of phase unwrapping. Finally, the effectiveness and feasibility of the proposed network are

收稿日期: 2021-05-31; 修回日期: 2021-06-27; 录用日期: 2021-07-15

基金项目: 国家自然科学基金(61901287, 62101364)、四川省重点研发专项(2021YFG0195, 2020YFG0112, 2020YFG0306)、四川省重大科技专项(2019ZDZX0039, 2018GZDZX0029)

通信作者: *zhoupei@scu.edu.cn

verified by comparative experiments.

Key words measurement; three-dimensional face modeling; phase unwrapping; multi-scale attention fusion; contextual feature information; encoder-decoder structure

1 引言

被动三维测量技术^[1]可用于室外自然光照环境下的三维人脸采集,但其建模精度较低,故在高精度、近距离的三维人脸采集中,其已无法满足需求。主动三维测量技术^[2]通过补充额外结构光源的方式丰富了人脸表面的特征信息,进而可有效解决被动三维测量技术精度不高的问题。其中,最具代表性的是基于条纹结构光投影的三维测量技术^[3-5],其通过对目标表面增加相位特征信息的方式进一步准确确定目标的深度信息。相位展开作为主动三维测量技术中的关键步骤,其精度对三维模型的精度起着决定性作用。

传统的相位展开算法包括空间相位展开算法和时间相位展开(TPU)算法。空间相位展开算法通常依据邻近点间的相位差展开相位,容易引入累计误差,其中空间不连续目标、遮挡等情况引起的相位错位问题一直是研究的难点。例如, Lu 等^[6]将伪相关图和最大相位梯度图作为质量评估标准,改善了相位展开时的错误传播问题; Goldstein 等^[7]提出了枝切(BC)法; 张妍等^[8]对 BC 法进行了改进,将分支的长度和作为基准来连接正负残差点,提高了相位展开的效率和成功率。TPU 算法利用多个频率的条纹相位信息进行逐级相位展开,对于静态目标的测量具有较好的适应性,如负指数拟合法^[9]、三频外差法^[10]、三频法^[11]以及双频法^[12]等。

近年来,一些基于深度学习的方法将相位展开问题抽象为分类任务,取得了较好的效果。Spoorthi 等^[13]利用全卷积神经网络 PhaseNet 来计算相位级次,得到的截断相位经过后处理后可得到连续相位。此外,一些研究是利用端到端的网络来实现相位展开的。Wang 等^[14]提出了 U-Net 网络,利用编码器-解码器结构学习像素到空间的映射关系,改善了低信噪比(低调制度)区域相位展开错误的问题。Wang 等^[15]又利用两个 U-Net 网络构建了 Y-Net 网络,在减少网络参数的同时,有效完成了相位展开任务。Zhang 等^[16]首先使用 DeepLabV3+ 结构提取了健壮的特征,并通过大量的标注工作得到了初步的相位图。然后,在网络中用优化层来修正错误展开的相位。由于端到端的相

位展开网络具有精度高、速度快的优点,故其在一定程度上突破了传统相位展开算法所面临的瓶颈,成为了解决复杂相位展开问题的新途径。

在不同场景下进行结构光三维人脸建模的过程中,由于人脸目标往往难以保持绝对静止,这就对测量精度和测量时间提出了较高的要求。然而,无论是空间相位展开算法还是 TPU 算法均很难兼顾测量精度和测量时间。本文提出了基于多尺度注意力机制的相位展开网络(multi-scale attention phase unwrapping network, MSAPUNet),将编码-解码网络与注意力机制相结合来融合多尺度特征。在编码网络中,5 个连续的下采样模块将特征图尺寸依次缩小为输入模块中特征图的 1/2。在解码网络中,上采样模块逐步恢复特征图尺寸,并将其与编码网络中相同尺寸的特征图进行连接。在解码网络中,第 2~4 个上采样模块得到的特征尺度依次为 $(1/8)H \times (1/8)W$ 、 $(1/4)H \times (1/4)W$ 和 $(1/2)H \times (1/2)W$,其中 H 和 W 分别为原始输入特征图的长和宽。然后,在上述 3 个模块之后增加注意力机制模块,将上采样模块的特征输出再次放入注意力机制模块中,可在多个尺度上实现全局上下文信息的捕获。在上述基础上进行空间位置信息的恢复,可实现快速准确的相位预测。所提网络的输入是截断相位,输出是连续相位,这表明即使在三维人脸采集过程中客观存在的欠抽样和相位不连续区域中,所提网络亦有较好的相位预测能力。

2 基本原理

基于条纹结构光的三维测量技术^[3,17]的关键在于连续相位信息的获取,主要包含相位提取和相位展开。根据条纹图像计算截断相位的过程称为相位提取。在截断相位的基础上进行相位展开得到连续相位信息的过程称为相位展开。

2.1 相移技术

恢复目标表面形貌的关键是从受目标表面形貌调制的变形结构光场中解析出截断相位^[18],并对截断相位进行相位展开得到连续相位。正弦条纹结构光的光强可表示为

$$I(x, y) = A(x, y) + B(x, y)\cos \varphi, \quad (1)$$

式中: $A(x, y)$ 表示背景光光强; $B(x, y)$ 表示调制度; φ 表示待求相位; x 和 y 表示图像中像素点的横坐标和纵坐标。以利用相移技术提取截断相位为例, 每相邻两幅条纹图像的相位差为 $2\pi/N$, 每幅条纹的光强可表示为

$$I_i(x, y) = A(x, y) + B(x, y) \cos\left(\varphi - \frac{2\pi i}{N}\right), \quad (2)$$

式中: i 表示第 i 步相移, $i=1, 2, \dots, N$ 。

采用四步相移法 ($N=4$), 结合 (2) 式可提取截断相位, 即

$$\varphi = -\arctan\left(\frac{\sum_{i=0}^3 I_i \sin \frac{2\pi i}{4}}{\sum_{i=0}^3 I_i \cos \frac{2\pi i}{4}}\right), \quad (3)$$

由于相位 φ 由 \arctan 函数求得, 故其取值范围为 $[-\pi, \pi]$ 。

2.2 相位展开

截断相位呈现为不连续的锯齿状, 不能真实反映物体表面形貌的相位信息。为了得到连续的物体表面形貌, 需要恢复出实际相位。通过在截断相位的基础上加或减 2π 的整数倍来实现相位展开^[19]

$$\varphi_{\text{unwrap}}(x, y) = \varphi_{\text{wrap}}(x, y) + 2k\pi, \quad (4)$$

