一种分组并行的轻量化实时微观三维形貌重建方法 *

闫涛[1,2,3], 高浩轩[1,2], 张江峰[1,2], 钱宇华[1], 张临垣[4]

[1](山西大学 大数据科学与产业研究院,山西 太原 030006)

[2](山西大学 计算机与信息技术学院,山西 太原 030006)

[3](哈尔滨工业大学 重庆研究院,重庆 401151)

[4](北京中钞钞券设计制版有限公司,北京 100070)

通讯作者: 钱宇华, E-mail: jinchengqyh@sxu.edu.cn

摘 要: 微观三维形貌重建作为精密制造领域生产制造的关键环节,其重建过程依赖于高分辨率稠密图像的采集.而面对复杂应用场景的高时效性需求,高分辨率稠密图像的输入会导致运算量与计算复杂度呈几何倍增加,无法实现高效率低延时的实时微观三维形貌重建.针对上述现状,本文提出一种分组并行的轻量级实时微观三维形貌重建方法 GPLWS-Net, GPLWS-Net 以 U 型网络为基础构造轻量化主干网络,以并行分组式查询加速三维形貌重建过程,并针对神经网络结构进行重参数化设计避免重建微观结构的精度损失.另外,为弥补现有微观三维重建数据集的缺失,本文公开了一组多聚焦微观三维重建数据集(Micro 3D),其标签数据利用多模态数据融合的方式获取场景高精度的三维结构.结果表明,本文提出的 GPLWS-Net 网络不仅可以保证重建精度,而且在三组公开数据集中相比于其他五类深度学习方法平均耗时降低 39.15%,在 Micro 3D 数据集中平均耗时降低 50.55%,能够实现复杂微观场景的实时三维形貌重建.

关键词: 微观三维形貌重建;轻量化神经网络;分组并行

中图法分类号: TP391

中文引用格式: 闫涛,高浩轩,张江峰,钱宇华,张临垣.一种分组并行的轻量化实时微观三维形貌重建方法.软件学报. http://www.jos.org.cn/1000-9825/7013.htm

英文引用格式: Yan T, Gao HX, Zhang JF, Qian YH, Zhang LY. A grouping parallel lightweight real-time microscopic 3D shape reconstruction method. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7013.htm

A grouping parallel lightweight real-time microscopic 3D shape reconstruction method

YAN Tao[1,2,3], GAO Hao-Xuan[1,2], ZHANG Jiang-Feng[1,2], QIAN Yu-Hua[1], ZHANG Lin-Yuan[4]

[1](Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China)

[2](School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)

[3](Chongqing Research Institute of Harbin Institute of Technology, Harbin Institute of Technology, Chongqing 401151, China)

[4](Beijing Zhongchao Banknote Designing and Plate-making Co., Ltd., Beijing 100070, China)

Abstract: Microscopic three-dimensional (3D) shape reconstruction is a crucial step in the field of precision manufacturing. The reconstruction process relies on the acquisition of high-resolution and dense images. However, in the face of high efficiency requirements in complex application scenarios, inputting high-resolution dense images will result in geometrically increased computation and complexity, making it difficult to achieve efficient and low-latency real-time microscopic 3D shape reconstruction. In response to this situation, this paper proposes a grouping parallelism lightweight real-time microscopic 3D shape reconstruction method GPLWS-Net. The

* 基金项目: 国家自然科学基金重点项目(62136005); 科技创新 2030-重大项目(2021ZD0112400); 国家自然科学基金(62006146); 中央引导地方科技发展资金项目(YDZJSX20231C001, YDZJSX20231B001)

收稿时间: 2023-05-14; 修改时间: 2023-07-07; 采用时间: 2023-08-24; jos 在线出版时间: 2023-09-11

GPLWS-Net constructs a lightweight backbone network based on a U-shaped network and accelerates the 3D shape reconstruction process with parallel group-querying. In addition, the neural network structure is re-parameterized to avoid the accuracy loss of reconstructing the microstructure. Furthermore, to supplement the lack of existing microscopic 3D reconstruction datasets, this article publicly releases a set of multi-focus microscopic 3D reconstruction dataset called Micro 3D. The label data uses multi-modal data fusion to obtain a high-precision 3D structure of the scene. The results show that the GPLWS-Net network can not only guarantee the reconstruction accuracy, but also reduce the average time of 39.15% in the three groups of public datasets and 50.55% in the Micro 3D dataset compared with the other five types of deep learning-based methods, which can achieve real-time 3D shape reconstruction of complex microscopic scenes.

Key words: microscopic 3D shape reconstruction; lightweight neural network; group parallelism

1 引言

微观三维形貌重建作为三维重建领域的重要分支,广泛应用于精密制造质量控制、新材料结构分析、生物观测鉴别等领域[1].现有的微观三维形貌重建方法包括主动光学与被动光学两大类,典型的主动光学方法包括激光共聚焦与白光干涉等,但这类方法需要昂贵的硬件设备支撑,难以进行大规模工业应用.被动光学以多聚焦图像三维形貌重建为代表,主要通过微米级光学成像技术从多聚焦图像序列中恢复场景的三维结构,较高的重建效率与较低的硬件成本使其广受学术与工业界关注[2].

现有的多聚焦图像三维形貌重建主要分为模型设计与数据驱动两大类[3],模型设计类方法旨在通过设计聚焦测量算子评价图像序列的聚焦水平,然后选择图像序列中聚焦水平最大值所在帧聚合为场景的深度信息.因此聚焦测量算子设计的优劣是决定模型类设计方法是否有效的关键,而现有的聚焦测量算子更擅于解决富纹理场景的重建问题,无法实现弱纹理或低对比度场景的精确重建,其场景偏向性导致模型设计类方法普遍缺乏良好的场景适应性.数据驱动类方法以基于深度学习的多聚焦图像三维形貌重建为代表,可直接通过多聚焦图像序列学习得到场景的深度信息[4].但现有的深度学习类方法主要围绕宏观场景展开,由于宏观场景通常具有低分辨率与稀疏采样的特点,加之这类场景的数据规模较小,针对这类深度网络模型的研究通常难以解决微观场景高分辨率稠密数据产生的计算负担和受限资源条件下网络推理时间增多等问题.

现阶段,构建更深更大的卷积神经网络(CNNs)逐渐成为多聚焦图像三维形貌重建领域的发展趋势[5].目前主流的深度网络模型通常有上百层卷积操作和数千个通道进行运算,这些网络的运算量(FLOPs)通常达到数百万甚至几千万次,从输入图像序列到三维结构的一次推理过程往往需要较长时间.图 1 为五种先进的深度学习多聚焦图像三维形貌重建算法 FVNet(2022/CVPR)[5] ,DFVNet(2022/CVPR)[5],DDFF(2018/ACCV)[6],DefocusNet(2020/CVPR)[7]和 AiFDepthNet(2021/ICCV)[8]分别在 128×128×10,256×256×10,512×512×10 与 1024×1024×10 四种不同尺度的输入数据中运算耗时比较.由图 1 可知,上述所有方法的推理耗时均随着输入数据量的增加而增多,这种高耗时导致其在解决高分辨率稠密数据的微观场景重建问题时会出现推理时间增大与计算复杂度增加等问题.因此迫切需要从网络模型的轻量化角度探索实时微观三维形貌重建新模型.

