多景深图像聚焦信息的三维形貌重建：数据集与模型

张江峰[1),2)] 闫 涛[1),2),3),4)] 王克琪[1),2)] 钱宇华[1),3),5)] 吴 鹏[1),5)]

[1)](山西大学大数据科学与产业研究院 太原 030006)

[2)](山西大学计算机与信息技术学院 太原 030006)

[3)](山西省机器视觉与数据挖掘工程研究中心(山西大学) 太原 030006)

[4)](哈尔滨工业大学重庆研究院 重庆 401151)

[5)](计算智能与中文信息处理教育部重点实验室(山西大学) 太原 030006)

摘 要 受限于数据采集方式的多源异性与三维重建结果的昂贵标注，现有基于多景深图像聚焦信息的三维形貌重建方法通常需要根据具体应用场景设计，缺乏场景适应性. 本文提出一种多景深图像数据集构建的理论与方法，并在此基础上设计具有良好鲁棒性的深度网络模型. 构建的多景深图像数据集(MDFI Datasets)旨在剥离图像实际语义与深度信息的强关联性，通过联合输入图像序列的富纹理特性与三维形貌固有的同质与阶跃特性，提出形貌核函数非线性空间映射方法扩展数据集的多维性与多样性. 设计的深度三维形貌重建网络模型(DSFF-Net)以 U-Net 为基础网络，添加可变形卷积模块(Deformable ConvNets v2)增强网络的特征提取能力，全新设计的局部-全局关系耦合模块(LGRCB)有助于提升模型全局聚焦信息的聚合能力. 为验证 MDFI Datasets 的跨场景适用性和 DSFF-Net 模型的鲁棒性与泛化性，本文从四个不同方面进行实验对比分析. 实验结果表明，相较于最先进的鲁棒聚焦体积正则化的聚焦形貌恢复算法(RFVR-SFF)和全聚焦深度网络(AiFDepth-Net)，本文提出的 DSFF-Net 模型在 *RMSE* 指标上分别下降 15%和 29%；大景深场景实验表明，本文提出的数据集构建方法能够适应实际应用场景.

关键词 三维形貌重建；深度学习；图像序列数据集；多聚焦图像；核函数

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2023.01734

3D Shape Reconstruction from Multi Depth of Field Images: Datasets and Models

ZHANG Jiang-Feng[1),2)] YAN Tao[1),2),3),4)] WANG Ke-Qi[1),2)] QIAN Yu-Hua[1),3),5)] WU Peng[1),5)]

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

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

[3)](*Engineering Research Center for Machine Vision and Data Mining of Shanxi Province, Shanxi University, Taiyuan* 030006)

[4)](*Chongqing Research Institute of Harbin Institute of Technology, Chongqing* 401151)

[5)](*Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan* 030006)

Abstract Limited by the multi-source heterogeneity of data acquisition and the expensive annotation of 3D reconstruction results, the existing 3D shape reconstruction methods based on multi-depth of field image focus information usually need to be designed according to specific application scenes, resulting in a lack of scene adaptability. This paper proposed a theory and

收稿日期：2022-04-05；在线发布日期：2023-04-10. 本课题得到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)资助. **张江峰**，硕士研究生，中国计算机学会(CCF)会员，主要研究方向为深度学习、三维重建. E-mail：zjf_8099@163.com. **闫 涛**(通信作者)，博士，副教授，中国计算机学会(CCF)会员，主要研究方向为深度学习、三维重建. E-mail：hongyanyutian@sxu.edu.cn. **王克琪**，硕士，中国计算机学会(CCF)会员，主要研究方向为深度学习、图像增强. **钱宇华**，博士，教授，中国计算机学会(CCF)专业会员，主要研究领域为人工智能、机器学习. **吴 鹏**，博士，讲师，主要研究方向为区块链、机器学习.

method of constructing multi-depth-of-field image datasets, and designed a robust deep network model on this basis. The constructed Multi-Depth-of-Field Image Datasets (MDFI Datasets) aimed at stripping the strong correlation between the actual semantics of images and the depth information. The shape kernel function nonlinear spatial mapping method was proposed to extend the multidimensionality and diversity of the datasets by combining the texture-rich characteristics of the input image sequences with the inherent homogeneity and step properties of the 3D shape. The Deep Shape from Focus Net (DSFF-Net) was designed with U-Net as the base network, and Deformable ConvNets v2 was added to enhance the feature extraction capability of the network, and the newly designed Local-Global Relationship Coupling (LGRCB) module helped to improve the aggregation capability of the global focus information of the model. To verify the cross-scene applicability of MDFI Datasets and the robustness as well as generalization of DSFF-Net model, this paper conducted experimental comparative analysis from four different aspects. The results of the experiments show that compared with the state-of-the-art Robust Focus Volume Regularization in Shape from Focus (RFVR-SFF) and All-in-Focus Depth Net (AiFDepth-Net), the DSFF-Net model proposed in this paper decreases 15% and 29% in the Root Mean Square Error (*RMSE*) index, while the experiments on large depth-of-field scenes show that the datasets construction method proposed in this paper can adapt to real application scenes.

Keywords　3D shape reconstruction; deep learning; image sequence datasets; multi-focus images; kernel function

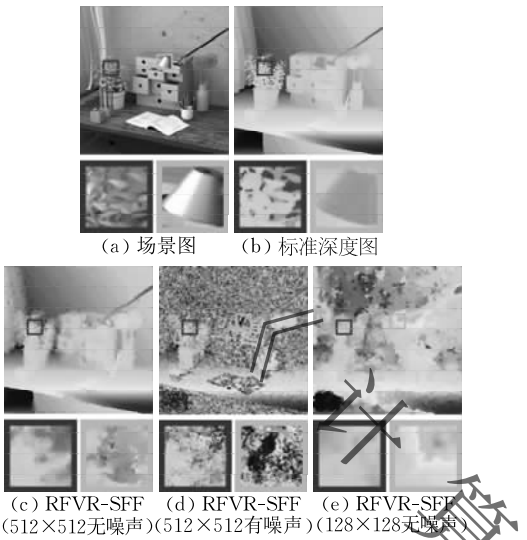
1　引　言

深度学习模型作为一种典型的数据驱动类方法,以标注数据建立因果关系,通过海量标注数据集训练得到高鲁棒性模型.目前,数据驱动的深度模型作为人工智能的研究热点已被广泛应用于计算机视觉、自然语言处理和语音识别等领域[1].由此可见,新数据集的构建往往会催生相关领域研究的新动能.对于三维重建领域而言,现有的三维重建数据集研究大多聚焦在单目深度估计[2]、点云[3]和4D光场[4]等方向.其中单目深度估计使用单张图像进行深度估计,其深度线索的单一性难以实现高精度的场景重建;点云重建主要通过采集待测场景中的点云数据进行重建[5],其稀疏特性给精细化重建带来一定挑战;4D光场数据中记录了光线的位置和方向,重建精度变相依赖于设备精度[4].基于多景深图像聚焦信息的三维形貌重建通过图像序列的聚焦信息获取场景深度,其较低的场景敏感度和易于高分辨率成像等特点使其广受学术与工业界重视[6].而目前基于多景深图像聚焦信息的三维形貌重建领域由于各类数据的采集方式、成像原理、宏微观应用场景存在异同,诸多因素的语义难统性导致统一的多景深图像数据集构建困难.因此如何构建剥离实际语义的标准多景深图像数据集对实现高鲁棒性与良好泛化性的深度模型具有重要意义.

基于多景深图像聚焦信息的三维形貌重建即聚焦形貌恢复(Shape from Focus,SFF)主要包括如下步骤:首先对单帧图像采用聚焦测量算子进行聚焦水平估计,然后聚合图像序列中聚焦水平最大值所在帧的位置信息,构成初始三维形貌结构;最后采用形貌近似函数或深度修复算法得到最终三维形貌结果.根据上述步骤可知,聚焦测量算子作为重建过程的关键步骤,其对图像聚焦信息判断的准确性直接影响重建结果的精度,而在真实场景中,聚焦测量算子通常会受到图像本身的尺度和噪声干扰,容易导致重建结果的精度下降.

图1为不同尺度与噪声干扰对三维形貌重建结果的影响,以鲁棒聚焦体积正则化的聚焦形貌恢复算法(Robust Focus Volume Regularization in Shape from Focus,RFVR-SFF)[7]为例.其中图1(a)为场景的实际语义,图像大小为512×512;图1(b)为该场景对应的标准深度图;图1(c)为RFVR-SFF算法

的三维重建结果;图 1(d)为图像 1(a)中加入高斯噪声后 RFVR-SFF 算法的三维重建结果;图 1(e)为尺度变为 128×128 后 RFVR-SFF 算法的三维重建结果.由图 1 可知,现有三维形貌重建方法侧重于聚焦评价算子的选择,尤其是特定应用场景设计的重建算法无法应对尺度和噪声的影响.



(a) 场景图 (b) 标准深度图

(c) RFVR-SFF (512×512无噪声) (d) RFVR-SFF (512×512有噪声) (e) RFVR-SFF (128×128无噪声)

图 1 图像尺度与噪声对三维形貌重建方法的影响

视觉类深度学习算法可从大规模图像数据中学习抽象的层次化特征,进而建立图像到具体类别之间的映射[8],但现有的基于深度学习的三维形貌重建仅限于小样本数据[9-11],且数据来源具有较强的实际语义约束,在此基础上训练的深度学习模型可能会导致其泛化性能下降.

综上所述,为了应对传统方法的鲁棒性问题与深度学习类方法的泛化性问题,本文提出一套多景深图像数据集自动生成的理论与方法,并构建了一种具有良好场景适用性和高鲁棒性的深度学习模型.本文的主要贡献如下:

(1)提出的多景深图像数据集(MDFI Datasets, Multi-depth of Field Image Datasets)将富纹理图像与深度形貌进行映射,最大限度剥离实际语义对深度信息的影响,有利于提升深度模型的泛化性能;

(2)利用初等函数与进化算法的测试函数模拟深度图像中同质性缓慢变化区域,分段函数则实现深度图像中异质区域变化的模拟,上述两类函数经过形貌核函数的非线性变化可实现 MDFI Datasets 在深度信息多样性方面的有效覆盖;

(3)可变形卷积模块的引入与局部-全局关系耦合模块的提出,使得构建的深度三维形貌重建网络模型 DSFF-Net(DSFF-Net, Deep Shape from Focus Net)具有较强的特征提取与全局聚焦信息聚合能力.

本文第 2 节主要介绍三维形貌重建的原理及研究进展;第 3 节阐述基于深度学习的三维形貌重建方法,并从框架、数据集、深度网络模型和深度图生成四方面进行分析;第 4 节与现有的三维形貌重建算法在不同数据集中进行对比分析;最后对本研究进行总结和展望.

2 相关工作

基于多景深图像聚焦信息的三维形貌重建方法利用相机聚焦原理来还原待测物体的三维形貌,其基本思想是图像中一点到相机感光成像面的相对距离对应该点最大聚焦水平的深度信息[12].通过调整相机与待测物体之间的距离产生不同区域的聚焦图像序列,根据聚焦信息估计场景中每个位置的相对距离.像距 v 取决于镜头焦距 f 与物距 u.这三个变量之间的关系由成像公式定义:

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

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.图 2 为薄透镜模型在图像中聚焦的效果.当相机与聚焦平面的距离为 δ 时,点 Q 投影到直径为 k_σ 的圆,仅当相机位于聚焦平面时点 Q 聚焦.由于物距 u 和像距 v 之间存在一一对应关系,因此在特定物距下,场景中会出现局部区域聚焦的情形.

