

基于神经网络的编码光场深度值估计

杨成卓^{1,2}, 向森^{1,2*}, 邓慧萍^{1,2}, 吴谨^{1,2}

¹武汉科技大学信息科学与工程学院, 湖北 武汉 430081;

²武汉科技大学冶金自动化与检测技术教育部工程研究中心, 湖北 武汉 430081

摘要 针对传统光场深度值估计中的测量对象纹理不足导致深度值精度低的问题和光场高维数据带来的高计算负荷的问题, 提出一个基于轻量级卷积神经网络对相位编码的光场进行深度值估计的方法, 并提出相应的相位编码光场数据集。所提方法利用水平和垂直视点流的多角度信息, 利用全卷积网络和逐级深化的平均池化充分提取特征, 并由中心视图引导, 将两个方向的特征流融合, 最终得到中心视点的深度图。实验结果表明, 所提方法可生成高精度深度图, 而且网络参数量和推理时间仅为典型光场深度值估计网络的 27.4% 和 41.2%, 具有更高的效率和实时性能。

关键词 深度估计; 光场; 卷积神经网络; 相位编码; 结构光

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP221145

Depth Estimation for Phase-Coding Light Field Based on Neural Network

Yang Chengzhuo^{1,2}, Xiang Sen^{1,2*}, Deng Huiping^{1,2}, Wu Jing^{1,2}

¹School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;

²Engineering Research Center for Metallurgical Automation and Measurement Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China

Abstract In this study, we propose a depth estimation method for phase-coding light field based on a lightweight convolutional neural network. This method aims to solve the problems of low accuracy for depth values caused by the insufficient texture of a measured object in traditional light field depth value estimation and high computational loads caused by high-dimensional light field data. In addition, a new phase-coding light field dataset is proposed. This novel method exploits the information of horizontal and vertical perspectives in phase-coding light field to extract the features using full convolutional networks and deepening average pooling. Furthermore, the central view is used as a guide to fuse the horizontal and vertical features and acquire the depth map. The experimental results demonstrate that the proposed method can generate high-accuracy depth maps, while number of parameters and computation time in generating such maps are, respectively, 27.4% and 41.2% of those of a typical light field depth estimation network. Thus, the proposed method has a higher efficiency and real-time performance than the traditional approach.

Key words depth estimation; light field; convolutional neural network; phase-coding; structured light

1 引言

空间中的任一光线都可由 4D 光场描述, 它同时记录了光线的强度和方向^[1], 其高维特性催生了一系列应用, 如视图渲染^[2-3]、超分辨率^[4-5]、重聚焦^[6-7]、三维重建^[8-10]、虚拟现实 (VR)、增强现实 (AR) 和混合现实 (MR), 而估计场景的深度值是这些应用中的关键和基础性任务。

在 4D 光场中, 深度值可以通过极平面图像 (EPI)^[11-13]或数字重聚焦图像^[14-16]来测算, 然而这些方法面临着两方面的问题。首先, 在无纹理和重复纹理等缺乏唯一视觉特征的情况下, 这两种方法会产生错误深度值; 其次, 这两种方法要测试大量的候选深度值, 需要相当繁重的计算量。为了应对这两方面的不足, 结合了结构光和光场的编码光场近年来受到广泛关注^[17]。编码光场通过投影仪向场景投射编码模板,

收稿日期: 2022-03-29; 修回日期: 2022-05-17; 录用日期: 2022-06-14; 网络首发日期: 2022-06-24

基金项目: 国家自然科学基金 (61702384, 62001180, 61871437)

通信作者: *xiangsen@wust.edu.cn

赋予场景丰富的特征,从而提升深度值估计的准确性和效率。具体来说,深度值可通过视点间匹配获得,也可利用光场高维特性生成。

视图匹配的方法遵循立体匹配或传统结构光深度计算的流程。Cai等^[17-18]校准了结构光场系统中的每条光线,建立了相位到深度值的直接转换模型。基于该模型,相位值可以直接转换为世界坐标。Zhang等^[19]沿用了传统的结构光深度值测量的框架,利用光场的包裹相位得出条纹序号和绝对相位,再计算深度值。Zhou等^[20-21]提出了一个由浅到深的框架,用相位光场估计深度值,首先根据EPI的相位一致性进行粗略估计,之后用投影仪和光场相机的一致性对场景点的空间位置进行优化。Wang等^[22]将工业相机加入结构光场系统,利用光场的角块辅助工业相机进行相位解包裹。总之,这些方法虽然利用了结构光的丰富特征来辅助深度值估计,但并没有结合4D相位光场的极平面几何特性。