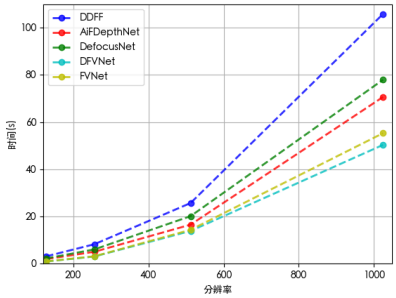

图 1 五种典型的深度学习多聚焦图像三维形貌重建算法在不同尺度输入数据中的运算耗时结果

现有轻量化网络如 MobileNets[9],ShuffleNet[10]和 GhostNet[11]等通常利用深度可分离卷积 DWConv(depthwise convolution)或者分组卷积 GConv(group convolution)来降低网络模型的计算复杂度.除上述纯卷积

神经网络之外,许多研究也开始设计更快更小的 ViT(vision transformer)和多层感知机 MLP(multilayer perceptron)架构[12]降低网络的计算复杂度.然而现有的轻量化网络大多基于二维图像问题设计,如将其直接扩展至三维场景,不仅会增加计算负担,而且也无法有效利用多聚焦图像序列间特有的邻域序列关联关系.

综上可知,现有的基于深度网络的多聚焦图像三维形貌重建主要侧重于宏观场景,较小的输入数据量使其更加关注网络模型设计的有效性.而轻量级网络大多侧重于二维图像问题的设计,并没有对三维数据进行针对性优化.除此之外,宏观场景数据具有典型的物体遮挡和大景深特性,与微观场景的缓慢过渡与小景深存在一定的数据鸿沟.因此现有的深度网络设计模式在解决实时微观三维形貌重建问题主要面临如下挑战:

(1) 数据量陡增:已有的基于深度网络模型的多聚焦图像三维重建方法会随着输入图像序列数据量的增加而导致模型的推理时间显著增大,无法满足实际微观场景中的高时效性需求;

(2) 模型不适用:现有的轻量级网络大多针对二维图像问题设计,而三维数据需要更多的计算资源与时间,导致模型无法有效兼顾低延时与高精度,且现有轻量级网络无法有效利用多聚焦图像序列间的关联关系;

(3) 宏微观鸿沟:现有深度网络模型大多采用宏观场景中的合成数据集进行训练,加之宏微观数据内蕴结构的差异性,导致采用这类合成数据训练的网络会出现过拟合,无法准确推断微观场景的三维结构变化.

针对上述挑战,本文提出一种分组并行的轻量化实时微观三维形貌重建网络模型 GPLWS-Net,主要贡献如下:

(1) 从神经网络各组件时间能耗的角度,重现审视现有微观三维形貌重建网络的性能瓶颈问题,提出轻量化、低延迟的网络主干;

(2) 从多域并行处理多聚焦特征的角度,设计与多聚焦图像序列三维形貌重建理论相契合的分组并行模块,并采用结构重参数化进行模型压缩,将原有多卷积层恒等映射为单卷积层,保持三维形貌精度的同时有效降低网络推理延迟;

(3) 针对微观三维场景数据匮乏的现状,公开了一组微细加工场景的微观三维数据集(Micro 3D).该数据集标签采用“激光共聚焦+多景深合成+手工微调”等方式生成,弥补了现有微观领域数据集缺乏的不足.

本文第 2 节主要介绍了多聚焦图像三维形貌重建方法与轻量化网络模型的研究进展;第 3 节提出了基于分组并行的轻量级实时微观三维形貌重建方法 GPLWS-Net;第 4 节与现有深度学习类方法和模型设计类方法在公用数据集和无标签真实数据集中进行比较分析;最后对本研究进行总结和展望.

2 相关工作

多聚焦图像三维形貌重建通过等间隔调整相机与待测场景之间的焦距,获取可以覆盖场景全部景深范围的多幅不同焦距的图像序列 $\{X_i\}_{i=1}^{N}$,采用聚焦测量算子 FM(focus measure)评价图像序列中各图像的聚焦水平,然后将同一区域聚焦水平最大值所在位置聚合为场景的初始深度 D_{Init},最后采用迭代修复、正则化等后处理方法对初始深度图进行精炼得到场景最终的三维形貌重建结果 D [13-14].

$$D = P\left(D_{Init}\right), D_{Init} = \arg\max_{1\le i\le N}\left\{FM * X_i\right\}_{i=1}^{N} \tag{1}$$

其中 X_i 为图像序列中第 i 幅图像, N 为图像序列总数, $P(\cdot)$ 为后处理函数.

2.1 多聚焦图像三维形貌重建研究进展

多聚焦图像三维形貌重建主要分为模型设计与数据驱动两大类方法.模型设计类方法主要围绕图像序列的聚焦评价与深度图精炼两个关键步骤展开,其中聚焦评价旨在通过设计聚焦测量算子评价一幅图像的聚焦水平,然后延伸至整个图像序列,进而获得场景的初始深度图.这些聚焦测量算子大体可以分为时域和频域两大类.时域类算子主要侧重局部图像聚焦水平的刻画,代表性方法有环状差分算子 RDF(ring difference Filter) [15],多方向拉普拉斯算子 MDML(multidirectional modified Laplacian) [16]等;频域类算子更加关注图像的全局聚焦信息,典型的频域类算子有非降采样小波变换[17]与 Curvelet 变换[18]等.深度图精炼主要通过对初始深度图添加

约束条件改善算法的重建效果,如非凸正则优化[19]、数据保真项[20]等.然而模型设计类方法在聚焦测量算子的设计过程中存在一定的场景偏向性,无法保障算法对未知场景的鲁棒性.除此之外,深度图精炼过分依赖于初始深度图的质量,低质量初始深度图在精炼过程中容易引发错误深度信息蔓延.因此,以深度学习为代表的数据驱动类方法逐渐引起学者们的关注.

近年来,已有一些研究从深度网络模型构建角度解决多聚焦图像三维形貌重建问题.但这类方法属于典型的有监督学习,模型的性能依赖于数据集本身.如 Yang 等[5]提出一种基于差分体积的聚焦和散焦网络 FVNet 和 DFVNet,该网络主要模拟模型设计类方法的聚焦评价过程;Hazirbas 等[6]提出一种深度卷积神经网络 DDFF,该网络利用光场和RGB-D相机对室内场景进行数据采集,构建了DDFF-12数据集,并对场景的聚焦信息和深度信息进行端到端学习;Wang 等[8]利用深度图像和全聚焦图像之间的关联关系设计了一个可共享的卷积神经网络 AiFDepthNet,该网络引入一个可以被共享的中间注意力图,用于预测场景深度和全聚焦图像;Maximov 等[7]提出一种利用散焦图像训练的聚焦与散焦对齐网络 DefocusNet.尽管上述网络模型为深度学习类多聚焦图像三维形貌重建提供一些有益的思路,但在解决实时微观三维形貌重建问题时需要考虑如下问题.首先,随着输入图像序列分辨率的提升,网络处理数据量的倍增会导致收敛速度变慢;其次,上述网络训练的数据集主要集中在宏观场景,且多数训练集为合成数据,基于这类数据集设计的网络可能无法有效刻画微观场景中缓慢的深度变化与噪声干扰等情况.因此,如何针对微观场景特有的数据特点设计轻量化网络模型是解决实时微观三维形貌重建问题的关键.

