JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
12
VOL.25

ISSN1006-8961
CN11-3758/TB

数字浮雕建模 *P2647*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第12期（总第296期）

2020年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

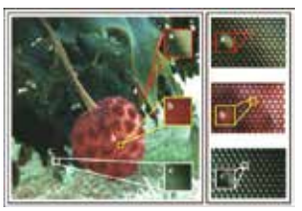
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

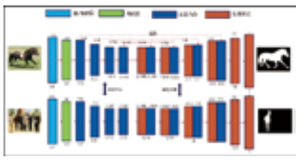
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年12月 第25卷 第12期（总第296期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

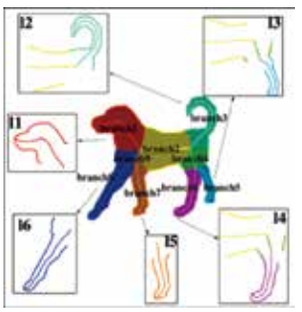
目次



光场显著性检测研究综述(第2465页)



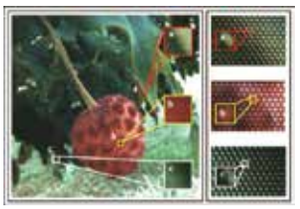
融合注意力机制与知识蒸馏的孪生网络压缩(第2563页)



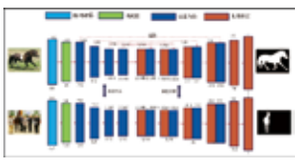
面向可解释性的物体拓扑结构骨架表征方法(第2587页)

Volume 25, Number 12
Published December 16, 2020

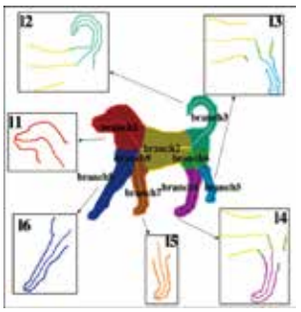
CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Review of saliency detection on light fields(P2465)



Combining attention mechanism and knowledge distillation for Siamese network compression (P2563)



Skeleton characterization of object topology toward explainability (P2587)

E-mail: jig@radi.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2020)12-2630-17

论文引用格式:Wang C, Zhang J and Gao J. 2020. Anti-specular light-field depth estimation algorithm. Journal of Image and Graphics, 25(12):2630-2646(王程,张骏,高隽. 2020. 抗高光的光场深度估计方法. 中国图象图形学报,25(12):2630-2646)[DOI:10.11834/jig.190526]

抗高光的光场深度估计方法

王程,张骏,高隽

合肥工业大学计算机与信息学院,合肥　230601

摘　要:**目的**　光场相机一次成像可以同时获取场景中光线的空间和角度信息,为深度估计提供了条件。然而,光场图像场景中出现高光现象使得深度估计变得困难。为了提高算法处理高光问题的可靠性,本文提出了一种基于光场图像多视角上下文信息的抗高光深度估计方法。**方法**　本文利用光场子孔径图像的多视角特性,创建多视角输入支路,获取不同视角下图像的特征信息;利用空洞卷积增大网络感受野,获取更大范围的图像上下文信息,通过同一深度平面未发生高光的区域的深度信息,进而恢复高光区域深度信息。同时,本文设计了一种新型的多尺度特征融合方法,串联多膨胀率空洞卷积特征与多卷积核普通卷积特征,进一步提高了估计结果的精度和平滑度。**结果**　实验在3个数据集上与最新的4种方法进行了比较。实验结果表明,本文方法整体深度估计性能较好,在4D light field benchmark合成数据集上,相比于性能第2的模型,均方误差(mean square error,MSE)降低了20.24%,坏像素率(bad pixel,BP)降低了2.62%,峰值信噪比(peak signal-to-noise ratio,PSNR)提高了4.96%。同时,通过对CVIA(computer vision and image analysis) Konstanz specular dataset合成数据集和Lytro Illum拍摄的真实场景数据集的定性分析,验证了本文算法的有效性和可靠性。消融实验结果表明多尺度特征融合方法改善了深度估计在高光区域的效果。**结论**　本文提出的深度估计模型能够有效估计图像深度信息。特别地,高光区域深度信息恢复精度高、物体边缘区域平滑,能够较好地保存图像细节信息。

关键词:深度估计;光场;抗高光;上下文信息;卷积神经网络

Anti-specular light-field depth estimation algorithm

Wang Cheng, Zhang Jun, Gao Jun

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei　230601, China

Abstract: **Objective** Image depth, which refers to the distance from a point in a scene to the center plane of a camera, reflects the 3D geometric information of a scene. Reliable depth information is important in many visual tasks, including image segmentation, target detection, and 3D surface reconstruction. Depth estimation has become one of the most important research topics in the field of computer vision. With the development of sensor technology, light field cameras, as new multi-angle image acquisition devices, have increased the convenience of acquiring optical field data. These cameras can simultaneously acquire the spatial and angular information of a scene and show unique advantages in depth estimation. At present, most of the available methods for light field depth estimation can obtain highly accurate depth information in many scenes. However, these methods implicitly assume that objects are on a Lambertian surface or a uniform reflection coefficient surface. When specular reflection or non-Lambertian surfaces appear in a scene, depth information cannot be accurately obtained. Specular reflection is commonly observed in real-world scenes when light strikes the surface of an object, such as metals, plastics, ceramics, and glass. Specular reflection tends to change the color of an object and obscure its tex-

收稿日期:2019-10-15;**修回日期**:2020-01-14;**预印本日期**:2020-01-21

基金项目:国家自然科学基金项目(61876057,61403116);国家留学基金项目(201806695016)

Supported by:National Natural Science Foundation of China (61876057, 61403116); National Scholarship Foundation of China (201806695016)

ture, thereby leading to local area information loss. Previous studies have shown that the specular region changes along with angle of view. Furthermore, we can speculate on the location of the specular area based on the context information of its surroundings. Inspired by these principles, we propose an anti-specular depth estimation method based on the context information of the light field image. In this way, this method can improve the reliability of the algorithm in handling problems associated with specular reflection. **Method** Based on the changes in the change of an image with the angle of view, we design our network by considering the light field geometry, select the horizontal, vertical, left diagonal, and right diagonal dimensions, and create four independent yet identical sub-aperture image processing branches. In this configuration, the network generates four directional independent depth feature representations that are combined at a later stage. We also use a fixed light direction, due to the obstruction of the front object or the incident angle of the light, smooth surface at the same depth level, not all areas will appear as highlights. In addition, the degree of reflection of specular on the smooth surface is different, indirectly showing the geometric characteristics. Therefore, we process each sub-aperture image branch via dilated convolution, which expands the network receptive field. Our constructed network obtains a wide range of image context information and then restores the specular region depth information. To improve the depth estimation accuracy in the specular area, we apply a novel multi-scale feature fusion method where the multi-rate dilated convolution feature is connected to a multi-kernel common convolution feature to obtain the fusion features. To enhance the robustness of our depth estimation, we use a series of residual modules to reintroduce part of the feature information that is lost by the previous layer convolution in the network, learn the relationship among the fusion features, and encode such relationship into higher-dimension features. We use Tensorflow as our training backend, the Ker as programming language to build our network, Rmsprop as our optimizer, and set the batch size to 16. We initialize our model parameters by using the Glorot uniform distribution initialization and set our initial learning rate to 1E-4, which decreases to 1E-6 along with the number of iterations. We use the mean absolute error (MAE) as our loss function given its robustness to outliers. We use an Intel i7-5820K@ 3.30 GHz processor with GeForce GTX 1080Ti as our experimental machine. Our network trains 200 epochs for approximately 2 to 3 days. **Result** 4D light field benchmark synthetic scene dataset was used for quantitative experiments, and the computer vision and image analysis (CVIA) Konstanz specular synthetic scene dataset and real scene dataset captured by Lytro Illum were used for the qualitative experiments. We used three evaluation criteria in our quantitative experiment, namely, mean square error (MSE), bad pixel (BP), and peak signal-to-noise ratio (PSNR). Experiment results show that our proposed method has an improved depth estimation. Our quantitative analysis on 4D light field benchmark synthetic dataset shows that our proposed method reduces the MSE value by 20.24%, has a BP value (0.07) that is 2.62% lower than that of the second-best model, and a 4.96% PSNR value. Meanwhile, in our qualitative analysis of the CVIA Konstanz specular synthetic dataset and the real scene dataset captured by Lytro Illum, our proposed algorithm achieves ideal depth estimation results, thereby verifying its effectiveness in recovering depth information in the specular highlight region. We also perform an ablation experiment of the network receptive field expansion and residual feature coding modules, and we find that the multi-scale feature fusion method improves the effect of depth estimation in the highlight areas and greatly improves the residual structure. **Conclusion** Our model can effectively estimate image depth information. This model achieves a high recovery accuracy in recovering highlight region depth information, has a smooth object edge region, and can efficiently preserve image detail information.