基于光场高维特性的方法利用光场深度值估计的框架,依靠光场重聚焦图像或极平面图像作为深度线索。Cai等^[23]通过对相位编码光场(PLF)进行重聚焦来估计深度值,其中相位解包裹在角度域进行,此后他们进一步提出用原始的包裹相位光场来估计深度值^[24]的方法。Liu等^[25]用“斜率下降”算法找到相位EPI中的最佳线斜率,并从中获得深度值。然而这些基于光场的深度值估计方法存在计算负荷重和局部最小值问题,尤其是基于光场重聚焦图像的方法需要剪切所有视点,引入了沉重的计算负担,不能高效率地估计准确的深度值。

神经网络因具有强大的特征提取能力而广泛应用于图像处理,近年来利用神经网络进行光场深度值估计的研究逐渐增多。Shin等^[26]提出了一个全卷积神经网络,利用光场的极平面几何特性进行快速而准确的深度值估计,并针对光场的多维特点进行数据增广。Tsai等^[27]提出了一个基于注意力的视图选择模块,按照对深度值估计潜在贡献的不同,给予孔径视图赋予不同的权重以提升深度值的准确性。Huang等^[28]设计了一个轻量级的深度值估计模型,通过构建多视差的代价体估计深度值,且设计了专门的边缘引导网络来提升场景边缘深度值的准确性。罗姚翔^[29]提出了一种以星型EPI Patch为输入的卷积神经网络,获取Patch所对应像素的视差概率分布,并以此获得最终视差。潘志伟^[30]提出了一种具有双子结构的网络模型,利用中心视图中像素点对应的水平和垂直方向的EPI块,将深度计算问题转换为分类问题。马浩鑫^[31]提出了一套基于卷积神经网络的光场深度估计系统,实现了从输入神经网络的光场多视点图像到得出视差图的可行方案。但上述方法中场景特征不足的情况依然存在,而光编码技术可有效丰富场景特征,提供像素级的码字,用于提升方法的精度和效率。例如Li等^[32]提出了

一种基于深度学习的单目双频复合条纹投影轮廓测量方法,该方法从单张条纹图像中恢复绝对相位,并避免了频谱混叠问题,实现高精度的动态3D形状测量。Yin等^[33]提出了一个端到端的立体匹配网络,对单个散斑投影图案进行主体形状测量,方法集成了一个简单快速的显著性检测网络,以避免预测遮挡和背景区域中的无效像素。而上述方法针对单目视差,数据维度远不如光场密集,导致场景中携带的信息不如光场丰富。

本文结合光场的高维数据特性和光编码带来的丰富特征,提出一种基于相位编码光场的深度值估计网络。该网络使用原始的包裹相位作为输入,利用光场的水平和垂直特性进行双支路特征提取并融合,结合中心视图的特征得到连续变化的深度图。所提方法具有如下特点。1)综合考虑高精度和高效率的需求,使用水平和垂直方向的“十字形”视点流作为网络的输入,较经典深度学习方法,减少了视点数量和网络参数量,在保持深度图精度与经典方法相近的同时加快了推理效率。2)考虑了相位图像和深度图的一致性,设计了第三条分支网络并利用中心视图为引导信息,结合空间注意力机制实现对场景整体结构和细节的差异性处理,使预测结果更加准确。3)使用相位编码的光场数据进行训练,与传统RGB光场相比,相位编码光场具有像素级码字,使得像素点具有更唯一和稳定的特征,避免了普通光场因缺乏视觉特征导致的深度值估计错误等问题。

综合以上特点,所提轻量级网络利用十字形视点流提取足够丰富的水平和垂直视差,可在获取高质量深度图的同时大大降低计算开销,具有理论上的可行性。而本文的实验结果也验证了所提方法的可行性,即所提方法的精度和公认的深度学习标准十分接近,而推理时间减少了58.8%,证明了所提方法的性能。

2 三支路网络深度值估计

为高精度且高效率地进行深度值估计,本文设计了一个基于卷积神经网络的三支路光场深度值估计模型,从特征明确的相位光场图像中估计深度图。网络总体框架如图1所示,由三部分组成:三支路特征提取,用于获取垂直和水平方向图像堆的特征流以及中心视图的多尺度特征;特征融合,对中心视图多尺度特征向两方向特征流进行映射,建立融合特征和目标场景的对应关系;深度值预测,逐像素计算场景深度,以生成完整深度图。

2.1 三支路特征提取

光场图像是包含空间分辨率和角度分辨率的4D数据,在带来更多场景信息的同时又具有庞大的数据量。4D光场图像表示为 $L(x, y, u, v)$,其中 (x, y) 是空间坐标, (u, v) 是角度坐标,二者可以决定光线 L 的强