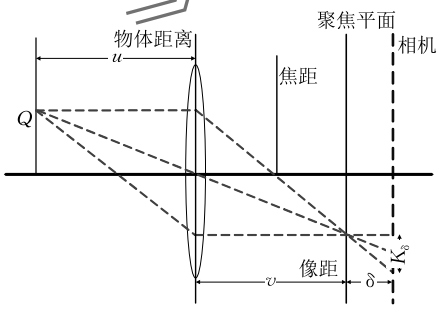


图 2 图像聚焦信息指导深度信息示意图

然后,采用聚焦测量算子(Focus Measure, FM)对图像序列进行聚焦评价,当某点聚焦水平达到最大时就可通过其所在图像序列位置推断出该点的深度信息.将图像中深度信息逐点融合,得到该场景对应的初始深度图[13]:

$$D[X,Y]=\underset{1\le i\le n}{\arg\max}\{FM_i[X,Y]\},1\le X\le W,1\le Y\le H \tag{2}$$

其中,图像序列总数为 N,图像帧大小为 $H\times W$,$FM_i[X,Y]$表示图像 i 中$[X,Y]$位置的焦点测量结果,$D[X,Y]$表示$[X,Y]$位置对应的深度信息;最后,通过形貌近似函数或深度修复算法对初始深度图进行处理,得到待测场景的三维形貌[14].

2.1　多景深图像聚焦信息三维形貌重建研究进展

现阶段,基于多景深图像聚焦信息的三维形貌重建方法可分为模型设计与数据驱动两大类.

2.1.1　模型设计类三维形貌重建方法

模型设计类三维形貌重建方法旨在通过像素处理的方式计算深度信息.这类重建方法可分为时域与频域两大类.在时域模型方面:有研究提出 Sum-modified-Laplacian[12]算子检测图像的聚焦信息,随后陆续出现多种有效的聚焦测量模型,如 Diagonal Laplacian[15]、Variance of Laplacian[16]和环形差分滤波算子(Ring Difference Filter,RDF)[17]等.但这类方法仅对二维图像局部区域进行聚焦测量,未考虑图像的全局结构信息.因此,出现了利用图像分割将重建问题转化为凸函数优化问题的图像分割算法(Graph Cut,GC)[18],但该方法严重依赖于图像分割算法的准确性.在频域模型方面:利用快速离散曲波变换[19]、非降采样轮廓波变换[20]和 3D 离散小波变换[21]等时频变换工具检测图像序列中的高频分量,然后将高频分量映射为深度信息进而得到场景的三维形貌重建结果;也有研究借鉴多粒度融合思想实现复杂场景的三维形貌重建[22];还有研究采用非降采样剪切波变换同时实现全聚焦图像与三维形貌结果的生成[23].但上述方法缺乏先验信息的补充,重建性能提升有限.

深度图修复算法则采用强制平滑性约束改善深度图的同质性.例如中值滤波器[24]、马尔科夫随机场[25]、数据保真项[26]、引导滤波[27]和非凸正则化器优化的 RFVR-SFF[7]等.此类方法极度依赖初始深度估计,假使初始深度估计出现偏差,将会导致深度信息的误差蔓延.

综上所述,传统基于多景深聚焦信息的三维形貌重建方法主要关注聚焦测量模型与深度图修复两类核心步骤,但由于不同场景的成像原理、噪声干扰各异,各类重建方法通常需要根据特定应用领域进行设计,缺乏场景适应性.

2.1.2　数据驱动类三维形貌重建方法

数据驱动类三维形貌重建方法以深度学习为代表.如基于焦点深度的网络模型(Deep Depth From Focus Net,DDFF-Net)[28]利用 RGB-D 数据构建了多景深数据集并提出深度网络模型,该网络利用卷积关联像素信息抽取场景中的聚焦特征;离焦网络模型(Defocus Net,Defocus-Net)[11]提出利用散焦图像作为监督信号,然后通过聚焦网络和深度网络分别生成全聚焦图和深度图.聚焦体积网络(Focus Volume Net,FV-Net)模拟深度立体匹配方法[9],利用深度差分体积结合焦点和上下文进行深度估计.上述方法将特定场景的多景深图像集和深度图经过神经网络抽象学习聚焦特征并自动拟合聚焦区域.相比传统方法,深度学习类方法具有精准高效的特点,但与此同时,重建性能也受制于数据集本身.因此,全聚焦深度网络模型(All-in-Focus Depth Net,AiFDepth-Net)[10]利用全聚焦图像监督或全聚焦图像和深度图共同引导网络得到聚焦栈,实现无监督的三维形貌重建.

上述基于深度学习的三维形貌重建方法极度依赖于数据集本身,加之不同数据集之间存在语义鸿沟,导致采用特定场景数据集训练的深度学习模型可能在求解其他场景重建问题时出现性能下降.因此,如何构造"基数据集"消弥不同场景间的语义鸿沟对设计具有良好泛化性的深度网络模型具有重要意义.

2.2　三维形貌重建数据集的研究进展

本节从单目深度估计数据集、点云数据集、4D 光场数据集和仿真数据集四种不同类型的三维重建数据说明构建多景深图像数据集的必要性.

2.2.1　单目深度估计数据集

针对室内和室外不同类型、不同深度范围的单目三维重建数据集已经较为成熟,例如 Make3D 数据集[29]和 KITTI 数据集[30]等.单目深度估计算法主要侧重宏观场景中的空间关系,但仅通过单张图像恢复场景相对深度,缺乏丰富的深度线索,导致深度信息估计欠准确;DDFF12-Scene 数据集[28]是第一个提供深度线索的单目三维重建数据集,该数据集使用 RGB-D 深度相机获得场景真实深度图,使用彩色相机获得 10 张不同聚焦水平的图像.但该数据集存在以下两个问题:图像内容来自于室内场景,场景内容较为欠缺;场景深度图的质量取决于 RGB-D 深度相机传感器的精度,而 RGB-D 相机在深度成像过程中极易受到光照、噪声等因素影响.FoD500

数据集[11]使用 Blender 渲染合成数据集，来自 400 个 CAD3D 模型，已公布 500 组数据，每组数据集包含 5 幅散焦图和 1 幅深度图. 但由于该数据集利用图像离焦信息作为深度线索，所以该样本的深度图无法有效覆盖场景的深度范围. 以上两个携带深度线索的单目三维重建数据集均缺乏聚焦区域与离焦区域之间的显著变化，同样无法刻画缓慢过渡的深度信息场景.

2.2.2 点云数据集

3D 扫描设备(激光雷达)产生的数据通常以点云形式呈现[31]. ModelNet[32]作为合成数据集的代表，提供了全面且清晰的 CAD 模型，搭建了虚拟与真实世界的桥梁；ScanNet[33]使用 RGB-D 相机对场景进行三维重建并且提供了密集的语义分割标注. 点云作为物体表面采样点的空间坐标集合，保留着最重要的空间三维几何信息. 但由于点云数据的稀疏特性导致基于点云的三维重建模型的效率和精度难以兼顾.

2.2.3 4D 光场数据集

4D 光场数据将特有的角度信息作为深度线索，可有效估计被遮挡物体的边缘信息. Honauer 等人[34]提供了密集采样的光场基准数据集和对应的视差图；Heber 等人[35]基于光线追踪 POV-Ray 软件合成光场数据，并提供了浮点级精度的深度图；Shi 等人[36]创建了两个光场数据集：稀疏采样的 SLFD 和密集采样的 DLFD. 4D 光场数据提供了更加丰富的深度线索，对于物体遮挡可有效进行深度估计，但此类方法具有较强的硬件设备依赖性，昂贵的使用成本使其难以开展大规模应用.

2.2.4 仿真数据集

在现阶段三维形貌重建模型的质量评价中，需要评价预测深度图与真实场景深度图之间的差异性来评判模型的优劣. 由于不同场景条件下存在多种影响因素，导致新模型的效果不易验证. 现阶段，新模型的提出首先需要通过仿真数据验证模型的聚焦识别能力，其次再通过特定场景实验以验证模型的泛化能力. 如 Subbarao 等人[37]验证点扩散函数在图像仿真离焦的有效性；随后，Pertuz 等人[14]改进了点扩散函数的离焦过程，将仿真深度图中每个点的离焦权重比值加入离焦过程中，统计空间中每个点的离焦量以产生最终的仿真深度图. 以上两种算法仅通过简单的形貌函数仿真，无法模拟实际场景中多景深图像序列内含深度信息的多变性. 韦号等人[38]通过三维建模软件来调整曲面轮廓，可有效模拟场景中的复杂表面，但不利于批量生产，不适合作为深度学习模型的数据集.

随着工业制造领域的精细化三维建模、消费电子领域多聚焦图像增强等多样化应用需求逐渐增多，现有的三维重建数据集无法满足上述多样化场景的应用需求，迫切需要构建标准的多聚焦图像数据集，实现跨场景与自适应的三维形貌重建方法.

3 DSFF-Net：基于深度学习的多聚焦图像三维形貌重建算法

3.1 基于 DSFF-Net 的三维形貌重建框架

本文提出的基于 DSFF-Net 的三维形貌重建方法，主要分为以下三个关键步骤：

(1)数据集构建. 鉴于多景深图像采集过程中会受到多种因素的协同制约，构造多景深图像数据集时在图像纹理和深度信息生成方面需要保持高度随机性，除此之外，还需要尽可能的排除图像实际语义对深度信息的影响. 因此，本文将富纹理图像与深度形貌进行随机抽样组合经过点扩散函数的仿真模拟生成多景深图像序列和深度图；

(2)网络训练. 多景深图像数据集作为深度网络模型的驱动力，结合神经网络稳健的特征泛化能力，通过提取多景深图像集与深度图之间的几何特征及抽象语义关系，进而泛化多场景的空间结构信息，实现在跨场景深度图像的自适应预测；

(3)深度图生成. 由于高质量深度图像需要兼顾聚焦信息和场景信息，本文利用多景深图像序列生成的融合图像[17]辅助修复深度图，其不仅对初始深度图中的离散噪声可有效滤除，而且对场景边缘信息能够实现有效增强. 最终获得聚焦信息和场景信息互补的高精度深度图. 框架示意图如图 3 所示.

3.2 多景深图像数据集构建

现有研究表明：针对特定场景(如室内场景[28]、微缩模型[35])构建的三维形貌数据集，均与图像的实际语义有强关联关系，这种设定将会影响深度模型的泛化性能. 因此，本文构建的 MDFI Datasets 更加侧重图像深度信息的覆盖性和完备性，最大限度消除图像实际语义对深度信息的影响，力求构建基础的多景深图像数据集. MDFI Datasets 构建主要分为初始形貌创建、形貌数据增广与多景深图像序列生成三部分.

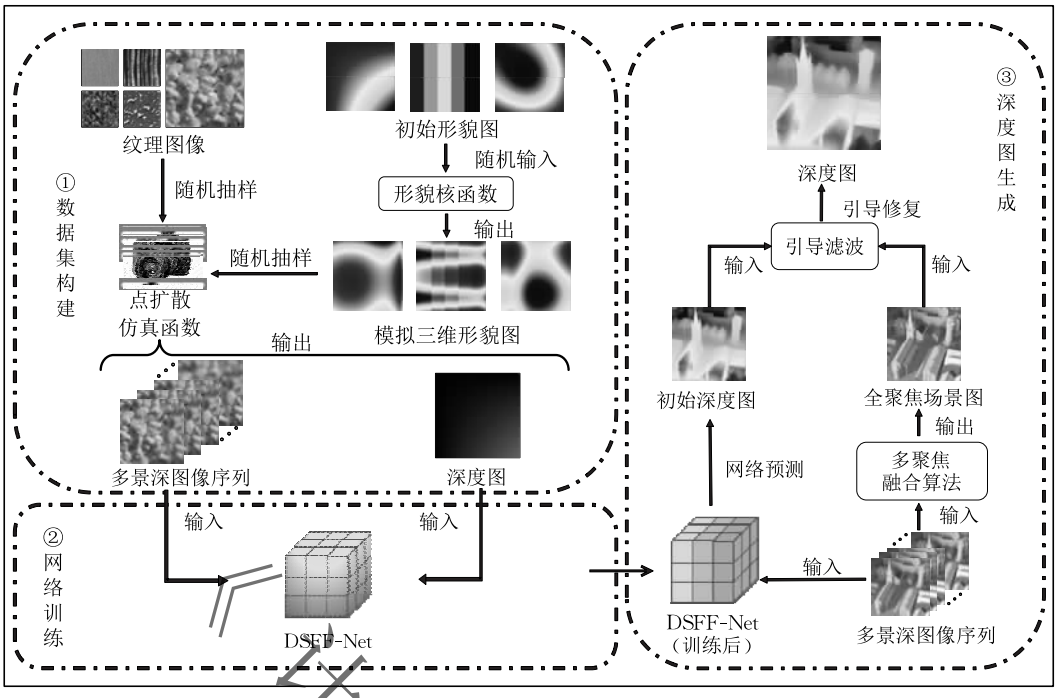


图 3　基于 DSFF-Net 的三维形貌重建框架示意图

3.2.1　创建初始形貌

由于聚焦评价算子主要通过像素之间的关系评估图像的聚焦水平,图像内容对三维形貌重建模型的影响尤为重要.通常在弱纹理场景下,聚焦测量算子无法准确估计场景的深度信息[39].为保证数据集对神经网络模型的正确引导,本文使用富纹理图像作为场景背景,并在此基础上生成多景深图像序列.本文使用BRODATZ's TEXTURE DATABASE[40]和 Kylberg TEXTURE Datasets V.1.0[41]作为数据集场景的来源,这两类数据集包含了不同背景强度的纹理图像.