2.2 轻量化神经网络相关研究进展

近年来,随着深度神经网络在计算机视觉领域取得巨大成功,越来越多场景提出了智能化应用需求,然而在实际的资源受限应用场景中通常无法满足神经网络的算力需求.为权衡神经网络的精度与性能,轻量化神经网络应运而生.如 ShuffleNet[10]、SqueezeNet[21]与 MobileViT[12]主要是对网络模型的参数进行优化;而 MobileNets[9]、MobileNeXt[22]、GhostNet[11]、Xception[23]和 IGCNets[24]等模型则侧重于优化 FLOPs;EfficientNet[25]和 TinyNet[26]在优化 FLOPs 的同时研究了网络的深度、宽度和输入图像分辨率的复合缩放;仅有少数网络如 ShuffleNetV2[27]、MobileNetV3[28]、FasterNet[29]和 MobileOne[30]等对网络推理时间进行优化,ShuffleNetV2[27]表明 FLOPs 和网络参数量与网络推理时间并没有呈现很好的相关性,MobileOne[30]则发现推理时间与 FLOPs 适度相关,与参数量弱相关.针对轻量化 ViT 的研究主要试图通过减少注意力操作的复杂度实现网络精度与推理时间的平衡.如 MobileFormer[31]和 MobileViT[12]针对参数和 FLOPs 进行优化,其表现已经超越了低 FLOPs 的高效卷积神经网络,尽管这些模型在精度上取得了显著提升,但推理时间并未随之缩短.因此,仅仅拥有低的 FLOPs 并不能必然导致推理时间的降低.

综上可知,现有的轻量级神经网络大多针对二维图像任务设计,而对于三维数据而言,更高的输入数据量可能导致网络的计算量成倍增加.因此,从参数量优化和 FLOPs 优化的视角并不能有效降低网络的推理时间和增加推理精度,需要从网络设计的全链条环节并结合三维数据特有的邻域序列关系重新进行轻量化网络模型的设计.本文首先从理论上分析多聚焦图像序列子域数据分组并行的可行性,并根据该理论设计了分组并行模块,可有效提升深度信息的寻找过程;其次摒弃原有二维卷积提取单帧图像局部聚焦特征的操作,转为三维立体卷积精确跟踪多聚焦图像序列间的差异性,进而充分利用图像序列间的邻域关系实现高可靠性的深度信息判断;最后采用结构重参数方法将三分支稀疏特征提取矩阵变为单分支密集特征提取矩阵,加速网络三维结构预测.

3 GPLWS-Net:基于分组并行的轻量化实时微观三维形貌重建方法

3.1 多聚焦图像序列的子域数据分组并行理论分析

基于多聚焦图像序列的微观三维形貌重建方法利用光学设备成像过程的有限景深判定聚焦区域与相机间的相对距离,进而还原待测场景的三维形貌.根据高斯成像公式可得,物距的倒数与像距的倒数之和等于焦距的倒数:

$$\frac{1}{f}=\frac{1}{u}+\frac{1}{v} \tag{2}$$

其中 f 表示镜头焦距,u 表示物体到透镜的距离,v 表示成像到透镜的距离.

上述理论表明多聚焦图像三维形貌重建的本质是从一系列不同聚焦水平的图像序列中挖掘深度信息,其中单帧图像中的聚焦与散焦变换可以通过全聚焦图像的滤波得到.因此,局部聚焦图像 I_L 可以通过全聚焦图像 I 与点扩散函数 h 的卷积得到,

$$I_L = I * h \tag{3}$$

其中点扩散函数在光学成像模式中可以简化成如下高斯函数:

$$h(i,j)=\frac{1}{2\pi\sigma_h}\exp\left(-\frac{i^2+j^2}{2\sigma_h}\right) \tag{4}$$

其中 σ_h 用来刻画一幅图像的模糊水平,研究表明在 $I_{i,j}$ 位置的模糊水平 σ_h 与场景深度 u 存在如下关系[32]:

$$\sigma_h(i,j)=\frac{\kappa}{A}\frac{f^2|u-u_f|}{u(u_f-f)} \tag{5}$$

其中,u_f 为相机设置的聚焦位置,κ 为相机参数,A 为相机焦距与透镜直径的比值.假定 $M=\kappa f^2/A(u_f-f)$,对公式(5)求导可得:

$$\frac{\partial\sigma_h(i,j)}{\partial u}=M\frac{|u-u_f|}{u}=\begin{cases}\dfrac{u_f}{u}-1 & u<u_f\\ 1-\dfrac{u_f}{u} & u\geq u_f\end{cases} \tag{6}$$

图 2 为模糊水平 σ_h 对场景深度 u 的一阶导数曲线,可以表明模糊水平 σ_h 为连续函数.

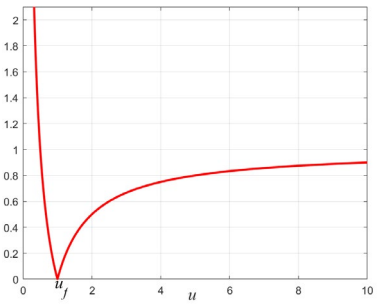

图 2 模糊水平 σ_h 对场景深度 u 的一阶导数曲线

根据连续函数的最大值定理可知,待测场景某点的深度对应于多聚焦图像序列内该点聚集的最大值,并根据模糊水平 σ_h 的一阶导数曲线可知该点有且仅有一处.而现有三维形貌重建大都从全局视角出发,通过依次遍历图像序列搜索全局聚焦信息的最大值,未能通过子域划分进行并行检索,导致出现运行效率瓶颈.

3.2 总体框架

本文提出的基于分组并行的轻量化实时微观三维形貌重建方法主要包括以下 4 个关键环节:

(1) 微观数据集生成:鉴于现有研究普遍缺乏微观场景数据集,通过自研的微米级超景深微观数据采集装置实现微观场景多聚焦图像序列的采集,采用 3D TFT 算法[3]得到场景初步的三维重建结果,联合亚微米级精度

的激光共聚焦显微镜得到场景的三维结构信息.采用 ALI-Net[33]对上述两类多模态重建结果进行配准,最后通过人工筛选微调获得高精度的标签数据.

(2) 轻量网络主干:为降低深度神经网络的内存访问成本和计算开销,利用网络结构设计的优化和网络模型之间的参数融合等原理对已有的 U 型网络架构进行重构,在减少网络运算的内存访问频次的同时保障其性能表现,从而实现更低的计算开销.

(3) 分组并行加速:为了最大程度提高图像序列的处理速度和效率,使用分组卷积对多聚焦图像序列中的图像进行特征处理以找出局部极大值,然后再利用更为高效 1×1 卷积进行全局特征处理找到极值点,分组卷积操作具有的可并行特性可有效提升计算效率,进而减少整个操作的时间成本.

(4) 结构重参数化:由于神经网络中存在大量冗余参数,这些参数可以融合为新参数进而赋予新的结构,以提高网络的效率.因此本文使用结构重参数化对网络架构从训练到推理进行解耦合,在不影响网络精度的同时降低推理时间.

本文提出的基于分组并行的轻量化实时微观三维形貌重建方法示意图如图 3 所示.