Key words: depth estimation; light field (LF); anti-specular; context information; convolutional neural network (CNN)

0 引 言

光场(light field,LF)是矢量函数,描述了任意点处光的位置、方向、波长和时间信息,能够解释场景3维结构(Adelson 和 Bergen,1991),现已应用于3D场景重建(Kim 等,2013;Perra 等,2016)、材质识别(Wang 等,2016)、超分辨率重建(Bishop 等,2009;Wang 等,2018;Zhang 等,2020)、虚拟/增强现实技术(Huang 等,2015)、显著性检测(Li 等,2014;Zhang 等,2015,2017)和深度估计(Wanner 和 Goldluecke,2014;Jeon 等,2015;熊伟 等,2017)。

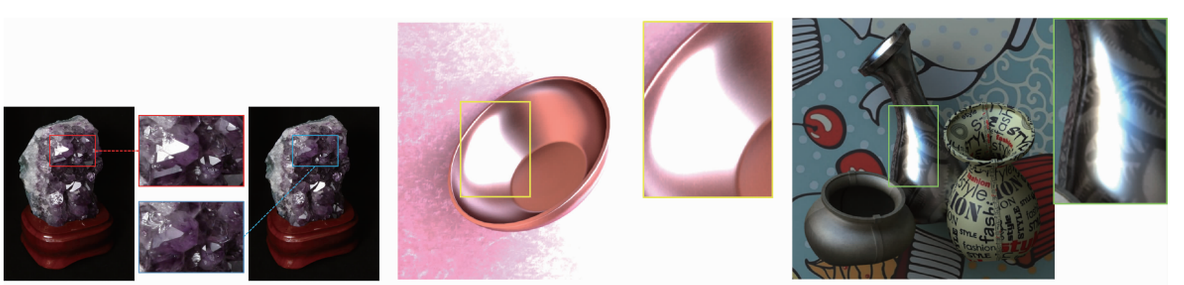
随着 Lytro(Ng,2006)和 Raytrix(Perwass 和 Wietzke,2012)等光场相机的商业化使用,光场数据

的获取更加便捷,设计有效且鲁棒的深度估计算法得到了更多的关注(Zhang 等,2016;Shin 等,2018;Wang 等,2015)。目前,多数先进的光场深度估计算法集中于对极平面图像(epipolar plane images,EPIs)特征的提取(Wanner 和 Goldluecke,2012a,2013a;Zhang 等,2016)。这些算法利用极线斜率、空间和角度方差(Tao 等,2013)等极平面图像几何特征,获取鲁棒的初始深度估计结果。然而,2 维 EPIs 难以充分表示高维光场信息,造成此类方法初始深度估计结果整体精确度不高。为得到较理想的结果,一般仍要进行后期优化处理,如 Heber 等人(2017)和 Feng 等人(2018)对已获取的初始深度图使用变分方法进行优化。此外,在 4D 光场双平面表示模型(Levoy 和 Hanrahan,1996)中,连续变化方向坐标,便可以得到该 4D 光场空间下的所有子孔径图像,其提供了目标场景密集采样的多方位视角(Adelson 和 Wang,1992;Dansereau 等,2013)。基于多视角的子孔径图像利用传统立体匹配原理可获取光场图像的深度图(Jeon 等,2015),然而此类方法的性能受到光场图像的窄基线问题和解码过程中光学畸变的影响(Feng 等,2018)。

卷积神经网络(convolutional neural networks,CNNs)在物体识别(Simonyan 和 Zisserman,2014)和语义分割(Long 等,2015)等领域取得了较大的成功。CNNs 具有强大的多层次特征学习能力,有助于探求图像数据的内部关系,学习到更具代表性的特征。Heber 等人(2016,2017)和 Feng 等人(2018)将其应用到光场图像的深度估计中,构建 CNNs 网络学习 4D 光场与在 2D 超平面上对应 4D 深度场的映射来预测深度值。然而,直接将高维光场图像作为 CNNs 输入具有较高复杂性,且 CNNs 训练需要足够多的训练数据,而现有光场数据集的数据量较少,限制了基于 CNNs 的光场图像深度估计研究(Wu 等,2017)。针对上述问题,Shin 等人(2018)根据光场图像的几何结构,设计了一种多支路全卷积神经网络,编码 EPIs 图像立体块实现深度估计,并通过提出光场特定的数据增强方法,弥补了训练数据的不足,从而快速且准确地估计深度。但该方法在训练数据中剔除了高光和无纹理区域,且卷积层卷积核大小均为 2 × 2,网络感受野较小,获取和使用图像的特征信息十分有限,限制了网络对图像更大区域信息的学习。因此,当图像场景中出现高光反射现象时,该方法无法得到精确的深度估计。

高光反射在现实场景中是一种常见的现象。当光线照射到某些材质(如金属、塑料、陶瓷和玻璃等)的物体表面时,会引起镜面反射,并在物体表面形成高光区域。物体表面的高光反射往往会改变物体表面的颜色、破坏物体的轮廓、遮挡物体表面的纹理,饱和的高光更是直接导致了局部区域信息的丢失(Shafer,1985)。

本文基于图像中高光区域随视角变化的原理(Tao 等,2016),受 Shin 等人(2018)方法的启发,利用光场图像多视角特性,选择水平、垂直、左对角线和右对角线 4 个方向,创建了 4 条独立但相同的子孔径图像处理支路。在这种结构下,网络生成了 4 个方向独立的深度特征表示,如图 1(a)所示。此外,若固定某个光照方向,则处于同一深度层面的光滑表面会因前置物的遮挡或光线入射角度等原因,并非全部区域均表现为高光,如图 1(b)所示。这些光滑表面不同区域表现出的高光程度,间接展现了光滑表面的几何特性(Shafer,1985),如图 1(c)所



(a) 不同视角下高光区域表征　(b) 同一深度平面并非全部表现为高光　(c) 光滑表面不同区域对高光的表现不同

图 1　高光反射现象举例

Fig. 1　Examples of highlight reflections ((a) characterization of highlight regions from different perspectives; (b) not all highlights appear at the same depth plane; (c) different areas of glossy surfaces behave differently in highlight)

示。本文在每条子孔径图像处理支路上，利用多膨胀率大小的空洞卷积获取多尺度特征，增大了网络感受野，网络获取到更大范围的图像上下文信息，进而恢复与正常区域在同一深度层面中的高光区域深度信息，提升了高光区域的深度估计质量。

本文与 Shin 等人(2018)方法的主要区别在于：1)在多支路输入的基础上，本文利用空洞卷积替换 Shin 等人(2018)采用的普通卷积的方式处理子孔径图像，达到扩大网络感受野的目的；2)本文设计了一种新型多特征融合方法，将多膨胀率空洞卷积特征与多卷积核普通卷积特征串联形成融合特征，提高了模型在高光区域的预测精度，增加了模型的泛化性；3)本文采用残差结构替换 Shin 等人(2018)网络后层的直连结构，融合网络深层特征与浅层特征，进一步提高了全局深度估计性能。

1　相关工作

自引入光场相机以来，基于光场图像的深度估计方法取得了较大发展。光场相机能够同时捕获空间光线的位置信息和角度信息，因此早期的方法扩展了传统多视角和立体视觉算法，利用多视角子孔径图像完成光场深度估计。Yu 等人(2013)将 3 维线性约束编码进光场图像中，然后通过子孔径图像之间的线匹配计算深度图，但它在视差范围较小的数据集上表现不佳。为了解决这一问题，Heber 和 Pock(2014)提出了一种新的多视图立体重建主成分分析(principal component analysis，PCA)匹配项。Jeon 等人(2015)对子孔径图像进行校正，并通过计算绝对颜色和梯度差之和，构建亚像素精度的成本量进行多视角立体匹配。尽管该方法深度信息提取精度较高，但需要对每个多视角图像对分别构造成本函数，时间复杂度高，且无法处理高光问题。

然而，将光场子孔径图像简单视为多视角图像有其局限性(Feng 等，2018)。首先，与传统多视角图像相比，光场图像的基线非常窄。其次，光场相机获取的光场图像包含由主透镜和微透镜引起的光学畸变。因此，由于空间域内的亚像素位移较小，亚孔径图像分辨率较低，传统的基于立体匹配的光场深度估计方法往往效果不佳。

深度学习技术已经应用到光场图像超分辨率重建(Bishop 等，2009；Wanner 等，2012a)、视角合成(Kalantari 等，2016)、单幅图像到光场图像的转换(Srinivasan 等，2017)，以及材料识别(Wang 等，2016)等任务中。对于深度估计，Heber 和 Pock(2016)、Heber 等人(2017)相继提出了一种端到端的编解码结构的深度估计网络和一种结合 CNN 和变分优化的方法，通过训练卷积神经网络来预测 EPI 极线的方向，利用高阶正则化的全局优化方法来优化网络输出结果。在此基础上，Heber 等人(2016)提出了 U 型全卷积网络，然而训练 U 型网络需要所有视角子孔径图像的视差图，这对现有光场数据集是无法实现的。Shin 等人(2018)根据光场图像的几何结构，设计了一种多支路全卷积网络编码 EPI 图像立体块的光场深度估计算法。

上述方法可以在某些特定场景中获取较为准确的深度信息。然而，这些方法均包含了一个隐含的假设：图像物体处于朗伯表面或者均匀反射系数表面(Cui 等，2017)。当场景中存在反射高光或非朗伯表面时，上述方法往往不能获得精确的深度信息。

于是，针对光场深度估计中的高光问题，一些学者提出了不同的解决方法。已有算法尝试通过阴影形状(shape-from-shading)的方法恢复这些区域的深度信息(Wu 等，2011；Langguth 等，2016；Oxholm 和 Nishino，2014)，然而这些方法的前提是需要已知或预估场景的光照情况(Cui 等，2017)。Johannsen 等人(2016)提出稀疏光场编码，将光场图像的高光表面分解为不同的叠加层，通过能量函数若干次优化迭代，结合超像素平滑和几何一致性约束，可以在无纹理区域高精度恢复深度。Wanner 和 Goldluecke(2013a)通过二阶结构检测极线斜率用来进行多层深度图的重建，成功精确地估计了镜面和透明物体的深度。然而上述算法没有对图像高光区域与其上下文信息的联系进行探索，忽略了图像上下文信息对解决高光问题的作用，缺乏对光场图像几何特性的利用，使得算法精度受限。同时，传统基于计算的方法计算量较大，计算速度较慢。

针对上述问题，本文提出了一种基于图像上下文信息的抗高光光场深度估计网络，利用图像中高光区域随视角改变而改变的原理，基于光场图像的几何结构，通过构建多视角分支路输入，获取高光区域在不同视角下的深度信息。同时，利用空洞卷积的形式扩大网络感受野，获取了更大范围的图像上下文信息，进而恢复与正常区域在同一深度层面中

高光区域深度信息。网络通过学习更大范围的多角度特征信息,有效缓解了高光现象对光场深度估计性能的影响。此外,为进一步提高高光区域预测精度,本文设计了一种新型的多尺度特征结合方式。对比于其他先进算法,本文网络有效缓解了高光现象的影响,在合成场景数据集上和真实场景数据集上取得了更优的结果。

