

# 基于数字孪生与迁移学习的结构光条纹图像分析(特邀)

金子衡<sup>1,2,3</sup>, 徐可<sup>1,2,3</sup>, 张宁远<sup>1,2,3</sup>, 邓潇<sup>1,2,3</sup>, 左超<sup>1,2,3</sup>, 陈钱<sup>1,3</sup>, 冯世杰<sup>1,2,3\*</sup>

<sup>1</sup>南京理工大学电子工程与光电技术学院智能计算成像实验室, 江苏 南京 210094;

<sup>2</sup>南京理工大学智能计算成像研究院, 江苏 南京 210019;

<sup>3</sup>南京理工大学江苏省光谱成像与智能感知重点实验室, 江苏 南京 210094

**摘要** 近年来,深度学习技术广泛应用于计算光学三维成像的研究中。在条纹投影轮廓术中,通过训练深度学习网络,可从单幅条纹图像中恢复高精度的相位信息。然而,为了训练神经网络模型,通常需要耗费大量的时间成本和人力成本来采集训练数据集。为了解决该问题:首先,建立数字孪生条纹投影系统,并利用域随机化技术对虚拟照明光栅进行增强,使用计算机进行虚拟扫描,生成大量仿真光栅条纹图像;其次,利用仿真光栅图像对 U-Net 神经网络进行预训练;最后,引入迁移学习,采用少量真实光栅条纹图像对神经网络进行参数微调。由于 U-Net 的结构特殊性,提出并分析了“从左至右”“从上至下”“全局微调”等 3 种 U-Net 神经网络微调策略。实验结果表明,采用“从上至下”策略微调 U-Net“瓶颈”网络模块的方法可获得最佳的迁移学习结果,神经网络的相位预测精度可得到显著提升。相比于使用大量真实数据进行训练,所述方法仅利用 20% 的数据就可训练神经网络获得高精度的相位重建结果。

**关键词** 计算成像; 条纹投影; 深度学习; 迁移学习; 条纹分析

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP232430

## Structured Illumination Fringe-Pattern Analysis Based on Digital Twin and Transfer Learning (Invited)

Jin Ziheng<sup>1,2,3</sup>, Xu Ke<sup>1,2,3</sup>, Zhang Ningyuan<sup>1,2,3</sup>, Deng Xiao<sup>1,2,3</sup>, Zuo Chao<sup>1,2,3</sup>,  
Chen Qian<sup>1,3</sup>, Feng Shijie<sup>1,2,3\*</sup>

<sup>1</sup>Smart Computational Imaging Laboratory, School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China;

<sup>2</sup>Smart Computational Imaging Research Institute of Nanjing University of Science and Technology, Nanjing 210019, Jiangsu, China;

<sup>3</sup>Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China

**Abstract** In recent years, deep learning techniques have been widely applied in computational optical three-dimensional imaging. Fringe projection profilometry uses a trained deep neural network to recover high-precision phase information from a single fringe image. However, collecting the training dataset for a neural network expends a considerable amount of time and human resources. To mitigate this problem, we establish a digital twin-fringe projection system that enhances virtual fringe patterns using domain randomization techniques. A U-Net neural network is pretrained using a large number of simulated fringe-pattern images generated through virtual scanning. Next, transfer learning is introduced and the neural network parameters are fine-tuned using a small number of real fringe-pattern images. Targeting fringe analysis applications, this study proposes and analyzes three U-Net neural network fine-tuning schemes: “from left to right” “from

收稿日期: 2023-11-03; 修回日期: 2023-12-09; 录用日期: 2023-12-13; 网络首发日期: 2023-12-23

基金项目: 国家重点研发计划(2022YFB2804600, 2022YFB2804603)、国家自然科学基金(62075096, 62005121, U21B2033)、江苏省基础研究计划前沿引领技术(BK20192003)、江苏省“333工程”科研项目资助计划(BRA2016407)、中央高校科研专项资助项目(30921011208, 30919011222, 30920032101)、中央高校基础科研业务费专项资金资助项目(2023102001)、中国博士后科学基金项目(2023T160318)、江苏省光谱成像与智能感知重点实验室开放基金(JSJP202105, JSJP202201)

通信作者: \*shijiefeng@njjust.edu.cn

top to bottom” “global fine-tuning”. The experimental results demonstrate that fine-tuning the bottleneck network module of the U-Net under the “from top to bottom” strategy optimizes the transfer learning results, largely improving the phase prediction accuracy of the neural network. The proposed method achieves high-precision phase reconstruction results after training the neural network on only 20% of the real data, thus avoiding the need for a large real dataset.

**Key words** computational imaging; fringe projection; deep learning; transfer learning; fringe analysis

## 1 引言

三维信息获取与处理已广泛应用于工业制造、航空航天、高端装备等领域<sup>[1-3]</sup>。随着计算机技术、光学和光电技术的发展,以光信号为载体的光学三维传感技术融合了光电子学、图像处理、计算机视觉与现代信号处理等多学科领域,逐渐发展成为光学计量和信息光学的最重要的研究领域和研究方向之一。在工业设计应用中,基于三维数字化模型的逆向设计方法可快速获得现有成熟产品的准确和完整的数字模型,大大缩短产品或模具的研发周期。在虚拟现实领域,大量景物的三维彩色模型化数据被用于模拟训练、科学试验、3D动画的建构中。在医学整形领域,三维数字化技术广泛用于面部软组织形态修复、外科检测,以及假牙假肢的量身定做。文物保护领域中,三维彩色数字化技术能以无损测量的方式获得文物的三维信息和纹理信息,便于长期保存与再现文物的重要信息。

立体视觉是一种模拟人类视觉原理的被动光学三维传感方法<sup>[4]</sup>。系统用两个或多个相机从不同角度获取同一场景的多幅图像,通过对同一物点在各幅图像上进行检测和匹配,根据立体视差进行测距,得到物体的深度信息。立体视觉测量系统硬件结构简单,易于实现。然而在实际应用中由于遮挡或阴影的影响,景物中的某些点有可能只出现在一幅图中,且视觉信息不足容易导致误匹配。另外,当处理无明显表面特征的物体时,难以从两个或者多个视角中找到相同的对应点,因而无法进行准确的三维重建。

条纹投影轮廓术是一种主动三维形貌测量方法,其通过将传统立体视觉中的摄像机替换为光源发生器(如投影仪)来实现<sup>[5-6]</sup>。测量过程中,光源以特定规则和模式投影编码的图案到被测物体上,从而实现主动式三维测量。这些编码图案会受到物体表面形状的调制而产生形变,然后由另外一台相机从不同角度捕捉到这些带有形变的结构光。通过分析相机、光源之间的位置关系,以及结构光的形变程度,可以精确测量出物体的三维坐标。与传统立体视觉方法相比,条纹投影轮廓术的最大优点在于求解物体的相位时进行点对点的运算,即每个点的相位值不受其相邻点光强值的影响。这有助于避免由于物体表面反射率不均匀或观察角度偏差引起的误差,从而使测量精度达到微米级别。随着数字技术的不断发展,现代结构光条纹投影方法主要采用数字光栅投影技术,这一技术利用数字设备代替传统机械

装置生成投影面上的结构光光栅,使编码过程更加灵活和准确。

条纹投影轮廓术因其非接触、高精度、全场测量、点云重建效率高等优点,已成为目前广泛使用的一种三维传感方法<sup>[7-8]</sup>。根据结构光条纹图像的使用数量,条纹投影轮廓术中的条纹分析方法通常可分为单帧法和多帧法。相移法是使用最为广泛的一种多帧相位解调方法<sup>[9]</sup>。通过相机拍摄具有相对空间相位移动的一系列正弦光栅图像,可准确恢复测量对象的相位信息。该类方法的优势是测量精度高,但不足在于测量效率较低,不适用于运动场景的测量<sup>[10]</sup>。对于单帧法而言,代表性的基于傅里叶变换的条纹分析方法仅需要一幅光栅图像即可实现相位解调,因此适合动态场景的测量<sup>[11-12]</sup>。但由于空域滤波的影响,场景中的高频细节信息难以得到高精度还原。