式中: φ_{unwrap} 和 φ_{wrap} 分别表示连续相位和截断相位; k 为相位级次值。

在基于深度学习的相位展开方法^[14, 16, 20]中, 首先利用构建的神经网络模型提取相位特征。然后, 利用大量截断相位数据对网络模型进行训练, 在网络模型学习到特征信息后, 将网络输出与真值进行比较。最后, 通过反向传播路径对网络模型参数进行不断迭代优化, 并继续将网络输出与真值进行比较, 直至收敛。在训练完成后, 将截断相位图输入到训练好的网络模型中, 经过一次前向传播可得到连续相位图像。在基于深度学习的相位展开方法中, 神经网络模型的特征提取能力、用于网络模型训练的数据量的大小和神经网络模型中反向传播路径对网络参数的调整优化能力在很大程度上决定了该方法的连续相位推测能力。

2.3 注意力机制

注意力机制^[21]能迅速捕捉图像中的重点区域, 并会为该区域分配更多的注意力以获取更多有效信息。注意力机制是从编码器-解码器结构发展而来的, 在编码器-解码器结构中生成不同目标输出时使用的都是相同的语义编码, 而在如相位

展开等实际任务中不同输入对同一目标信息的作用是不同的, 故需根据输入对输出的作用程度不同赋予输入不同的权重。注意力机制的本质就是对特征信息在不同维度进行重新编码。在计算出权重矩阵后, 将特征图与其融合可得到新的特征。新的特征中包含了更丰富的上下文信息, 进而有助于准确地展开相位, 尤其是在欠抽样和相位不连续区域中。

3 多尺度相位展开网络

虽然在空间相位展开算法中仅需单幅截断相位, 但是该算法中易出现误差传播现象。TPU 算法虽然避免了误差传播, 但是在相位展时需要额外的图像, 进而完成一次测量所需的时间较长。因此, 只有在保持待测物体静止时, TPU 算法才可准确解算相位信息。此外, 虽然一些基于深度学习的方法可以通过网络预测相位的级次值, 并且还有一些端到端的网络能够直接输出连续相位图, 但是后续均需进行修正和优化。

本文利用条纹结构光投影的方式进行三维人脸建模, 构建了 MSAPUNet, 其输入为通过高频相移条纹计算所得的截断相位, 输出为可直接用于三维建模的连续相位。与 U-Net 不同的是, MSAPUNet 利用注意力子网络来恢复空间位置信息, 可快速准确地预测相位, 即使在欠抽样和相位不连续区域中也有较好的相位预测能力。由于 DeepLabV3+ 结构是分类网络, 故其在处理数据方面的工作量很大, 而 MSAPUNet 无需对数据进行过多处理且无需进行后续的修正和优化操作。

3.1 基于多尺度注意力机制的相位展开网络

MSAPUNet 主要包含编码结构和解码结构两部分, 如图 1 所示。编码网络由 5 个下采样模块组成, 每个下采样模块将特征图压缩至输入模块中特征图的 $1/2$ 。同时, 通过降低分辨率的方式减小网络负担使特征图学习到像素定位信息, 这有助于在展开相位时准确地获取相位级次值。解码网络主要包含上采样模块和注意力子网络。上采样模块利用反卷积层来恢复特征图尺寸, 并将其与编码网络中相同尺度的特征图进行连接, 从而可更精确地进行像素定位。注意力子网络使用空间注意力模块^[22]来捕获全局上下文信息, 增强了网络模型在欠抽样和相位不连续区域中的连续相位预测能力。

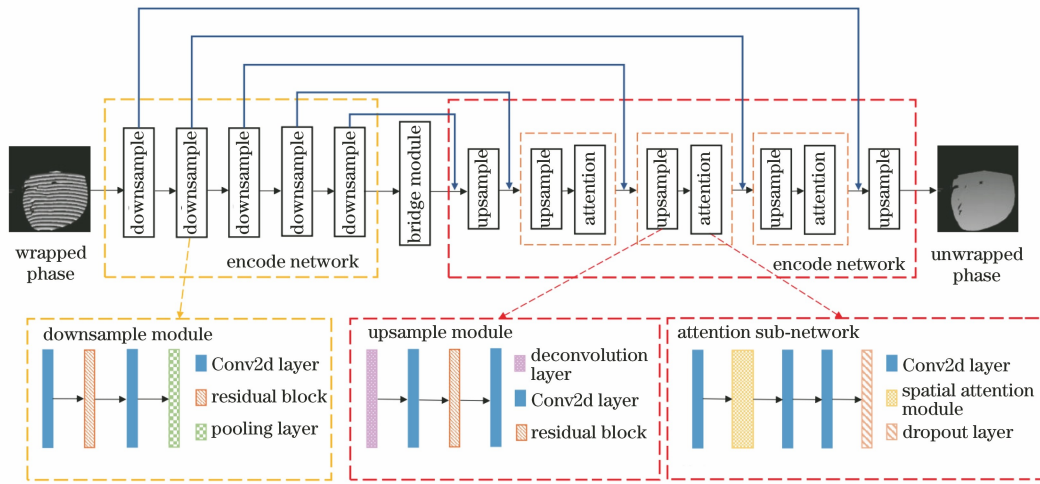


图 1 MSAPUNet 的网络架构

Fig. 1 Network architecture of MSAPUNet

3.1.1 编码网络

编码网络共包含 5 个下采样模块,在每个下采样模块中,输入特征图均会经过 1 个大小为 3×3 的卷积层,进而实现特征图维度的扩大。在第 1 个下采样模块中第 1 个大小为 3×3 的卷积层将特征图维度由 3 扩大至 8,后 4 个下采样模块均将特征图的维度扩大为输入模块中特征图维度的 2 倍。然后,输出的特征图会经过 1 个残差模块,该残差模块由 2 个大小为 3×3 的卷积层和输出特征图维度与前一层输出特征图维度相同的卷积层组成。随后,输出的特征图会经过 1 个大小为 3×3 的卷积层,此时特征图的大小和维度均保持不变。上述操作旨在对特征进行初步提取。在提取特征后,输出的特征图会进入 1 个步长为 2 的池化层,该池化层会将输入特征图的大小降为原来的 $1/2$,这实现了在减少网络参数的同时,提取特征信息,将网络中的特征图进行了抽象化。

在编码网络中,输入特征图的尺寸为 $H \times W$,维度为 3,在经过 5 个下采样模块后,特征图的大小和维度均得到了改变,得到的 5 个特征图的大小依次为 $H/2 \times W/2, H/4 \times W/4, H/8 \times W/8, H/16 \times W/16, H/32 \times W/32$,维度依次为 8, 16, 32, 64, 128。在编码网络后紧跟一个桥接模块,该模块同样包含 2 个大小为 3×3 的卷积层和 1 个残差模块。在进一步提取特征后,桥接模块可将特征图的维度扩大至 256,并将扩大后的特征图传入解码网络。