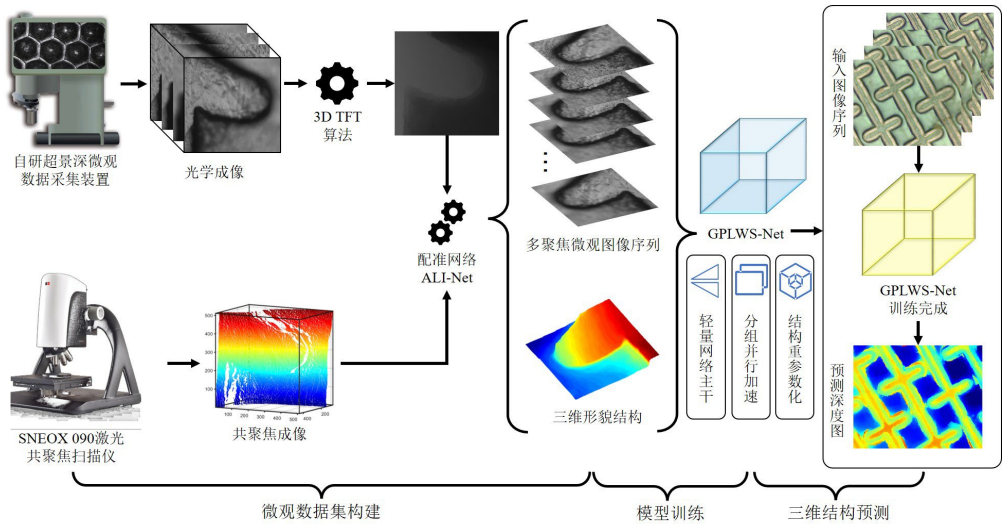

图 3 基于分组并行的轻量化实时微观三维形貌重建方法示意图

3.3 GPLWS-Net轻量化网络模型

受到 U 型网络的高效率特征融合方式和低成本参数运算的启发,本文以 U-Net 作为网络主干基础. 由于相同聚焦设置下不同尺度的图像序列对于同一聚焦算子的敏感度不同,其关键在于聚焦算子对于像素信息的抽取会伴随图像尺度的改变发生变化. 而 U-Net 网络主干通过多次下采样操作降低特征图分辨率,可获得多聚焦图像序列不同尺度的特征. U 型网络中收缩模块和扩张模块在特征信息尺寸和神经网络宽度方面保持相互对称,在收缩模块中逐步缩小特征信息的尺度并增加特征信息的维度,而收缩模块恰好相反, 二者间使用跳跃连接(skip connection)有效耦合图像的表层特征和深层特征. 而跳跃连接在神经网络中具有的缺点对于网络的轻量化设计是不容忽视的:首先跳跃连接会增加网络参数量,导致模型的复杂性和训练难度增加;其次,跳跃连接需要额外的计算操作处理连接路径,增加了计算资源的需求和时间成本;此外,如果底层特征和上层特征存在冗余,跳跃连接可能会引入特征冗余.而残差连接通过将前一层的输入添加至后续层,有效缓解了梯度消失问题,使得整个网络更易于优化和训练.其次残差连接加速了网络优化过程,使得网络更快收敛,有效减少了训练时间和算力需求.此外,残差连接允许底层特征直接传递至较深层并进行特征融合的操作可有效提升网络性能,增强对数据细节信息和复杂关系的捕捉能力.综上所述,残差连接不仅可有效改善神经网络的训练优化过程,而且能够在提高模型性能的同时减少训练时间和计算资源需求,非常有利于轻量化网络模型的构建.因此,本文设计的网络主干将跳跃连接改为残差连接, 既能有效耦合前后特征信息的梯度传递, 同时也能降低网络参数量. 图 4

为本文构建的 GPLWS-Net 网络结构示意图.

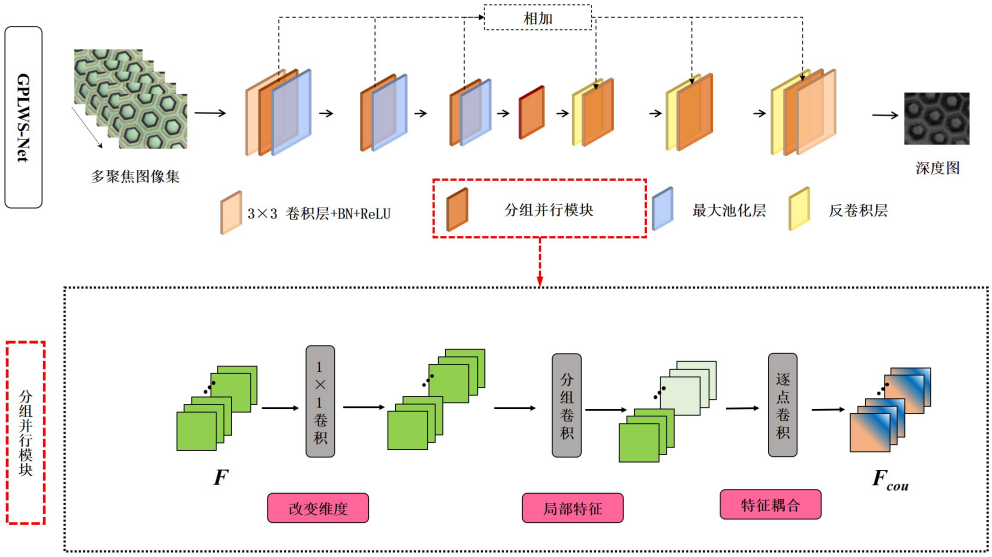

图 4 GPLWS-Net 网络模型结构

3.4 分组并行模块

基于多聚焦图像序列的三维形貌重建方法主要通过光学成像设备对场景进行等间隔扫描,由于场景中各点深度信息的唯一性可得出聚焦最大值点存在唯一性.分组卷积的子域并行特性完美契合聚焦判定过程中局部极值点导出全局极值点的特性.本文重新审视分组卷积的固有特性,验证多聚焦图像单峰时序信号中分组卷积的可用性和高效性.

该模块借鉴深度可分离卷积[34],依据聚焦曲线特性进行重新设计,用于贴合三维形貌重建过程.深度可分离卷积作为常规卷积的流行变体,其核心在于通过拆分空间维度和通道维度的相关性,减少卷积计算所需的参数.深度可分离卷积由两层卷积代替原有普通卷积层,其中包含对单通道数据进行特征提取的逐深度卷积(depthwise convolution)和对单维度进行特征融合的逐点卷积(pointwise convolution).对于输入序列 $I\in\mathbb{R}^{C\times H\times W}$ (C,H,W 分别为通道,图像高和宽),深度卷积对每个输入通道应用单个卷积核 $W\in\mathbb{R}^{k\times k}$ (k 为卷积核大小)计算输出的序列 $O\in\mathbb{R}^{C\times H\times W}$,然后逐点卷积应用 1×1 的卷积核将深度卷积的输出进行线性组合.与具有运算量的常规卷积相比,深度可分离卷积的运算量低至 $k^2\times c\times H\times W+H\times W\times c^2$,约为常规卷积的 $1/k^2$.尽管深度可分离卷积显著提高了模型的运行速度和减少了计算成本,但仍面临分组过多导致内存访问量过大与单层特征抽取等问题.

本文将原有的逐通道卷积改为分组卷积平衡精度与效率,可有效契合三维形貌重建过程聚焦最大值的并行查找过程.此外,本文仍保留 1×1 卷积对全域聚焦最大值进行整合.研究表明引入通道重洗(channel shuffle)[11]可实现通道组之间的信息流不再受到限制(图 5b),进而有效提升特征的表示能力,但通道重洗操作也会带来相应的缺点:即需要大量的指针跳转和内存空间.同时通道重洗操作又特别依赖于实现细节,导致实际运行速度不理想.对于线性计算的卷积层来说,通道重洗操作使得网络无法进一步优化.而本文使用 1×1 卷积加强各通道间的信息流动(图 5c),可增强网络的特征表达能力.通道重洗仅能互换部分特征区域,而 1×1 卷积可有效选择全域聚焦信息.