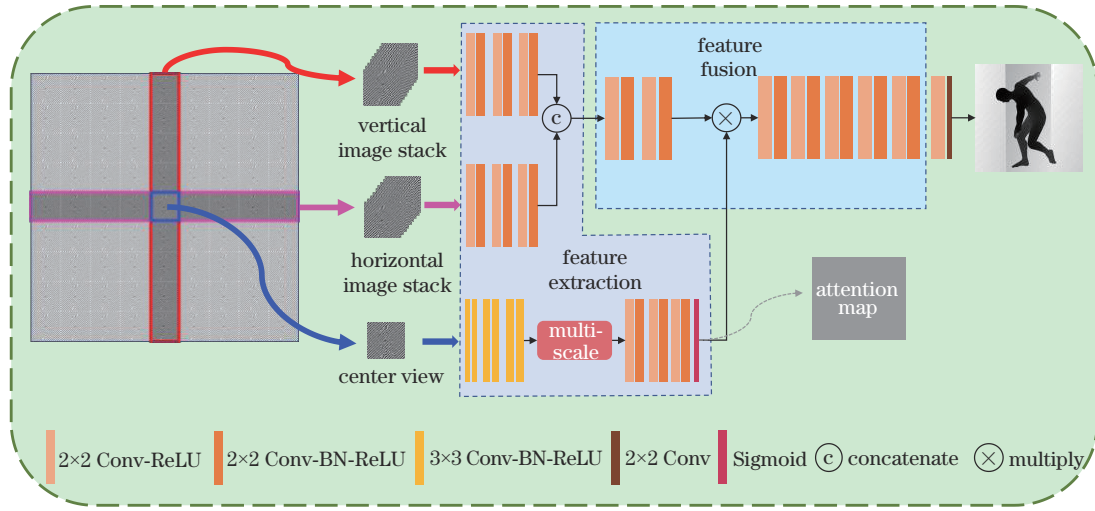


图1 三支路光场深度估计网络

Fig. 1 Three-branch light field depth estimation network

度。同一根光线在中心视图和其他视点之间的关系为 $L(x, y, 0, 0) = L[x + d(x, y) * u, y + d(x, y) * v, u, v]$, (1)

式中: $d(x, y)$ 是中心视点中的像素 (x, y) 与其相邻视点中的对应像素的视差。对于角度方向 θ ($\tan \theta = v/u$), 可以将关系重新表述为

$$L(x, y, 0, 0) = L[x + d(x, y) * u, y + d(x, y) * u \tan \theta, u, u \tan \theta]. \quad (2)$$

式(2)表明, 通过不同视点之间的强度一致性约束, 可从四维光场数据中求取视差 d 。理论上说, 视点增加会带来更强的相关性和约束性, 从而生成更精准深度图。然而, 过多的视点对深度值精度的贡献递减, 复杂度却迅速增加, 制约了在速度和延时敏感场景下的应用。为解决这个问题, 本文提出一种轻量级的编码光场深度值估计网络, 采用光场图像中垂直方向和水平方向的极平面图像堆(EPI-volume)和中心视图, 在尽量减少计算量的前提下充分利用光场的多视角特性, 高效率且高精度地生成深度信息。

如图1特征提取部分所示, 采用三个支路, 分别提取水平方向视图、垂直方向视图和中心视图的特征。水平和垂直方向的图像各自叠加, 形成图像堆, 分别提供了互补的水平和垂直深度值线索。考虑要生成稠密的深度图, 采用全卷积网络进行像素级密集预测^[34], 在垂直和水平图像堆的双支路上, 采用结构相同、相互独立的全卷积模块进行光场单方向的特征提取, 模块由三个 Conv-ReLU-Conv-BN-ReLU 块组成。考虑到光场的窄基线问题并要减少计算量, 该模块使用尺寸为 2×2 的卷积。

在水平和垂直图像堆采用全卷积结构的基础上, 进一步考虑纹理图和深度图之间的相关一致性, 采用中心视图引导深度图的生成。具体来说, 本文采用了

空间注意力机制实现网络对整体结构和边缘细节的差异性处理, 从而提升深度图尤其是边缘区域深度值的准确性。在中心视图特征提取支路上, 首先由逐渐加深的 3×3 卷积块提取浅层特征, 然后将浅层特征送入多尺度特征提取模块以获取中心视图上下文信息, 最后将所有尺度特征堆叠并使用 Sigmoid 函数生成特征图, 以引导后面的特征融合。多尺度特征提取模块如图2所示, 它由 2×2 的全局平均池化串联实现, 模块共提取 4 个不同尺度的特征, 最后通过双线性插值方式将各尺度特征上采样至原始尺寸并进行堆叠。