为有效筛选富纹理图像,本文对纹理图像随机裁取 55×55 的图像块作为整个图像纹理代表,计算该图像块的图像信息熵和图像梯度累加和作为图像纹理评价指标,去除图像纹理评价值低于 0.3 的图像,经过多次迭代最终保留富纹理图像:

$$E=-\sum p\times\log_2 p \tag{3}$$

$$G=sum\left(\left|\begin{bmatrix}1&0&-1\\2&0&-2\\1&0&-1\end{bmatrix}\times I\right|+\left|\begin{bmatrix}1&2&1\\0&0&0\\-1&-2&-1\end{bmatrix}\times I\right|\right) \tag{4}$$

$$V=0.4\times E/8+0.6\times G/(3\times 10^6) \tag{5}$$

其中,E 代表图像信息熵,p 代表图像中各像素统计值,G 代表图像梯度累加和,V 代表图像纹理评价指标.图像纹理评价指标中,0.4 和 0.6 分别为图像信息熵和图像梯度累加和的权重,8 代表所裁取图像块最大的图像信息熵近似值,10^6 代表所裁取图像块设定的图像梯度累加和阈值.

图 4 富纹理图像示意中展示来自 ImageNet 数据集的 5 张图像和来自 Kylberg TEXTURE Datasets V.1.0 和 BRODATZ's TEXTURE DATABASE 的 4 张图像,并分别展示图像归一化后大小为 55×55 图像块的灰度直方图.表 1 图像纹理信息对比中展示图 4 所有框内图像块的图像信息熵、图像梯度累加和与图像纹理评价值.由图 4 的灰度直方图可以直观看出图像(a)~(i)框内的灰度分布,其中图像(a)~(e)灰度分布较为集中,不利于模拟生成多景深图像序列,而图像(f)~(i)利于模拟生成多景深图像序列.由表 1 图像纹理信息对比可以得出富纹理图像块的图像纹理评价值皆大于 0.3,本文以此为标准筛选富纹理图像.

表 1　图像纹理信息对比

图像序号	图像信息熵	图像梯度累加和	图像纹理评价值
(a)	4.7215	184546	0.2730
(b)	3.8150	12202	0.1932
(c)	2.2197	9844	0.1130
(d)	1.9754	6204	0.1000
(e)	4.1140	57050	0.2171
(f)	5.5841	162600	**0.3117**
(g)	5.5958	839720	**0.4477**
(h)	5.3799	308214	**0.3306**
(i)	6.5973	567870	**0.4434**

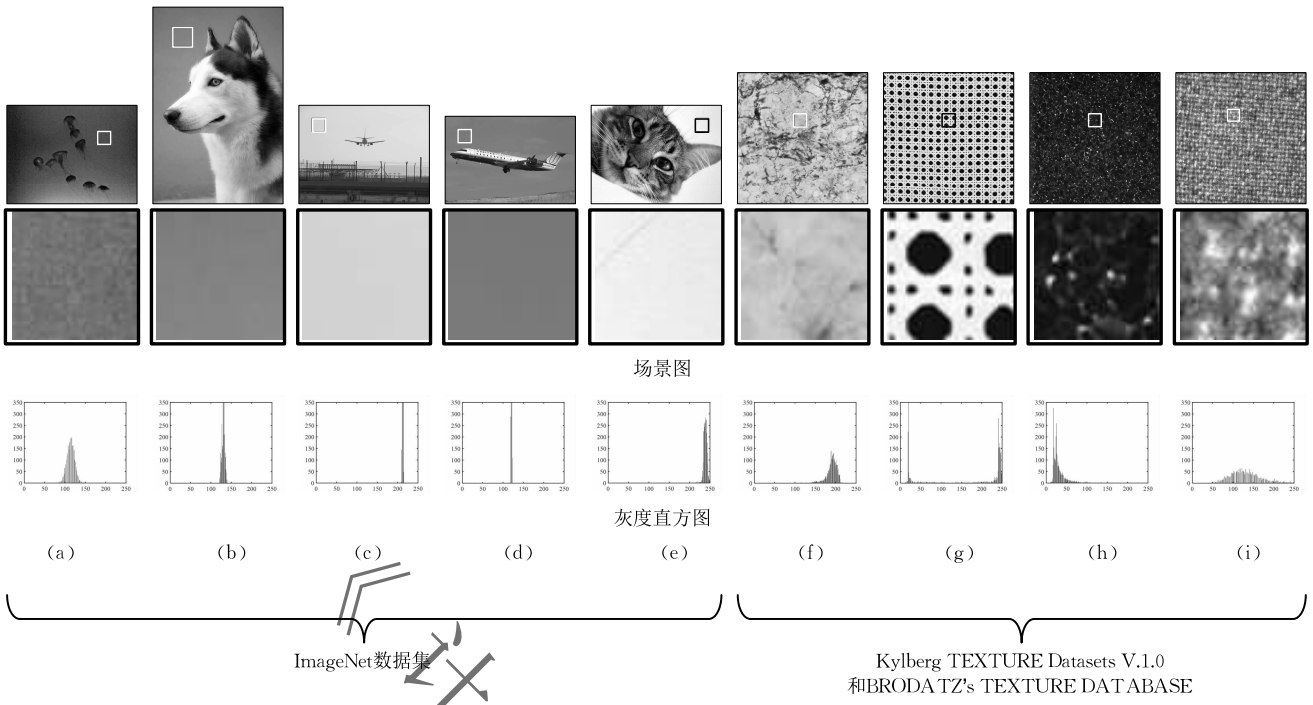


图 4 富纹理图像示例

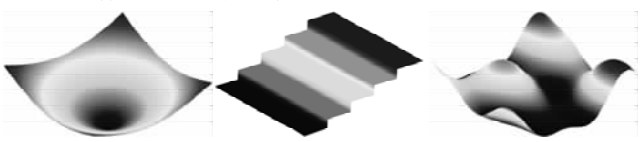
为模拟真实场景的深度信息,需要在数据集生成过程中满足形貌随机化的约束.因此,本节利用初等函数、分段函数和进化算法的测试函数三类初始形貌函数搭建初始深度图.初等函数使三维形貌具有基础场景属性,如平面、圆锥和余弦等;分段函数则是模拟真实场景中不连续区域,如遮挡、深度信息变化明显的异质区域等;进化算法的测试函数则拟合真实场景深度信息的多样性变化.

图 5 中展示了三类初始形貌函数的三维形貌样例,表 2 中展示了部分初始形貌函数.初始形貌图具体构造方法如下:

初等函数:以图像宽高为限制,分别以随机的参数截取任意两段,随后以矩阵乘法扩充为图像大小,从而生成初始形貌图;

分段函数:以图像宽度为限制,分割为若干段,每一段使用随机参数的初等函数,再通过矩阵扩展与图像高度相同,从而生成初始形貌图;

进化算法的测试函数:以图像宽高为限制,使用随机参数生成初始形貌图.



(a) 初等函数 (b) 分段函数 (c) 进化算法的测试函数

图 5 三类初始形貌函数的三维形貌样例

表 2 初始形貌函数

类别	数学表达式	参数
初等函数	$f(x)=k$	k
	$f(x)=kx+b$	k,x,b
	$f(x)=\sin x+b$	x,b
	$f(x)=a^x+b$	a,x,b
	$f(x)=\log_a x$	a,x
分段函数	$f(x,y)=\begin{cases}k_1 f_1(x,y), 0\le x<a,\\ \cdots\\ k_n f_n(x,y), b\le x\le c\end{cases}$	$k_1,\cdots,k_n,$ $a,b,c,\cdots$ $1\le n\le 10$
进化算法的测试函数	$f(x,y)=k_1x^2+k_1y^2$	k_1,k_2,x,y
	$f(x,y)=-k\sin(\sqrt{x}-y\sin(\sqrt{y})$	k,x,y
	$f(x,y)=-k\dfrac{1+\cos(12\sqrt{x^2+y^2})}{(x^2+y^2)+2}$	k,x,y
	$f(x,y)=k(y-x^2)^2+(1-x)^2$	k,x,y

为确保生成的初始形貌图可以适应生成的图像序列,对所有的初始形貌结果进行拉伸以满足图像序列大小.

3.2.2 形貌数据增广

鉴于初始形貌的数据量较小,因此需要对场景中深度图像的同质和异质区域进行多样性扩展.本文使用核函数增强数据集的多维性和多样性.即通过对上述三种形貌函数进行随机选择以模拟现实场景的复杂性.核函数通过两向量之间内积运算实现空间映射,本节在核函数的基础上提出形貌核函数 $K_m()$:

$$D=K_m(d_p,d_q) \tag{6}$$

其中,D 是新生成的深度图,而 d_p 和 d_q 则是初始形

貌函数生成的初始深度图中随机选择的两幅深度图.表 3 中列举了五种基础形貌核函数,其中参数为随机变量.通过随机选择形貌核函数的参数,深度图不仅可以满足数据集的多维性和多样性,同时可以有效降低生成数据集的成本.

表 3　部分形貌核函数

函数名	数学表达式	参数
add	$D_{(i,j)}=d_{p(i,j)}+Rd_{q(i,j)}$	R
linear	$D_{(i,j)}=Rd_{p(i,j)}d_{q(i,j)}$	R
ploy	$D_{(i,j)}=(d_{p(i,j)}d_{q(i,j)}+C)^R$	R、C
sigmoid	$D_{(i,j)}=\tanh(Rd_{p(i,j)}d_{q(i,j)}+C)$	R、C
gaussian	$D_{(i,j)}=Ae^{-0.5\frac{(d_{p(i,j)}+d_{q(i,j)})^2}{R}}$	A、R

3.2.3　多景深图像序列生成

多景深图像序列的仿真过程是模拟光学系统在相对于场景移动过程中图像序列的采集过程.通过连续调整相机与场景之间的距离,可实现相机景深范围内局部区域清晰成像,景深范围外是离焦的模糊图像.

局部聚焦图像可以视为全聚焦图像经过滤波后的图像,因此局部聚焦图像 I_d 可以描述为全聚焦图像 I 与点扩散函数 h 的卷积:

$$I_d=I\times h \tag{7}$$

其中,点扩散函数 h 是相机单位点源的响应[42].在具有非相干光照明的衍射限制光学器件中,点扩散函数可以简化为如下高斯公式[42]:

$$h(x,y)=\frac{1}{2\pi\sigma_h}\exp\left(-\frac{x^2+y^2}{2\sigma_h}\right) \tag{8}$$

其中,σ_h 与图像散焦程度成正比,$[x,y]$为图像坐标.