近年来,深度学习技术已广泛应用于全息成像、无透镜成像、散射介质成像、显微成像等多个计算光学成像领域<sup>[13-16]</sup>。在光学计量领域,Feng等<sup>[17-19]</sup>证明了通过训练卷积神经网络,可实现基于单帧条纹图像的相位解调。相比于传统基于傅里叶变换的单帧相位解调方法,深度学习技术的应用能够有效提升复杂物体的相位测量精度。Yin等<sup>[20]</sup>通过引入具有相位解调频域先验知识的LeFTP模块,建立了基于物理先验的条纹图像分析模型,提高了神经网络处理罕见样本的能力。相位展开方面,Guo等<sup>[21]</sup>利用深度神经网络构建了一个面向时间相位展开的统一框架,将传统孤立的多频法、多波长法与数论法融合到一个基于人工神经网络的统一分析框架,有效克服了相位噪声的影响,提升了相位展开的鲁棒性。三维重建方面,Li等<sup>[22]</sup>提出一种基于深度学习的单帧双频复合条纹投影轮廓术,经过大量数据集的学习,通过适当训练的神经网络即可直接从一幅双频复合条纹图像中恢复出绝对相位,从而实现高精度的单帧三维测量。

对于上述基于深度学习的结构光三维成像方法而言,为保证深度神经网络的优异表现,需要从真实场景中采集大量的数据集。然而,大量数据集的采集需要耗费大量的人力、物力和时间成本。为此,Zheng等<sup>[23]</sup>构建了数字孪生条纹投影系统,并使用计算机渲染工具Blender进行虚拟扫描,利用生成的虚拟数据集训练深度神经网络,节省了时间成本和劳动力。Wang等<sup>[24]</sup>利用Blender构建了一个虚拟条纹投影系统,为了提高输出深度图像的准确性,还提出了一种结合SSIM索引和拉普拉斯算子的损失函数。此外,Zhu

等<sup>[25]</sup>也提出了基于Blender模拟的光场结构光照明数据生成方法,并公开了利用该方法生成的数据集。该数据集包含光场结构光图像、标准深度信息和视差信息。尽管利用计算机生成的虚拟数据集可避免采集真实数据带来的时间和人力成本,但虚拟仿真数据与真实数据之间仍存在差异,尤其是当实际的条纹投影系统存在噪声、环境光、畸变等干扰因素时,神经网络预测的可靠性将受到影响。

为了解决上述问题,本文提出一种基于数字孪生与迁移学习的结构光条纹图像分析方法。首先,利用Blender对真实条纹投影系统进行模拟,建立相对应的虚拟数字孪生条纹投影系统。其次,利用该数字孪生系统开展虚拟扫描,运用域随机化技术随机改变虚拟照明光栅的属性,生成大量、多样化的仿真光栅训练数据对U-Net神经网络进行预训练。最后,利用真实条纹投影系统拍摄少量的真实条纹图像数据,对预训练的U-Net神经网络进行微调。值得注意的是,由于U-Net结构的特殊性,对于U-Net神经网络的微调并不能通过简单地冻结浅层和微调深层来实现,实际上存在多种不同的方法<sup>[26]</sup>。然而,在现有基于深度学习的条纹分析研究工作中,尚未有文献讨论如何微调U-Net来实现最佳的条纹图像分析。为此,提出“从左至右”“从上至下”“全局微调”等3种策略。通过实验验证,发现采用“从上至下”策略微调U-Net“瓶颈”网络模块时可获得最佳的迁移学习结果。同时,得益于迁移学习,神经网络在处理复杂表面时的预测可靠性有了显著提升。相比于使用大量真实数据,所提方法可将真实数据集的使用减少80%,即仅利用20%的真实数据就能实现高精度的单帧光栅图像相位解调。所提方法有助于降低深度神经网络对于大量真实条纹数据集的依赖,并且对全息成像、显微成像等其他光学成像研究领域也有借鉴意义。

## 2 基本原理

### 2.1 条纹投影三维成像模型

条纹投影轮廓术的原理如图1所示。光源向被测物体投影正弦结构光图像,编码图案受到物体表面形状的调制而产生形变,而带有形变的光栅条纹被另外位置的相机拍摄,通过相机投影光源之间的位置关系和结构光形变的程度可以确定出物体的三维形貌。对于相机而言,空间点 $(x_w, y_w, z_w)$ 投影至相机像素 $(u, v)$ 的成像过程可以描述为

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \mathbf{K} [\mathbf{R} \quad \mathbf{T}] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \quad (1)$$

式中: $s$ 为缩放因子; $\mathbf{P} = \mathbf{K} [\mathbf{R} \quad \mathbf{T}]$ ,为投影矩阵, $\mathbf{K}$ 为相机的内参矩阵, $[\mathbf{R} \quad \mathbf{T}]$ 为系统的外参矩阵, $\mathbf{R}$ 和 $\mathbf{T}$ 分

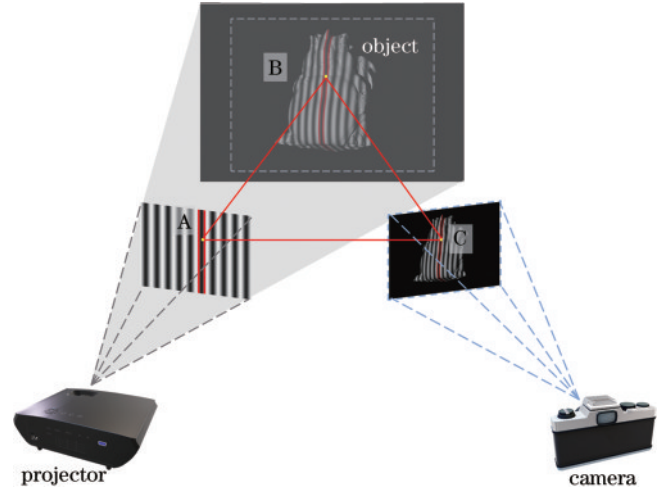


图1 条纹投影系统示意图

Fig. 1 Diagram of fringe projection profilometry

别代表旋转矩阵和平移矩阵。内参矩阵 $\mathbf{K}$ 和外参矩阵 $[\mathbf{R} \quad \mathbf{T}]$ 可进一步表示为

$$\mathbf{K} = \begin{bmatrix} f_u & \alpha & \mu_0 \\ 0 & f_v & \nu_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

$$[\mathbf{R} \quad \mathbf{T}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}, \quad (3)$$

式中: $\mathbf{K}$ 包含 $f_u, f_v, \alpha, \mu_0, \nu_0$ 等5个参数, $f_u, f_v$ 为相机沿着 $u$ 和 $v$ 方向的焦距, $\alpha$ 为相机坐标轴之间的倾斜系数, $(\mu_0, \nu_0)$ 表示光学中心在成像平面的坐标。

对于投影仪而言,可将其当作“反相机”来处理<sup>[27]</sup>,因此也可以使用式(1)来表示该空间点与相应照明像素 $(u^p, v^p)$ 之间的投影关系:

$$s^p \begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = \mathbf{P}^p \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \mathbf{K}^p [\mathbf{R}^p \quad \mathbf{T}^p] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \quad (4)$$

式中:上标 $p$ 表示与投影仪相关的参数矩阵。利用三角法,联立式(1)和式(4),可以计算出点 $(x_w, y_w, z_w)$ 的空间坐标:

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} p_{11} - p_{31}u & p_{12} - p_{32}u & p_{13} - p_{33}u \\ p_{21} - p_{31}v & p_{22} - p_{32}v & p_{23} - p_{33}v \\ p_{21}^p - p_{31}^p u^p & p_{22}^p - p_{32}^p u^p & p_{23} - p_{33}u^p \end{bmatrix}^{-1} \begin{bmatrix} p_{14} - p_{34}u \\ p_{24} - p_{34}v \\ p_{24} - p_{34}u^p \end{bmatrix}. \quad (5)$$

### 2.2 数字孪生条纹投影系统的构建

利用相机的标定算法<sup>[28]</sup>,可获得相机和投影仪的内外参数。利用这些参数,可在计算机虚拟环境中建立与真实系统一一对应的数字孪生系统。以开源三维



图形图像软件 Blender 为例,阐述数字孪生条纹投影系统的构建过程。首先,在对真实系统标定后,可获得相机的感光芯片尺寸、靶面尺寸与焦距等参数,这些参数可直接输入 Blender 中。

对于外参,它提供了相机坐标系与世界坐标系的变换关系。将(0, 0, 0)坐标点设置为坐标系的原点,利用标定关系,可以得到:

$$(x_c, y_c, z_c) = -\mathbf{R}_c^{-1}\mathbf{T}, \quad (6)$$

式中: $(x_c, y_c, z_c)$ 为相机投影中心在世界坐标系下的坐标。同样地,可以用该方法得到投影仪的投影中心在世界坐标系下的坐标。接下来,利用旋转矩阵 $\mathbf{R}$ 来计算法线方向与世界坐标系 $x, y, z$ 轴的夹角。由于旋转矩阵 $\mathbf{R}$ 可以看作是围绕3个轴的矩阵变换乘积,因此,旋转矩阵 $\mathbf{R}$ 可以拆解为

$$\mathbf{R} = \mathbf{R}_z(\phi)\mathbf{R}_y(\theta)\mathbf{R}_x(\psi) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}. \quad (7)$$

因此,可以获得对应方向的夹角:

$$\psi = \arctan(r_{32}, r_{33}), \quad (8)$$

$$\theta = \arctan\left(-r_{31}, \sqrt{r_{32}^2 + r_{33}^2}\right), \quad (9)$$

$$\phi = \arctan(r_{21}, r_{11}). \quad (10)$$

随后,将计算得到的夹角导入 Blender 实现数字孪生相机的构建。关于虚拟投影仪的构建,在 Blender 中使用了聚光灯进行投影仪的构建。对于外部参数而言,可按照与相机类似的方法进行参数的设置,即基于式(6)将参数分解并直接输入 Blender 中。虚拟相机与投影仪设置完成后,即完成了数字孪生条纹投影系统

的配置。

### 2.3 基于数字孪生系统生成训练数据集

选用 Thingi10K 模型数据集<sup>[29]</sup>。这个库包含 10000 个不同形貌的三维模型。在生成数据集时,用仿真投影仪向虚拟待测物体投影光栅条纹,然后记录下这些光栅条纹。这些条纹将作为输入数据训练神经网络。训练时,神经网络输入相位的正弦项和余弦项(即反正弦函数的分子项和分母项)<sup>[17]</sup>。为计算虚拟光栅图像对应的训练标签,采用 12 步相移法计算正弦项和余弦项的标准值。

由于构建的数字孪生系统难以完全复制真实环境中的系统,为了增加网络的推理能力,采用域随机化技术<sup>[30-31]</sup>,对仿真训练数据集的建立进行增强。具体而言,利用 $N$ 步相移光栅图像对虚拟物体进行照明,条纹图像的光强分布可以表示为

$$I_n(u, v) = A(u, v) + B(u, v) \cos[2\pi fu + 2\pi n/N] + \delta, \quad (11)$$

式中: $n$ 为相位索引; $A$ 为背景光强; $B$ 为条纹的对比度; $f$ 为条纹的空间频率; $\delta$ 为噪声(通常假设服从高斯分布)。通过改变背景光强 $A$ 、条纹的对比度 $B$ 、条纹的空间频率 $f$ 和噪声 $\delta$ ,可生成不同种类的照明光栅图像。此外,在虚拟相机拍摄光栅图像时,还可以改变其视场角(FOV)的大小(即调整相机的焦距),进而生成更为丰富的条纹图像。具体而言,这些参数的调整范围如表 1 所示。基于初始值,在一定范围内对这些参数随机取值,生成对应的光栅图像,然后投影至虚拟物体。通过虚拟相机进行图像的采集,生成仿真训练集的光栅图像数据。

表 1 利用域随机化技术生成不同的虚拟照明光栅

Table 1 Generating different virtual fringe patterns using domain randomization

Index	Background intensity	Contrast	Frequency	Noise	FOV
Initial value	$A_0$	$B_0$	$f_0$	$\delta_0$	$F_0$
Random range	$\pm 10\%$	$\pm 10\%$	$\pm 10\%$	0-2.4629	$\pm 2\%$

### 2.4 深度神经网络结构

为实现结构光条纹图像的分析,构建了基于 U-Net 的网络架构<sup>[32]</sup>。U-Net 是一种具有编码器-解码器架构的全卷积网络,广泛用于图像分割等各种图像处理任务。基于 U-Net 的条纹图像分析原理如图 2 所示。输入图像为一幅结构光条纹图像,利用有监督的训练方式,训练 U-Net 从该条纹图像中解调相位的正弦项 $M(u, v)$ 和余弦项 $D(u, v)$ :

$$\Phi(u, v) = \arctan \frac{M(u, v)}{D(u, v)} = \arctan \frac{cB(u, v) \sin \Phi(u, v)}{cB(u, v) \cos \Phi(u, v)}. \quad (12)$$

在神经网络预测得到 $M(u, v)$ 和 $D(u, v)$ 之后,利用式(12)所示的反正切函数,即可计算得到包裹相

位 $\Phi(u, v)$ 。在 U-Net 的编码器中,输入的条纹图像首先经过第一个卷积块 Block 1(在图中 Block 缩写为 B, Block 1 即对应 B1),经过池化处理之后,沿 $x$ 和 $y$ 方向提取的特征分辨率减小 1/2。然后,这些特征分别经过 Block 2 到 Block 5 的处理,以获得更多不同尺寸规模的图像特征。接下来,与编码器相反, U-Net 解码器子网络执行上采样操作以恢复输入图像的原始大小。对于解码器的每一个 Block,都有一个跳跃连接使卷积层的输出与来自相同级别编码器的特征图连接起来。这种跳跃连接的结构有助于同时提取条纹图像中的低级和高级信息,并缓解深度卷积网络的梯度消失问题,提升相位预测的准确性。U-Net 的最后一层是由线性激活函数激活的卷积层。输出数据包含两个通道,分别对应正弦项 $M(u, v)$ 和余弦项