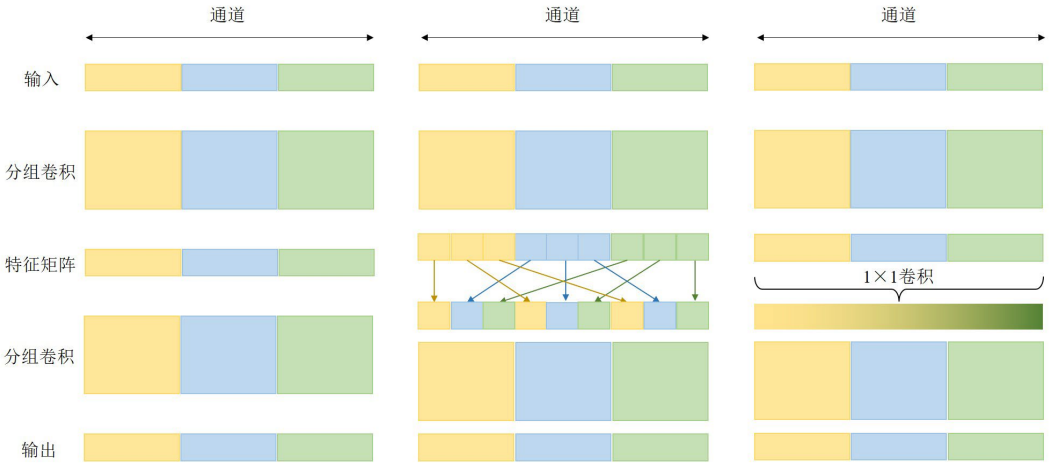

图 5 (a) 分组卷积 (b) 通道重洗卷积 (c) 逐点卷积

如图 6 所示,在分组并行模块中,首先将输入的特征张量按照原先设定的分组数进行等分;然后针对分组张量并行抽取特征,得到分组子域内聚集极大值;最后 1×1 卷积层将每组卷积的聚焦结果进行合并,并选择全域内聚焦最大值.相较于传统卷积操作,分组卷积将输入特征在通道维度进行划分以减少卷积提取区域,可降低模型训练及推理的时间.

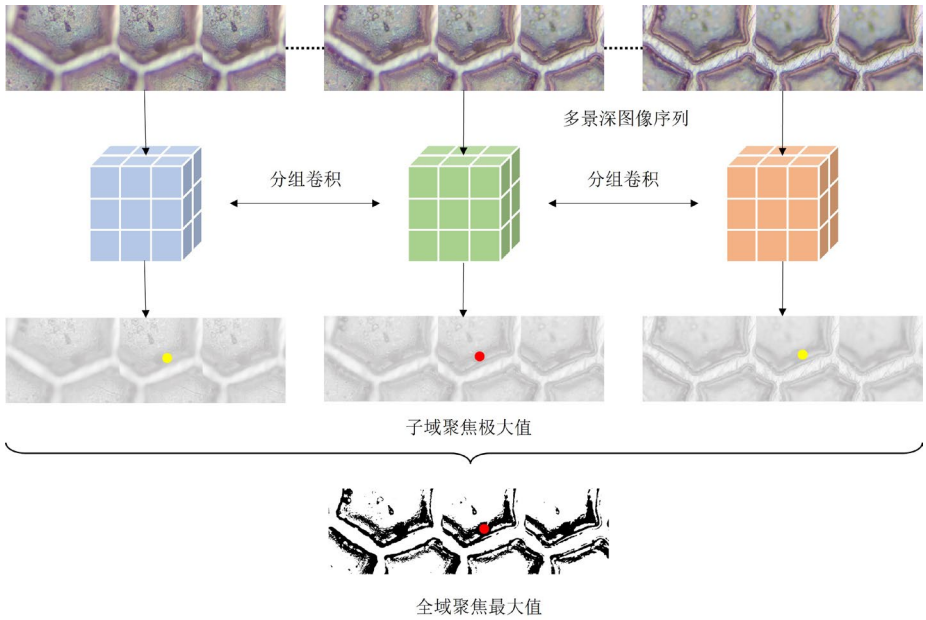

图 6 本文采用的分组并行策略

3.5 结构重参数化

基于神经网络的微观三维形貌重建模型使用更深更宽或多分支的卷积层抽取高维特征矩阵,可有效增加神经网络模型的特征表达能力.在三维形貌重建过程中,单分支串型结构易于陷入局部最优解,而多分支并行结构则额外增加内存访问成本.因此,本文在训练时采用多分支结构,推理时将多分支等价转换为单分支结构,通过网络结构重参数化实现优势互补.具体操作如下:对于卷积核大小为 K 的卷积层而言,输入通道的维度是 C_{in},

输出通道的维度是 C_{out},则卷积核的权重矩阵可以表示为 $W' \in \mathbb{R}^{C_{in} \times C_{out} \times K^2}$, 偏置项 $b' \in \mathbb{R}^{D}$.归一化层中包含累积得到的均值 μ,标准差 α 和学到的缩放因子 γ 以及偏置项 β.由于卷积层和归一化层均属于线性运算.因此在推理时可合并为一个新卷积操作.在新卷积层中,卷积核权重可以表示为 $\hat{W} = W'\gamma / \alpha$,偏置项为 $\hat{b} = (b - \mu)\gamma / \alpha$.神经网络中多分支结构组成的多角度特征空间有利于挖掘三维形貌重建过程的最大聚焦点.

4 实验分析

4.1 实验设置

本文 Micro 3D 数据集共采集 488 组 20 张大小为 256×256 的多景深微观图像序列,并通过聚焦形貌测量构建初始形貌、激光逐点扫描细化形貌和手工微修场景先验深度等步骤生成标准的三维形貌图作为监督信息,其中部分数据集形貌如图 7 所示.

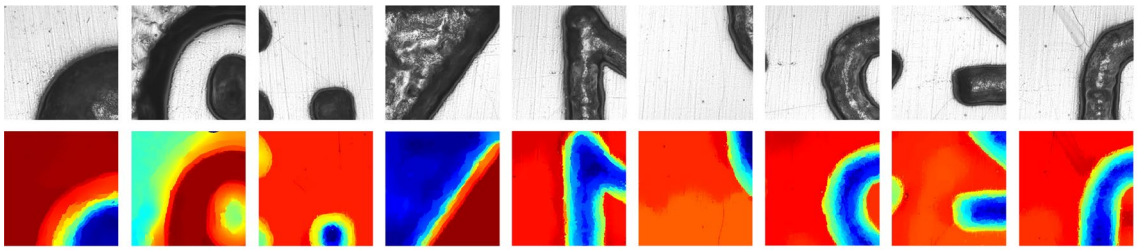

图 7 Micro 3D 数据集中部分数据

本文使用 382 组多聚焦微观图像序列训练网络模型,训练过程使用图像旋转、场景翻转、伽马变换和区域裁剪等数据增广方式.多聚焦图像序列作为网络模型的输入,三维形貌图作为网络模型的监督信息,在 106 组数据进行测试,共进行 720 次迭代训练.在训练过程中使用 Adam 优化器,初始学习率设置为 0.001,批处理大小设置为 4,其余参数皆为默认参数.本文使用 Pytorch 框架搭建网络模型,采用 TITAN XP GPU 训练并测试 GPLWS-Net 与其他网络模型的性能. Micro 3D 数据集和网络模型 GPLWS-Net 已公布在 GitHub: https://github.com/jiangfeng-Z/Multi-focus_Microscopic_3D_reconstruction_Datasets.