Pentland[43]推导出图像 $I_{(x,y)}$ 位置处模糊参数 σ_h 和场景深度 u 之间的联系:

$$\sigma_h(x,y)=\frac{kf^2|u-u_f|}{Au(u_f-f)} \tag{9}$$

其中,u_f 是相机设置的对焦位置,k 为相机常数,A 是相机焦距 f 与镜头直径 d 的比值.

为避免合成数据时产生光晕现象,通过对每个场景点(x,y)与其对应的点扩散函数值进行卷积得到模糊图像 B:

$$B_{x,y}=I_{x,y}\times h_{x,y} \tag{10}$$

其中,点扩散函数值 $h_{x,y}$ 与图像像素点 $I_{(x,y)}$ 一一对应.同理,大小为 $H\times W$ 离焦图像 I_d 中点(x_0,y_0)将像素中每个点的散焦程度进行累加:

$$I_{d(x_0,y_0)}=\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}B_{i,j}(i-x_0,j-y_0) \tag{11}$$

在多景深图像序列生成中,借助场景形貌图的多维性和多样性联合富纹理图像集可生成具有一定规模形态随机化的多景深图像序列数据集.

3.3　DSFF-Net 深度网络模型

3.3.1　整体网络结构

受到U-Net 精准定位和全局信息耦合的启发[44],本文使用 U-Net 作为网络结构的主干,主要包括收缩路径、瓶颈模块和扩张路径三部分构成.在基于多景深图像聚焦信息的三维形貌重建中,由于相同聚焦设置下不同尺度的图像序列对于聚焦算子的敏感程度不同.U-Net 主干网络通过池化操作降低特征分辨率以获得不同尺度的图像信息.随着特征维度的增加,U-Net 通过跳跃连接(Skip Connection)保留图像的浅层细节特征以明确聚焦区域的边界范围.相比其他具有深度线索的三维重建,基于聚焦信息的三维形貌重建输入图像容量大,U-Net 网络主干不仅可以确保深度预测精确,而且可以减少网络参数量.

为保证每一层特征都集中在聚焦区域,在收缩路径中加入可变形卷积(Deformable ConvNets v2)[45]作为特征增强.在收缩路径中逐渐缩小输入特征矩阵的空间维度,并提取多景深图像序列的高层特征.其可以分为大致相同的三组卷积操作,其中每一组卷积包含 2 层卷积核数目相同且连续的 3×3 卷积操作.由浅至深,三组卷积操作的频道数从 128 依次成倍扩增至 512.为防止网络在训练过程中出现梯度弥散现象,采用 LeakyReLU 激活操作和 BN(BatchNorm)归一化层穿插于卷积层之间,使每层的特征分布相对稳定并加快模型的收敛速度.可变形卷积层和最大池化层附在每组卷积操作之后,作为特有的特征增强模块凸出高权重的聚焦特征信息.

在传统的瓶颈模块嵌入全新设计的局部-全局关系耦合模块(Local-Global Relationship Coupling Block,LGRCB)[46]增强模型对全局聚焦信息的聚合能力.在两层卷积核为 3×3 频道数为 512 的卷积层中间插入局部-全局关系耦合模块,以应对收缩路径收集的多景深图像序列的高级语义信息,将代表性的语义信息传播到后续的扩张路径中.

扩张路径与收缩路径在输出特征矩阵的尺寸和卷积层频道数相互对称,同样也可以分为三组卷积操作.首先,自底向上由反卷积层还原特征矩阵的尺寸,提供来自低分辨率的高层聚焦特征.然后,将上一层输出的特征矩阵连接来自收缩路径中对应的高分辨率浅层细节特征.最后通过三层卷积核数目相同且连续的 3×3 卷积操作融合网络中高层和浅层特征矩阵.

图 6 所示为本文构建的 DSFF-Net 结构图.

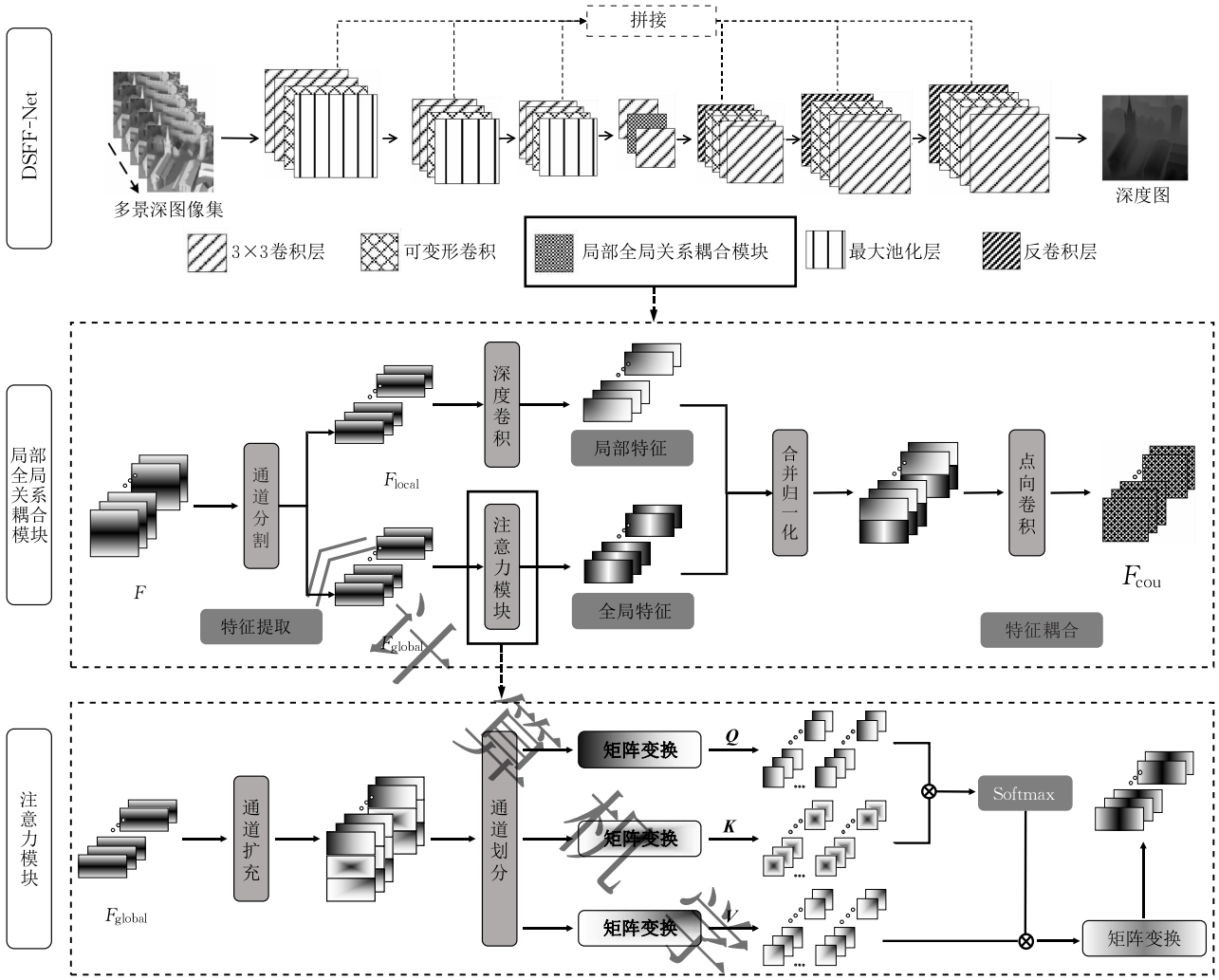


图 6 DSFF-Net 模型结构

3.3.2 可变形卷积

场景中不同区域的聚焦形状会呈现间断性和多变性,尤其对于设备采集数据时相对位置的随机性,使得聚焦区域的查找和定位困难.卷积神经网络具有平移不变性,能够对简单区域的聚焦信息进行有效提取,但对于场景信息未知和聚焦区域扭曲的情形,特征提取和泛化的能力则会下降.目前针对该问题主要有如下两种解决方案[47]:(1)通过扩充数据集以提升网络模型归纳偏置的能力;(2)利用几何不变特征的提取算法以解决特征抽取问题.

本文同时从两方面着手:

(1)在数据集构建之初便通过上述步骤生成足量的多景深图像数据集;

(2)传统的几何特征不变算法(如 SIFT)为手工设计,无法应对复杂场景.本文使用可变形卷积应对过度复杂的聚焦区域形变.可变形卷积的优越性在于精准定位局部感兴趣区域,适应于多景深图像序列中聚焦区域位置不同和形态各异的情况.可变形卷积在卷积点采样过程中引入偏移量,使网络可以拟合不规则的聚焦区域,重点关注聚焦区域的变化.

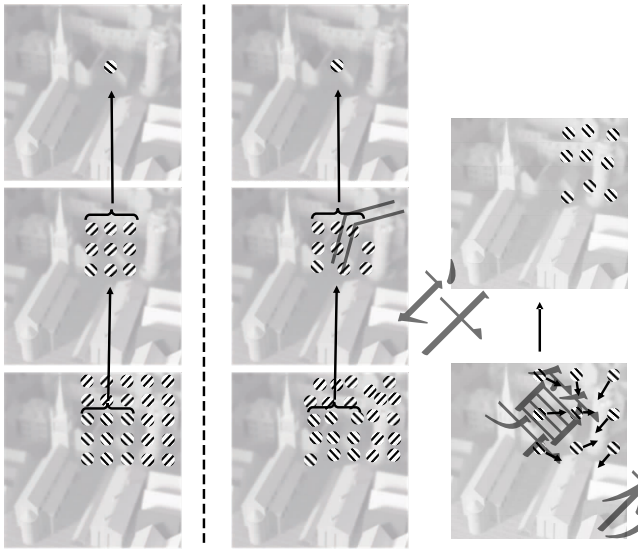
在多景深图像序列中,聚焦区域尺度、姿态和形变的判定是三维形貌重建的主要挑战.可变形卷积使用额外偏移量增加卷积核的空间采样位置,并调整不同位置空间的特征权重.设采样大小为 K 的可变形卷积核,W_k、P_k 分别表示第 k 个位置的权重和预先指定的偏移量,$\boldsymbol{X}(p)$ 和 $\boldsymbol{Y}(p)$ 分别表示输入特征矩阵 $\boldsymbol{X}$ 和输出特征矩阵 $\boldsymbol{Y}$ 位置 p 处的特征,可变形卷积可以表示为

$$\boldsymbol{Y}(p)=\sum_{k=1}^{K}W_k\cdot\boldsymbol{X}(p+P_k+\Delta P_k)\cdot\Delta M_k \quad (12)$$

其中,ΔP_k 与 ΔM_k 分别是采样位置 k 可学习的偏移量和调剂标量.可变形卷积通过动态调整感受野有助于增强网络区域聚焦的感知能力.

图 7 中展示传统卷积和可变形卷积的对比,其中图 7(a)展示传统卷积采样过程;图 7(b)展示可变形卷积采样过程.传统卷积对输入的特征张量在固

定位置采样,缺乏处理几何变换的内在机制.可变形卷积对卷积核添加水平和垂直的偏移量,使卷积核的采样点发生偏移,集中在神经网络的感兴趣区域.图 7(c)展示可变形卷积示意图,其中箭头代表神经网络自学习的偏移量.可变形卷积具体实现过程如下:首先,可变形卷积学习输入的特征张量得到特征张量的偏移位置坐标;随后,对特征张量按照偏移位置坐标进行调整得到已变形的特征张量;最后,传统卷积对已变形的特征张量提取特征值.