3.1.2 解码网络

解码网络主要由 5 个上采样模块组成。随着网络的加深,特征中包含的空间位置信息随之减少,因此在第 2~4 个上采样模块中融入了注意力子网络。

注意力子网络通过空间注意力模块深度学习并获取全局上下文特征信息,以减少下采样带来的信息丢失,这有助于所提网络从问题区域的特征图中获取更多有效的相位特征。此外,在编码网络和解码网络之间使用跳跃层连接,将具有相同尺度的特征图融合,以实现浅层网络和深层网络中定位信息与分类信息的融合,这有利于对像素进行准确分类,得到精准的相位级次值 k ,进而准确地预测出连续相位。

在上采样模块中,首先利用一个步长为 2 的反卷积层将特征图分辨率恢复至输入特征图的 2 倍,同时捕获相关的细节信息。然后,输出的特征图依次经过一个大小为 3×3 的卷积层、一个残差模块和一个大小为 3×3 的卷积层,将特征图的维度降低为输入的 $1/2$,进一步将抽象特征解码为对输出目标有利的信息。解码网络通过上采样模块逐步恢复特征图尺寸,将输入的 5 个上采样模块的特征图尺寸与编码网络中 5 个下采样模块的输出尺寸依次对应,其中输入的 5 个上采样模块的特征图尺寸分别为 $H/32 \times W/32, H/16 \times W/16, H/8 \times W/8, H/4 \times W/4, H/2 \times W/2$ 。将尺寸相同的特征图拼接共同作为解码网络中上采样模块的输入,5 个模块的最终输出尺寸依次为 $H/16 \times W/16, H/8 \times W/8, H/4 \times W/4, H/2 \times W/2, H \times W$ 。

注意力子网络的作用是利用空间注意力模块在网络深层中捕获空间信息,尤其在欠抽样和相位不连续区域中。由于解码网络中第 1 个上采样模块恢复特征不足,融合注意力机制效果不明显,第 5 个上采样模块的特征尺寸过大,融合注意力机制后计算量增大,将占用更大的内存空间,因此选择在第 2~4 个上采样模块后融合注意力机制模块,尺寸分别为

$H/8 \times W/8, H/4 \times W/4, H/2 \times W/2$ 。上述 3 个上采样模块的特征输出经过注意力机制模块处理,再连接至下一个上采样模块,以此实现多尺度注意力机制融合。多尺度注意力模块可捕获全局上下文信息,恢复空间位置信息,进而可快速准确地预测相位。

图 2 为空间注意力模块中的计算过程,其中 transform 为转置模块。首先,注意力子网络的输入经过卷积层 conv1 后得到尺寸为 $C \times H \times W$ 的输出 X_1 ,其中 C 为输入特征图的维度。然后,利用两个分支对 X 进行池化(pool)和卷积(conv)操作得到两个特征图,分别是尺寸为 $M \times C$ 的 query 和尺

寸为 $C \times M$ 的 key,其中 $M=H \times W$ 。对于 query 中某个特定的特征,捕获其与 key 中所有具有相似特征的位置上的特征依赖,经 softmax 层处理后,可得到两者的相关性,并构成维度为 $M \times M$ 的权重矩阵 W 。最后,将权重矩阵 W 作用在经过卷积的输入特征图 X_1 上,并将其输出再与输入特征图 X_1 直接融合便可得到注意力模块的特征输出 Y_1 。将某一位置的特征与其他位置间的相关性作为权重,使得融合后的新特征中包含更多有效的空间上下文信息,实现了欠抽样和相位不连续等相位梯度变化较大区域中相位展开错误问题的改善。

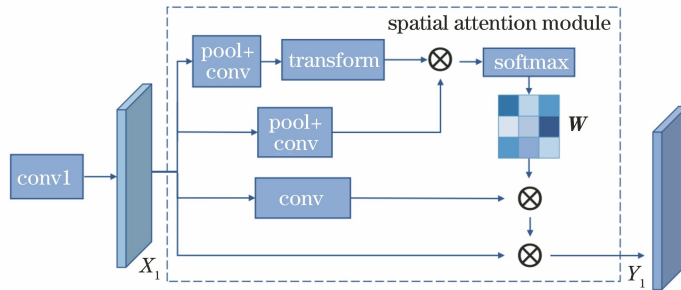


图 2 空间注意力模块中的计算过程

Fig. 2 Calculation process in spatial attention module

3.2 损失函数

本文将 Smooth_L1 函数 $[L(x)]^{[23]}$ 作为损失函数,其表达式为

$$L(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (5)$$

图 3 为损失函数曲线。可以发现,Smooth_L1 函数在训练过程中可以很好地拟合离群点,并且收敛速度较快(仅需训练 100 个 epoch,模型即可收敛)。

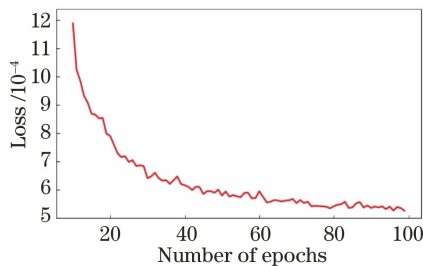


图 3 损失函数曲线

Fig. 3 Loss function curve

4 实验及结果分析

为验证 MSAPUNet 生成的相位图的精度,构建了 FACE 数据集和 MASK 数据集,共设计了 3 组实验,分别测试了 MSAPUNet 在欠抽样区域、相

位不连续区域以及动态目标情况下的相位展开效果。MSAPUNet 采用 python3.7 编码,使用 torch 版本为 1.2.0 的 pytorch 框架,算法在 Intel Core i7-7820X、16 GB-RAM、NVIDIA RTX 1080Ti 的 Ubuntu 20.04 的环境下运行。

4.1 数据集构建

FACE 数据集和 MASK 数据集的采集系统如图 4 所示,包括两台分辨率为 1280 pixel \times 1024 pixel 的工业相机和一台分辨率为 912 pixel \times 1140 pixel 的投影仪。待测目标与采集系统之间的距离为 600~1000 mm。

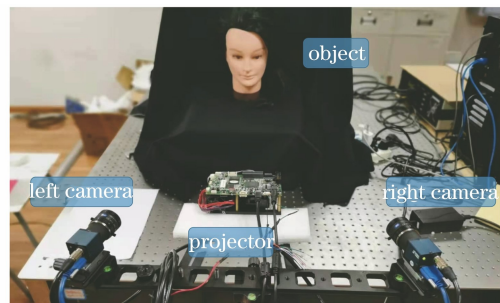


图 4 采集系统

Fig. 4 Acquisition system

在实验中,利用三频四步相移 TPU 算法获得真值的连续相位并构建数据集,具体过程如下。

1) 在有效测量区域内放置待测目标,利用光场投影模块将三频(对应条纹数为 1,8,64)四步正弦条纹结构光投射到待测表面,并使用工业相机同步采集条纹图像。

2) 获取变形条纹有效区域对应的二值掩码图像,去除变形条纹图的背景。对于人脸数据,使用 MTCNN (multi-task cascaded convolutional networks)人脸检测方法检测矩形框和 81 个关键点。对于面具数据,采用人工手动的方式标记有效区域中的矩形框。最后,使用 OpenCV(open source computer vision library)中的 FindContours 函数提取边界轮廓,进而得到二值掩码图像。