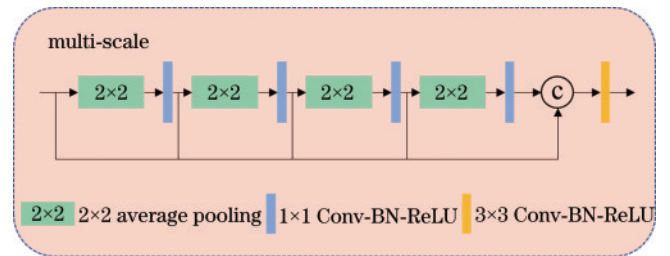


图2 多尺度特征提取模块

Fig. 2 Multi-scale feature extraction module

2.2 多角度特征流融合

多视点的优势给光场图像带来丰富的场景信息, 视图间的差异使场景深度更完整, 这些差异是光场图像深度值估计的关键。本文提取中心视图的多尺度特征并将其映射到垂直和水平方向特征流中, 针对视点间的视差关系建立与场景深度的联系。因此, 在对极平面图像堆和中心视图进行特征提取后, 需要对两方向特征流与中心视图进行融合。如图1特征融合部分所示, 对于水平和垂直两个方向提取的特征图, 首先使用两个 2×2 的 Conv-ReLU-Conv-BN-ReLU 模块对两个角度方向的特征图进行聚合, 再将中心视图支路提取出的特征图作为引导与聚合后的特征图相乘。最后的融合网络包含 5 个 Conv-ReLU-Conv-BN-ReLU

模块(用来进一步探寻垂直流、水平流和中心视图三个支路的特征图之间的联系)及 1 个 Conv-ReLU-Conv 模块(用于逐像素深度预测得到完整深度图)。

2.3 数据增广

针对神经网络需要大量数据驱动训练但实际数据集有限的问题,采用一种适用于光场图像的数据增广方法。在数据增广中,对图像进行旋转是一种常用的手段,但常规旋转不考虑光场图像的方向特性,并不能在光场任务下直接使用。在网络的双流特征提取部分,本文分别提取了水平视点流和垂直视点流的极平面几何特征,考虑到光场图像的方向特性,首先在空间维度上旋转子孔径图像,然后重新排列视点并将其送入对应的网络支路中。由于极平面图像保留了三维空间的几何结构,所以这种旋转的数据增广方式仍保留了光场数据隐含的深度值线索。如图 3 所示,在空间域对子孔径图像逆时针旋转 90° 后,角度域的水平视点流实际反映空间中垂直方向的视差关系,因此需将这时的水平视点流送入网络中对应的垂直特征流提取的支路中。根据这种 EPI 辅助的方式,采取旋转(90° 、 180° 、 270°)和翻转的方式对随机采样产生的训练数据进行扩充。

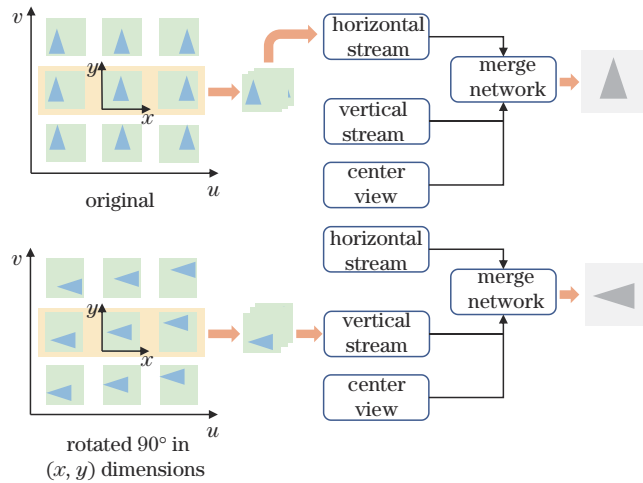


图 3 通过旋转进行数据增强
Fig. 3 Data augmentation by rotation

3 实验与分析

3.1 实施细节

为了对所提三支路网络进行训练,构建了一个相位光场数据集。该数据集包含 80 个角度分辨率为 9×9 、空间分辨率为 512×512 的场景,其中训练场景和测试场景各 40 个。首先由仿真软件 3DsMax 生成由被测场景、相机阵列和投影仪组成的结构光场系统,其次通过投影仪依次投射三个相位差为 $2/3\pi$ 的正弦条纹模板到被测场景上,由相机阵列捕获场景的编码光场数据,最后利用三步相移法^[35]计算出包裹相位数据。场景由前景和背景组成,前景为不同的主体模型,背景

由三个角度和深度值渐变的平面构成以增加场景的复杂度。数据增广使训练样本扩增 8 倍。实验采用 patch-wise 的训练方式,角度域不变,从 40 个场景的空间域中随机采样大小为 32×32 的图像块进行训练。网络使用平均绝对误差(MAE)作为损失函数,公式为

$$E_{MA} = \frac{1}{N} \sum_{i=1}^N |Y_i - \bar{Y}_i|, \quad (3)$$

式中: Y_i 和 \bar{Y}_i 分别为真值和预测值; N 为像素点数量。网络训练采用 RMSProp 优化器, batch size 设置为 16, 学习率为 0.0001。网络在 NVIDIA RTX 2070 上训练了 120 h, 使用 TensorFlow 框架搭建。网络损失函数收敛曲线如图 4 所示。