2 本文网络结构

本文网络结构如图 2 所示。如前文所述,要考虑从多个角度观察高光区域,又要尽可能地扩大网络感受野,使得网络可以获取高光区域不同角度的特征信息和图像上下文信息,进而恢复出高光区域的深度信息。本文利用光场图像的多视角特性,选择不同视角的子孔径图像序列作为输入数据,创建了部分视角图像分支路输入,避免密集的视角图像造成的计算冗余。同时,高光现象是因为镜面反射产生的,即图像中高光区域处于同一深度平面,深度值连续,因此可以通过图像高光区域的上下文信息推断因高光现象损失的信息。本文受到空洞卷积(dilated convolution)在图像语义分割任务(Yu 和 Koltun,2015)中有效应用的启发,空洞卷积可以在不增加计算量的前提下,扩大网络感受野,使得网络能够处理更大范围的图像信息。本文采用多膨胀率空洞卷积的方式扩大网络感受野,获取包含较大感受野的网络特征,并设计一种多尺度特征新型融合方式,进一步提高高光区域深度估计精度。

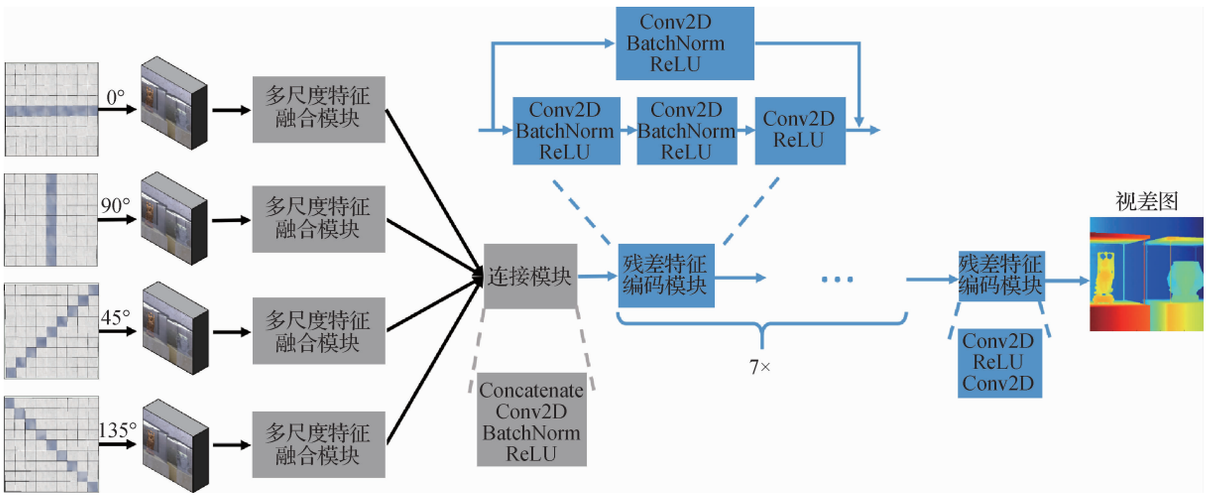


图 2 本文网络结构概览

Fig. 2 Network structure overview

2.1 多视角分支路输入模块

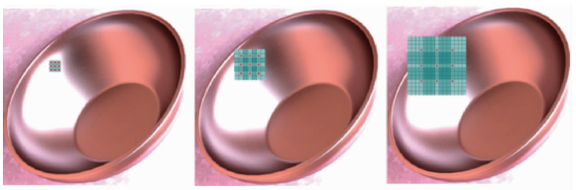
Shin 等人(2018)验证了多视角分支路输入对深度估计任务的有效性。如图 2 所示,本文保留 Shin 等人(2018)采用的多视角分支路输入,在角度分辨率为 9 ×9 的光场子孔径图像中,选择 0°、45°、90°和 135°方向的各 9 幅子孔径图像整合成序列,以 4 条支路分别输入网络。多视角分支路输入形式利用光场图像的多视角特性,可以提取高光区域的多角度特征信息,且避免输入全部视角导致数目太多造成的计算冗余。

2.2 网络感受野扩大模块

Shin 等人(2018)指出,卷积层的卷积核大小为 2 × 2,能够预测 ±4 个像素大小的图像视差范围,解决了光场图像基线较窄的问题。然而,卷积核较小会导致网络感受野较小,获取和使用图像的特征信息十分有限,限制了网络对图像更大区域信息的学习。针对该问题,本文使用空洞卷积的形式扩大网络感受野,使得网络可以获取到更大范围的图像信息。

图 3 是不同膨胀率大小的空洞卷积对高光区域的作用示意图。假设卷积核大小为 3 × 3,红点为实际卷积作用处。图 3(a)为膨胀率为 1 的空洞卷积,即普通卷积。可以看到卷积作用到整个高光区域。图 3(b)对应膨胀率为 2 的空洞卷积,实际的卷积核大小仍为 3 × 3,作用于 7 × 7 像素大小的图像块(对应图中 9 个红点)。此时,感受野实际大小为 7 × 7,换句话说,相当于卷积核的大小为 7 × 7,

可看到实际卷积点部分作用在高光区域外未发生高光区域，也就是高光区域邻域，进而获取高光区域的上下文信息，以此推断高光区域深度信息。图 3(c)为膨胀率为 4，感受野大小为 15 × 15 的空洞卷积。同上，通过进一步扩大感受野，获取高光区域的上下文信息，进而推断整个卷积作用图像块的深度信息。



(a) 膨胀率为 1　(b) 膨胀率为 2　(c) 膨胀率为 4

图 3　不同膨胀率大小的空洞卷积作用高光区域示意图

Fig. 3　Schematic diagram of the highlight areas of the dilated convolution with different dilation rates

((a) dilation rate 1; (b) dilation rate 2; (c) dilation rate 4)

本文设计了一种新型多尺度特征融合结构，进一步提高网络的预测精度。如图 4 所示，4 条输入支路通过卷积核大小为 2 ×2，膨胀率大小分别为 1、2、4、6 的空洞卷积层，每条输入支路也通过卷积核大小分别为 1 × 1、2 × 2、4 × 4、6 × 6 普通卷积层。随后，串联对应尺度的空洞卷积层特征与普通卷积层特征，得到融合特征。最后，融合特征经过卷积核大小为 1 × 1 的普通卷积层，目的是增加特征非线性，减少特征图通道数，进而减少了模型大小和计算量。同时，为保证串联过程中特征图大小一致，本模块所有卷积层使用了填 0 的方式。

2.3　残差特征编码模块

He 等人(2016)指出，残差结构可以在网络中重新引入因前层卷积丢失的部分特征信息，确保网络深层特征在细节上不弱于浅层特征，有效提升网络性能。受此启发，为了进一步提升预测精度，本文网络对串联后的多尺度深度特征，使用一系列残差模块学习特征之间的关系，将其编码为更高维的特征。如图 2 所示，残差特征编码模块共有 7 个残差块，每个残差块有两条支路，一条是单一普通卷积层的残差映射支路，卷积核大小为 3 × 3，保证了特征图连接时保持大小一致；另一条是 3 个连续普通卷积层组成的直连映射支路，前两层的卷积核大小均为 2 × 2，最后一层卷积核大小为 1 × 1，该层的作用是为了减少残差特征连接时通道数，控制计算量。为加快网络训练速度，残差特征编码模块所有卷积层使用不填 0 的方式。使用残差模块的另一个优势是，随着卷积层层数的大量增加，可以避免梯度弥散和梯度爆炸问题。

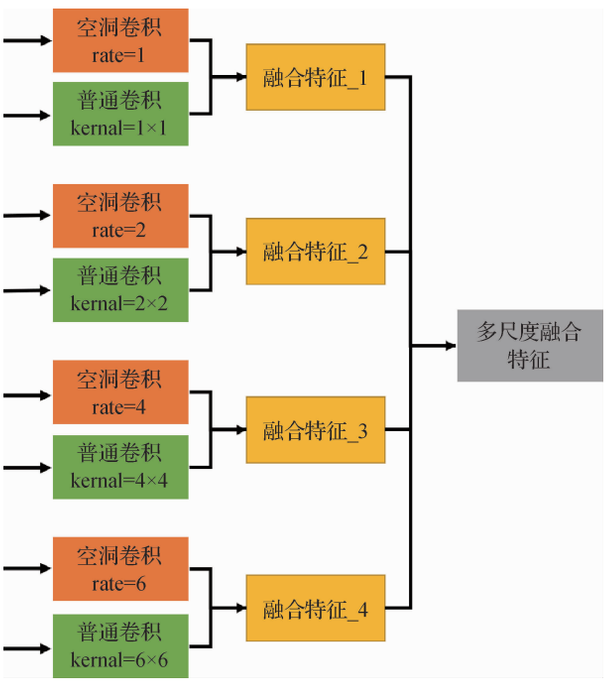


图 4　新型多尺度特征融合结构

Fig. 4　Multi-scale feature fusion

2.4　预测模块

为了估计视差信息，网络使用 Conv-ReLU-Conv 的结构作为预测模块，本模块普通卷积层的卷积核大小为 2 × 2，采用不填 0 的方式。

由 4D 光场双平面表示模型可知，4D 光场可表示为平行于普通像平面的若干视角的针孔视图的集合，如图 5 所示。平面 $\boldsymbol{\Pi}$ 为相机平面，平面 $\boldsymbol{\Omega}$ 为像平面，点 $\boldsymbol{P}$ 为任一空间点 $\boldsymbol{P}=(X,Y,Z)$。根据光场成像原理中视差与深度之间的几何关系(Levoy, 2006)，可以快速计算深度值，进而获取深度图。具体计算为

$$\Delta x = -\frac{f}{Z}\Delta u \tag{1}$$

式中，f 为双平面参数模型之间的距离(光场相机焦距)，Z 为 $\boldsymbol{P}$ 点深度值，Δx 表示 $\boldsymbol{P}$ 点在不同视角下的位移量(视差)，Δu 为光场相机微透镜阵列相邻两个微透镜中心点之间的距离。

2.5　训练细节

本文网络各层具体参数设置如表 1 所示，除预测模块，各卷积层后均有一个批处理归一化层

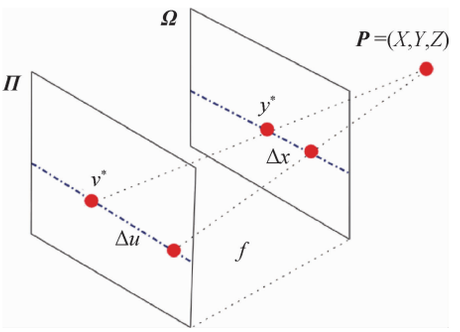


图 5　4 维光场双平面参数模型示意图

Fig. 5　4D light field two-plane parameterization model

(batch normalization, BN)和一个 ReLU 激活层。所有卷积层步长均为 1。同时,网络中多次使用了卷积核大小为 1×1 的卷积层,目的是增加特征非线性,降低参数,减少特征图通道数,限制模型大小和计算量。