4.2 消融实验

为进一步验证本文 GPLWS-Net 网络模型结构的合理性,消融实验中主要使用公共数据集 4D Light Field[33] 对消融模型进行交叉验证,从模型性能(MSE)、模型参数和延迟进行多角度分析.为避免设备自身算力波动对实验结果的影响,消融实验在同一环境下重复验证 50 次并计算其均值.以下分别从残差连接与跳跃连接、分组并行模块与结构重参数化三个模块的有效性展开.

根据 U 型主干网络中相同尺度特征连接方式的不同,分别使用“拼接式”的跳跃连接和“相加式”的残差连接进行对比,表 1 结果表明:跳跃连接和残差连接两种连接方式并不影响模型的性能,但残差连接的网络延迟显著优于跳跃连接,由于跳跃连接之后网络变宽与卷积核参数增加,使得运算量激增导致网络预测时间增加.

表 1 残差连接与跳跃连接消融实验结果

连接方式	模型参数/个	延迟/s	MSE
跳跃连接	546321	2.419	0.0275
残差连接	**540273**	**0.437**	0.0275

分组并行模块的消融实验如表 2 所示,根据 U 型网络收缩模块和扩张模块的连接方式不同可分为“拼接式”和“相加式”,其中又分别将常规卷积、深度可分离卷积和分组卷积进行对比,因此共得到六组对比实验.消融模型 1 和消融模型 4 对比可知,相加连接方式不仅有助于提升模型精度,而且可降低网络预测延迟;消融模型 4 和消融模型 5 对比可知:深度可分离卷积尽管可以降低网络参数量,但由于分组数过多导致内存访问频繁,进而导致网络预测时长增加;消融模型 5 和消融模型 6 对比可知:分组并行模块仅牺牲部分存储空间,但可有效降低网络延迟,并保证网络预测精度.因此,本文采用消融模型 6 中分组并行模块和相加连接方式.

表 2 分组并行模块消融实验结果

消融模型	基础模块	模型参数/个	延迟/s	模型性能 MSE	连接方式
1	常规卷积	1213872	0.457	0.0516	
2	深度可分离卷积	200400	0.463	0.0473	拼接
3	分组并行模块	444048	0.399	0.0534	
4	常规卷积	754020	0.353	0.0468	
5	深度可分离卷积	**106344**	0.394	0.0479	相加
6	分组并行模块	260853	**0.350**	**0.0372**	

结构重参数化的消融实验结果如表 3 所示,使用与推理阶段相同配置的普通卷积层代替结构重参数化模块,结果表明在延迟不发生改变的情况下,使用结构化重参数的模型参数量更低,性能更佳.

表 3 结构重参数化与常规卷积消融实验结果

卷积方式	模型参数/个	延迟/s	MSE
结构重参数化	**540273**	0.437	**0.0275**
常规卷积	540936	0.437	0.0339

4.3 实验分析

本节根据应用场景的差异分为对比实验、泛化实验和延时实验三类,分别对不同三维形貌重建方法进行定量对比和定性分析,其中对比实验通过对公共数据集学习验证网络模型的有效性,泛化实验可验证本文数据集在微观领域提出的必要性,而延时实验则可明确对比各模型在实际应用场景的效率.

本节主要选择 4D Light Filed[33]、DefocusNet[7]、FlyingThing3D[33]、Middlebury[33]和本文的 Micro 3D 数据集进行测试.通过对比先进的模型设计类三维形貌重建模型 RDF[15]和 RR[19]与基于深度学习的三维形貌重建模型(DDFF[6]、DefocusNet[7]、AiFDepthNet[8]、FVNet[5]、DFVNet[5])评估本文提出数据集 Micro 3D 数据集和 GPLWS-Net 算法的性能.其中 4D Light Field 数据集中包含了 25 种复杂现实场景,主要用于测试精细结构与弱纹理及光滑表面等情形;DefocusNet 数据集构建浮点级深度信息,主要验证网络模型对于复杂场景的拟合情况;FlyingThing3D 数据集共公布 1000 组三维场景数据.

实验中使用均方误差 MSE(mean squared error)、平均绝对误差 MAE(mean absolute error)、均方根误差 RMSE(root mean squared error)、绝对相关系数 AbsRel(absolute relative error)、平方相对误差 SqRel(square relative error)、颠簸性(bumpiness)和推理时间(secs)指标定量评估 GPLWS-Net 与其他三维形貌重建模型的性能.

4.3.1 对比实验分析

本节将 GPLWS-Net 与现有的深度学习类三维形貌重建方法进行对比,为确保实验结果的公正客观,本节在公共数据集 4D Light Field 与 DefocusNet 上进行网络模型的性能分析.具体定量指标对比如表 4 所示,其中部分数据来自论文[8,33].由表 4 可知,本文提出的 GPLWS-Net 模型在两类数据集上的重建精度显著优于其他算法.(注:性能最优用红色加粗字体标注,次优用蓝色加粗字体标注.)

表 4 4D Light Field 数据集和 DefocusNet 数据集的定量评价

Datasets/Metrics / Methods	4D Light Field[33]			DefocusNet[7]		
	MSE	RMSE	Bump	MSE	MAE	AbsRel
FVNet[5]	**0.0301**	**0.1537**	-	0.0189	-	0.1400
DFVNet[5]	0.0317	0.1549	-	0.0205	-	**0.1300**
DDFF[6]	0.1150	0.3310	2.95	0.0440	0.1312	0.3556
DefocusNet[7]	0.0593	0.2355	2.69	0.0175	0.0637	0.1386
AiFDepthNet[8]	0.0472	0.2014	**1.58**	**0.0127**	**0.0549**	**0.1115**
GPLWS-Net	**0.0275**	**0.1442**	**2.49**	**0.0106**	**0.0534**	0.1394

4D Light Field 可验证各模型对于精细结构的三维形貌重建.由图 8 所示:DDFF 模型和 DefocusNet 模型仅能分辨场景的相对深度,无法重建场景内的精细结构;AiFDepthNet 模型无法有效保持深度图的边缘细节,深度信息容易弥散;DFVNet 模型和 FVNet 模型对于场景内富纹理背景处理不佳.本文提出的 GPLWS-Net 模型在精细结构的表达和聚焦区域的判断方面表现良好,例如场景一的绳结和场景二的鞋身.

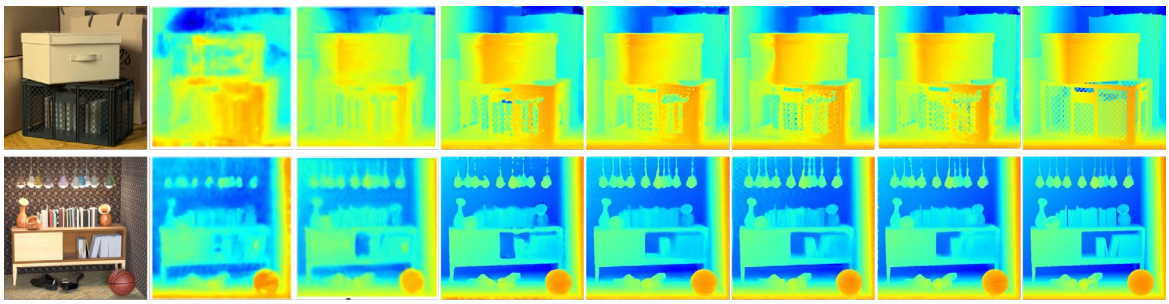

图 8 4D Light Field 数据集的定性比较

由表 5 可以看出,本文的 GPLWS-Net 模型在 FlyingThing3D 数据集上与 FVNet、DFVNet、DDFF 和 AiFDepthNet 相比在所有评价指标方面均有较大提升,其中 AiFDepthNet 在各项指标中也得到了第二优的结果.(注:由于 DefocusNet 未公布其训练模型和网络代码,因此下列表单中未列出该网络的定量评估结果以及预测图.)