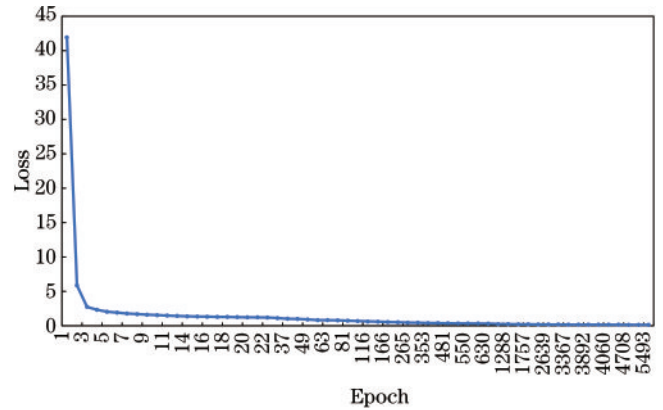


图 4 网络训练的收敛曲线
Fig. 4 Convergence curve of network training

3.2 实验结果

因目前主流方法多使用普通光场数据集而缺乏公开的相位编码光场数据集,为进行定量评估来说明所提方法的有效性,使用自建数据集中测试集部分计算这些场景的相对坏点率(BP)和 MAE。相对坏点率的计算公式为

$$R_{BP}(\epsilon) = \frac{\sum_{(i,j) \in N} \frac{|Y(i,j) - \bar{Y}(i,j)|}{Y(i,j)} > \epsilon}{N}, \quad (i,j) \in N, \quad (4)$$

式中: $Y(i,j)$ 和 $\bar{Y}(i,j)$ 分别为像素 (i,j) 处的真实深度值和预测深度值; ϵ 表示设定的相对误差阈值。为了更好地体现算法在复杂场景下预测的准确程度,采用相对误差阈值 7%、5% 和 3% 下的坏点率。对所提方法和主流算法进行性能对比,主流算法包括传统方法 CAE^[16]、OCC^[8]、SPO^[36]、REFOCUS^[23] 和深度学习方法 EPINet^[26],其中传统算法的深度分辨率设置为 140。CAE、OCC、REFOCUS 方法通过最小化 Cost 函数实现:

$$\min_p \text{Cost}(p, \partial), \quad (5)$$

式中: p 为空间坐标; ∂ 为待求深度。CAE 中 Cost 函数定义为

$$\text{Cost}(\mathbf{p}, \vartheta) = \frac{\left\{ -\sum_i \frac{g(i)}{|g|} \log[g(i)] \right\}_R + \left\{ -\sum_i \frac{g(i)}{|g|} \log[g(i)] \right\}_G + \left\{ -\sum_i \frac{g(i)}{|g|} \log[g(i)] \right\}_B}{3}, \quad (6)$$

式中： $g(i)$ 为自适应权重的强度概率分布； i 为角块中 \mathbf{p} 处的强度； R, G, B 分别表示三个颜色通道。OCC中Cost函数定义为

$$\text{Cost}(\mathbf{p}, \vartheta) = C_\vartheta(\mathbf{p}) + D_\vartheta(\mathbf{p}), \quad (7)$$

式中： $C_\vartheta(\mathbf{p})$ 为最小区域方差； $D_\vartheta(\mathbf{p})$ 为散焦响应。REFOCUS中Cost函数定义为

$$\left\{ \begin{aligned} \text{Cost}(\mathbf{p}, \vartheta) &= \frac{1}{|\mathbf{W}_p|} \sum_{p' \in \mathbf{W}_p} \sigma_\vartheta(\mathbf{p}) \\ \sigma_\vartheta^2(\mathbf{p}) &= \frac{1}{|\mathbf{N}_u|} \sum_{u' \in \mathbf{N}_u} \left\{ (|u'| + 1) \times [\phi_\vartheta(\mathbf{p}, u') - \bar{\phi}_\vartheta(\mathbf{p})] \right\}^2, \end{aligned} \right. \quad (8)$$

式中： \mathbf{W}_p 是空间坐标周围的空间窗口； \mathbf{N}_u 是角度分辨率， \mathbf{u} 表示角度坐标； ϕ 表示相位。

SPO通过寻找EPI斜率的方式获取深度，表达式为

$$\left\{ \begin{aligned} \max_{\theta} d_u(\mathbf{p}, \theta) \\ d_u(\mathbf{p}, \theta) &= c_{y,v^*}(x, u^*) d_{y,v^*}(x, u^*, \theta) + \\ & c_{x,u^*}(y, v^*) d_{x,u^*}(y, v^*, \theta) \end{aligned} \right. \quad (9)$$