实验使用 Tensorflow 作为训练后端,使用 Keras 语言搭建网络。优化器选择 Rmsprop 并设置批次处理大小为 16,初始化方式为 Glorot 均匀分布初始化,初始学习率设置为 0.000 1,并随着迭代次数减少至 0.000 001。本文使用平均绝对误差(mean absolute error, MAE)作为损失函数,其对异常值较为鲁棒(Goodfellow 等,2016)。实验机器配置为英特尔 i7-5820K@3.30 GHz 处理器,英伟达 GTX 1080Ti 显卡。网络共训练 200 epochs,2～3 d 时间。

表 1　网络各层具体参数

Table 1　Specific parameters of each layer of the network

网络结构	具体参数		输出大小
输入数据	0°/45°/90°/135°子孔径图像序列:空间分辨率为 512 × 512 像素,角度分辨率为 9 × 9		512×512×3×9
多视角分支路输入模块	多尺度特征融合结构(0°/45°/90°/135°)	dilated conv_1: filters=50, kernel_size=2×2, dilation_rate=(1, 1)	512×512×50
		dilated conv_2: filters=50, kernel_size=2×2, dilation_rate=(2, 2)	512×512×50
		dilated conv_3: filters=50, kernel_size=2×2, dilation_rate=(4, 4)	512×512×50
		dilated conv_4: filters=50, kernel_size=2×2, dilation_rate=(6, 6)	512×512×50
		conv2 d_1: filters=50, kernel_size=1×1	512×512×50
		conv2 d_2: filters=50, kernel_size=2×2	512×512×50
		conv2 d_3: filters=50, kernel_size=4×4	512×512×50
		conv2 d_4: filters=50, kernel_size=4×4	512×512×50
		concatenate_1: concatenate([dilated conv_1, conv2 d_1], axis=−1)	512×512×100
		concatenate_2: concatenate([dilated conv_2, conv2 d_2], axis=−1)	512×512×100
		concatenate_3: concatenate([dilated conv_3, conv2 d_3], axis=−1)	512×512×100
		concatenate_4: concatenate([dilated conv_4, conv2 d_4],axis=−1)	512×512×100
		concatenate_5: concatenate([concatenate_1, concatenate_2, concatenate_3, concatenate_4], axis=−1)	512×512×400
	concatenate_6: concatenate([0°/45°/90°/135°_concatenate_5], axis=−1)		512×512×1 600
连接模块	conv2 d__concate: filters=200, WKernel_size=1×1, padding='same'		512×512×200
残差特征编码模块(7 个残差块)	残差映射支路	conv2d_5: filters=100, kernel_size=3×3, padding='valid'	510×510×100
	直连映射支路	conv2d_6: filters=200, kernel_size=2×2, padding='valid'	511×511×200
		conv2d_7: filters=200, kernel_size=2×2, padding='valid'	510×510×200
		conv2d_8: filters=100, kernel_size=1×1, padding='valid'	510×510×100
	concatenate_7: concatenate([conv2d_5, conv2d_8], axis=−1)		510×510×200
	…		…
预测模块	conv2d_pre_1: filters=200, kernel_size=2×2, padding='valid'		496×496×200
	conv2d_pre_2: filters=1, kernel_size=1×1, padding='valid'		496×496×1

3　结果与分析

本文使用 4D light field benchmark 合成场景数据集(Honauer 等,2016)进行定量分析实验,分别使用 CVIA computer vision and image analysis Konstanz specular dataset (Alperovich 等,2018)合成场景数据集和 Lytro Illum 拍摄的真实场景数据集(Alperovich 和 Goldluecke,2017)进行定性分析实验。

定量分析实验中,本文采用 3 个评估标准:均方误差(mean square error,MSE)、坏像素率(bad pixel,BP)和峰值信噪比(peak signal-to-noise ratio,PSNR)。

均方误差用于描述估计结果的平滑度,均方误差越小,表明估计结果越好。计算为

$$MSE = \frac{1}{M}\sum_{m=1}^{M}(y_m - \hat{y}_M)^2 \tag{2}$$

式中,M 为测试样本总数,$\hat{y}_M$ 为预测值,y_m 为真实值。

坏像素率用于描述估计结果的准确度,其值越小,表示估计结果精确度越高。计算为

$$BP(t) = \frac{\{m \in M: |y_m - \hat{y}_M| > t\}}{M} \tag{3}$$

式中,t 表示阈值,本文采用的阈值 t 为 0.07。

峰值信噪比用于描述估计结果的视觉效果(图像质量),其值越大,表示估计结果视觉效果越好。计算为

$$PSNR = 10 \cdot \lg\left(\frac{MAX^2}{MSE}\right) \tag{4}$$

式中,MAX 为图像中最大像素值,针对浮点型数据,最大像素值为 1。本文 MAX 值为 1。

本文与 LF(Jeon 等,2015)、LF_OCC(occlusion-aware)(Wang 等,2015)、SPO(spinning parallelogram operator)(Zhang 等,2016)和 EPINET(epipolar network)(Shin 等,2018)等算法进行对比分析。值得一提的是,EPINET(Shin 等,2018)为避免高光导致的错误匹配,在网络训练阶段将包含反射高光和折射区域的训练数据,如玻璃、金属和无纹理区域剔除,同时移除了图像块中心像素和其他像素平均绝对误差小于 0.02 的无纹理区域。为公平比较,本文使用原训练数据对 Shin 等人(2018)的方法进行重新训练,其他参数设置与原方法保持一致,称为 EPINET_含高光。

由 2.3 节知,残差特征编码模块在卷积层使用不填 0 方式,网络输出结果大小逐层减少,使得最终本文网络输出深度图尺寸为 496 × 496 像素,比输入图像和真值深度图的尺寸(512 × 512 像素)要小,能够加快模型训练速度。在 4D light field benchmark 光场合成场景数据集中,提供分辨率大小为 482 × 482 像素的深度图用于定量分析。保证实验其他条件一致,该数据集提供工具(https://github.com/lightfield-analysis/evaluation-toolkit # 1-evaluate-light-field-algorithms)可以实现自动裁剪预测深度图和对应真值深度图的大小,提取相同位置相同大小(482 ×482 像素)的图像块进行对比分析。

3.1　训练数据

光场合成数据集有两个优点,一是具有视差/深度真值,二是光场中心视角图像与真值图像完全对齐。目前仅有两个公开可用的光场合成数据集:Wanner 等人(2013b)制作的 HCI(Heidelberg Collaboratory for Image Processing) light field benchmark 数据集和 Honauer 等人(2016)制作的 4D light field benchmark 数据集。本文采用 4D light field benchmark 数据集作为训练数据,该数据集图像的空间分辨率为 512 × 512 像素,角度分辨率为 9 × 9,包含了 24 种精心设计的场景并带有视差/深度真值图像。每种场景的物体、纹理均不相同。

本文遵循 Shin 等人(2018)的数据选择方式,从 4D light field benchmark 数据集中选择 16 幅光场图像用做训练,余下 8 幅光场图像作为测试和定量分析。训练图像随机采样为 32 × 32 像素大小的灰度图像块作为输入。由于训练图像只有 16 幅,本文使用数据增强方法,通过视角偏移、图像旋转、图像缩放、色彩值域变化、随机梯度变化、gamma 变换和翻转进行数据增强,使得训练数据从 16 幅增加到 4 608 幅,增大为原来的 288 倍。

3.2　消融实验

本文基于图像中高光区域随视角变化的原理(Tao 等,2016),利用光场图像多视角特性和上下文信息构建网络模型,研究在高光区域深度估计效果。其中,多视角分支路输入模块通过整合不同视角的子孔径图像输入序列,有利于网络获取高光区域多角度特征信息;网络感受野扩大模块通过获取图像高光区域更大范围的上下文信息,进而恢复该区域的深

度信息;同时,本文网络在网络后层使用一系列残差模块进一步学习深度特征之间的关系,探索该模块对模型性能的影响。下面对3个模块进行具体分析。

3.2.1 多视角分支路输入模块

本文保证实验各项条件与网络其余结构一致,研究输入子孔径图像(sub-aperture image,SAI)角度分辨率的不同对网络输出深度图质量的影响,选择角度分辨率分别为3 × 3、7 × 7、9 × 9进行对比分析。

表2是在4D light field benchmark合成数据集上,不同输入子孔径图像角度分辨率对网络输出深度图质量的影响。从表2可以看出,随着输入子孔径图像的角度分辨率的增加,MSE值和BP值逐渐减小,PSNR值逐渐增大,表示图像输出深度图的质量逐渐提高。

表2 不同输入SAI角度分辨率对网络输出深度图定量分析

Table 2 Quantitative analysis of network output depth map with different input SAI angular resolution

角度分辨率/像素	MSE	BP	PSNR/dB
3 × 3	3.974	12.83	17.989 7
7 × 7	2.156	7.34	20.469 8
9 × 9	**1.757**	**5.305**	**21.440 3**

注:加粗字体为每列最优值,加下划线字体为每列次优值。

图6是在Lytro Illum拍摄的真实场景数据集(Alperovich和Goldluecke,2017)owl场景下,不同角度分辨率输出深度图定性比较。可以看出,SAI角度分辨率越高,深度图预测效果越好。