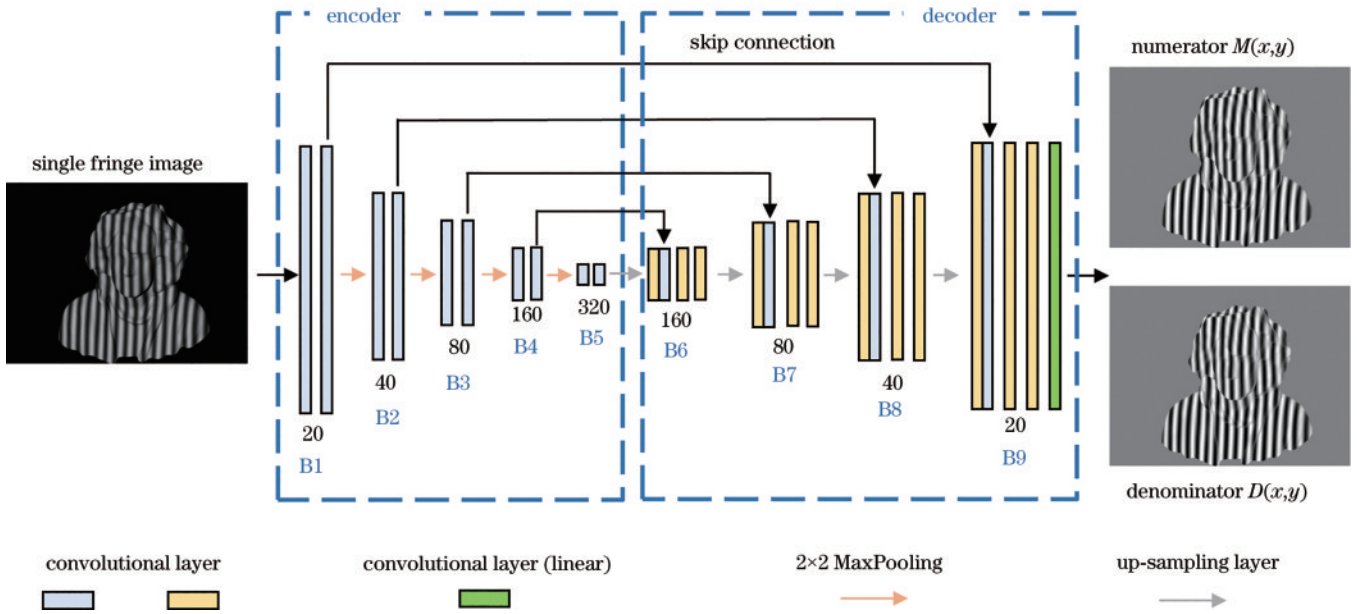


图 2 基于 U-Net 的单帧条纹图像分析原理图

Fig. 2 Diagram of single frame fringe image analysis using U-Net

$D(u, v)$ 。为了训练 U-Net, 构建了以下的损失函数:

$$L(\theta) = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W (y_{h,w} - \hat{y}_{h,w})^2, \quad (13)$$

式中:  $\theta$  表示 U-Net 模型的参数, 这些参数可在神经网络的训练过程中获得;  $H$  和  $W$  分别表示以像素为单位的图像高度和宽度;  $y$  是 U-Net 的预测结果, 它由正弦项  $M(u, v)$  和余弦项  $D(u, v)$  组成;  $\hat{y}$  是  $M(u, v)$  和  $D(u, v)$  的标准真值, 可由 12 步相移法计算得到。

### 2.5 微调与迁移

由于仿真数据无法完全表达真实数据的分布特征, 在利用仿真数据对神经网络进行预训练之后, 进一步利用迁移学习的方法对该神经网络进行微调。研究对象是 U-Net, 其是在自动编码器结构之上发展而来的。相比于自动编码器, U-Net 的主要特征体现在跳跃连接操作, 通过将编码器的浅层特征与解码器的深层特征相结合, 跳跃连接能够捕捉到图像的多尺度信息。这有助于模型更好地识别和区分不同大小和复杂度的物体。

由于跳跃连接的存在, 在 U-Net 结构中, “浅层”和“深层”的概念不像在典型的前馈式自动编码器网络中那么简单。根据 U-Net 的结构特征, 可以从两个方面来理解它的“浅层”和“深层”。1) “从左到右”的策略。可将 U-Net 视为具有跳跃连接的自动编码器。从这个角度出发, 可以沿用传统顺序神经网络的“浅层”和“深层”概念, 即第 1 层 (在编码器的开头) 是最浅的, 而最后一层 (在解码器的输出处) 是最深的。在第 1 种策略中, 可以设计 6 种微调的方式: 微调前 4 个 Block 和 Block 5 的前一半; 微调 Block 5 的后一半和最后 4 个 Block; 只微调 Block 1; 只微调 Block 9;

微调 Block 1 和 Block 2; 微调 Block 3 和 Block 4, 如图 3 所示。

2) “从上到下”的策略。由于跳跃连接的存在, 也可以认为 U-Net 的“U”形结构的顶部是最浅的部分, 而底部 (瓶颈位置) 是最深的部分。从这个角度来看, 通过神经网络层的路径越短, 它就越“浅”, 反之越“深”。此时, Block 1 和 Block 9 都是最浅的, 而 Block 5 是最深的。根据第 2 种策略, 也可以采取 6 种方式进行微调: 微调 Block 1 和 Block 9; 微调 Block 2 至 Block 8; 微调 Block 1、Block 2、Block 8、Block 9; 微调 Block 3 至 Block 7; 微调除最深的 Block 5 之外的所有网络; 只微调 Block 5, 如图 3 所示。

此外, “全局微调”也可用于 U-Net 的迁移学习, 如图 3 所示。第 3 种策略中, 对 U-Net 的所有神经网络层参数进行微调。分析不同微调方法的相位误差, 当误差最小时, 即确定其为针对 U-Net 的最佳微调方式。值得注意的是, 对预训练的 U-Net 进行微调时, 不同的真实数据集规模也将对迁移学习的效果造成影响, 因此需要在相同数据集规模条件下, 比较不同微调策略的效果。

## 3 实验

为了验证所提方法的有效性, 首先构建了一套真实的条纹投影系统, 包括一个 DLP 投影仪 (型号 LightCrafter 4500) 和一台工业相机 (型号 Basler acA640-750um)。相机镜头焦距为 12 mm, 系统工作距离为 1 m。其次, 对该系统进行标定, 确定相机和投影仪的内外参数。利用相关标定参数, 在 Blender 中构建一套与该真实系统具有相同内外参数的数字孪生系统。