3) 利用 2.1 节中的四步相移法计算不同频率条纹图像所对应的截断相位。

4) 使用三频四步 TPU 算法获得连续相位,将高频(对应条纹数为 64)连续相位作为真值,并将图像大小处理为 256 pixel \times 256 pixel。

图 5 为数据集构建示意图,展示了 FACE 数据集和 MASK 数据集的部分样本。两个数据集分别包含 5000 组和 100 组数据,每组数据均包含截断相位以及其对应的连续相位,图像是大小为 256 pixel \times 256 pixel 的 8 位图像。FACE 数据集中包含 GOOD 类型(条纹图像质量较好,不存在遮挡)和 GLASSES 类型(戴眼镜样本,存在局部遮挡,在眼镜框区域处条纹图像的质量相对较差),GOOD 类型与 GLASSES 类型样本的数量比例为 6:4。MASK 数据集包括人脸面具、十二星座面具和十二生肖面具模型样本。

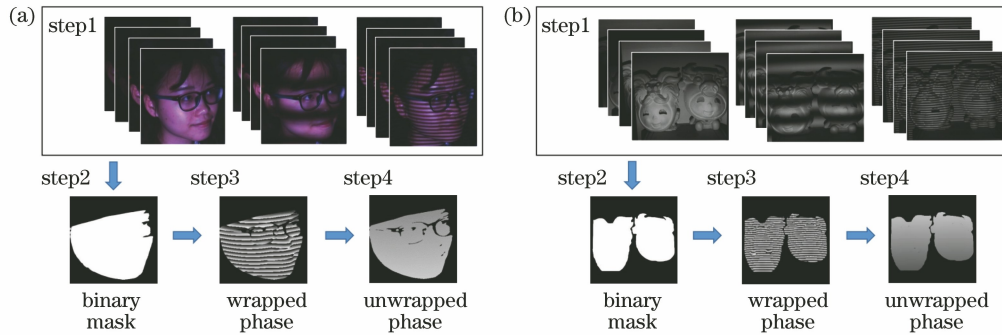


图 5 数据集构建示意图。(a) FACE 数据集;(b) MASK 数据集

Fig. 5 Schematic diagram of dataset construction. (a) FACE dataset; (b) MASK dataset

4.2 算法评估方法

在 MSAPUNet 迭代训练过程中,需要监控算法的相位解析性能,本文将均方根误差(RMSE, M_{RMSE})和结构相似性(SSIM, M_{SSIM})作为算法的评价指标。RMSE 的单位为 rad,SSIM 的表达式为

$$M_{RMSE} = \sqrt{|y^* - y|^2}, \quad (6)$$

式中: y^* 表示利用 TPU 算法得到的连续相位的真值; y 表示 MSAPUNet 预测的相位值。

SSIM 的作用是对两个样本间的图像强度、对比度和结构性质进行相似性度量,取值范围为 $[0,1]$,其值越大表明两图像的相似度越高^[24],此时图像失真小,相位展开效果好。SSIM 的表达式为

$$M_{SSIM} = \frac{(2u_{pre} \times u_{real} + C_1)(\omega_{pre,real} + C_2)}{(u_{pre}^2 + u_{real}^2 + C_1)(\omega_{pre}^2 + \omega_{real}^2 + C_2)}, \quad (7)$$

式中: u_{pre} 和 u_{real} 分别表示相位预测值和相位真值的均值; ω_{pre}^2 和 ω_{real}^2 表示相位预测值与相位真值的方差; $\omega_{pre,real}$ 表示相位预测值与相位真值的协方差; C_1 和 C_2 为常量。

4.3 训练、测试过程及参数

在 FACE 数据集上进行训练和测试,共训练 100 个 epoch,初始学习率为 0.001,每间隔 20 个 epoch,学习率减半。在模型训练完成后,向网络中输入截断相位图,可直接预测出一张连续相位图,算法耗时约 40 ms。为进一步评估 MSAPUNet 的性能,在 MASK 数据集上进行实验。共设计 3 组实验来检验 MSAPUNet 的有效性,分别为欠抽样实验、相位不连续实验和动态目标实验。欠抽样实验和动态目标实验在 FACE 数据集上进行,相位不连续实验在 MASK 数据集上进行。

4.4 基于多尺度注意力机制的相位展开网络的精度分析

为验证提出的 MSAPUNet 的精度,将其与多种传统的空间相位展开算法和基于深度学习的相位展开算法进行了比较。表 1 中统计了 MSAPUNet、QG(quality guided)^[25]算法、BC 算法以及 U-Net 在 FACE 数据集和 MASK 数据集上的 RMSE 和 SSIM。

表 1 在 FACE 数据集和 MASK 数据集中,不同方法的 RMSE 和 SSIM

Table 1 RMSE and SSIM of different algorithms in FACE dataset and MASK dataset

Dataset	QG		BC		U-Net		MSAPUNet	
	M_{RMSE}/rad	M_{SSIM}	M_{RMSE}/rad	M_{SSIM}	M_{RMSE}/rad	M_{SSIM}	M_{RMSE}/rad	M_{SSIM}
FACE	0.2048	0.7513	0.1888	0.7353	0.0504	0.9795	0.0387	0.9850
MASK	0.1112	0.7792	0.1031	0.8010	0.0551	0.9687	0.0273	0.9793

实验结果表明,在 FACE 数据集中,相较于传统的 QG 和 BC 算法,MSAPUNet 的 RMSE 均提高了 5 倍左右;与 U-Net 算法相比,其 RMSE 提高了 23.21%;MSAPUNet 的 SSIM 达到了 0.9850,优于其他对比算法。在 MASK 数据集中,虽然存在遮挡区域,即影响连续相位估计的干扰因素增加,但是 MSAPUNet 的精度仍大幅领先传统的 QG、BC 算法, RMSE 精度约提升 4 倍,同时也优于 U-Net。

4.5 基于多尺度注意力机制的相位展开网络的效率分析

为验证 MSAPUNet 的效率,将其运行时间与 QG、BC 和 U-Net 进行比较,结果如表 2 所示。

表 2 不同方法的效率

Table 2 Efficiency of different methods

Method	QG	BC	U-Net	MSAPUNet
Time	>10 s	>2 s	30 ms	40 ms

从表 2 中可以发现,MSAPUNet 相较于传统的

QG、BC 算法效率分别提升了 250 倍和 50 倍。然而,与 U-Net 相比,由于 MSAPUNet 中融入了多尺度注意力机制,此时计算量会增加,故其效率略有降低。