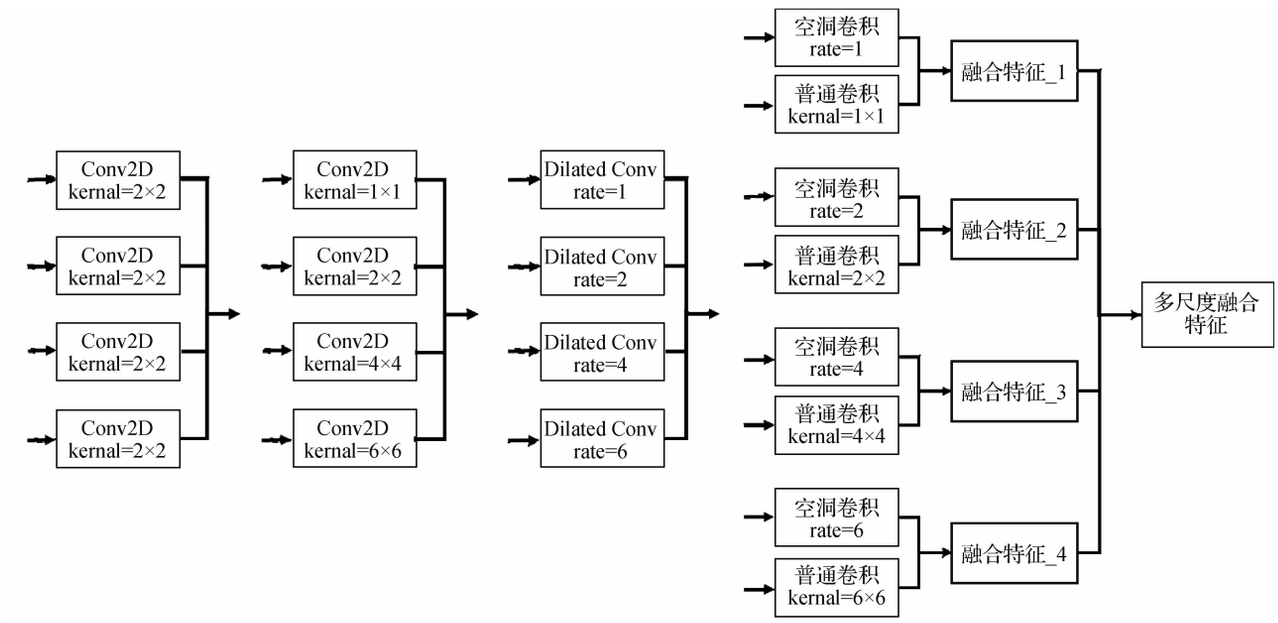
(a) 3 × 3 (b) 7 × 7 (c) 9 × 9

图6 不同输入SAI角度分辨率网络输出深度图定性比较

Fig. 6 Qualitative comparison of network output depth maps with different input SAI angular resolutions ((a) 3×3; (b) 7×7; (c) 9×9)

3.2.2 网络感受野扩大模块

如图7所示,本文保证实验各项条件与网络其余结构一致,研究不同网络感受野扩大方式对高光区域深度信息估计的影响,包括单一卷积核(卷积核大小均为2 × 2)的普通卷积、多卷积核(卷积核大小分别为1 × 1、2 × 2、4 × 4、6 × 6)的普通卷积、多膨胀率(膨胀率大小分别为1、2、4、6)的空洞卷积(卷积核大小2 × 2),以及多卷积核普通卷积与多膨胀率空洞卷积的多尺度特征融合结构。



(a) 单一卷积核普通卷积 (b) 多卷积核普通卷积 (c) 多膨胀率空间卷积 (d) 多尺度特征融合形式

图7 不同网络感受野扩大方式

Fig. 7 Different types of expansion methods of receptive fields ((a) single convolution kernel ordinary convolution; (b) multi-convolution kernel ordinary convolution; (c) multi-rate dilated convolution; (d) multi-scale feature fusion)

表 3 展示了不同感受野扩大方式在 4D light field benchmark 合成数据集上整体性能定量结果。从表中可以看到，特征融合的方式取得了最优性能。同时，随着网络感受野的扩大，评价标准 MSE 的值逐渐减小，评价标准 BP(0.07)逐渐升高，表明预测结果的平滑度提高了，但精确度下降了，预测结果视觉效果提升了。如前文所述，卷积核大小为 2 × 2 可以预测 ±4 个像素的视差范围，有效缓解了光场图像基线窄的问题，预测结果的精度高，故单一卷积(卷积核大小为 2 × 2)要比多卷积核普通卷积和多膨胀率空洞卷积的形式 BP(0.07)值要低。而随着网络感受野的扩大，每次卷积后的特征包含了更多的图像上下文信息，可推测在同一深度层面图像的任一处的深度信息，使得预测结果更为平滑，MSE 的值更低。特征融合的形式既包含了单一卷积核(2 × 2)普通卷积后的特征，可预测 ±4 个像素的视差范围这一优势，整体精度较高；又包含了多膨胀率空洞卷积特征，网络感受野较大，图像整体平滑度较高的优势，故在 BP(0.07)、MSE 和 PSNR 评价标准下取得最优。

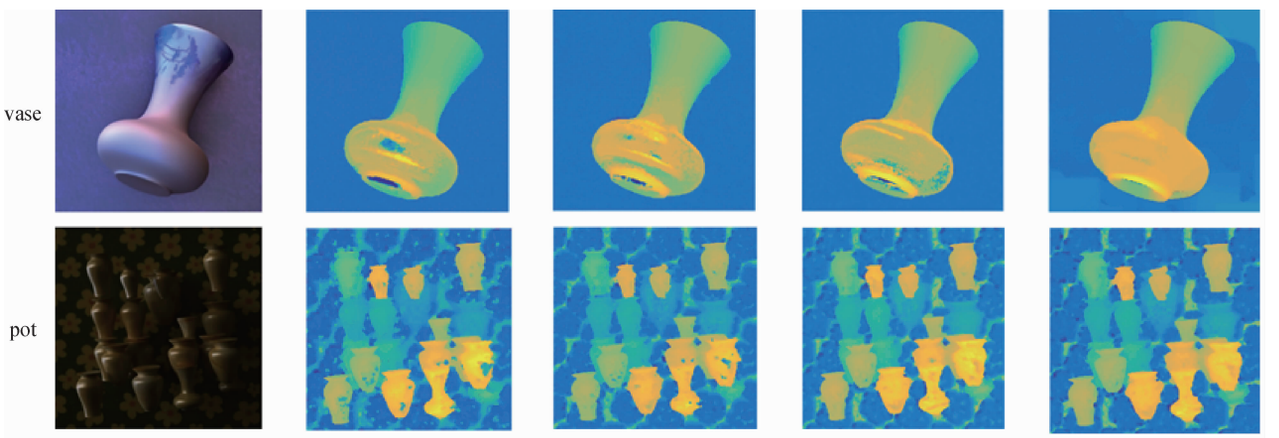
图 8 展示了不同感受野扩大方式在高光场景中的定性对比结果。本文选择高光区域小数量多和高光区域大数量少的两种高光场景进行对比。可以看出，随着感受野的扩大，预测结果更加平滑，尤其高光区域的预测结果更加理想。

表 3　不同感受野扩大方式性能对比

Table 3　Performance comparison of different expansion methods of receptive fields

感受野扩大方式	MSE	BP	PSNR/dB
单一卷积核普通卷积	2.038	5.709	20.830 7
多卷积核普通卷积	2.008	5.820	20.846 4
多膨胀率空洞卷积	1.988	6.078	20.889 6
多尺度特征融合形式	**1.757**	**5.305**	**21.440 3**

注：加粗字体为每列最优值，加下划线字体为每列次优值。



(a) 中心视角图像　(b) 单一卷积核普通卷积　(c) 多卷积核普通卷积　(d) 多膨胀率空间卷积　(e) 多尺度特征融合形式

图 8　不同感受野扩大方式在含有高光现象场景中的定性对比结果

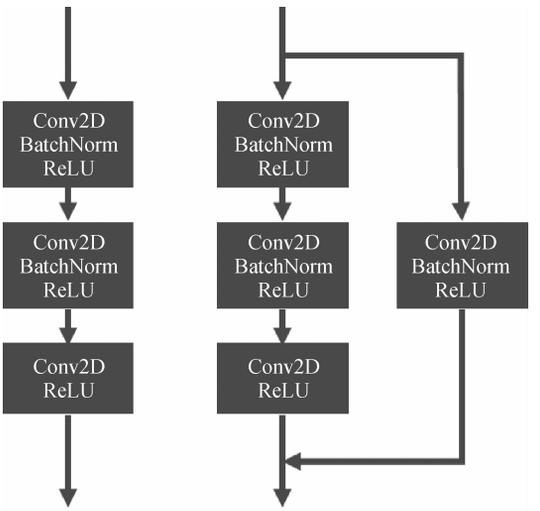
Fig. 8　Qualitative results of different receptive field expansion methods for scenes with highlights ((a) central view image; (b) single convolution kernel ordinary convolution; (c) multi-convolution kernel ordinary convolution; (d) multi-dilation rate dilated convolution; (e) multi-scale feature fusion construction)

3.2.3　残差特征编码模块

由上一节可知，网络感受野的扩大有助于图像平滑度的提升，在高光区域得到更理想的深度预测结果。然而表 3 数据显示，随着网络感受野的扩大，整体预测精度降低。因此，本文在网络后层使用残差模块，探究残差模块对网络预测精度的影响。针对直连映射与残差映射做具体分析，二者的结构对比如图 9 所示。

表 4 展示了直连映射和残差映射结构在 4D light field benchmark 数据集上整体性能结果。从中可以看出，残差映射的结构性能更优。残差结构在网络中重新引入因卷积丢失的部分特征信息，确保网络深层特征在细节上不弱于浅层特征，使得整体性能有较大提升。

图 10 展示了直连映射和残差映射在高光场景下的定性对比结果。从中可以看出，残差结构重新引入的浅层特征有助于提升预测精度，特别对图像中高光区域取得更理想的结果。



(a) 直连映射　(b) 残差映射

图 9　直连映射与残差映射的结构对比

Fig. 9　Comparison of direct connection mapping and residual mapping ((a) direct connection map; (b) residual map)

表 4　直连映射与残差映射整体性能结果

Table 4　Overall performance results of direct connection mapping and residual mapping