表 5 FlyingThing3D 数据集定量评估

Dataset/Metrics / Methods	FlyingThings3D			
	MAE	RMSE	AbsRel	SqRel
FVNet[5]	27.442	35.406	2.388	78.952
DFVNet[5]	27.951	36.098	2.334	79.705
DDFF[6]	17.182	27.077	1.654	45.132
AiFDepthNet[8]	**6.838**	**12.247**	**0.666**	**8.788**
GPLWS-Net	**6.053**	**11.045**	**0.631**	**7.575**

FlyingThing3D 数据集可验证各模型对于复杂遮挡场景的三维形貌重建.由图 9 可知,DDFF 模型仅能分辨场景的前后关系,无法识别场景的语义关系,且易发生聚焦扭曲;DFVNet、FVNet 和 AiFDepthNet 模型可表达场景的语义关系,但易受到场景纹理信息的干扰,无法捕获更加丰富的层次信息.而本文的 GPLWS-Net 模型在三维形貌的边缘保持和遮挡重叠区域有较好的重建效果.

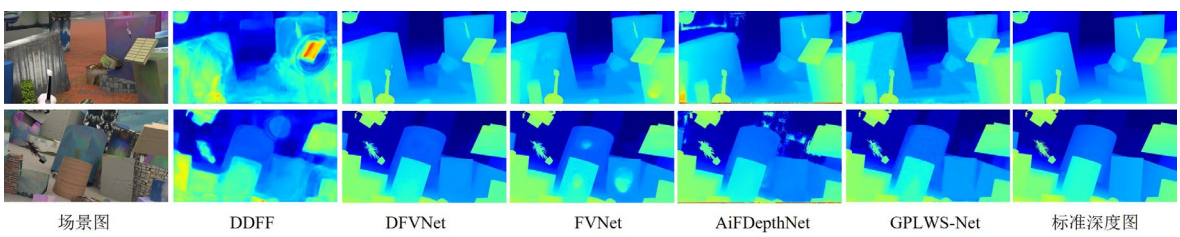

图 9 本文 GPLWS-Net 与其他四类网络在 FlyingThing3D 数据集上的深度预测

4.3.2 泛化实验分析

本节使用 FlyingThings3D 数据集训练本文 GPLWS-Net 与其他四种深度学习模型,并在 Middlebury 数据集、DefocusNet 数据集和 4D Light Field 数据集上进行测试,用以定量分析本文模型的泛化性能.如表 6 所示,除在 DefocusNet 数据集中的 AbsRel 和 SqRel 指标外,本文 GPLWS-Net 模型在三类测试集中的其他指标均能保持最优性能.

表 6 跨不同数据集的定量结果

模型	训练集	测试集	MAE	MSE	RMSE	AbsRel	SqRel
FVNet[5]	FlyingThings3D	Middlebury	11.276	177.649	13.147	0.360	5.011
DFVNet[5]			11.407	183.267	13.302	0.344	4.711
DDFF[6]			32.499	1480.444	37.544	1.197	52.169
AiFDepthNet[8]			**3.825**	**58.570**	**5.936**	**0.165**	**3.039**
GPLWS-Net			**2.539**	**17.497**	**4.062**	**0.100**	**0.642**

FVNet[5]	FlyingThings3D	DefocusNet	0.271	0.144	0.353	**0.555**	**0.198**
DFVNet[5]			0.271	0.152	0.360	**0.529**	**0.204**
DDFF[6]			89.351	9305.360	95.214	331.124	3357.152
AiFDepthNet[8]			**0.183**	**0.080**	**0.261**	0.725	0.404
GPLWS-Net			**0.126**	**0.053**	**0.234**	2.011	1.826
FVNet[5]	FlyingThings3D	4D Light Field	1.485	3.053	1.704	2.421	5.676
DFVNet[5]			1.352	2.432	1.552	1.780	3.377
DDFF[6]			94.106	9806.356	95.715	77.899	7464.454
AiFDepthNet[8]			**0.205**	**0.106**	**0.313**	**0.198**	**0.151**
GPLWS-Net			**0.021**	**0.042**	**0.036**	**0.010**	**0.036**

图 10 使用本文公开的 Micro 3D 数据集进行训练, 并在印辊微观场景(不同于本文的 Micro 3D 数据集)验证其泛化性,由此说明本文数据集的必要性和泛化性.其中 RR、AiFDepthNet 和 FVNet 方法易受离散噪声干扰,对于场景的微细纹理变化敏感; RDF、DDFF 和 DFVNet 方法都具备对噪声的抗干扰能力,但深度边缘保持不佳;而本文提出的 Micro 3D 数据集及 GPLWS-Net 模型可有效适应未知的微观场景,具有良好的抗噪性和鲁棒性.

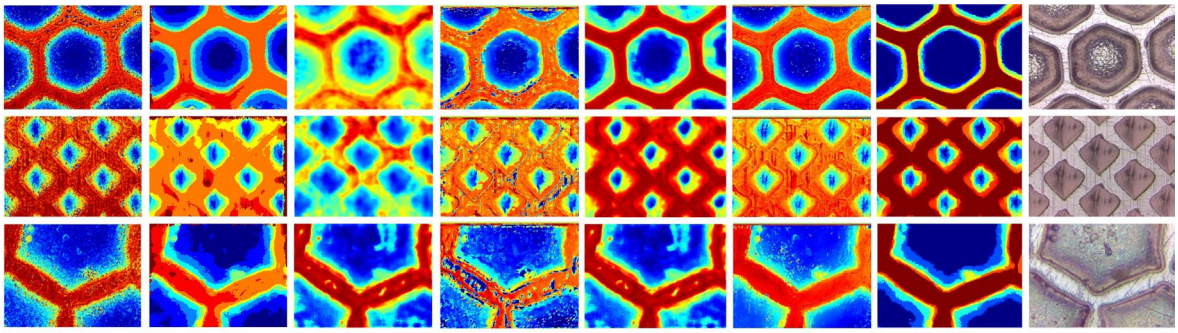

图 10 模型设计与深度学习类方法以及本文的 GPLWS-Net 在三组微观印辊数据集下的泛化性对比.

4.3.3 延时实验分析

本节将设计不同图像分辨率和采样频率的数据验证各模型的效率.图像分辨率中选定 256×256、512×512、600×800 和 540×960 测试各模型的尺度性能,采样频率则是从 5-100 等间隔验证各模型在不同采样频率的性能.由表 7 可以看出:相比于其他深度网络模型,本文提出的 GPLWS-Net 模型具有显著的速度优势,未来通过 GPU 加速可大幅降低运算时间,具备了工业化应用实时三维形貌重建的条件.