4.6 欠抽样实验

图像的欠抽样区域往往会给相位解析造成困难,在 MSAPUNet 中通过在解码网络中加入注意力子网络,从网络深层中捕获上下文空间位置信息,进而改善欠抽样区域中相位展开容易出错的问题。在 FACE 数据集中,GLASSES 类型样本在镜框区域处的图像质量较差,即镜框区域为欠抽样区域。为了验证所提网络在欠抽样区域中的相位解析效果,将其与 QG 算法、BC 算法以及 U-Net 在 GLASSES 样本上的相位展开情况进行对比。

图 6 为欠抽样实验中不同算法得到的相位图、点云模型和点云模型误差。从图 6(a)~(e)所示的相位图中可以看出,所有算法在脸部边缘和镜框周

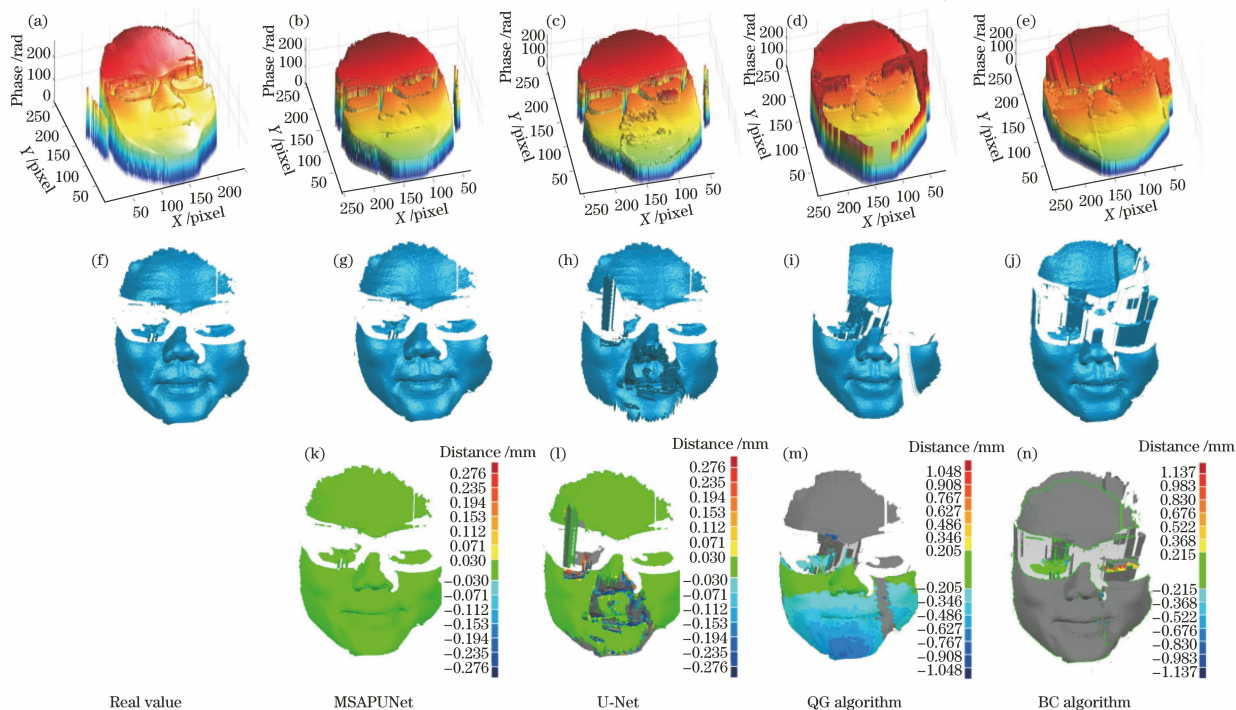
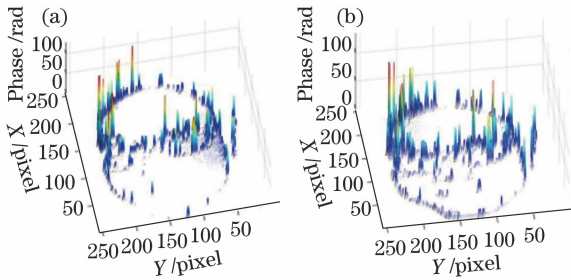


图 6 欠抽样实验中不同方法得到的相位图、点云模型和点云模型误差。(a)~(e)相位图;

(f)~(j)点云模型;(k)~(n)点云模型的误差

Fig. 6 Phase map, point cloud model and error of point cloud model obtained by different methods in undersampling experiment. (a)~(e) Phase map; (f)~(j) point cloud model; (k)~(n) error of point cloud model

边均未能实现相位展开。其中,传统的 QG 和 BC 算法在欠抽样区域中进行相位解析时出现了较大范围的错误,而基于深度学习的算法均能够在欠抽样区域中较为准确地进行相位解析。在利用 U-Net 得到的结果中鼻子和嘴巴区域出现了明显的错误,而 MSAPUNet 在欠抽样区域中的误差较小。从图 6(f)~(j) 所示的点云模型中可以看出,QG 和 BC 算法生成的点云模型效果不佳,这是因为二者输出的相位误差较大。U-Net 生成的点云模型同样在鼻子和嘴巴区域有明显错误。由于 MSAPUNet 的相位展开精度高,故其重建的点云模型中仅有少量



误差。

为了进一步验证所提网络在欠抽样区域的相位解析能力,欠抽样实验统计了样本在三维空间的误差分布。由图 6 可知,脸部边缘以及镜框周边是低质量区域,故误差也主要分布在这些区域。图 7 为欠抽样实验中不同方法得到的误差图。可以发现,利用 QG 和 BC 算法得到的结果在镜框和边缘区域出现了大面积误差,相位误差集中在 100~250 rad 范围内。然而,利用基于深度学习的算法得到的结果在相应区域仅有少量误差,并且与 U-Net 相比,利用 MSAPUNet 得到的结果的误差更小。

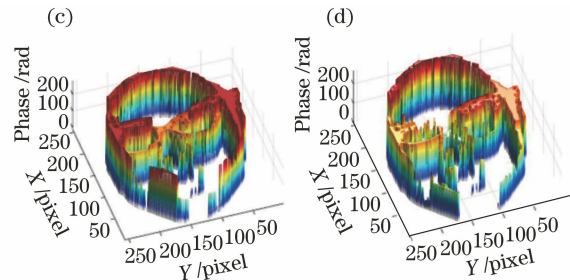


图 7 欠抽样实验中不同方法得到的误差图。(a) MSAPUNet;(b) U-Net;(c) QG 算法;(d) BC 算法

Fig. 7 Error maps obtained by different methods in undersampling experiment. (a) MSAPUNet; (b) U-Net; (c) QG algorithm; (d) BC algorithm