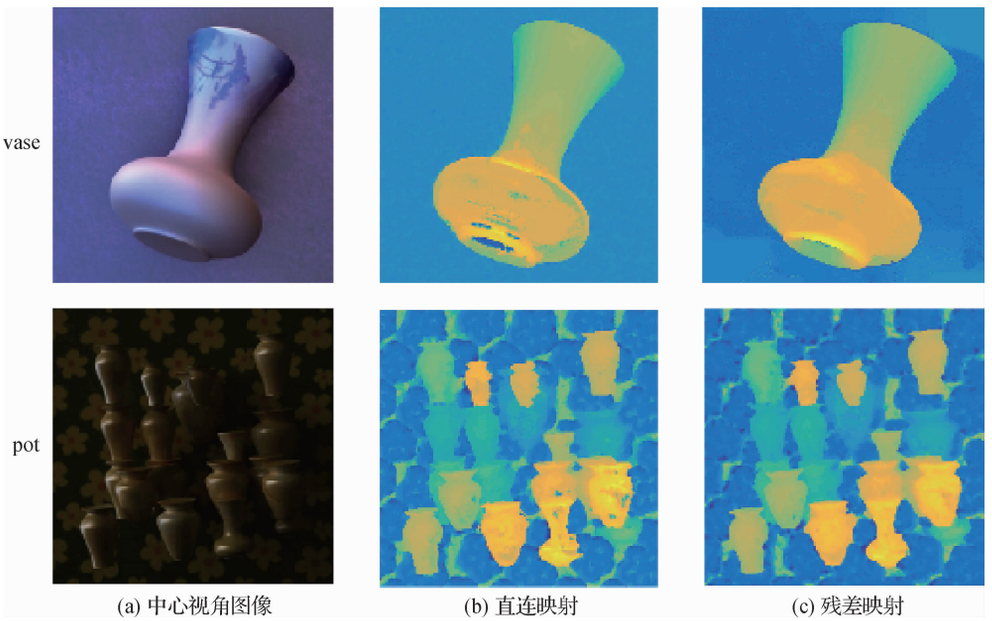
	MSE	BP	PSNR/dB
直连映射	**2. 034**	**5. 805**	20. 767 9
残差映射	1. 757	5. 305	**21. 440 3**

注:加粗字体为每列最优值。

3. 3　与先进算法的对比

3. 3. 1　在合成数据集上的结果

表 5、表 6 和表 7 分别为本文方法与其他先进算法在 4D light field benchmark 测试数据集上的定量分析比较。定量数据表明,本文网络的整体性能优于其他算法,在多数场景中取得最优,少数场景也取得了次优的性能。本文算法无需任何后续优化操作,在光场深度估计任务中存在性能优势。



(a) 中心视角图像　(b) 直连映射　(c) 残差映射

图 10　直连映射和残差映射在真实场景下的定性对比结果

Fig. 10　Qualitative results of direct connection mapping and residual mapping in specular-highlight scenes ((a) center view image; (b) direct connection map; (c) residual map)

表 5　实验结果在 MSE 评估标准下的对比

Table 5　Comparison of the MSE metric

算法	boxes	cotton	dino	sideboard	backgammon	dots	pyramids	stripes	整体性能
LF	17. 434	9. 168	1. 164	5. 071	13. 007	5. 676	0. 273	17. 454	8. 656
LF_OCC	9. 850	1. 068	1. 137	2. 304	21. 587	3. 301	0. 098	8. 131	5. 935
SPO	9. 107	1. 313	**0. 310**	1. 024	4. 587	5. 238	0. 043	6. 955	3. 572
EPINET_含高光	6. 440	**0. 270**	0. 940	0. 770	4. 700	3. 320	0. 020	1. 160	2. 203
本文	**5. 797**	0. 346	0. 508	**0. 737**	**3. 318**	**2. 546**	**0. 014**	**0. 793**	**1. 757**

注:加粗字体为每列最优值,加下划线字体为每列次优值。

表 6　实验结果在 BP(0.07)评估标准下的对比

Table 6　Comparison of the BP (0.07) metric

算法	boxes	cotton	dino	sideboard	backgammon	dots	pyramids	stripes	整体性能
LF	23.019	7.829	19.026	21.989	5.516	**2.900**	12.354	35.741	16.047
LF_OCC	36.520	6.218	14.913	18.495	19.006	5.822	3.172	18.408	15.319
SPO	15.889	2.594	**2.184**	9.297	3.781	16.274	0.861	14.987	8.233
EPINET_含高光	14.190	**0.810**	2.970	6.260	4.130	9.370	**0.540**	5.310	5.448
本文	**13.788**	0.882	2.989	**5.709**	**2.716**	10.807	0.664	**4.883**	**5.305**

注：加粗字体为每列最优值，加下划线字体为每列次优值。

表 7　实验结果在 PSNR 评估标准下的对比

Table 7　Comparison of the PSNR metric

/dB

算法	boxes	cotton	dino	sideboard	backgammon	dots	pyramids	stripes	整体性能
LF	7.586 0	10.377 3	19.340 5	12.949 1	8.858 2	12.459 6	25.638 4	7.581 1	13.098 8
LF_OCC	10.065 6	19.714 3	19.442 4	16.375 2	6.658 1	14.813 5	30.087 7	10.898 6	16.006 9
SPO	10.406 2	18.817 4	**25.086 4**	19.897 0	13.384 7	12.808 3	33.665 3	11.577 0	18.205 3
EPINET_含高光	11.911 1	**25.686 4**	20.268 7	21.135 1	13.279 0	14.788 6	36.989 7	19.355 4	20.426 8
本文	**12.368 0**	24.609 2	22.941 4	**21.325 3**	**14.791 2**	**15.941 4**	**38.538 7**	**21.007 3**	**21.440 3**

注：加粗字体为每列最优值，加下划线字体为每列次优值。

图 11 是在 4D light field benchmark 合成数据集下，本文网络的预测结果与其他先进算法预测结果的定性分析比较。从中可以看出，本文网络在各场景取得了更理想的预测结果，对比 EPINET_含高光(Shin 等，2018)算法更加精确。

如图 12 所示，在 boxes 场景中，本文结果保留了更加尖锐的边缘，更大程度上恢复箱子上网格形状，其他算法则在复杂的网格区域出现了不精确估计情况。此外，LF_OCC(Wang 等，2015)在 dots 场景中表现出最好的结果，原因可能是该算法使用离散标签表示深度进而取得了更好的结果。然而，此方法不适用于深度值连续的其他场景。

图 13 是本文方法在 CVIA Konstanz specular dataset 合成场景数据集上与其他算法预测结果的定性分析比较。本文选取高光区域小数量多(图 13 第 1 行)和高光区域大数量少(图 13 第 2 行)的两类典型高光场景进行比较。可以看出，本文方法在深度不连续区域有较强的鲁棒性，在纹理良好的区域和边缘区域具有较高的精度，在反射高光和无纹理区域较其他算法减少了预测误差，在深度边缘区域具有较高的平滑度，取得了更为理想的视差估计结果。

3.3.2　在真实数据集上的结果

图 14 展示了本文算法与 LF(Jeon 等，2015)、LF_OCC(Wang 等，2015)、SPO(Zhang 等，2016)和 EPINET_含高光(Shin 等，2018)算法在 Lytro Illum 拍摄的真实场景数据集(Alperovich 和 Goldluecke，2017)中的定性实验结果。本文选择了 cat、owl、koala 和 matreshka 等 4 种场景。从图 14 可以看出，本文算法能够准确地估计出物体的轮廓、平滑物体表面和背景上的视差结果。其中，EPINET_含高光(Shin 等，2018)受光照变化影响较大，SPO(Zhang 等，2016)仅估计了场景中物体的离散视差值，忽略了背景视差值。LF(Jeon 等，2015)正确估计了物体的轮廓，但视差图受噪音的影响较严重。LF_OCC(Wang 等，2015)虽然得出较为平滑的视差图，但忽略较多细节信息。在 owl 场景的预测结果中，本文网络正确预测了猫头鹰尖锐的耳朵形状和猫头鹰头与身体的平滑交界区域。在 matreshka 场景中，本文方法正确捕获了多个套娃轮廓的形状。在 koala 场

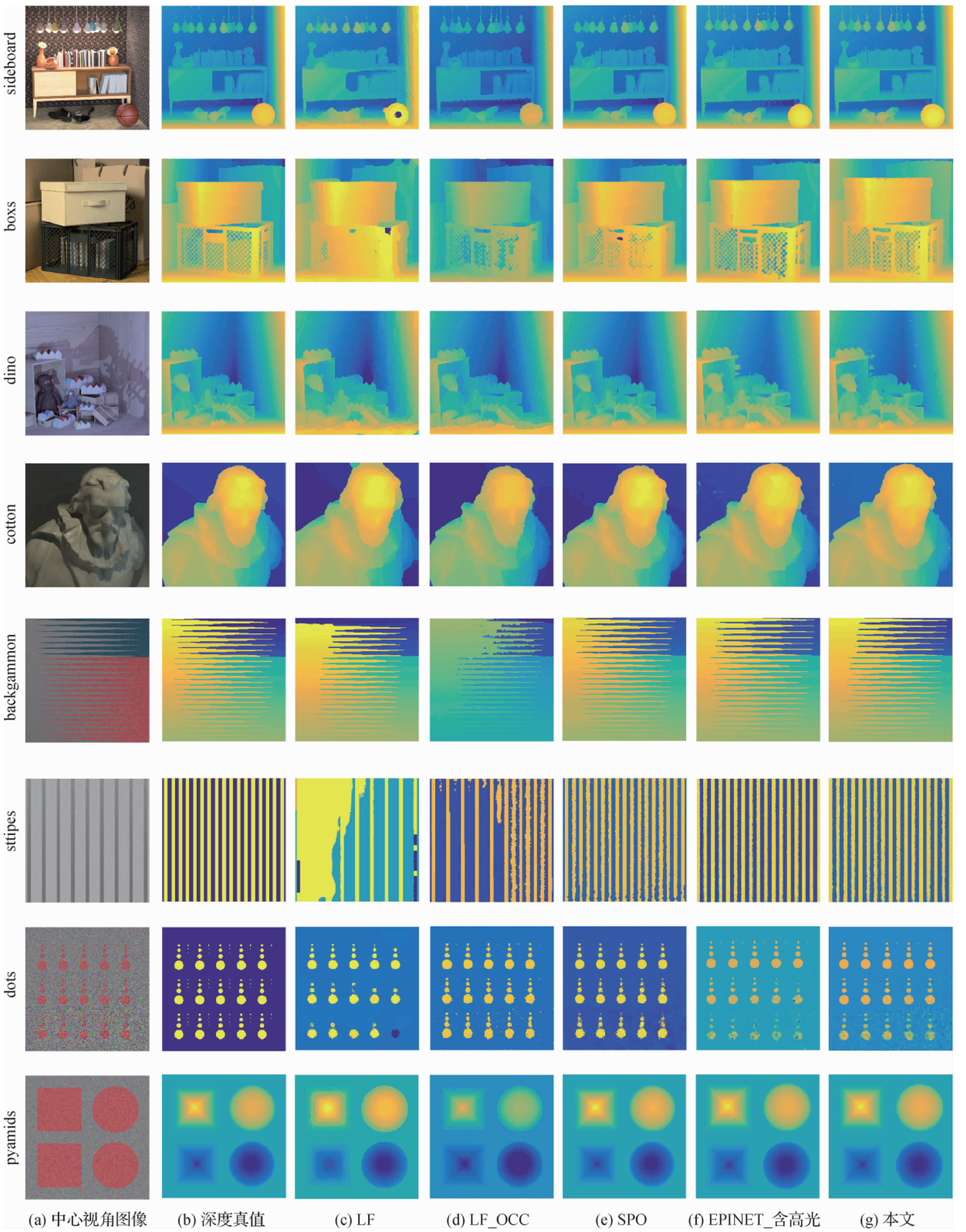


图 11　在 4D light field benchmark 数据集上的定性结果对比

Fig. 11　Qualitative results on 4D light field benchmark dataset((a) center view image; (b) ground truth; (c) LF; (d) LF_OCC; (e) SPO; (f) EPINET_highlight; (g) ours)