(a)传统卷积采样过程　(b)可变形卷积采样过程　(c)可变形卷积示意图

图 7　传统卷积与可变形卷积对比图

3.3.3　局部-全局关系耦合模块

当同一场景中不同位置处于同一深度时,聚焦区域的离散分布对网络模型的构建是一个新的挑战,亟需一种统筹局部和全局信息的模型来解决该问题.本文提出局部-全局关系耦合模块对聚焦信息的局部和全局进行建模,并将两部分特征有效耦合,从而增强模型的几何特征表达能力,提升三维形貌重建精度.设输入到局部-全局耦合模块的特征信息为 F,首先通过特征提取子模块得到新的特征 F_{local} 和 F_{global},再分别将 F_{local} 输入局部关系建模分支得到局部特征,F_{global} 输入全局关系建模分支得到全局特征.随后将局部特征和全局特征耦合得到局部-全局特征 F_{cou}:

$$F_{\text{cou}}=f(l(F_{\text{local}}),g(F_{\text{global}})) \tag{13}$$

其中,$l(\cdot)$表示局部关系建模过程,$g(\cdot)$表示全局关系建模过程,$f(\cdot)$表示局部-全局关系耦合过程.

在局部关系建模分支中本文采用深度卷积(Depth-Wise Convolution,DWConv),可保证网络模型性能的同时降低模型的参数量.深度卷积作为传统卷积的空间维度分解,具有更加高效与轻量化的特点.在全局关系建模分支中,本文借鉴文献[48]中多头自注意力机制[48]并对其进行重新设计.首先使用点向卷积(Point-Wise Convolution,PWConv)对输入特征进行通道扩张,再对其通道分割为 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 三部分矩阵,随后通过自注意力机制计算特征的全局关系.其中 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 计算方式如下式所示:

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{14}$$

其中,d_k 代表 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 向量维度.自注意力机制通过检索全局特征分配权重,擅长捕捉全局特征之间的相关性,进而实现多景深图像序列之间不同关系的特征互补,提升特征信息的丰富程度,进一步增强了网络模型的性能.

3.4　深度图生成

通过 DSFF-Net 网络预测得到的深度图 D_i 可能会存在噪声和边缘误差等问题,本文利用全聚焦图像作为场景先验信息指导修复深度图[49],其主要来源分为测试及评价数据集中的场景图、光场数据集的中心视图和 RDF 多聚焦融合算法[17]生成的全聚焦图像:

$$D=GT(D_i,I_f) \tag{15}$$

其中,I_f 是根据多景深图像序列得到的全聚焦图像,$GT(\cdot)$代表基于融合图像的滤波函数[50].

首先,构建初始深度图 D_i 的代价量 C.根据初始深度给每个像素点分配代价标签 l,代价量 C 存储为三维矩阵包含每个像素点 i 的代价标签 l.其次,通过给定的引导图 I_f 引导过滤代价量 C:

$$C'_{i,l}=\sum_j W_{i,j}(I_f)C_{j,l} \tag{16}$$

$$W_{i,j}(I_f)=\frac{1}{|\omega|^2}\sum_{k:(i,j)\in\omega_k}\left(1+\frac{(I_i-\mu_k)(I_j-\mu_k)}{\delta_k^2+\varepsilon}\right) \tag{17}$$

其中,C'表示总代价量,$W(I_f)$来自引导图的权重系数,i 和 j 表示像素索引值,μ_k 和 δ_k 分别为引导图 I_f 中心像素点为 k 的方形窗口 ω_k 的均值和方差,$|\omega|$ 表示窗口中像素个数,ε 表示平滑因子.最后,对总代价量 C'汇总最大值得到深度图 D:

$$D=\arg\max_l C'_{i,l} \tag{18}$$

4　实验与结果

4.1　实验设置

本文共收集富纹理图像数据 3600 张,通过旋转、翻转和缩放等数据增广方式共生成 12 000 张大小为 256×256 背景纹理图.通过初始形貌函数生成 100 张 256×256 基础深度图,对其进行随机组合辅

以形貌核函数的随机参数共生成 12 000 组深度图.背景纹理图与深度图一一组合生成 12 000 组多景深图像序列,每组数据中包含 100 张大小为 256×256 多景深图像、256×256 的深度图和 256×256 的全聚焦场景图,其中部分数据集形貌如图 8 所示.

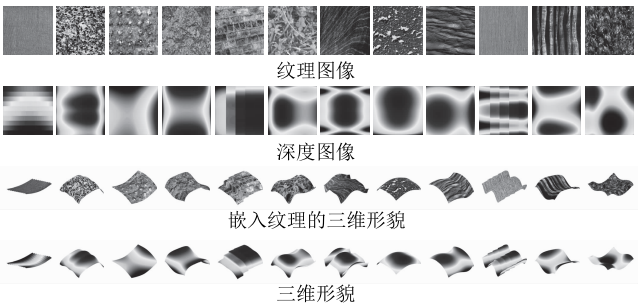


图 8 部分数据集

本文使用模拟生成的 10 000 组多景深图像序列训练网络,多景深图像序列作为网络模型输入,深度图作为网络模型的标签信息,共进行 100 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 10^{-4},其余参数皆为默认参数.训练过程中随机进行图像增强(整体旋转和图像序列倒转),批处理大小设置为 2.本文借助 pytorch 框架搭建网络模型,采用 3090 GPU 训练 DSFF-Net.

4.2 消融实验

为进一步探索 DSFF-Net 网络结构的合理性,本文在使用公共数据集 FoD500 中对模型进行消融测试,从模型性能(*MSE*)、模型参数量(*Param*)和推理时间(*Runtime*)三个方面进行分析.为移除设备自身算力影响,在相同环境中重复 100 次实验,并计算其均值.消融实验中依据消融方式不同主要分为"叠加式"和"替换式",二者之间根据是否对原始卷积层替换进行区分.

如表 4 所示,在消融实验 1(U)中仅使用 U 型主干网络结构进行训练;在消融实验 2(U+A)中使用局部-全局耦合模块插入 U 型主干网络的瓶颈模块;在消融实验 3(U+D)中使用可变形卷积添加在 U 型主干网络的每层收缩路径中;在消融实验 4(U+A+D)中结合消融实验 2 和消融实验 3 的改变;消融实验 5(U_A)中使用局部-全局耦合模块替换 U 型主干网络中瓶颈模块中间的卷积层;在消融实验 6(U_D)中使用可变形卷积替换位于 U 型主干网络中每层收缩路径最后的卷积层;在消融实验 7(U_A_D)中结合消融实验 5 和 6 的改变.

表 4 消融实验结果

模型名	结构			模型参数量/个	运行时间/s	模型性能	测试方式
	U 型主干	可变形卷积	局部-全局关系耦合模块			*MSE*	
U	√			26 855 681	**0.576**	0.0095	/
U+A	√		√	27 320 321	0.668	0.0081	
U+D	√	√		28 274 898	0.796	0.0072	叠加
U+A+D	√	√	√	30 634 706	0.855	**0.0067**	
U_A	√		√	**24 960 513**	0.632	0.0085	
U_D	√	√		27 072 594	0.788	0.0074	替换
U_A_D	√	√	√	25 177 426	0.800	0.0068	

通过实验结果可以看出,可变形卷积有助于提升网络精度;局部-全局关系耦合模块有效提升网络精度的同时可以有效降低参数量,但由于计算方式的复杂度高于普通卷积,所以推理时间较长.

本文采用模型(U_A_D),该模型同时使用可变形卷积和局部全局耦合模块达到性能次优,模型参数量整体达到次优,运行时间稍长.通过以上七组消融实验,可以看出本文提出的 DSFF-Net 网络模型可以得到较好网络性能,同时平衡了模型参数量和运行时间的问题.

4.3 对比实验结果及分析

本节从四个方面进行对现阶段三维形貌重建方法进行定性和定量分析,根据方法类型和适用场景分为传统模型对比、基于深度学习的三维形貌重建模型对比、数据集对三维形貌重建模型的引导实验和大景深场景实验四部分进行定性和定量对比.

本节主要选择 Base-Line Datasets[51]、4D Light Field Datasets[34]、The Synthetic POV-Ray datasets[35]、SLFD and DLFD datasets[36]、FoD500[11]、HCI 4D Light Field Datasets[52]、Mobile Depth datasets[53]和 MDFI Datasets 数据集进行三维形貌测试.通过对比传统三维形貌重建模型(GC[18]、RDF[17]、RFVR-SFF[7])和基于深度学习的三维形貌重建模型(DDFF-Net[28]、Defocus-Net[11]、AiFDepth-Net[10]、FV-Net[9])来对比评估本文所提出数据集 MDFI Datasets 和 DSFF-Net 算法的有效性.传统三维形貌重建对比模型的算法参数选择见表 5.

表 5　对比算法选择

算法模型	发表期刊、会议	发表年份	参数选择
GC	Journal of Visual Communication and Image Representation	2018	$thresSmooth=32$、$thresSmooth=32$、$mask=3$、$G=1$、$lambda=0.01$
RDF	IEEE Transactions on Image Processing	2020	$Tmad=0.1$、$Tbokeh=0.15$、FM($r_1=1$,$r_2=2$,$r_3=5$)
RFVR-SFF	IEEE Transactions on Image Processing	2021	$lambda=0.3$、$alpha=0.1$、$beta=1.5$、$gamma=2.5$、$itr=8$、$nei=2$

其中,Base-Line Datasets 选择不同纹理、形状和噪声水平进行模拟聚焦.该数据集中包含 15 种基础形貌的图像,例如:基础圆锥形貌、球体形貌和余弦形貌等等.此数据常被作为基础测试来检测聚焦测量算子的可用性;4D Light Field Datasets 中包含了 25 种复杂现实场景,其中该数据集主要用于测试精细结构,弱纹理及光滑表面等情形;The Synthetic POV-Ray Datasets 主要用于测试不同模型在前景、中景和背景之间的区分度,在该数据集上随机抽选 30 种场景作为测试;SLFD and DLFD Datasets 侧重弱纹理背景和镜面反射情况的测试,选择该数据集中 30 种场景作为测试.MDFI Datasets 随机抽样 15 组多景深图像序列数据作为实验补充,辅助验证各个三维形貌重建模型.以上数据集具有纹理、景深多样等特征,可用来验证传统三维形貌重建模型.另外,FoD500 数据集作为合成数据集,其样本多样且有准确的深度信息,适用于对比深度学习类三维形貌重建模型.而 HCI 4D Light Field Datasets 和 Mobile Depth datasets 则用来对比各个模型在多景深,大景深条件下的模型效果.

本节实验使用以下指标定量评估各三维形貌重建模型性能差异[28]:均方误差 *MSE*(Mean Square Error)、均方根误差 *RMSE*(Root Mean Squard Error)、峰值信噪比 *PSNR*(Peak Signal to Noise Ratio)、结构相似性 *SSIM*(Structural Similarity)、二维相关系数 *Correlation*(Two Dimensional Correlation Coefficient)、均方对数误差 *logMSE*(Mean Squared Logarithmic Error)、颠簸性(Bumpiness)、相对误差绝对值 *Abs.rel*(Absolute relative error)、相对误差平方值 *Sqr.rel*(Square relative error).

4.3.1　传统模型对比

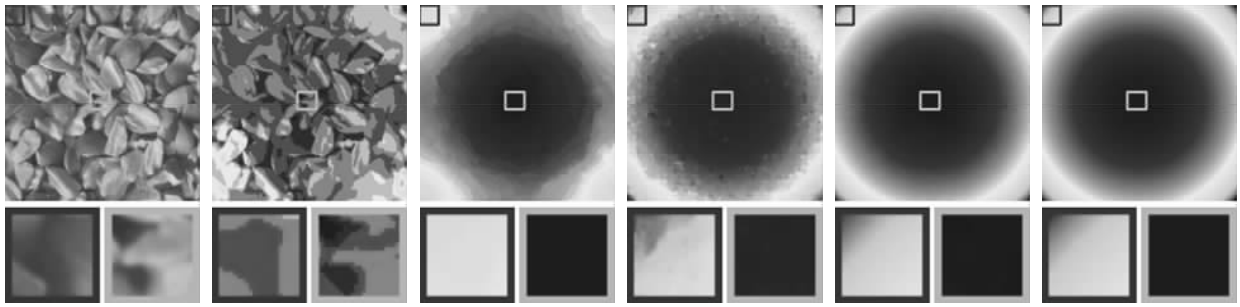
本文提出的网络模型在仅在 MDFI Datasets 进行训练,然后在五个数据集中进行客观和主观分析.