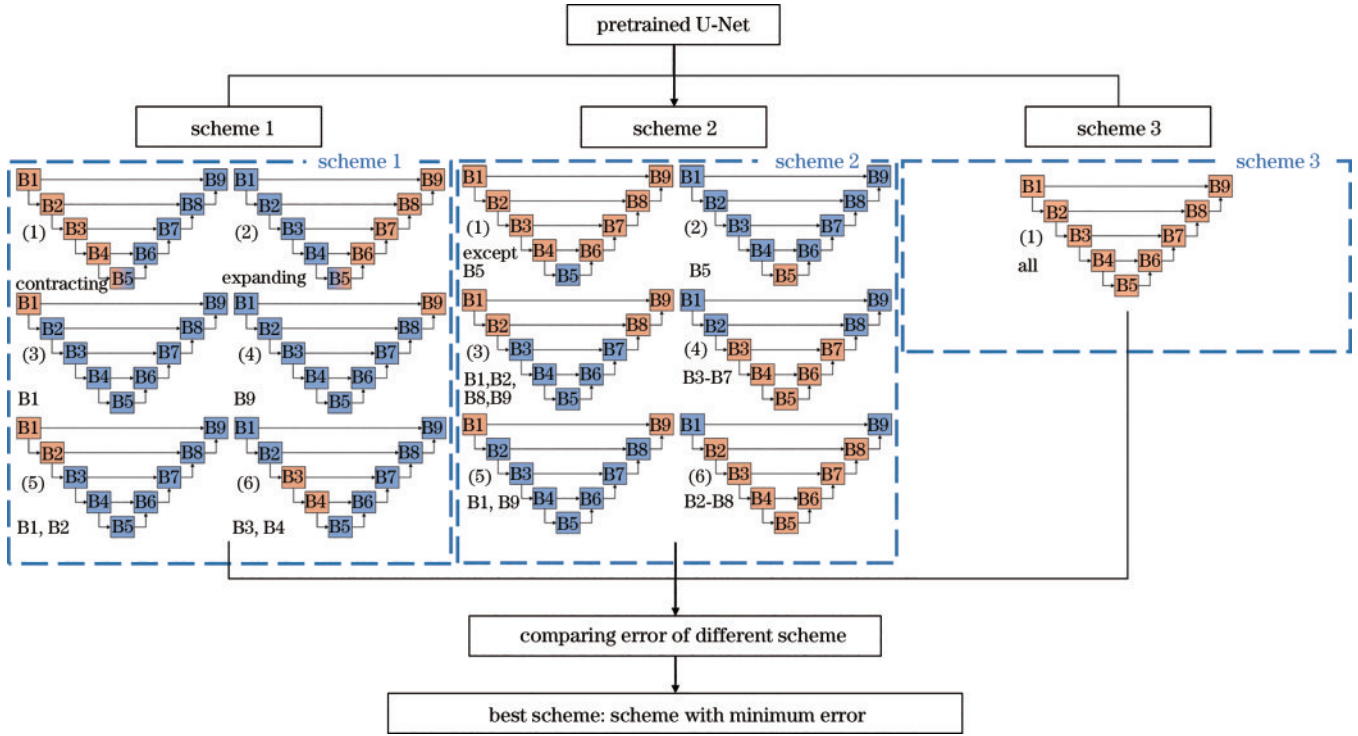


图 3 针对条纹分析 U-Net 神经网络的微调策略分析

Fig. 3 Fine-tuning strategy analysis for U-Net neural network in fringe pattern analysis

### 3.1 仿真数据生成与域随机化处理

从 Thingi10K 的模型中随机挑选 100 个几何形貌较为复杂的三维模型作为数字孪生条纹投影系统的测量模型<sup>[29]</sup>。对于每一个模型,使用数字孪生条纹投影系统对其进行虚拟扫描。为了高效利用每一个三维模型,对每一个模型进行角度的旋转,并在每个角度下分别对该物体进行扫描。旋转时  $z$  轴和  $y$  轴都以  $40^\circ$  旋转为一个间隔进行采样。在每个采样角度下,数字孪生条纹投影系统投影并通过 12 步相移法采集光栅,总计生成 16200 幅光栅图像,其中,渲染生成的图片像素分辨率与真实系统使用的相机分辨率保持一致。随后,使用 12 步相移法计算这些光栅图像对应的训练标准值。

为了进一步缩小仿真数据和真实系统拍摄图像的差异,利用域随机化技术对照明光栅进行增强。实验中以背景光强、对比度、条纹频率、噪声和视场角等参数的初始值为基础,在一定取值范围内随机生成新的照明光栅,如表 2 所示。对于视场角的变化,通过修改相机的焦距来完成。实验中,经过域随机化处理后,额外生成了随机化的新投影光栅图片,并生成 16200 幅光栅图像。然后将这些光栅图再次投影至不同的被测物体,生成相应的光栅测量图像并利用相移法计算训练标准值。图 4 的第 1 行与第 2 行分别为参数取初始值时的光栅图像与域随机化处理后对应的光栅图像。不难看出,调整照明光栅的参数后,光栅图像的多样性得到了有效增强。

表 2 Blender 中利用域随机化技术对照明光栅进行增强

Table 2 Enhancing fringe pattern using domain randomization in Blender

Index	Background intensity	Contrast	Frequency	Noise	FOV
Initial value	110	100	48	0	12
Random range	$\pm 10$	$\pm 10$	$\pm 5$	2.4629	$\pm 2$

### 3.2 神经网络的预训练与测试

在实验中,将构建的 U-Net 部署在 GeForce RTX 4090 显卡上,并基于 TensorFlow 框架对神经网络进行训练。预训练过程中,对所有数字孪生系统生成的数据集进行随机划分,分别将 80% 的数据和 20% 的数据作为训练集和验证集。使用 Adam 优化器,将 minibatch 设置为 2,训练 200 轮。初始学习率设置为  $1 \times 10^{-3}$ ,并进行迭代优化,即当损失函数在一定时间内无法再下降的时候,降低学习率。训练时的损失函数如式(13)所示。为了测试预训练网络的性能,利用其测量了两个未见场景的相位与三维形貌。结果如图 5 所示,对于复杂区域的轮廓,相位误差较为明显。同时,对于重建的三维模型,被测物体的头发、面部等细节轮廓存在较大测量误差。实验结果表明,完全依赖仿真数据训练的模型在处理真实场景下的复杂物体时测量误差仍然较大。

### 3.3 迁移与微调

为提升预训练神经网络的性能,利用迁移学习技术,对其进行了微调。首先,使用不同规模的真实数据集对 U-Net 进行训练。并将该训练结果作为迁移学习



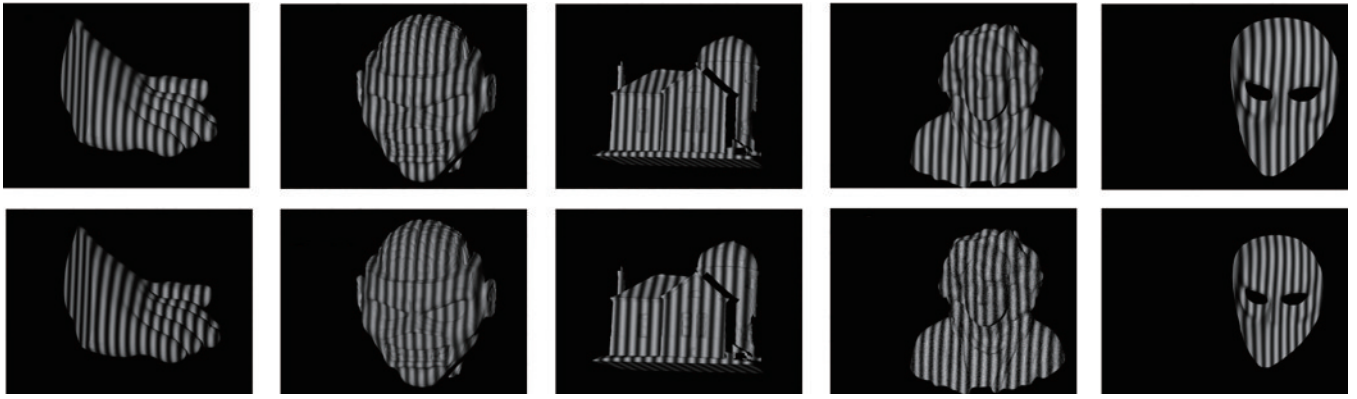


图 4 初始光栅与域随机化后的光栅图像

Fig. 4 The initial fringe pattern and the fringe pattern after domain randomization