景中,由于该场景反射高光区域较为密集,其他算法难以分割轮廓区域信息,很难进行有效估计,而本文网络精准预测了高光区域的形状和清晰的轮廓信息。

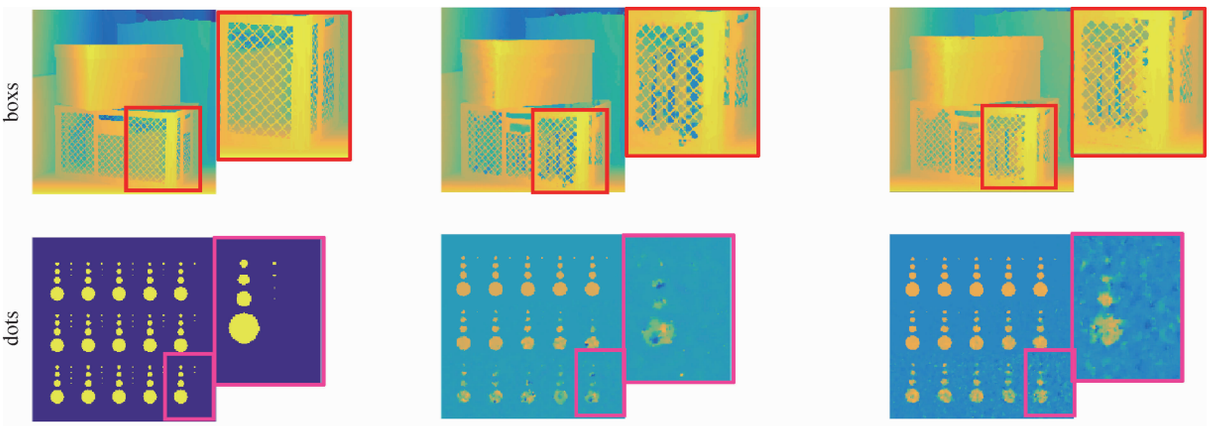


图 12　场景 boxs 和 dots 的定性结果

Fig. 12　Qualitative results of boxs and dots ((a) ground truth; (b) EPINET_highlight; (c) ours)

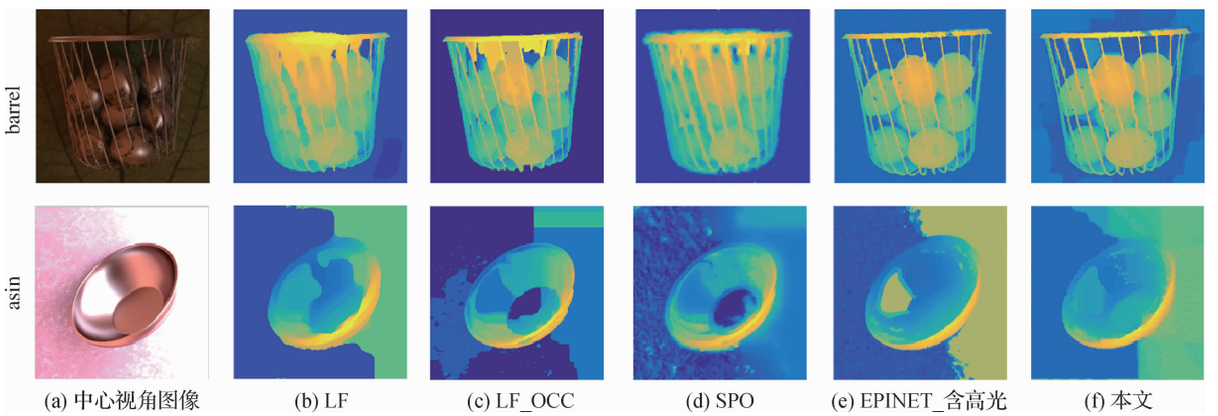


图 13　在 CVIA Konstanz specular dataset 数据集上的结果

Fig. 13　Results on the CVIA Konstanz specular dataset

((a) center view image; (b) LF; (c) LF_OCC; (d) SPO; (e) EPINET_highlight; (f) ours)

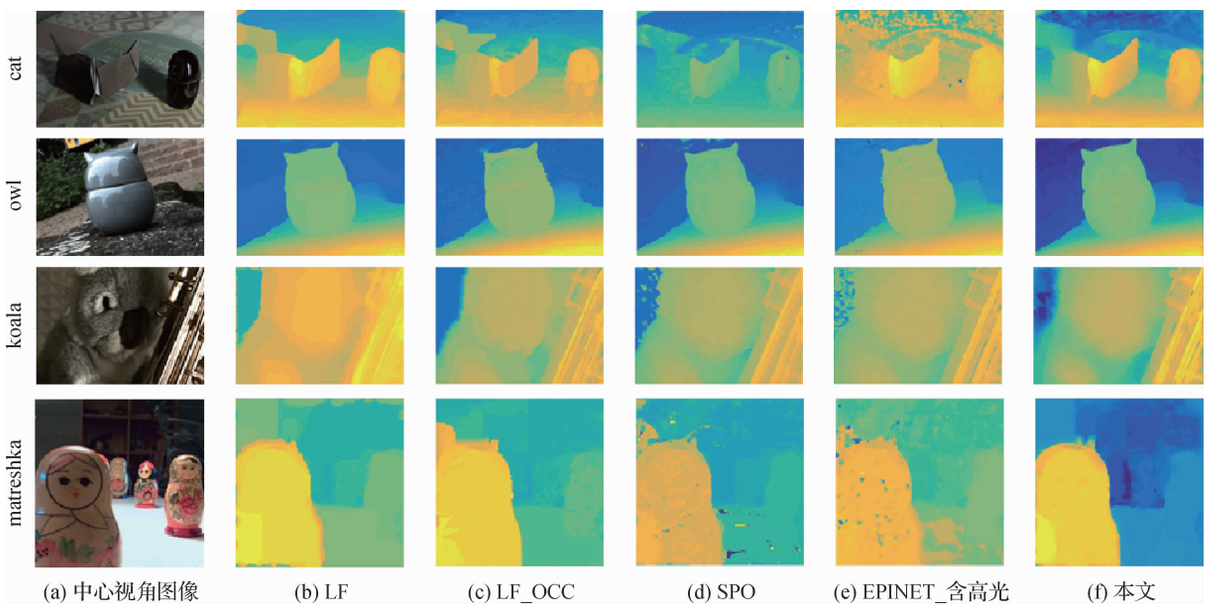


图 14　在 Lytro Illum 拍摄的真实场景数据集下的定性对比结果

Fig. 14　Qualitative results in real-world datasets taken by a Lytro Illum camera

((a) center view image; (b) LF; (c) LF_OCC; (d) SPO; (e) EPINET_highlight; (f) ours)

4 结 论

本文针对高光问题,提出了一种基于图像上下文信息的抗高光光场深度估计方法。该方法基于图像中高光区域随视角改变而改变的原理,建立全卷积神经网络模型。利用光场图像多视角特性,通过设置网络多视角分支路输入,获取高光区域在不同视角下的深度信息。同时,利用空洞卷积的形式扩大网络感受野,获取了更大范围的图像上下文信息。网络通过获取到图像更大范围的多角度特征信息,有效缓解了高光现象对深度估计的影响。此外,本文设计了一种新型的多尺度特征融合方式,串联多膨胀率空洞卷积特征与多卷积核普通卷积特征,进一步提高网络预测结果的精度和平滑度。

对比其他先进算法,本文网络有效缓解了光场图像深度估计任务中高光现象的影响,在合成场景数据集和真实场景数据集上均取得了更优的结果,具有较高的应用价值。

今后的工作将围绕以下两个方向开展。首先,本文网络训练数据量较少,可以通过增加更多含高光场景的训练数据,进一步提高网络在高光区域的深度估计效果。同时,模型训练过程可以增加物体材质等先验知识。其次,从实验结果可以看到,本文网络在背景复杂或背景包含高光区域的图像中,图像前景物体中高光区域深度信息获取较好,但无法准确估计背景的深度信息。因此,可以对训练数据进行背景和前景分割的预处理。

参考文献(References)

Adelson E H and Bergen J R. 1991. The plenoptic function and the elements of early vision//Landy M, Movshon J A, eds. Computational Models of Visual Processing. Cambridge, USA: MIT Press: 3-20

Adelson E H and Wang J Y A. 1992. Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2): 99-106 [DOI: 10.1109/34.121783]

Alperovich A and Goldluecke B. 2017. A variational model for intrinsic light field decomposition//Proceedings of the 13th Asian Conference on Computer Vision. Taipei, China: Springer: 66-82 [DOI: 10.1007/978-3-319-54187-7_5]

Alperovich A, Johannsen O, Strecke M and Goldluecke B. 2018. Light field intrinsics with a deep encoder-decoder network//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 9145-9154 [DOI: 10.1109/CVPR.2018.00953]

Bishop T E, Zanetti S and Favaro P. 2009. Light field superresolution//Proceedings of 2009 IEEE International Conference on Computational Photography. San Francisco: IEEE: 1-9 [DOI: 10.1109/ICCPHOT.2009.5559010]