式中： d 为沿垂直或水平方向由旋转平行四边形算子测算出的分布距离； c 为置信度。对于深度学习的方法，为体现模型在相位编码场景下的预测能力，利用所提数据集对EPINet的官方代码进行重新训练并对比结果。实验所得客观指标如表1和表2所示，其中BP3、BP5、BP7分别为相对误差阈值为3%、5%、7%的坏点率。

表1 不同算法在不同场景的客观指标对比
Table 1 Comparison of objective metrics of different algorithms in different scenes

Method	Scene 1				Scene 8				Scene 24			
	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %
CAE	34.86	1.371	1.666	1.959	29.09	0.668	0.864	1.122	78.29	4.769	5.187	5.952
OCC	77.36	2.498	3.923	5.282	69.50	1.295	2.362	4.051	88.43	4.131	5.618	9.697
SPO	145.38	1.002	18.430	18.430	194.36	1.581	28.213	53.186	90.13	0.947	2.634	11.222
REFOCUS	91.42	4.710	6.350	8.489	68.36	1.761	3.289	5.437	89.05	4.980	6.424	9.599
EPINet	16.99	0.021	0.052	0.193	15.26	0.013	0.056	0.206	19.79	0.191	0.383	0.858
Proposed method	17.66	0.043	0.091	0.091	18.11	0.015	0.055	0.183	20.72	0.187	0.380	0.815
Method	Scene 38				Scene 40				Scene 55			
	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %
CAE	31.63	1.127	1.290	1.544	25.76	0.927	1.108	1.388	31.74	0.886	1.094	1.415
OCC	82.35	2.218	4.930	7.082	68.27	2.990	4.712	5.893	78.64	1.630	3.252	5.793
SPO	140.89	0.785	19.203	32.404	110.02	1.025	19.219	28.477	198.86	1.006	28.920	55.598
REFOCUS	81.27	3.615	5.396	6.995	97.72	7.074	8.846	10.340	75.81	2.605	4.483	7.101
EPINet	22.34	0.071	0.156	0.409	19.02	0.055	0.147	0.370	17.30	0.024	0.066	0.252
Proposed method	22.76	0.091	0.170	0.384	16.29	0.057	0.145	0.400	19.43	0.027	0.078	0.239
Method	Scene 58				Scene 69				Scene 86			
	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %	MAE / 10 ²	BP7 / %	BP5 / %	BP3 / %
CAE	29.82	0.965	1.120	1.345	35.05	1.254	1.394	1.650	25.94	0.665	0.827	1.115
OCC	82.84	1.976	4.499	6.174	67.74	1.819	2.902	4.330	74.92	1.812	3.910	5.661
SPO	208.54	1.248	38.245	56.931	192.19	1.125	30.187	52.479	189.23	1.021	32.608	51.655
REFOCUS	77.55	3.202	5.779	7.803	73.69	3.289	5.175	7.853	82.28	3.397	5.688	8.005
EPINet	19.47	0.040	0.109	0.249	15.28	0.080	0.142	0.336	16.97	0.023	0.065	0.197
Proposed method	22.29	0.055	0.118	0.282	17.68	0.077	0.142	0.332	19.33	0.018	0.064	0.191

从表 1 和表 2 可以看出:所提方法的平均绝对误差分别是 CAE 的 56.7%、OCC 的 26.7%、SPO 的

表 2 不同算法的平均客观指标对比

Table 2 Comparison of average objective metrics of different algorithms

Method	MAE /10 ²	BP7 /%	BP5 /%	BP3 /%
CAE	34.9138	1.1345	1.3402	1.6896
OCC	74.1117	1.9221	3.5428	5.4535
SPO	161.6611	0.9002	22.5860	40.8998
REFOCUS	78.2058	3.4325	5.1551	7.2922
EPINet	18.1747	0.0648	0.1396	0.3362
Proposed method	19.7986	0.0885	0.1591	0.3551

12.2% 和 REFOCUS 的 25.3%, 相对坏点率也比传统方法低 1~2 个数量级; 对比深度学习方法 EPINet, 在平均绝对误差上, 所提方法的误差略高于 EPINet 的结果, 但非常接近, 相差 1.62 左右, 在相对坏点率上, 所提方法和 EPINet 在不同场景上性能相仿, 从各阈值相对坏点率的平均指标上看, 所提方法与 EPINet 的差距在 0.02% 左右。各类算法在所有场景上的预测表现如图 5 所示。

从图 5 可以看出, 所提方法在测试集上表现出最优或次优的结果, 且与 EPINet 表现相近。为体现所提方法与对比方法在主观视觉上的性能差异, 对各类方法的深度结果进行了主观比较, 部分测试场景结果如图 6 所示。