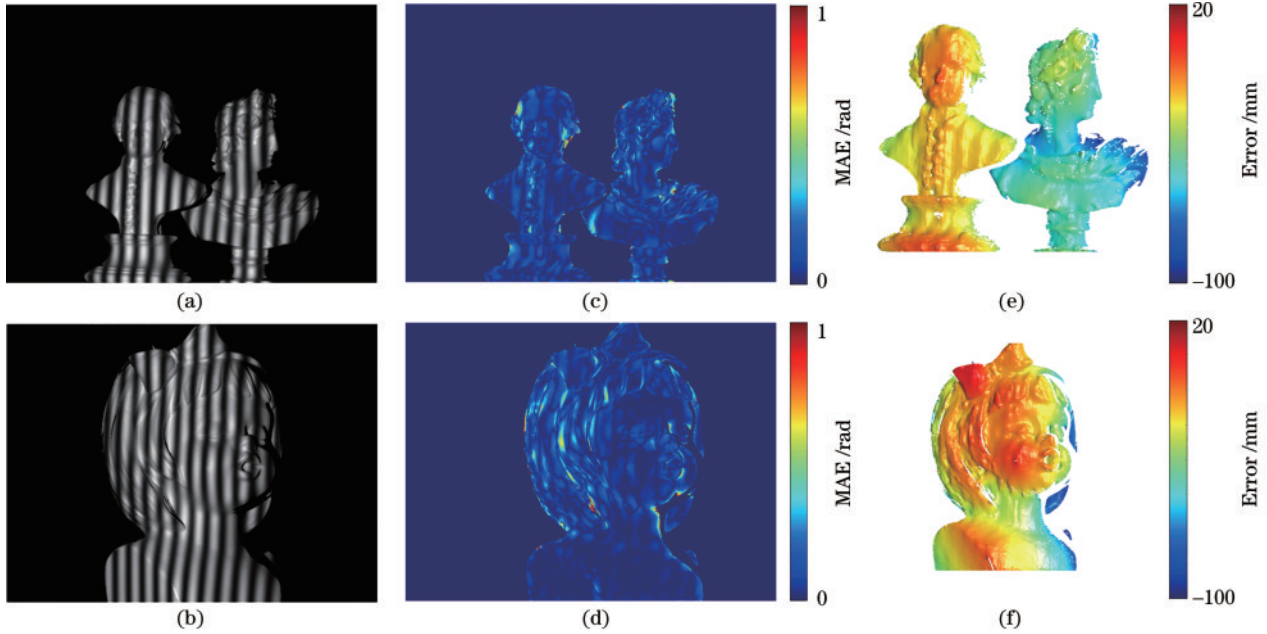


图 5 利用仿真训练网络在真实场景下的测试效果。(a)(b)输入的真实条纹图;(c)(d)计算得到的相位误差分布;(e)(f)三维重建结果

Fig. 5 The effect of simulated dataset trained network under real-world scenarios. (a) (b) The input real fringe pattern images; (c) (d) the phase error maps; (e) (f) the results of 3D reconstruction

效果比对的参考基准。在 100 个随机选择的不同真实场景下,对训练的网络进行测试,获得的平均相位误差如图 6 所示。可以看出,当真实光栅图像数量为 100 时,神经网络的相位误差较大,达到了 0.133 rad。当参与训练的真实数据不断增加时,神经网络的表现逐步提升。当真实数据集规模达到 800 时,相位误差已下降至 0.097 rad。

接下来,利用第 2.5 节所述迁移学习方法对预训练的 U-Net 进行微调。实验结果如图 7 和表 3 所示。首先,测试了“从左到右”策略,发现微调浅层(收缩)部分并冻结深层(扩张)部分的网络相较于冻结浅层而微调深层部分可以获得更好结果。尽管实际上浅层部分的参数要比深层部分的参数要少得多(小于一半),但仍然获得了更小的相位误差。尤其是在真实数据集规

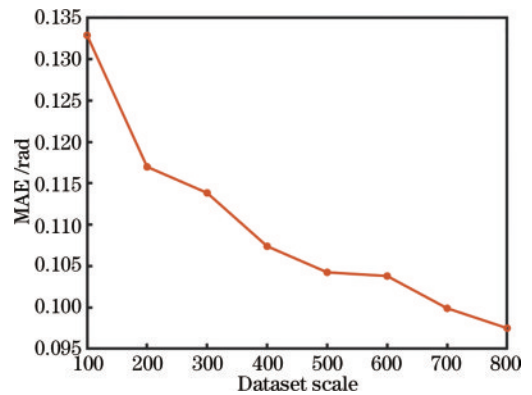


图 6 利用不同规模的真实数据训练 U-Net 后,测试得到的平均相位误差

Fig. 6 The average phase error of U-Net trained by real data of different scales

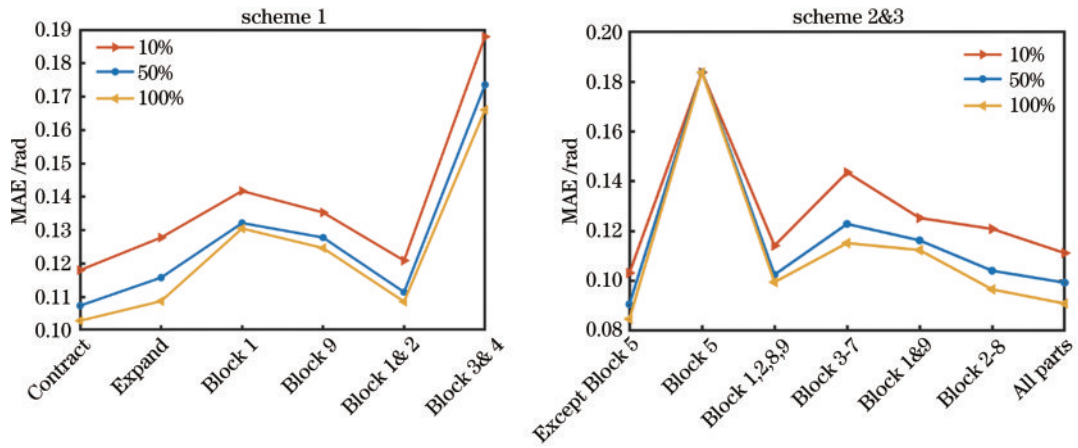


图 7 不同迁移学习策略下微调 U-Net 不同部分对应的相位误差,其中,左图为策略 1 相关微调方法相位误差,右图为策略 2 和策略 3 相关微调方法相位误差

Fig. 7 The phase errors corresponding to fine-tuning different parts of the U-Net under different transfer learning schemes, the left image shows the phase errors for fine-tuning methods related to scheme one, the right image displays the phase errors for fine-tuning methods related to scheme two and three

表 3 在不同迁移策略下利用不同规模真实训练集的迁移学习相位误差对比

Table 3 Comparison of transfer learning phase error using real datasets of different scales under different transfer schemes

Scheme	Parameters trained	Proportion of total parameters / %	Phase mean absolute error of different dataset scales		
			10%	50%	100%
Contracting part	$9.19 \times 10^5$	30.31	0.1181	0.1074	0.1029
Expanding part	$2.11 \times 10^6$	69.69	0.1278	0.1158	0.1088
Block 1	$3.82 \times 10^3$	0.13	0.1418	0.1322	0.1305
Block 9	$1.44 \times 10^4$	0.48	0.1353	0.1278	0.1247
Block 1-2	$2.55 \times 10^4$	0.84	0.1209	0.1115	0.1086
Block 3-4	$4.32 \times 10^5$	14.26	0.1880	0.1736	0.1661
All except Block 5	$1.65 \times 10^6$	54.39	0.1032	0.0906	0.0847
Block 5	$1.38 \times 10^6$	45.61	0.1838	0.1838	0.1838
Block 1-2, 8-9	$9.60 \times 10^4$	3.17	0.1142	0.1025	0.0995
Block 3-7	$2.94 \times 10^6$	96.83	0.1436	0.1229	0.1152
Block 1, 9	$1.82 \times 10^4$	0.60	0.1253	0.1163	0.1124
Block 2-8	$3.01 \times 10^6$	99.40	0.1209	0.1041	0.0966
All parts	$3.03 \times 10^6$	100.00	0.1112	0.0993	0.0909

模较小的时候,两者的差异会更加明显。接下来,通过对比微调 Block 1 和 Block 2,以及微调 Block 3 和 Block 4,发现微调 Block 1 和 Block 2 可获得更好的结果,但其误差仍高于收缩部分神经网络时的。然后,测试了微调 Block 1 和 Block 9 的实验效果。实验结果表明,在第 1 种策略下,微调最浅和最深的神经网络模块都难以获得高精度的相位预测结果。

随后,测试了第 2 种“从上到下”的策略。与第 1 种策略不同,此时“瓶颈”位置的 Block 5 被认为是最深的神经网络模块,而 Block 1 和 Block 9 均为最浅的神经网络模块。首先对比了仅微调 Block 5 和微调 Block 5 以外的模块这两种方法的效果,结果表明,仅微调 Block 5 会获得较大测量误差,而微调 Block 5 以外的模块可获得更好的相位测量精度。此外,当真实数据集规模增大时,相位预测结果会更加准确。接下来,对“深层”模块进行微调。测试了微调 Block 1、Block 2、