表 3 统计了欠抽样实验中各种方法的 RMSE 和 SSIM。MSAPUNet 的 RMSE 为 0.0135,相比于 U-Net 误差减少了 17.18%。与 QG 和 BC 算法相比,MSAPUNet 的准确度分别提升了 25 倍和 20 倍。MSAPUNet 的 SSIM 为 0.9976,相比于 QG 和 BC 算法分别提高了 44.58% 和 37.11%。由此可知,MSAPUNet 可实现更精准的相位解析,故其可显著改善欠抽样区域中相位展开易出错的问题。

表 3 欠抽样实验中不同方法的 RMSE 和 SSIM

Table 3 RMSE and SSIM of different methods in undersampling experiment

Index	QG	BC	U-Net	MSAPUNet
	algorithm	algorithm		
$M_{\text{RMSE}}/\text{rad}$	0.3392	0.2688	0.0163	0.0135
M_{SSIM}	0.6900	0.7276	0.9952	0.9976

欠抽样实验中不同算法得到的点云模型误差如表 4 所示。可以发现,利用传统的 QG 和 BC 算法生成的三维模型的标准偏差分别为 0.5569 mm 和 0.5977 mm,而利用 MSAPUNet 生成的三维模型的标准偏差为 0.0174 mm,这表明与 QG 和 BC 算法相比,所提网络较大地提升了三维模型的精度。

表 4 欠抽样实验中不同方法得到的点云模型误差

Table 4 Errors of point cloud model obtained by different methods in undersampling experiment unit: mm

Index	QG	BC	U-Net	MSAPUNet
	algorithm	algorithm		
Average distance	0.7816	0.7097	0.1956	0.0006
Standard deviation	0.5569	0.5977	0.6030	0.0174

4.7 相位不连续实验

在包含人脸鼻子或孤立物体(面具)的图像中,由于目标表面高度变化剧烈,条纹会出现截断错位,导致图像中出现相位不连续区域,故利用传统空间相位展开算法很难成功进行相位展开。为了验证 MSAPUNet 在该区域的解相效果,将 MASK 数据集中包含孤立图像的样本用于测试,并将 MSAPUNet 与 U-Net、QG 和 BC 算法进行了对比。

图 8 为相位不连续实验中利用不同算法得到的相位图、点云模型和点云模型误差,实验包含图 8(a)~(e)的右侧和图 8(f)~(n)的左侧所示的双鱼座模型与图 8(a)~(e)的左侧和图 8(f)~(n)的右侧所示的金牛座模型这两个孤立面具。从图 8(a)~(e)所示的相位图中可以看出,QG 和 BC

算法成功展开了双鱼座模型,但两个模型之间的区域高度变化较大即存在相位不连续区域,此时相位展开出错,经误差传递后在金牛座模型解相过程中出现了更多的错误。虽然 U-Net 的相位展开结果显著优于 QG 和 BC 算法,但在展开双鱼座模型左侧轮廓时仍出现了误差。然而,利用 MSAPUNet 得到的相位仅在边缘区域出现少量误差,即与其他

比方法相比,利用 MSAPUNet 得到的相位图更连续、更接近于真值。由于利用 QG 和 BC 算法得到的金牛座面具的相位值存在较大误差,因此在点云模型中未能得到正确的三维信息。利用 U-Net 得到的双鱼座面具的信息不完整,在左脸边缘存在明显误差。利用 MSAPUNet 生成的点云模型接近真实模型,仅在边缘区域存在少量误差。

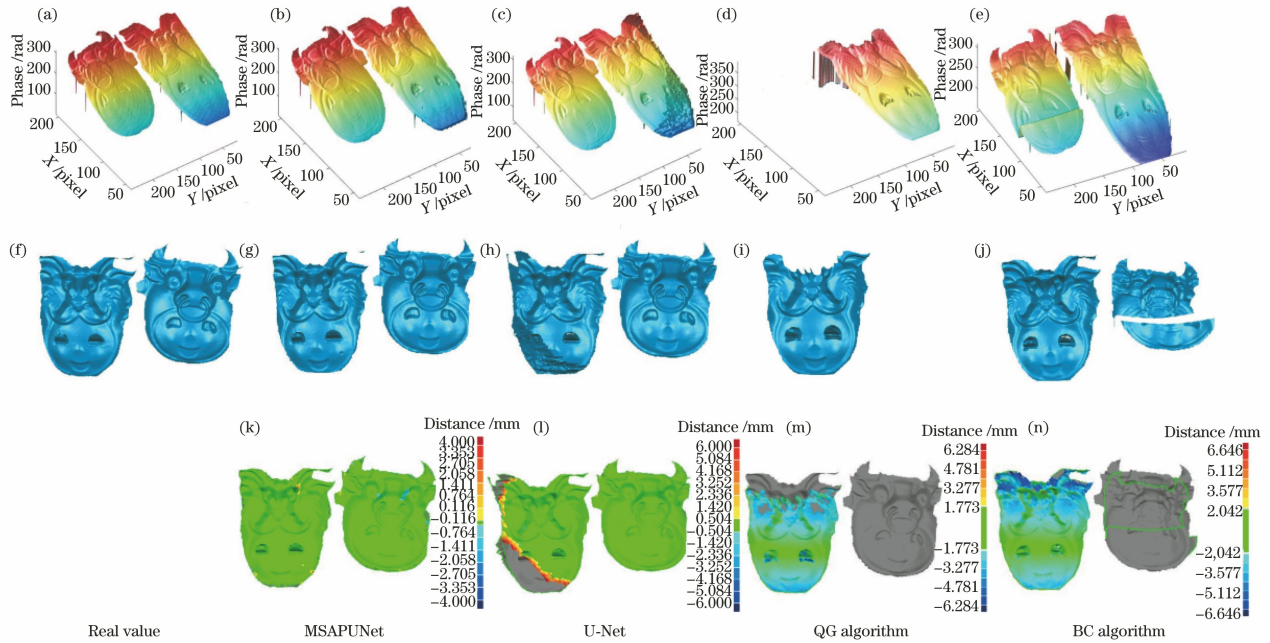


图 8 相位不连续实验中不同方法得到的相位图、点云模型和点云模型误差。(a)~(e)相位图;(f)~(j)点云模型;(k)~(n)点云模型的误差

Fig. 8 Phase maps, point cloud models and errors of point cloud model obtained by different methods in phase discontinuity experiment. (a)–(e) Phase map; (f)–(j) point cloud model; (k)–(n) error of point cloud model

图 9 为相位不连续实验中不同方法得到的相位误差图。同样,对 MSAPUNet、U-Net、QG 和 BC 算法在孤立目标上的 RMSE 和 SSIM 进行了统计,如表 5 所示。可以发现,MSAPUNet 的 RMSE 最小,与 U-Net 相比误差减小了 46.40%,并且 SSIM