客观指标评价:如表 6 所示,在与传统三维形貌重建模型对比中可以看出本文提出数据集可以有效引导模型测量聚焦区域,但部分数据集预测深度图的光滑性仅次于借助图像分割的 GC 模型.

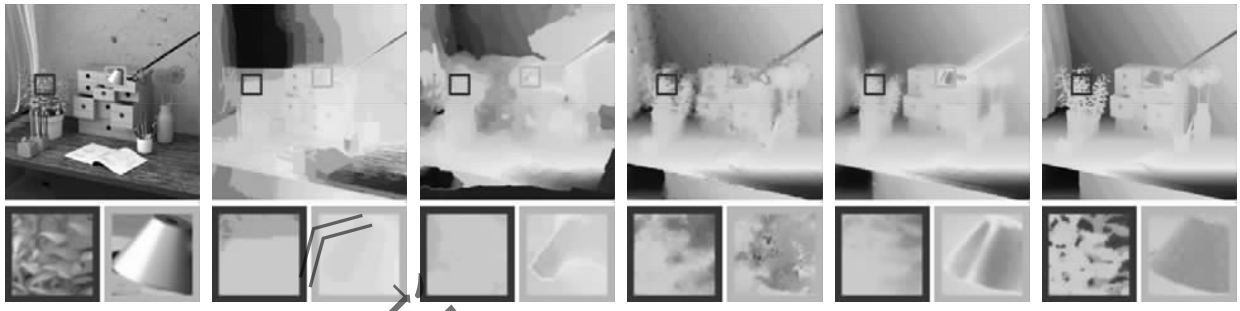
主观分析:图 9 中展示各个传统模型在不同数据集的三维形貌重建分析.Base-Line Datasets 和 MDFI Datasets 中仅考验各种模型聚焦区域测量和

表 6　传统模型对比

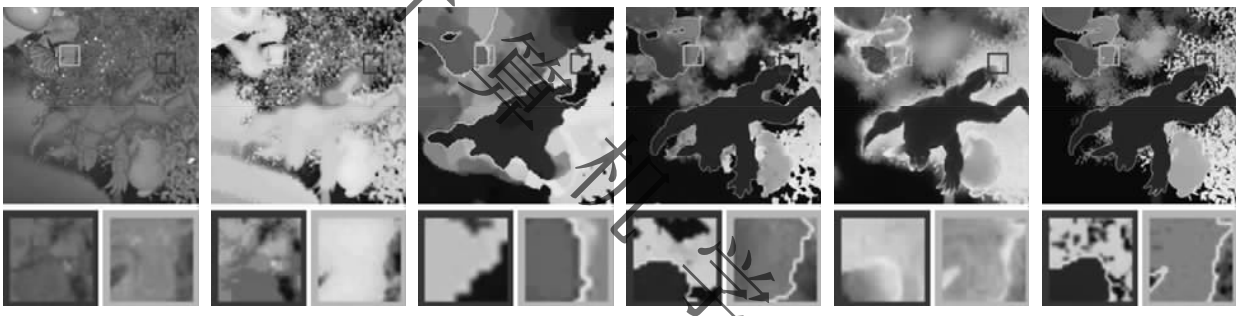
数据集	模型名	评价指标					
		RMSE	*PSNR*	*SSIM*	*Correlation*	*logMSE*	*Bumpiness*
Base-Line Datasets	GC	31.0549	18.6753	0.6945	0.7805	1.0583	4.4006
	RDF	7.5254	31.0699	0.9001	0.9617	0.1900	4.4377
	RFVR-SFF	2.6444	40.2357	0.9433	0.9958	0.0641	4.6793
	DSFF-Net	**1.2341**	**47.6830**	**0.9872**	**0.9998**	**0.0217**	**4.1893**
4D Light Field Datasets	GC	28.7857	19.1625	0.7880	0.6974	0.4862	**3.5328**
	RDF	14.1952	25.9345	0.8609	0.8165	0.4046	3.6308
	RFVR-SFF	7.6344	32.0967	0.8868	0.9346	0.0617	4.0753
	DSFF-Net	**5.7720**	**35.9345**	**0.9430**	**0.9558**	**0.0599**	4.1575
The synthetic POV-Ray Datasets	GC	36.1867	16.9871	0.4013	0.5383	1.0891	**4.1036**
	RDF	27.5430	19.3640	0.4631	0.6575	0.6163	4.3617
	RFVR-SFF	20.7392	21.9149	0.6148	0.8186	0.2310	4.4590
	DSFF-Net	**17.5819**	**23.3142**	**0.6176**	**0.8856**	**0.1553**	4.8628
SLFD and DLFD Datasets	GC	32.1132	18.1783	0.7691	0.7107	0.6034	**3.2031**
	RDF	16.7392	24.2084	0.8615	0.8065	0.2036	3.6453
	RFVR-SFF	9.6026	29.1287	0.8910	0.9300	0.0830	4.0774
	DSFF-Net	**8.1863**	**31.0837**	**0.9345**	**0.9524**	**0.0600**	4.2050
MDFI Datasets(our)	GC	32.4645	18.1829	0.7494	0.7065	0.7084	4.5445
	RDF	9.2067	29.5061	0.9222	0.8997	0.0997	4.6526
	RFVR-SFF	3.0771	38.7133	0.9523	0.9935	0.0084	4.6852
	DSFF-Net	**0.7557**	**50.8376**	**0.9962**	**0.9997**	**0.0026**	**4.5113**



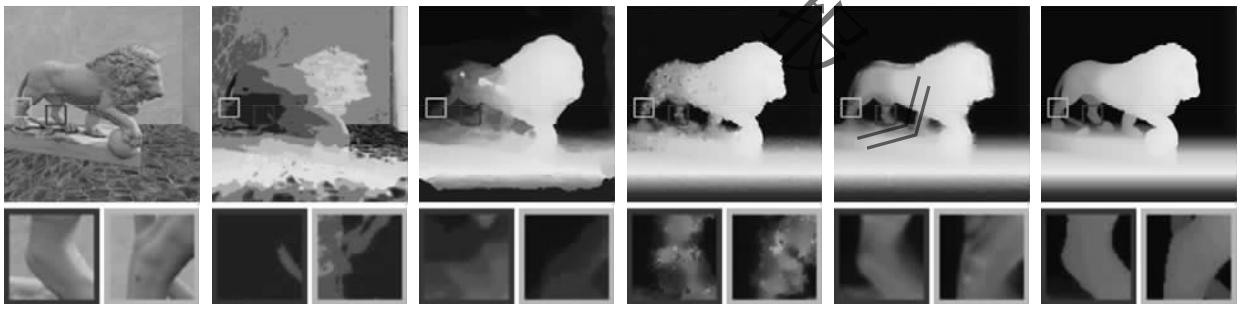
Base-Line Datasets



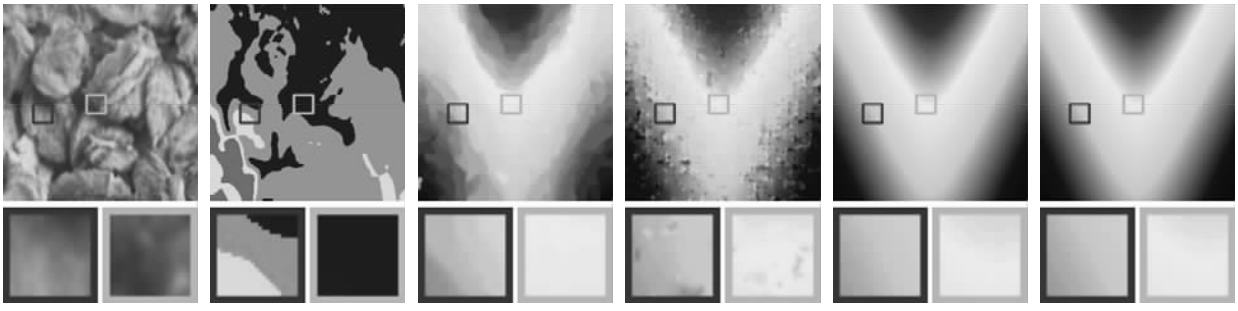
4D Light Field Datasets



The synthetic POV-Ray Datasets



SLFD and DLFD Datasets



MDFI Datasets (our)

(a) 场景图 (b) GC (c) RDF (d) RFVR-SFF (e) DSFF-Net (f) 标准深度图

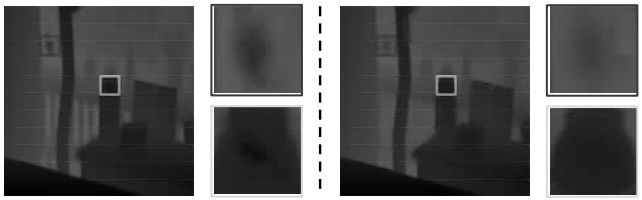
图 9 各传统模型在五个数据集的三维形貌重建分析

定位能力. GC 算法(图 9(b))关注场景中物体的分割结果,因此该算法易受到图像纹理影响;RDF 算法(图 9(c))和 RFVR-SFF 算法(图 9(d))在聚焦评价方面表现相对良好,但对于聚焦区域的定位不佳;本文提出的 DSFF-Net 模型(图 9(e))在聚焦区域评价和定位方面表现最佳.

4D Light Field Datasets 中可进一步验证各算法在实际场景中对于精细结构重建的表现. GC 算法(图 9(b))对于此类场景较为擅长,部分精细结构表现良好,但对于场景嵌套和弱纹理区域则表现较差;RDF 算法(图 9(c))在聚焦评价方面表现相对良好,但在实际场景中物体边缘保持性能较差;RFVR-SFF 算法(图 9(d))在聚焦区域评价和定位都表现良好,但对于整体而言存在部分区域噪点;本文提出的 DSFF-Net 模型(图 9(e))在聚焦评价和定位表现最好,尤其是对于精细结构的表达,但在部分弱纹理区域仍有提升空间.

The synthetic POV-Ray Datasets 用以验证各种算法在实际场景中对于微小结构和场景层次的表达. GC 算法(图 9(b))虽然可以捕捉到部分微小结构,但对于场景层次并不能有效分辨;RDF 算法(图 9(c))仅能判断出整体趋势,对于场景层次和细微纹理无法事先精确重建;RFVR-SFF 算法(图 9(d))在聚焦区域评价和定位均表现良好,尤其在深度图边缘保持方面;本文提出的 DSFF-Net 模型(图 9(e))尽管可以分辨某些细微结构,但在深度图的边缘保持方面尚有一定的提升空间,其主要原因在于引导滤波未能修复场景中颜色同质区域的边缘结构.

在 SLFD and DLFD Datasets 中可以进一步验证各种算法在实际场景中对于弱纹理背景的表现. GC 算法(图 9(b))在部分数据中表现良好,但对于场景中弱纹理背景表现不佳;RDF 算法(图 9(c))可以分辨出大区域的弱纹理背景,但对于整体而言分辨性较差;RFVR-SFF 算法(图 9(d))在弱纹理背景和深度细节方面表现良好,但存在部分噪声;本文提出的 DSFF-Net 模型(图 9(e))得益于数据集构建过程中对于场景前景和背景间隔的模拟和深度平滑的模拟,使得其在弱纹理背景区域表现较好;引导滤波有助于对弱纹理区域的深度信息进行误差修复;数据集的多样性和多维性使得其在场景结构有上佳表现. 图 10 展示了深度图修复效果.



(a) 初始深度图　　(b) 修复深度图

图 10　引导滤波修复效果

综上所述,本文提出的 DSFF-Net 模型无论在聚焦测量还是聚焦信息定位中均有良好性能,尤其对于精细结构和弱纹理区域的重建方面具有较大优势.

与传统三维形貌重建模型对比中设置噪声与尺度鲁棒性实验,以验证不同模型的鲁棒性. 本节的鲁棒性实验使用公共数据集 Base-Line Datasets、4D Light Field Datasets、The Synthetic POV-Ray datasets 和 SLFD and DLFD datasets.