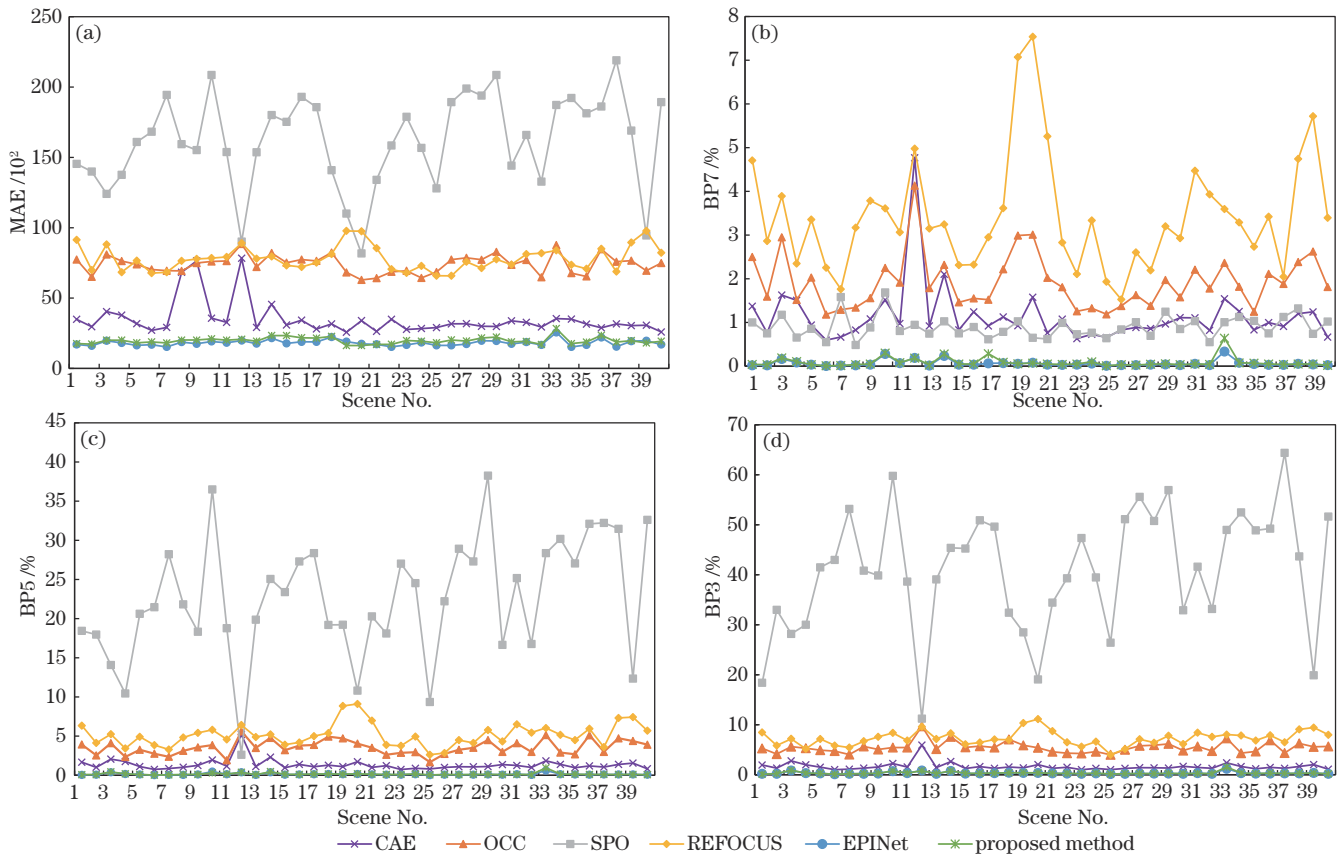


图 5 各类方法在测试集的数据结果对比。(a)MAE;(b)BP7;(c)BP5;(d)BP3

Fig. 5 Comparison of the data results of each method on the test set. (a) MAE; (b) BP7; (c) BP5; (d) BP3

测试结果中的 10 个场景由图 6 所示, 其中最右侧为 ground-truth, 其余部分从左到右分别为 CAE、OCC、SPO、REFOCUS、EPINet 以及所提方法得到的结果。从图 6 可以看出: 传统方法得到的结果与 ground-truth 差距较大, 深度断层明显且边缘不够平滑; 深度学习方法得到的结果中深度值连续, 边缘清晰, 更加接近 ground-truth, 但出现了与相位数据结构相似的轻微条纹状误差; 对比 EPINet, 所提方法与其差别不明显, 图 7 展示了两种深度学习方法的误差图对比。