Cui Z P, Gu J W, Shi B X, Tan P and Kautz J. 2017. Polarimetric multi-view stereo//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 369-378 [DOI: 10.1109/CVPR.2017.47]

Dansereau D G, Pizarro O and Williams S B. 2013. Decoding, calibration and rectification for lenselet-based plenoptic cameras//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE: 1027-1034 [DOI: 10.1109/CVPR.2013.137]

Feng M T, Wang Y N, Liu J, Zhang L, Zaki H F M and Mian A. 2018. Benchmark data set and method for depth estimation from light field images. IEEE Transactions on Image Processing, 27(7): 3586-3598 [DOI: 10.1109/TIP.2018.2814217]

Goodfellow I, Bengio Y and Courville A. 2016. Deep Learning. Cambridge: MIT Press

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Heber S and Pock T. 2014. Shape from light field meets robust PCA//Proceedings of the 13th European Conference on Computer Vision. Switzerland: Springer: 751-767 [DOI: 10.1007/978-3-319-10599-4_48]

Heber S and Pock T. 2016. Convolutional networks for shape from light field//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 3746-3754 [DOI: 10.1109/CVPR.2016.407]

Heber S, Yu W and Pock T. 2016. U-shaped networks for shape from light field//Proceedings of British Machine Vision Conference. York: BMVA Press: #5 [DOI: 10.5244/C.30.37]

Heber S, Yu W and Pock T. 2017. Neural EPI-volume networks for shape from light field//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 2271-2279 [DOI: 10.1109/ICCV.2017.247]

Honauer K, Johannsen O, Kondermann D and Goldluecke B. 2016. A dataset and evaluation methodology for depth estimation on 4 d light fields//Proceedings of the 13th Asian Conference on Computer Vision. Taipei, China: Springer: 19-34 [DOI: 10.1007/978-3-319-54187-7_2]

Huang F C, Luebke D P and Wetzstein G. 2015. The light field stereoscope//Proceedings of ACM SIGGRAPH 2015 Emerging Technologies. Los Angeles: ACM: #24 [DOI: 10.1145/2782782.2792493]

Jeon H G, Park J, Choe G, Park J, Bok Y, Tai Y W and So Kweon I.

2015. Accurate depth map estimation from a lenslet light field camera//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 1547-1555 [DOI: 10.1109/CVPR.2015.7298762]

Johannsen O, Sulc A and Goldluecke B. 2016. What sparse light field coding reveals about scene structure//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 3262-3270 [DOI: 10.1109/CVPR.2016.355]

Kalantari N K, Wang T C and Ramamoorthi R. 2016. Learning-based view synthesis for light field cameras. ACM Transactions on Graphics, 35(6): #193 [DOI: 10.1145/2980179.2980251]

Kim C, Zimmer H, Pritch Y, Sorkine-Hornung A and Gross M. 2013. Scene reconstruction from high spatio-angular resolution light fields. ACM Transactions on Graphics, 32(4): #73 [DOI: 10.1145/2461912.2461926]

Langguth F, Sunkavalli K, Hadap S and Goesele M. 2016. Shading-aware multi-view stereo//Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer: 469-485 [DOI: 10.1007/978-3-319-46487-9_29]

Levoy M and Hanrahan P. 1996. Light field rendering//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New Orleans: ACM: 31-42 [DOI: 10.1145/237170.237199]

Levoy M. 2006. Light fields and computational imaging. Computer, 39(8): 46-55 [DOI: 10.1109/mc.2006.270]

Li N Y, Ye J W, Ji Y, Ling H B and Yu J Y. 2014. Saliency detection on light field//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE: 2806-2813 [DOI: 10.1109/CVPR.2014.359]

Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]

Ng R. 2006. Digital Light Field Photography. Stanford: Stanford University: 1-203

Oxholm G and Nishino K. 2014. Multiview shape and reflectance from natural illumination//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE: 2163-2170 [DOI: 10.1109/CVPR.2014.277]

Perra C, Murgia F and Giusto D. 2016. An analysis of 3D point cloud reconstruction from light field images//Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications. Oulu: IEEE: 1-6 [DOI: 10.1109/IPTA.2016.7821011]

Perwass C and Wietzke L. 2012. Single lens 3D-camera with extended depth-of-field//Proceedings of the SPIE 8291, Human Vision and Electronic Imaging XVII. Burlingame: SPIE: #829108 [DOI: 10.1117/12.909882]

Zhang J, Liu Y, Zhang S, Poppe R and Wang M. 2020 Light field saliency detection with deep convolutional networks. IEEE Transactions on Image Processing, 29: 4421-4434 [DOI: 10.1109/TIP.2020.2970529.]

Shafer S A. 1985. Using color to separate reflection components. Color Research and Application, 10(4): 210-218 [DOI: 10.1002/col.5080100409]

Shin C, Jeon H G, Yoon Y, So Kweon I and Joo Kim S. 2018. EPINET: a fully-convolutional neural network using epipolar geometry for depth from light field images//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4748-4757 [DOI: 10.1109/CVPR.2018.00499]

Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-01-10]. https://arxiv.org/pdf/1409.1556.pdf

Srinivasan P P, Wang T Z, Sreelal A, Ramamoorthi R and Ng R. 2017. Learning to synthesize a 4 d RGBD light field from a single image [EB/OL]. [2020-01-10]. https://arxiv.org/pdf/1708.03292.pdf

Tao M W, Hadap S, Malik J and Ramamoorthi R. 2013. Depth from combining defocus and correspondence using light-field cameras//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney: IEEE: 673-680 [DOI: 10.1109/ICCV.2013.89]

Tao M W, Su J C, Wang T C, Malik J and Ramamoorthi R. 2016. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(6): 1155-1169 [DOI: 10.1109/TPAMI.2015.2477811]

Wang T C, Efros A A and Ramamoorthi R. 2015. Occlusion-aware depth estimation using light-field cameras//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 3487-3495 [DOI: 10.1109/ICCV.2015.398]

Wang T C, Zhu J Y, Hiroaki E, Chandraker M, Efros A A and Ramamoorthi R. 2016. A 4D light-field dataset and CNN architectures for material recognition//Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer: 121-138 [DOI: 10.1007/978-3-319-46487-9_8]

Wang Y L, Liu F, Zhang K B, Hou G Q, Sun Z N and Tan T N. 2018. LFNet: a novel bidirectional recurrent convolutional neural network for light-field image super-resolution. IEEE Transactions on Image Processing, 27(9): 4274-4286 [DOI: 10.1109/TIP.2018.2834819]

Wanner S and Goldluecke B. 2012a. Spatial and angular variational super-resolution of 4D light fields//Proceedings of the 12th European Conference on Computer Vision. Florence: Springer: 608-621 [DOI: 10.1007/978-3-642-33715-4_44]

Wanner S and Goldluecke B. 2012b. Globally consistent depth labeling of 4D light fields//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE: 41-48 [DOI: 10.1109/CVPR.2012.6247656]

Wanner S and Goldluecke B. 2013a. Reconstructing reflective and transparent surfaces from Epipolar plane images//Proceedings of the 35th German Conference on Pattern Recognition. Saarbrücken: Springer: 1-10 [DOI: 10.1007/978-3-642-40602-7_1]

Wanner S and Goldluecke B. 2014. Variational light field analysis for disparity estimation and super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(3): 606-619 [DOI: 10.1109/TPAMI.2013.147]

Wanner S, Meister S and Goldluecke B. 2013b. Datasets and benchmarks for densely sampled 4D light fields//Bronstein M, Favre J and Hormann K, eds. Vision, Modeling and Visualization. The Eurographics Association: 225-226 [DOI: 10.2312/PE.VMV.VMV13.225-226]

Wu C L, Wilburn B, Matsushita Y and Theobalt C. 2011. High-quality shape from multi-view stereo and shading under general illumination//Proceedings of CVPR 2011. Providence: IEEE: 969-976 [DOI: 10.1109/CVPR.2011.5995388]

Wu G C, Masia B, Jarabo A, Zhang Y C, Wang L Y, Dai Q H, Cjai T Y and Liu Y B. 2017. Light field image processing: an overview. IEEE Journal of Selected Topics in Signal Processing, 11(7): 926-954 [DOI: 10.1109/JSTSP.2017.2747126]

Xiong W, Zhang J, Gao X J, Zhang X D and Gao J. 2017. Anti-occlusion light-field depth estimation from adaptive cost volume. Journal of Image and Graphics, 22(12): 1709-1722 (熊伟, 张骏, 高欣健, 张旭东, 高隽. 2017. 自适应成本量的抗遮挡光场深度估计算法. 中国图象图形学报, 22(12): 1709-1722) [DOI: 10.11834/jig.170324]

Yu F and Koltun V. 2015. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2020-01-20]. https://arxiv.org/pdf/1511.07122.pdf

Yu Z, Guo X Q, Lin H B, Lumsdaine A and Yu J Y. 2013. Line assisted light field triangulation and stereo matching//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney: IEEE: 2792-2799 [DOI: 10.1109/ICCV.2013.347]

Zhang J, Wang M, Gao J, Wang Y, Zhang X D and Wu X D. 2015. Saliency detection with a deeper investigation of light field//Proceedings of the 24th International Conference on Artificial Intelligence. Aires: ACM: 2212-2218

Zhang J, Wang M, Lin L, Yang X, Gao J and Rui Y. 2017. Saliency detection on light field: a multi-cue approach. ACM Transactions on Multimedia Computing, Communications, and Applications, 13(3): #32 [DOI: 10.1145/3107956]

Zhang S, Sheng H, Li C, Zhang J and Xiong Z. 2016. Robust depth estimation for light field via spinning parallelogram operator. Computer Vision and Image Understanding, 145: 148-159 [DOI: 10.1016/j.cviu.2015.12.007]

作者简介

王程,1993 年生,男,硕士研究生,主要研究方向为计算机视觉、光场技术。
E-mail: wangcheng93@mail.hfut.edu.cn

张骏,通信作者,女,副研究员,主要研究方向为计算机视觉、图像处理与分析、机器学习。
E-mail:zhangjun@hfut.edu.cn

高隽,男,教授,主要研究方向为智能信息处理。
E-mail:gaojun@hfut.edu.cn