Block 8、Block 9 和微调 Block 3 到 Block 7 两种方案,结果表明,微调 Block 3 到 Block 7 能取得更好的结果。其次,继续让更多的深层模型保持冻结,仅微调浅层部分,相比于微调 Block 1 和 Block 9,微调 Block 2 到 Block 8 可带来更高的相位预测精度。

接下来,对于第 3 种全局微调方案,发现训练的神经网络可以获得较好的神经网络预测结果,但综合 3 种迁移学习策略来看,发现微调 Block 5 以外的模块可以带来最佳的相位预测精度(相位误差为 0.0847 rad),并且此时需要重新训练的参数只有总参数的 54.39%,即可以实现高精度与高效率的神经网络微调。

最终,将微调 Block 5 以外的模块确定为后续实验中预训练网络的迁移学习方法,并进一步验证了不同规模真实数据集下微调预训练网络的效果。相位误差分布如图 8 所示,其中,横坐标为使用的数据集规模,纵坐标为相位测量误差。从该图可以看出,



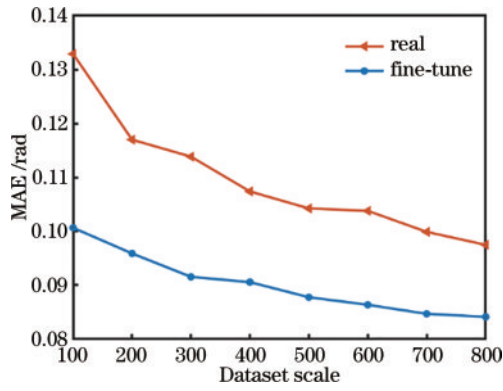


图 8 在不同规模真实数据集下,预训练 U-Net 与随机初始化 U-Net 的相位误差对比  
 Fig. 8 Comparison of phase errors between pre-trained U-Net and randomly initialized U-Net on real datasets of different scales

微调神经网络时,随着真实数据集规模不断增大,训练模型的相位预测精度也不断升高。值得注意的是,当仅采用 20% 的真实数据(160 幅左右的光栅图像)时,此时的相位误差为 0.0972 rad。该测量精度达到了与 800 幅真实数据训练网络相当的精度。实验结果表明,尽管真实的训练数据减少了 80%,但借助数字孪生系统和迁移学习方法,神经网络依然可以进行高精度的相位测量。此外,还发现在迁移学习的过程中当真实数据规模继续增加时,神经网络表现会继续提升,直至超过大量真实数据训练的神经网络。

图 9 为利用 20% 的真实数据进行迁移学习的 U-Net 神经网络相位预测结果,分别测试了石膏模型、塑料物体,以及金属工件,测量结果分别如图 9(a)~(d)、图 9(e)~(f),以及图 9(g)~(h)所示。实验结果表明,

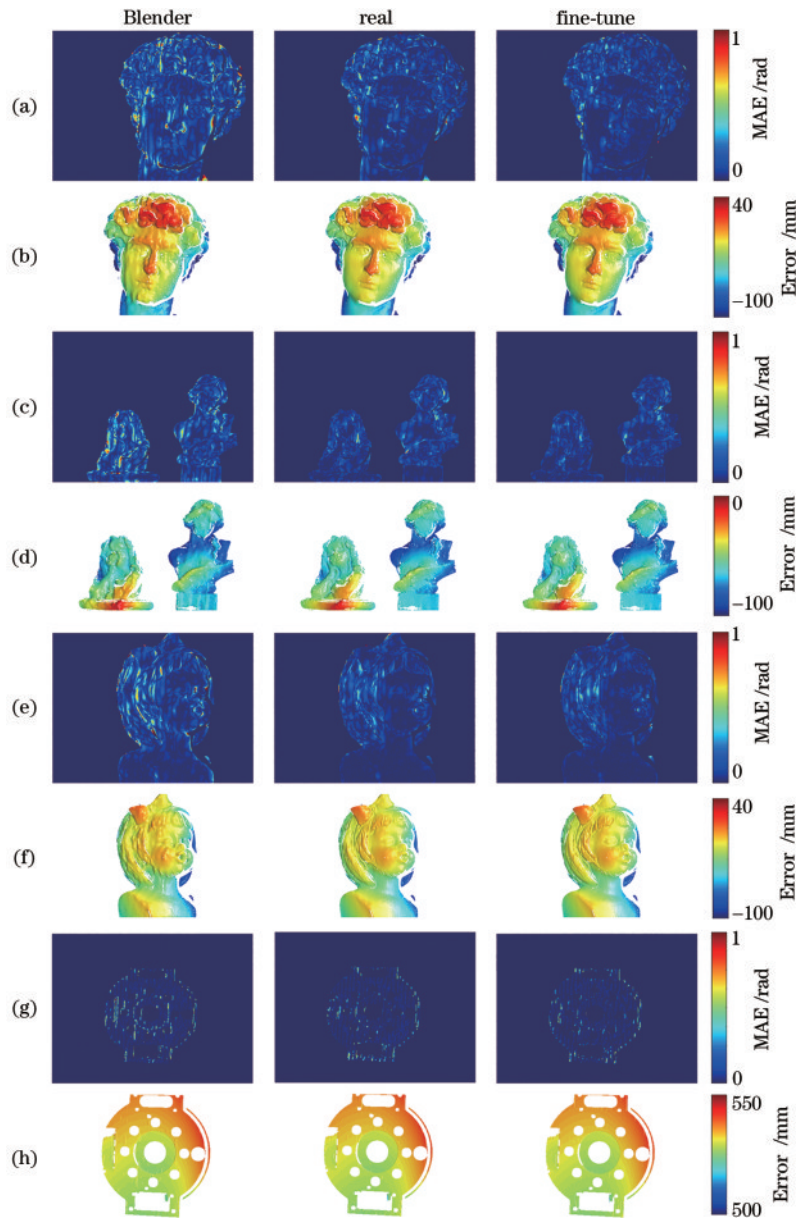


图 9 相位误差与三维重建结果对比  
 Fig. 9 Comparison of phase error and 3D reconstruction results

对于不同反射特性的塑料表面和金属表面,利用所提方法都可以实现高精度的相位测量和三维重建。需要注意的是,实验中预训练生成的条纹图像与拍摄的真实数据均为不饱和的光栅图像,即不存在高光情况。所以,如果测试表面的反射率过高,比如存在高光时,所提方法将难以处理这种表面。在后续的研究中,会进一步探索利用迁移学习处理高反光物体的方法。

关于畸变对于迁移学习的影响,在基于Blender的数字孪生条纹投影系统建模中,所提方法并未考虑投影成像畸变。但在上述实验中,构建的真实条纹投影三维成像系统其实存在畸变。利用张正友标定法,在

对系统进行标定后,得到了相机的畸变参数。其中, $k_1 \sim k_3$ 为径向畸变, $p_1 \sim p_2$ 为切向畸变。由于投影仪可被当成一个“反相机”处理,因此利用相同的标定方法,可得到投影仪的畸变参数,如表4所示。需要指出的是,对于拍摄的真实光栅图像,并没有进行图像畸变矫正的处理。从图9可以看出,当使用Blender生成的无畸变数据集训练神经网络时,由于神经网络没有见过带有畸变的图像数据,因此该模型在处理真实场景时,测量误差较大。但当使用了迁移学习方法后,神经网络对于图像畸变有了较好的适应,神经网络的相位预测精度有了明显提升。