噪声鲁棒性分析:将多景深图像序列加上不同程度的噪声,分别设置了 7 种不同的高斯噪声水平 0、0.04、0.08、0.16、0.2、0.5 和 1,选定 *RMSE* 为评价指标对比预测深度图和标准深度图之间的偏差;尺度鲁棒性分析:将数据中场景图像分别缩放或扩张为 128×128、256×256 和 512×512,以测试不同模型在各个尺度方面的性能,选定 *SSIM* 作为评价指标来判断多尺度下各个模型是否可以保留原始场景结构信息.

Base-Line Datasets、4D Light Field Datasets 和 The synthetic POV-Ray Datasets 中,本文提出的 DSFF-Net 模型在噪声鲁棒性实验中优于其他方法;多尺度鲁棒性实验中 128、256 和 512 尺度下优于其他模型. SLFD and DLFD Datasets 中,本文提出的 DSFF-Net 模型在噪声逐渐增强中仍表现出较高的预测精度,仅在 128 尺度下略逊色于 RFVR-SFF 算法.

如表 7 所示,本文提出的 DSFF-Net 模型在多场景下噪声与尺度鲁棒性高于相比的其他传统方法,而且随着噪声的增加或尺度的降低,该模型性能下降较为缓慢.

4.3.2　基于深度学习的三维形貌重建模型对比

本节将 DSFF-Net 与现有的基于深度学习的三维形貌重建方法进行对比,为使实验结果公平客观,本节使用公共数据集 FoD500 作为各个模型的数据驱动. 具体结果如表 8 所示,各个模型预测深度图共使用六组指标进行评价. 其中部分数据来自最新论

表 7 传统模型对比补充实验

数据集	模型名	抗噪性实验(*RMSE*)							多尺度实验(*SSIM*)		
		no-noise	0.04-noise	0.08-noise	0.16-noise	0.2-noise	0.5-noise	1-noise	128	256	512
Base-Line Datasets	GC	31.0549	29.2851	28.5833	27.1055	26.6324	26.2313	26.2603	0.5824	0.6945	0.4791
	RDF	7.5254	6.6158	6.8185	7.7201	6.5616	7.2413	7.2407	0.7809	0.9001	0.9333
	RFVR-SFF	2.6444	3.8838	4.9665	7.4996	9.0484	15.4434	21.1444	0.8856	0.9433	0.7627
	DSFF-Net	**1.2341**	**1.5431**	**1.7449**	**2.2026**	**2.4867**	**3.0353**	**3.9368**	**0.9262**	**0.9872**	**0.9843**
4D Light Field Datasets	GC	28.7857	31.9851	31.5419	33.8365	35.2053	33.6488	35.4327	0.5024	0.6656	0.7880
	RDF	14.1952	15.0633	15.9339	16.6255	17.0182	18.1281	19.6769	0.5592	0.7459	0.8609
	RFVR-SFF	7.6344	18.5505	23.1902	25.9138	26.6327	29.6114	31.8186	0.7153	0.7971	0.8868
	DSFF-Net	**5.7720**	**8.0515**	**9.8961**	**11.2622**	**11.7334**	**13.5621**	**15.0008**	**0.7351**	**0.8666**	**0.9430**
The synthetic POV-Ray Datasets	GC	36.1867	40.4512	39.6273	38.2517	36.8516	35.5131	35.8048	0.3641	0.4013	0.4431
	RDF	27.5430	31.3696	32.7742	33.4892	33.3854	33.9465	34.8914	0.3631	0.4631	0.6397
	RFVR-SFF	20.7392	25.8936	27.4615	28.7213	29.1467	30.2773	31.7073	0.5120	0.6148	0.7340
	DSFF-Net	**17.5819**	**22.0349**	**23.7014**	**24.8537**	**24.9478**	**25.8358**	**26.5744**	**0.5518**	**0.6176**	**0.8082**
SLFD and DLFD Datasets	GC	32.1132	32.3257	32.7613	33.9787	33.8596	35.4630	35.5295	0.4916	0.6555	0.7691
	RDF	16.7392	18.8880	20.5748	21.4399	21.3474	22.6572	23.6360	0.5379	0.7459	0.8615
	RFVR-SFF	9.6026	15.0272	19.4253	22.7574	23.8952	28.0351	31.0361	**0.6928**	0.7734	0.8910
	DSFF-Net	**8.1863**	**13.0562**	**15.9887**	**18.0589**	**18.8054**	**21.0741**	**22.4507**	0.6912	**0.8707**	**0.9345**

表 8 基于深度学习的三维形貌重建对比实验

模型名	*MSE*	*RMSE*	*Abs.rel*	*Sqr.rel*	*Bumpiness*	参数量	发表期刊、会议
DDFF-Net	0.0334	0.1670	0.17	0.0356	1.74	39806222	18ACCV
Defocus-Net	0.0218	0.1340	0.15	0.0359	2.52	**1508047**	20CVPR
AiFDepth-Net	0.0127	0.1043	0.06	0.3523	/	16533873	21ICCV
FV-Net	0.0188	0.1250	**0.14**	**0.0243**	1.45	15963225	22CVPR
DSFF-Net	**0.0068**	**0.0743**	0.49	0.0350	**0.56**	25177426	/

文[9]公布的数据.

由实验结果表明:本文DSFF-Net在*MSE*、*RMSE*和*Bumpiness*更具优势,从侧面验证了本文DSFF-Net模型可有效预测新场景中的深度信息的变化趋势,引入的全局聚焦信息使得预测深度变化较为平滑.与此同时,本文DSFF-Net模型在深度值预测的误差较大,后续需要持续优化.

4.3.3 数据集对三维形貌重建模型的引导实验

将本文提出的DSFF-Net模型分别在MDFI Datasets和FoD500数据集进行训练并在Base-Line Datasets中进行测试评价.

如表9所示,传统三维重建模型RDF和RFVR-SFF对于验证聚焦和离焦性能高于FoD500训练的DSFF-Net,而低于使用MDFI Datasets训练的DSFF-Net.由此可见:本文提出的MDFI Datasets相较于FoD500数据集更加关注场景中聚焦和离焦本身,而不是适应于特定场景的深度关联关系,因此可适用于更广泛的场景.

表 9 数据集对三维形貌重建模型的引导性实验

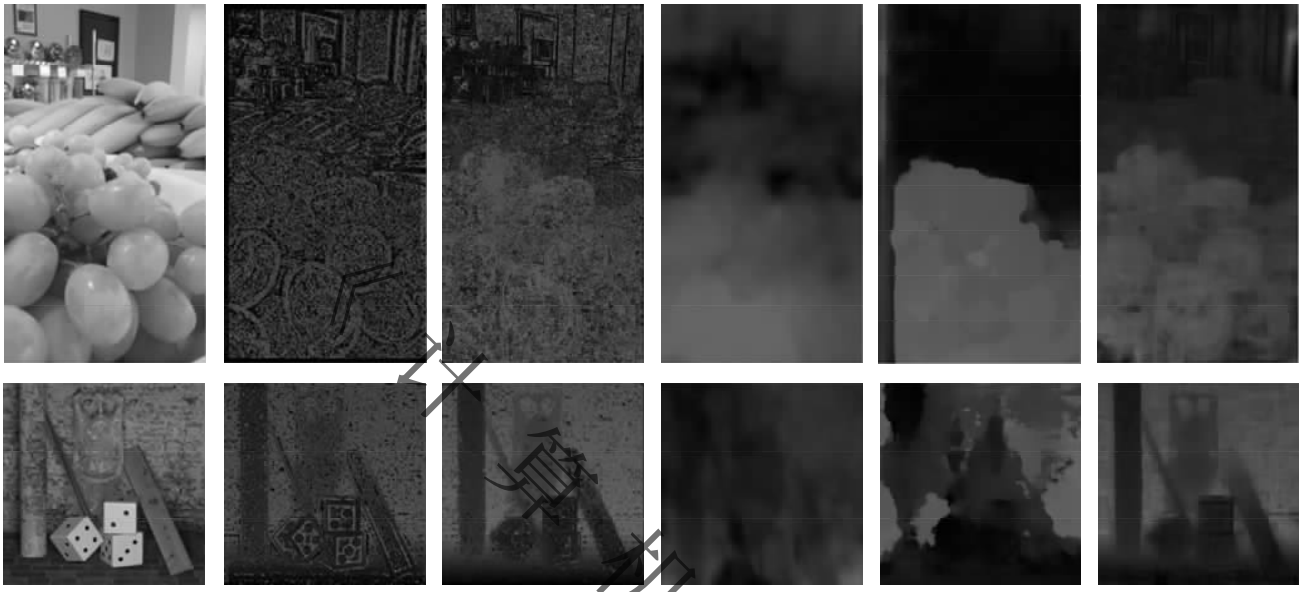
模型名	评价指标					
	RMSE	*PSNR*	*SSIM*	*Correlation*	*logRMSE*	*Bumpiness*
GC	0.0034	49.7775	0.9609	0.7086	0.3874	0.1379
RDF	0.0015	56.8204	0.9928	0.9225	0.1830	0.0564
RFVR-SFF	0.0016	55.7972	0.9948	0.9241	0.1937	0.2237
DSFF-Net(FoD500)	0.0086	41.6050	0.8332	-0.0368	0.8022	0.1167
DSFF-Net(MDFI Datasets)	**0.0010**	**59.1663**	**0.9960**	**0.9569**	**0.1321**	**0.0484**

4.3.4 大景深场景实验

将本文提出的MDFI Datasets数据集和DSFF-Net模型在HCI 4D Light Field Datasets和Mobile Depth Datasets进行评测,以验证DSFF-Net模型是否可以应用于大景深场景.为便于对比,本文将输入数据进行切片对齐和复制扩充,并在对比过程中删除了初始深度图修复过程,仅使用各个模型的聚焦区域测量方法,另外基于深度学习的模型算法仅在标注数据集训练未在测试过程中做任何域适应.其中DDFF-Net模型在DDFF12-Scene数据集训练,AiFDepth-Net在FoD500数据集训练,而本文DSFF-Net模型在本文提出的数据集MDFI Data-

sets 训练.

图 11 展示了本文提出的 DSFF-Net 模型与传统三维形貌重建模型(RDF、RFVR-SFF)和基于深度学习的三维形貌重建模型(DDFF-Net、AiFDepth-Net)在 HCI 4D Light Field Datasets 和 Mobile Depth datasets 数据集上预测的深度图. RDF 模型(图 11(a))可以有效判定场景中的边缘信息,但初始的聚焦点测量表现一般;RFVR-SFF(图 11(b))在聚焦测量及边缘检测过程中表现良好,但不可避免存在噪点;DDFF-Net 模型(图 11(c))和 AiFDepth-Net(图 11(d))模型可以简单鉴别场景中的聚焦趋势,但由于缺乏场景的先验知识,无法准确预测;本文提出模型(图 11(e))不仅可以精确判断场景的聚焦信息,且深度结果中对噪点有一定的抑制作用.



(a) 场景图　(b) RDF　(c) RFVR-SFF　(d) DDFF-Net DDFF12-Scene　(e) AiF Depth-Net FoD500　(f) DSFF-Net MDFI Datasets

图 11　各模型在大景深场景的三维形貌重建分析

5　总　结

多景深图像数据集(MDFI Datasets)的构建不仅为图像聚焦信息恢复三维形貌类方法提供基础测试数据集,而且最大限度地剥离了图像实际语义对深度信息的影响. 与此同时,本文提出的 DSFF-Net 网络架构具有兼顾局部聚焦信息定位与全局聚焦信息耦合的特性,实现了基于图像聚焦信息的三维重建方法由基准数据集训练到跨场景与跨领域应用,尤其是在噪声与尺度鲁棒性方面,DSFF-Net 方法相比于当前最先进的传统方法与深度学习类重建算法具有优异的抗噪性、良好的跨尺度适应性和较好的场景泛化性. 未来研究主要聚焦于如下两方面:

(1) 如何在基准数据集(MDFI Datasets)基础上根据不同应用领域构建场景适配的多种类型数据集,用于消除模拟数据与真实场景数据之间的数据鸿沟,进一步提升应用场景的重建精度.