达到了 0.9754。由相位不连续实验可知,MSAPUNet 在相位不连续区域中可准确展开相位。此外,相位不连续实验所采用的样本并不在网络的训练集中,这进一步证明了所提网络具有较好的泛化性。

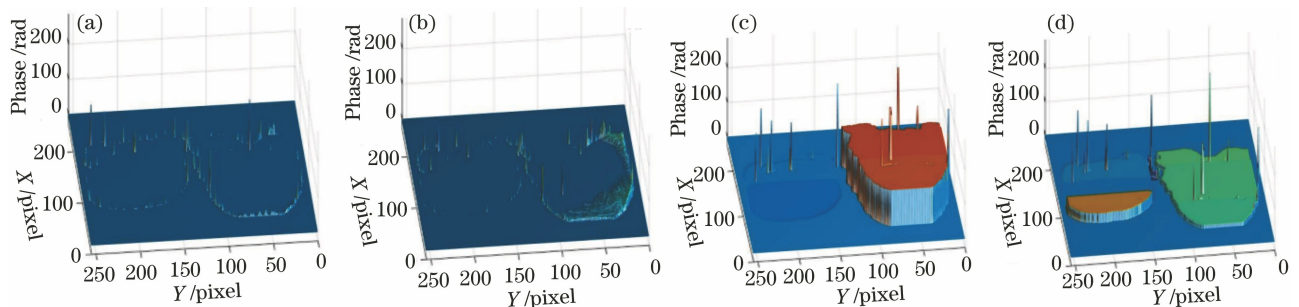


图 9 相位不连续实验中不同方法得到的误差图。(a) MSAPUNet;(b) U-Net;(c) QG 算法;(d) BC 算法

Fig. 9 Error maps obtained by different methods in phase discontinuity experiment. (a) MSAPUNet; (b) U-Net; (c) QG algorithm; (d) BC algorithm

表 5 相位不连续实验中不同方法的 RMSE 和 SSIM
Table 5 RMSE and SSIM of different methods in phase discontinuity experiment

Index	QG	BC	U-Net	MSAPUNet
	algorithm	algorithm		
$M_{\text{RMSE}}/\text{rad}$	0.1364	0.1180	0.0543	0.0291
M_{SSIM}	0.7192	0.7888	0.9650	0.9754

相位不连续实验中不同方法得到的点云模型误差如表 6 所示。可以发现,利用 MSAPUNet 生成的三维模型的标准偏差最小,这表明所提网络可实现高精度的三维建模。

表 6 相位不连续实验中不同方法得到的点云模型误差
Table 6 Errors of point cloud model obtained by different methods in undersampling experiment unit: mm

Index	QG	BC	U-Net	MSAPUNet
	algorithm	algorithm		
Average distance	1.7725	2.0420	0.1083	0.0065
Standard deviation	1.1066	1.4077	0.6376	0.0915

4.8 动态目标实验

为了进一步验证所提网络的鲁棒性与适用性,

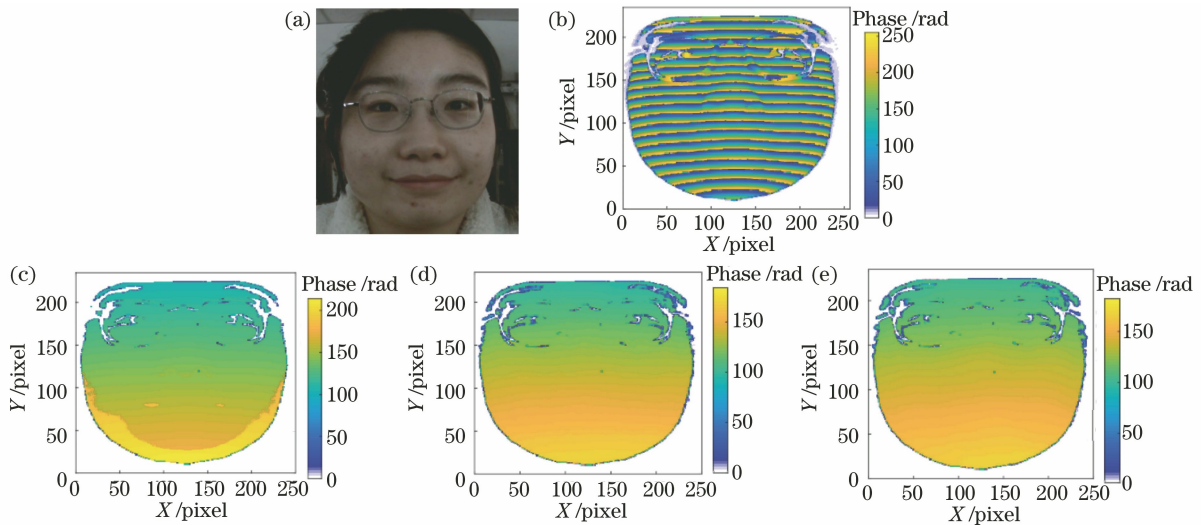


图 10 动态目标实验结果。(a)纹理图;(b)截断相位;(c)利用 TPU 算法生成的相位;

(d)利用 U-Net 生成的相位;(e)利用 MSAPUNet 生成的相位

Fig. 10 Experimental results of dynamic target. (a) Texture map; (b) wrapped phase; (c) phase generated by TPU algorithm; (d) phase generated by U-Net; (e) phase generated by MSAPUNet

5 结 论

针对复杂场景三维人脸建模中传统算法及深度学习算法在相位展开时存在的问题,提出了基于多尺度注意力机制的相位展开网络(MSAPUNet)。在所提网络中,利用编码器-解码器结构实现了多尺

将其与 TPU 算法和 U-Net 进行对比,验证其在动态人脸目标情况下的展开效果。其中,TPU 算法采用三频(对应条纹数为 1,8,64)三步 TPU。

图 10 为在人脸大笑情境下所拍摄样本的纹理图及计算出的截断与连续相位。由于目标在大笑时未保持静止,故利用 TPU 算法得到的结果在下颌附近出现了大面积的相位不连续区域。利用 U-Net 生成的相位存在部分误差,而利用 MSAPUNet 生成的相位较为连续,无截断现象,且相较于 U-Net 误差更小。

相位解析所得到的连续相位在行方向或列方向上应满足单调性,即某一行(列)的相位值单调递增(减)。图 11 展示了 MSAPUNet、U-Net 和 TPU 算法在第 60 列的绝对相位值。可以发现,MSAPUNet 利用高频截断相位预测出的连续相位单调递减,可较准确地拟合出连续相位。然而,TPU 算法需要低频连续相位辅助展开高频相位,且在高频位置出现了相位跳变,进而容易受到人脸运动的影响。由此可知,相较于 TPU 算法,在处理非静止目标时 MSAPUNet 具有更高的鲁棒性。