表4 系统畸变  
Table 4 Distortion of the system

Module	$k_1$	$k_2$	$k_3$	$p_1$	$p_2$
Camera	-0.0843	0.7848	0	0.0016	$-4.9974 \times 10^{-4}$
Projector	-0.0030	-0.2706	0	$1.5148 \times 10^{-4}$	$-9.9825 \times 10^{-4}$

## 4 结 论

提出一种基于数字孪生结构光照明系统与迁移学习的条纹图像分析方法。首先,利用Blender对真实条纹投影系统进行模拟,建立相对应的虚拟数字孪生条纹投影系统。其次,利用该数字孪生系统开展虚拟扫描,运用域随机化技术随机改变虚拟照明光栅的属性,生成大量、不同分布的仿真光栅训练数据对U-Net神经网络进行预训练。最后,利用真实条纹投影系统拍摄少量的真实条纹图像数据,对预训练的U-Net神经网络进行微调。在对U-Net神经网络进行微调时分析和对比了“从左至右”“从上至下”“全局微调”等3种策略,发现对于条纹分析的任务而言,微调除“瓶颈”(最底层)网络模块外的所有其他模块可带来最佳的相位预测精度。同时,这几种微调方法具有一定的普适性。针对与U-Net结构相似的网络,可采取上述的这些微调策略进行尝试,通过实验结果进一步确定最佳的微调策略。针对与U-Net结构不同的顺序连接神经网络,可采用“从左至右”的策略,通过比较冻结深层或者浅层的实验结果找到最佳的微调方式。实验结果表明,相比于使用大量真实数据训练得到的神经网络,所提方法仅使用20%的真实数据就能达到相当的相位解调精度,其对使用少量真实数据获得高精度相位解调研究具有重要意义。

## 参 考 文 献

[1] 丁少闻, 张小虎, 于起峰, 等. 非接触式三维重建测量方法综述[J]. 激光与光电子学进展, 2017, 54(7): 070003.  
Ding S W, Zhang X H, Yu Q F, et al. Overview of non-contact 3D reconstruction measurement methods[J]. Laser & Optoelectronics Progress, 2017, 54(7): 070003.

[2] 尹维, 李明雨, 胡岩, 等. 基于VCSEL投影阵列的散斑

结构光三维成像技术及其传感器设计[J]. 激光与光电子学进展, 2023, 60(8): 0811014.

Yin W, Li M Y, Hu Y, et al. Speckle structured-light-based three-dimensional imaging technology and its sensor design using VCSEL projection array[J]. Laser & Optoelectronics Progress, 2023, 60(8): 0811017.

- [3] Leach R. Optical measurement of surface topography[M]. Berlin: Springer Berlin Heidelberg, 2011.
- [4] Lazaros N, Sirakoulis G C, Gasteratos A. Review of stereo vision algorithms: from software to hardware[J]. International Journal of Optomechatronics, 2008, 2(4): 435-462.
- [5] 吴周杰, 张启灿. 基于条纹投影的高速三维形貌测量技术发展综述[J]. 激光与光电子学进展, 2023, 60(8): 0811001.  
Wu Z J, Zhang Q C. High-speed 3D topography measurement based on fringe projection: a review[J]. Laser & Optoelectronics Progress, 2023, 60(8): 0811001.
- [6] Xu J, Zhang S. Status, challenges, and future perspectives of fringe projection profilometry[J]. Optics and Lasers in Engineering, 2020, 135: 106193.
- [7] Zhang Z H. Review of single-shot 3D shape measurement by phase calculation-based fringe projection techniques[J]. Optics and Lasers in Engineering, 2012, 50(8): 1097-1106.
- [8] Wu Z J, Guo W B, Zhang Q C. High-speed three-dimensional shape measurement based on shifting Gray-code light[J]. Optics Express, 2019, 27(16): 22631-22644.
- [9] 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.
- [10] Feng S J, Zuo C, Tao T Y, et al. Robust dynamic 3-D measurements with motion-compensated phase-shifting profilometry[J]. Optics and Lasers in Engineering, 2018, 103: 127-138.
- [11] Takeda M, Mutoh K. Fourier transform profilometry for the automatic measurement of 3-D object shapes[J]. Applied Optics, 1983, 22(24): 3977-3982.

- [12] Qian K M. Two-dimensional windowed Fourier transform for fringe pattern analysis: principles, applications and implementations[J]. *Optics and Lasers in Engineering*, 2007, 45(2): 304-317.
- [13] Barbastathis G, Ozcan A, Situ G H. On the use of deep learning for computational imaging[J]. *Optica*, 2019, 6(8): 921-943.
- [14] Wang F, Bian Y M, Wang H C, et al. Phase imaging with an untrained neural network[J]. *Light: Science & Applications*, 2020, 9: 77.
- [15] Xue Y J, Cheng S Y, Li Y Z, et al. Reliable deep-learning-based phase imaging with uncertainty quantification[J]. *Optica*, 2019, 6(5): 618-629.
- [16] Zuo C, Qian J M, Feng S J, et al. Deep learning in optical metrology: a review[J]. *Light: Science & Applications*, 2022, 11: 39.
- [17] Feng S, Chen Q, Gu G, et al. Fringe pattern analysis using deep learning[J]. *Advanced Photonics*, 2019, 1(2): 025001.
- [18] Feng S J, Xiao Y L, Yin W, et al. Fringe-pattern analysis with ensemble deep learning[J]. *Advanced Photonics Nexus*, 2023, 2(3): 036010.
- [19] Feng S J, Zuo C, Hu Y, et al. Deep-learning-based fringe-pattern analysis with uncertainty estimation[J]. *Optica*, 2021, 8(12): 1507-1510.
- [20] Yin W, Che Y X, Li X S, et al. Physics-informed deep learning for fringe pattern analysis[J]. *Opto-Electronic Advances*: 230034[2023-11-12]. <https://www.oejournal.org//article/doi/10.29026/oea.2024.230034>.
- [21] Guo X M, Li Y X, Qian J M, et al. Unifying temporal phase unwrapping framework using deep learning[J]. *Optics Express*, 2023, 31(10): 16659-16675.
- [22] Li Y X, Qian J M, Feng S J, et al. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement [J]. *Opto-Electronic Advances*, 2022, 5(5): 210021.
- [23] Zheng Y, Wang S D, Li Q, et al. Fringe projection profilometry by conducting deep learning from its digital twin[J]. *Optics Express*, 2020, 28(24): 36568-36583.
- [24] Wang F Z, Wang C X, Guan Q Z. Single-shot fringe projection profilometry based on deep learning and computer graphics[J]. *Optics Express*, 2021, 29(6): 8024-8040.
- [25] Zhu X J, Zhang Z Z, Hou L P, et al. Light field structured light projection data generation with Blender [C]//2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), May 20-22, 2022, Changchun, China. New York: IEEE Press, 2022: 1249-1253.
- [26] Amiri M, Brooks R, Rivaz H. Fine-tuning U-net for ultrasound image segmentation: different layers, different outcomes[J]. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020, 67(12): 2510-2518.
- [27] Zhang S, Huang P S. Novel method for structured light system calibration[J]. *Optical Engineering*, 2006, 45(8): 083601.
- [28] Zhang Z. A flexible new technique for camera calibration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [29] Zhou Q N, Jacobson A. Thingi10K: a dataset of 10, 000 3D-printing models[EB/OL]. (2016-05-16)[2023-11-12]. <https://arxiv.org/abs/1605.04797>.
- [30] Tobin J, Fong R, Ray A, et al. Domain randomization for transferring deep neural networks from simulation to the real world[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 24-28, 2017, Vancouver, BC, Canada. New York: IEEE Press, 2017: 23-30.
- [31] Tremblay J, Prakash A, Acuna D, et al. Training deep networks with synthetic data: bridging the reality gap by domain randomization[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1082-10828.
- [32] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.