表 7 GPLWS-Net 与其他网络在不同分辨率下的延时对比

测试数据集	描述		模型	时间(s)	降低百分比↓	CPU
	焦点堆栈	分辨率				
DefocusNet	5	256×256	FVNet[5]	3.100	15.56%	Intel(R) Xeon(R) Silver 4210 CPU
			DFVNet[5]	**3.020**	13.34%	
			DDFF[6]	4.491	41.73%	
			DefocusNet[7]	5.906	55.69%	
			AiFDepthNet[8]	4.405	40.60%	
			GPLWS-Net	**2.617**	/	
4D Light Field	10	512×512	FVNet[5]	14.330	33.71%	
			DFVNet[5]	13.850	31.42%	
			DDFF[6]	23.489	59.57%	
			DefocusNet[7]	18.624	46.59%	
			AiFDepthNet[8]	**11.987**	20.77%	
			GPLWS-Net	**9.498**	/	
FlyingThings3D	15	540×960	FVNet[5]	**42.614**	28.56%	
			DFVNet[5]	42.776	28.83%	
			DDFF[6]	79.165	61.55%	
			DefocusNet[7]	70.076	56.57%	
			AiFDepthNet[8]	64.441	52.77%	
			GPLWS-Net	**30.440**	/	
Micro 3D	10	600×800	FVNet[5]	39.580	55.02%	

			DFVNet[5]	36.980	51.87%	
			DDFF[6]	38.780	54.11%	
			DefocusNet[7]	34.328	48.15%	
			AiFDepthNet[8]	**31.567**	43.62%	
			GPLWS-Net	**17.800**	/	

5 总 结

微观三维形貌重建作为微纳级显微设备的核心技术,可对精密制造领域的数据建模、产品加工以及质量控制的全链条环节提供保障.而现有的三维形貌重建方法无法应对微观场景中的高分辨率稠密数据的处理,给实时微观三维形貌重建带来挑战.本文从多聚焦图像序列特有的聚焦曲线连续性特点出发,分割一维时序景深数据进行多分支并行,通过网络结构的重参数化保障重建精度,可有效兼顾网络的效率与精度,为微观三维形貌重建方法的多场景部署应用提供解决思路.除此之外,本文公开的微观三维形貌数据集 Micro 3D,可有效缓解现阶段微观领域数据集缺失的问题,为设计高效的深度网络提供数据基础.未来研究主要从标签数据的自动标注和微观三维重建大模型的设计方面展开.

References:

[1] Huang B, Wang W, Bates M, Zhuang X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science, 2008,319(5864):810–813.

[2] Nayar S, Nakagawa Y. Shape from focus. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1994,16(8):824–831.

[3] Yan T, Qian YH, Li FJ, et al. Intelligent microscopic 3D shape reconstruction method based on 3D time-frequency transformation. Sci Sin Inform, 2023,53:282-308. (in Chinese with English abstract). [doi: 10.1360/SSI-2021-0386]

[4] Zhang JF, Yan T, Wang KQ, Qian YH, Wu P. 3D shape reconstruction from multi depth of filed images: datasets and models. Chinese journal of computers, 2023,46(8):1734-1752. (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2023.01734]

[5] Yang F, Huang X, Zhou Z. Deep depth from focus with differential focus volume. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2022. 12642-12651.

[6] Hazirbas C, Soyer SG, Staab MC, et al. Deep depth from focus. In: Proc. of the Asian Conference on Computer Vision. 2018: 525-541.

[7] Maximov M, Galim K, Leal-Taixé L. Focus on defocus: bridging the synthetic to real domain gap for depth estimation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2020:1071-1080.

[8] Wang NH, Wang R, Liu YL, et al. Bridging unsupervised and supervised depth from focus via all-in-focus supervision. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2021:12621-12631.

[9] Howard AG, Zhu M, Chen B et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv Preprint arXiv:1704.04861, 2017.

[10] Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2018:6848-6856.

[11] Han K, Wang Y, Tian Q, Guo J, Xu C. GhostNet: More features from cheap operations. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2020:1577-1586.

[12] Mehta, S, Mohammad R. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. arXiv Preprint arXiv:2110.02178, 2021.

[13] Lee JY, Park RH. Complex-valued disparity: Unified depth model of depth from stereo, depth from focus, and depth from defocus based on the light field gradient. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021,43(3):830–841.

[14] Muhammad M, Choi TS. Sampling for shape from focus in optical microscopy. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(3):564–573.

[15] Jeon HG, Surh J, Im S, et al. Ring difference filter for fast and noise robust depth from focus. IEEE Trans. on Image Processing, 2020,29:1045-1060.

[16] Yan T, Hu Z, Qian, YH, Qiao ZW, Zhang LY. 3D shape reconstruction from multifocus image fusion using a multidirectional modified laplacian operator. Pattern Recognition, 2020,98:107065.

[17] Yan T, Wu P, Qian YH, Hu Z, Liu FX. Multiscale fusion and aggregation pcnn for 3D shape recovery. Information Sciences, 2020, 536:277-297.

[18] Minhas R, Mohammed AA, Wu QM. Shape from focus using fast discrete curvelet transform. Pattern Recognition, 2011,44(4): 839-853.

[19] Ali U, Muhammad TM. Robust focus volume regularization in shape from focus. IEEE Trans. on Image Processing, 2021,30: 7215-7227.

[20] Moeller M, Benning M, Schnlieb C, Cremers D. Variational depth from focus reconstruction. IEEE Trans. on Image Processing, 2015,24(12):5369–5378.

[21] Hu, J, Li S, Gang S. Squeeze-and-excitation networks. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2018:7132-7141.

[22] Daquan Z, Hou Q, Chen Y, et al. Rethinking bottleneck structure for efficient mobile network design. arXiv Preprint arXiv:2007.02269, 2020.

[23] F Chollet. Xception: Deep learning with depthwise separable convolutions. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017:1800-1807.

[24] Zhang T, Qi GJ, Xiao B, Wang J. Interleaved group convolutions for deep neural networks. arXiv Preprint arXiv:1707.02725,2017.

[25] Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv Preprint arXiv:1905.11946,2019.

[26] Han K, Wang Y, Zhang Q, Zhang W, Zhang T. Model rubik's cube: Twisting resolution, depth and width for TinyNets. arXiv Preprint arXiv:2010.14819,2020.

[27] Ma N Zhang X, Zheng H, Sun J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In: Proc. of the European Conference on Computer Vision. 2018:122-138.

[28] Andrew H, Mark S, Grace C, et al. Searching for MobileNetV3. In: Proc. of the IEEE/CVF International Conference on Computer Vision. 2019:1314-1324.

[29] Chen J, Kao S, He H, Zhuo W et al. Run, Don't Walk: chasing higher FLOPS for faster neural networks. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2023.

[30] Pavan K, Vasu A, Gabriel J, et al. MobileOne: An improved one millisecond mobile backbone. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2023.

[31] Chen Y, Dai X, Chen D, Liu M, et al. Mobile-Former: Bridging MobileNet and transformer. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2022:5260-5269.

[32] Pentland AP. A new sense for depth of field. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1987,4:523–531.

[33] Won C, Jeon H. Learning depth from focus in the wild. In: Proc. of the European Conference on Computer Vision. 2022,1-18.

[34] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017,60(624):84-90.

附中文参考文献:

[3] 闫涛,钱宇华,李飞江,等.三维时频变换视角的智能微观三维形貌重建方法.中国科学:信息科学,2023,53:282-308. [doi: 10.1360/SSI-2021-0386]

[4] 张江峰,闫涛,王克琪,钱宇华,吴鹏.多景深图像聚焦信息的三维形貌重建:数据集与模型.计算机学报,2023,46(8):1734-1752. [doi: 10.11897/SP.J.1016.2023.01734]