度特征融合,并通过引入注意力子网络的方式恢复空间位置信息,实现了快速准确的相位预测。针对欠抽样、相位不连续及动态目标实验,在构建的 FACE 数据集和 MASK 数据集中,所提网络的相位解析精度与 QG 和 BC 算法相比分别至少提高了 5 倍和 4 倍。在欠抽样和相位不连续实验中,所提

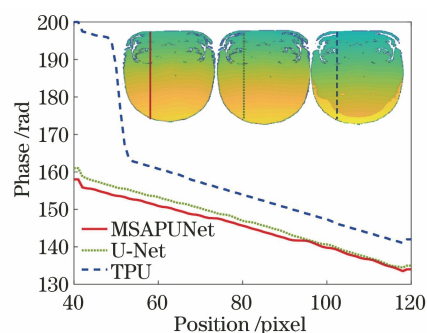


图 11 动态目标实验中的第 60 列相位

Fig. 11 The 60th column phase in dynamic target experiment

网络的误差与 U-Net 相比分别减小了 17.18% 和 46.40%，较基于深度学习的方法也有明显优势。虽然所提网络的相位展开结果较为理想，但其涉及的样本均是同光照、同频率、同设备下采集所得，故后续将丰富样本采集环境以验证网络在不同情况下的相位展开性能，进一步提高网络的鲁棒性。

参 考 文 献

- [1] Tariq R U, Ming Y, Zhi C, et al. Passive intermodulation measurement of Radiofrequency interference shielding gasket [C]//2019 International Applied Computational Electromagnetics Society Symposium-China (ACES), August 8-11, 2019, Nanjing, China. New York: IEEE Press, 2019: 19514251.
- [2] Liu C, Ju W G, Zhang G, et al. A SDN-based active measurement method to traffic QoS sensing for smart network access [J]. *Wireless Networks*, 2021, 27(5): 3677-3688.
- [3] Zuo C, Feng S J, Huang L, et al. Phase shifting algorithms for fringe projection profilometry: a review[J]. *Optics and Lasers in Engineering*, 2018, 109: 23-59.
- [4] Guo W B, Zhang Q C, Wu Z J. Real-time three-dimensional imaging technique based on phase-shift fringe analysis: a review [J]. *Laser & Optoelectronics Progress*, 2021, 58(8): 0800001. 郭文博, 张启灿, 吴周杰. 基于相移条纹分析的实时三维成像技术发展综述 [J]. *激光与光电子学进展*, 2021, 58(8): 0800001.
- [5] Liu L, Xi D D, Chen Z J, et al. Three-dimensional shape measurement based on hybrid dual-frequency fringe projection [J]. *Laser & Optoelectronics Progress*, 2021, 58(12): 1212001. 刘路, 奚冬冬, 陈志健, 等. 混合双频条纹投影三维测量方法 [J]. *激光与光电子学进展*, 2021, 58(12): 1212001.
- [6] Lu Y G, Wang X Z, Zhong X H, et al. A new quality map for quality-guided phase unwrapping [J]. *Chinese Optics Letters*, 2004, 2(12): 698-700.
- [7] Goldstein R M, Zebker H A, Werner C L. Satellite radar interferometry: two-dimensional phase unwrapping [J]. *Radio Science*, 1988, 23(4): 713-720.
- [8] Zhang Y, Feng D Z, Qu X N, et al. Application of a novel branch-cut algorithm in phase unwrapping [J]. *Journal of University of Electronic Science and Technology of China*, 2013, 42(4): 555-558. 张妍, 冯大政, 曲小宁, 等. 改进的枝切法在相位展开中的应用 [J]. *电子科技大学学报*, 2013, 42(4): 555-558.
- [9] Huntley J M, Saldner H O. Error-reduction methods for shape measurement by temporal phase unwrapping [J]. *Journal of the Optical Society of America A*, 1997, 14(12): 3188-3196.
- [10] Ohtsuka Y. Real-time electro-optical heterodyne signal processors [J]. *Opto-electronics*, 1974, 6(3): 235-241.
- [11] Chen L, Deng W Y, Lou X P. Phase unwrapping method base on multi-frequency interferometry [J]. *Optical Technique*, 2012, 38(1): 73-78. 陈玲, 邓文怡, 娄小平. 基于多频外差原理的相位解包裹方法 [J]. *光学技术*, 2012, 38(1): 73-78.
- [12] Li J L, Su H J, Su X Y. Two-frequency grating used in phase-measuring profilometry [J]. *Applied Optics*, 1997, 36(1): 277-280.
- [13] Spoorthi G E, Gorthi S, Gorthi R K S S. PhaseNet: a deep convolutional neural network for two-dimensional phase unwrapping [J]. *IEEE Signal Processing Letters*, 2019, 26(1): 54-58.
- [14] Wang K Q, Li Y, Qian K M, et al. One-step robust deep learning phase unwrapping [J]. *Optics Express*, 2019, 27(10): 15100-15115.
- [15] Wang K Q, Dou J Z, Qian K M, et al. Y-Net: a one-to-two deep learning framework for digital holographic reconstruction [J]. *Optics Letters*, 2019, 44(19): 4765-4768.
- [16] Zhang T, Jiang S W, Zhao Z X, et al. Rapid and robust two-dimensional phase unwrapping via deep learning [J]. *Optics Express*, 2019, 27(16): 23173-23185.
- [17] Chen F. Phase unwrapping in dynamic 3-D profilometry [D]. Chengdu: Sichuan University, 2005. 陈锋. 动态三维面形测量中相位展开问题的研究 [D]. 成都: 四川大学, 2005.
- [18] Srinivasan V, Liu H C, Halioua M. Automated phase-measuring profilometry of 3-D diffuse objects

- [] . Applied Optics, 1984, 23(18): 3105-3108.
- [19] Zuo C, Huang L, Zhang M L, et al. Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review [J]. Optics and Lasers in Engineering, 2016, 85: 84-103.
- [20] Qian J M, Feng S J, Tao T Y, et al. Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement [J]. APL Photonics, 2020, 5(4): 046105.
- [21] Niu Z Y, Zhong G Q, Yu H. A review on the attention mechanism of deep learning [J]. Neurocomputing, 2021, 452: 48-62.
- [22] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups [J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [23] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [24] Sara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR: a comparative study [J]. Journal of Computer and Communications, 2019, 7(3): 8-18.
- [25] Zhu Y J, Liu L R, Luan Z, et al. A hybrid phase-unwrapping method for optical interferometry based on new parameter map and local plane approximation [J]. Proceedings of SPIE, 2006, 6292: 62921C.