(2) 现有的 DSFF-Net 网络架构在精细化细节重建方面尚有提升的空间,如何将图像丰富的语义信息加入网络结构中提升场景边缘区域的重建精度是下一步主要考虑的方向之一.

参　考　文　献

[1] Zhang Shun, Gong Yi-Hong, Wang Jin-Jun. The development of deep convolutional neural networks and its applications on computer vision. Chinese Journal of Computers, 2019, 42(3): 453-482(in Chinese)
(张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用. 计算机学报, 2019, 42(3): 453-482)

[2] Song Wei, Zhu Meng-Fei, Zhang Ming-Hua, et al. A review of monocular depth estimation techniques based on deep learning. Journal of Image and Graphics, 2022, 27(2): 292-328(in Chinese)
(宋巍, 朱孟飞, 张明华等. 基于深度学习的单目深度估计技术综述. 中国图象图形学报, 2022, 27(2): 292-328)

[3] Li Hai-Sheng, Wu Yu-Juan, Zheng Yan-Ping, et al. A survey

of 3D data analysis and understanding based on deep learning. Chinese Journal of Computers, 2020, 43(1): 41-63(in Chinese)
(李海生, 武玉娟, 郑艳萍等. 基于深度学习的三维数据分析理解方法研究综述. 计算机学报, 2020, 43(1): 41-63)

[4] Han Lei, Xu Meng-Xi, Wang Xin, et al. Depth estimation from multiple cues based light-field cameras. Chinese Journal of Computers, 2020, 43(1): 107-122(in Chinese)
(韩磊, 徐梦溪, 王鑫等. 基于光场成像的多线索融合深度估计方法. 计算机学报, 2020, 43(1): 107-122)

[5] Wang Jun, Zhu Li. 3D building facade reconstruction based on image matching-point cloud fusing. Chinese Journal of Computers, 2012, 35(10): 2072-2079(in Chinese)
(王俊, 朱利. 基于图像匹配-点云融合的建筑物立面三维重建. 计算机学报, 2012, 35(10): 2072-2079)

[6] Yan Tao, Qian Yu-Hua, Li Fei-Jiang, et al. Intelligent microscopic 3D shape reconstruction method based on 3D time-frequency transformation. SCIENTIA SINICA Informationis, 2023, 53(2): 282-308(in Chinese)
(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(2): 282-308)

[7] Ali U, Mahmood M T. Robust focus volume regularization in shape from focus. IEEE Transactions on Image Processing, 2021, 30: 7215-7227

[8] Zhang Rui, Li Jin-Tao. A survey on algorithm research of scene parsing based on deep learning. Journal of Computer Research and Development, 2020, 57(4): 859-875(in Chinese)
(张蕊, 李锦涛. 基于深度学习的场景分割算法研究综述. 计算机研究与发展, 2020, 57(4): 859-875)

[9] Yang Fengting, Huang Xiaolei, Zhou Zihan. Deep depth from focus with differential focus volume//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 12642-12651

[10] Wang Ning-Hsu, Wang Ren, Liu Yu-Lun, et al. Bridging unsupervised and supervised depth from focus via all-in-focus supervision//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021: 12621-12631

[11] Maximov M, Galim K, Leal-Taixé L. Focus on defocus: Bridging the synthetic to real domain gap for depth estimation //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 1071-1080

[12] Nayar S K, Nakagawa Y. Shape from focus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(8): 824-831

[13] Muhammad M, Choi T-S. Sampling for shape from focus in optical microscopy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 564-573

[14] Pertuz S, Puig D, Garcia M A. Analysis of focus measure operators for shape-from-focus. Pattern Recognition, 2013, 46(5): 1415-1432

[15] Thelen A, Frey S, Hirsch S, et al. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. IEEE Transactions on Image Processing, 2008, 18(1): 151-157

[16] Pech-Pacheco J L, Cristóbal G, Chamorro-Martinez J, et al. Diatom autofocusing in brightfield microscopy: A comparative study//Proceedings of the 15th International Conference on Pattern Recognition(ICPR-2000). Barcelona, Spain, 2000: 314-317

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

[18] Ribal C, Lermé N, Le Hégarat-Mascle S. Efficient graph cut optimization for shape from focus. Journal of Visual Communication and Image Representation, 2018, 55: 529-539

[19] Minhas R, Mohammed A A, Wu Q M J. Shape from focus using fast discrete curvelet transform. Pattern Recognition, 2011, 44(4): 839-853

[20] Yan Tao, Chen Bin, Liu Feng-Xian, et al. Multi-focus Image Fusion Model for Micro 3D Reconstruction. Journal of Computer-Aided Design and Computer Graphics, 2017, 29(9): 1613-1623 (in Chinese)
(闫涛, 陈斌, 刘凤娴等. 基于多景深融合模型的显微三维重建方法. 计算机辅助设计与图形学学报, 2017, 29(9): 1613-1623)

[21] Ali U, Mahmood M T. 3D shape recovery by aggregating 3D wavelet transform-based image focus volumes through 3D weighted least squares. Journal of Mathematical Imaging and Vision, 2020, 63(1): 54-72

[22] Yan Tao, Wu Peng, Qian Yuhua, et al. Multiscale fusion and aggregation PCNN for 3D shape recovery. Information Sciences, 2020, 536: 277-297

[23] Yan Tao, Hu Zhiguo, Qian Yuhua, et al. 3D shape reconstruction from multifocus image fusion using a multidirectional modified Laplacian operator. Pattern Recognition, 2020, 98: 107065

[24] Minhas R, Mohammed A A, Wu Q M, et al. 3D shape from focus and depth map computation using steerable filters// Proceedings of the 6th International Conference on Image Analysis and Recognition. Halifax, Canada, 2009: 573-583

[25] Gaganov V, Ignatenko A. Robust shape from focus via Markov random fields//Proceedings of the GraphiCon Conference. Moscow, Russia, 2009: 74-80

[26] Boshtayeva M, Hafner D, Weickert J. A focus fusion framework with anisotropic depth map smoothing. Pattern Recognition, 2015, 48(11): 3310-3323

[27] Mahmood M T. MRT letter: Guided filtering of image focus volume for 3D shape recovery of microscopic objects. Microscopy Research and Technique, 2014, 77(12): 959-963

[28] Hazirbas C, Soyer S G, Staab M C, et al. Deep depth from focus//Proceedings of the Asian Conference on Computer Vision. Perth, Australia, 2018: 525-541

[29] Saxena A, Schulte J, Ng A Y, et al. Depth estimation using monocular and stereo cues//Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad, India, 2007: 2197-2203

[30] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 3354-3361

[31] Cai Qin-Yi, Chen Zhong-Gui, Cao Juan. High-quality point cloud resampling method based on optimal transport theory. Chinese Journal of Computers, 2022, 45(1): 135-147 (in Chinese)
(蔡钦镒,陈中贵,曹娟. 基于最优传输理论的高质量点云重采样方法. 计算机学报, 2022, 45(1): 135-147)

[32] Wu Z, Song S, Khosla A, et al. 3D ShapeNets: A deep representation for volumetric shapes//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1912-1920

[33] McCormac J, Handa A, Leutenegger S, et al. Scenenet RGB-D: Can 5M synthetic images beat generic ImageNet pre-training on indoor segmentation?//Proceedings of the IEEE International Conference on Computer Vision. Boston, USA, 2017: 2678-2687

[34] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields//Proceedings of the Asian Conference on Computer Vision. Taipei, China, 2016: 19-34

[35] Heber S, Pock T. Convolutional networks for shape from light field//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 3746-3754

[36] Shi J, Jiang X, Guillemot C. A framework for learning depth from a flexible subset of dense and sparse light field views. IEEE Transactions on Image Processing, 2019, 28(12): 5867-5880

[37] Subbarao M, Lu M-C. Image sensing model and computer simulation for CCD camera systems. Machine Vision and Applications, 1994, 7(4): 277-289

[38] Wei Hao, Cui Hai-Hua, Cheng Xiao-Sheng, et al. Image defocus simulation technology applied to evaluation of focused morphology recovery algorithm. Acta Optica Sinica, 2019, 39(11): 140-148(in Chinese)
(韦号,崔海华,程筱胜等. 一种用于评价聚焦形貌恢复算法的图像离焦仿真技术. 光学学报, 2019, 39(11): 140-148)

[39] Sundaram H, Nayar S. Are textureless scenes recoverable?// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Juan, USA, 1997: 814-820

[40] Abdelmounaime S, He Dong-Chen. New Brodatz-based image databases for grayscale color and multiband texture analysis. International Scholarly Research Notices, 2013, 2013: 1-14

[41] Https://Kylberg.Org/Datasets/

[42] Subbarao M, Choi T-S, Nikzad A. Focusing techniques. Optical Engineering, 1993, 32(11): 2824-2836

[43] Pentland A P. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987, (4): 523-531

[44] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany, 2015: 234-241

[45] Zhu X, Hu H, Lin S, et al. Deformable ConvNets v2: More deformable, better results//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 9308-9316

[46] Wang Ke-Qi, Qian Yu-Hua, Liang Ji-Ye, et al. Local-global coupling relationship based low-light image enhancement. SCIENTIA SINICA Informationis, 2022, 52(3): 443-460(in Chinese)
(王克琪,钱宇华,梁吉业等. 局部-全局关系耦合的低照度图像增强. 中国科学: 信息科学, 2022, 52(3): 443-460)

[47] Xie Juan-Ying, Lu Yin-Yuan, Kong Wei-Xuan, et al. Butterfly species Identification from natural environment based on improved RetinaNet. Journal of Computer Research and Development, 2021, 58(8): 1686-1704(in Chinese)
(谢娟英,鲁银圆,孔维轩等. 基于改进 RetinaNet 的自然环境中蝴蝶种类识别. 计算机研究与发展, 2021, 58(8): 1686-1704)

[48] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 6000-6010

[49] Ali U, Lee I H, Mahmood M T. Guided image filtering in shape-from-focus: A comparative analysis. Pattern Recognition, 2021, 111: 107670

[50] Hosni A, Rhemann C, Bleyer M, et al. Fast cost-volume filtering for visual correspondence and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(2): 504-511

[51] Pertuz S, Puig D, Garcia M A. Reliability measure for shape-from-focus. Image and Vision Computing, 2013, 31 (10): 725-734

[52] Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for densely sampled 4D light fields//Proceedings of the Vision Modeling and Visualization. Lugano, Switzerland, 2013: 225-226

[53] Suwajanakorn S, Hernandez C, Seitz S M. Depth from focus with your mobile phone//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3497-3506

ZHANG Jiang-Feng, M.S. candidate. His research interests include deep learning, 3D shape reconstruction.

YAN Tao, Ph. D., associate professor. His research interests include deep learning, 3D shape reconstruction.

WANG Ke-Qi, M. S. His research interests include deep learning, image enhancement.

QIAN Yu-Hua, Ph. D., professor. His research interests include artificial intelligence, machine learning.

WU Peng, Ph. D., lecturer. His research interests include block chain, machine learning.

Background

The three-dimensional shape reconstruction based on the focus information from multi depth of field images studied in this paper belongs to the field of machine vision, and is widely used in the process of three-dimensional modeling and quantitative analysis in medical, biological, precision manufacturing and other fields. In response to such problems, it is common in the world to design field-adapted algorithm models for specific application scenarios. This paper takes the lead in researching from the perspective of deep learning, and proposes a construction method for multi-depth of field image focus information datasets and a robust depth network model. As one of the main research contents of the National Natural Science Foundation of China "Dynamic Three-Dimensional Reconstruction of Microscopic Images", this paper aims to provide a general datasets and a scene-adaptive depth model for 3D shape reconstruction. The previous research results of this research group were published in *Pattern Recognition*, *Information Sciences*, *SCIENTIA SINICA Informationis*.