图 7 中颜色越深代表误差越大。从图 7 可见, 所提方法出现条纹状误差的频率和场景的背景部分误差略高于 EPINet, 但在场景的主体部分, 所提方法的准确率更高, 误差出现的范围小于 EPINet。除了准确性对比, 还比较了各方法的效率, 如表 3 所示。

从表 3 可以看出: 所提方法推理得到深度图的时间为 0.2282 s, 远低于传统深度值估计方法, 为其中最快速算法 OCC 的 1.2%。神经网络在经过训练后可进行多次推理, 在大规模预测任务中推理效率远超传统深度值估计算法。两种深度学习的方法中, 所提方法

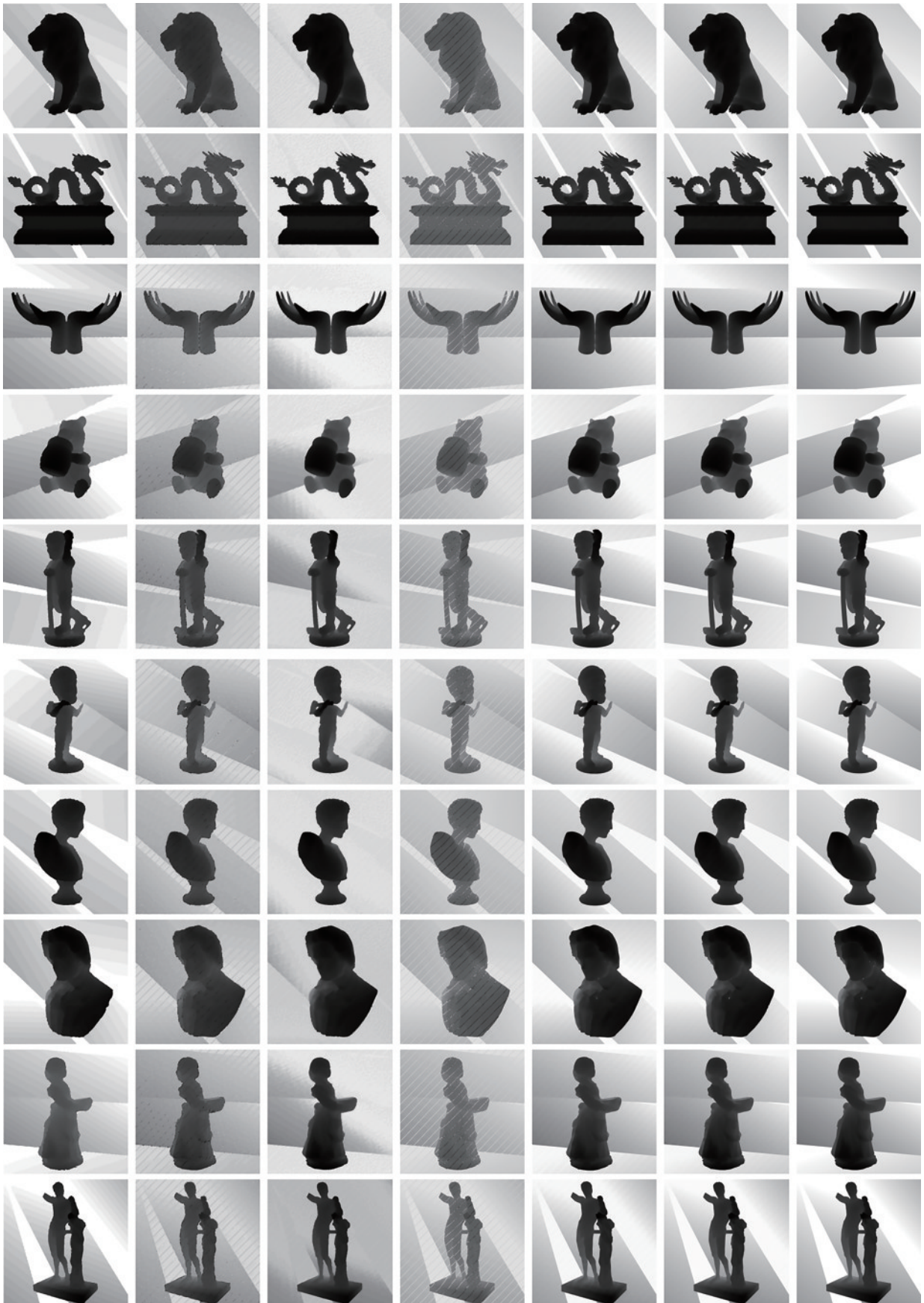


图6 各种方法的深度图结果对比。(a)CAE;(b)OCC;(c)SPO;(d)REFOCUS;(e)EPINet;(f)所提方法;(g)ground-truth
 Fig.6 Comparison of depth map results of various methods. (a) CAE; (b) OCC; (c) SPO; (d)REFOCUS; (e) EPINet; (f) proposed method; (g) ground-truth

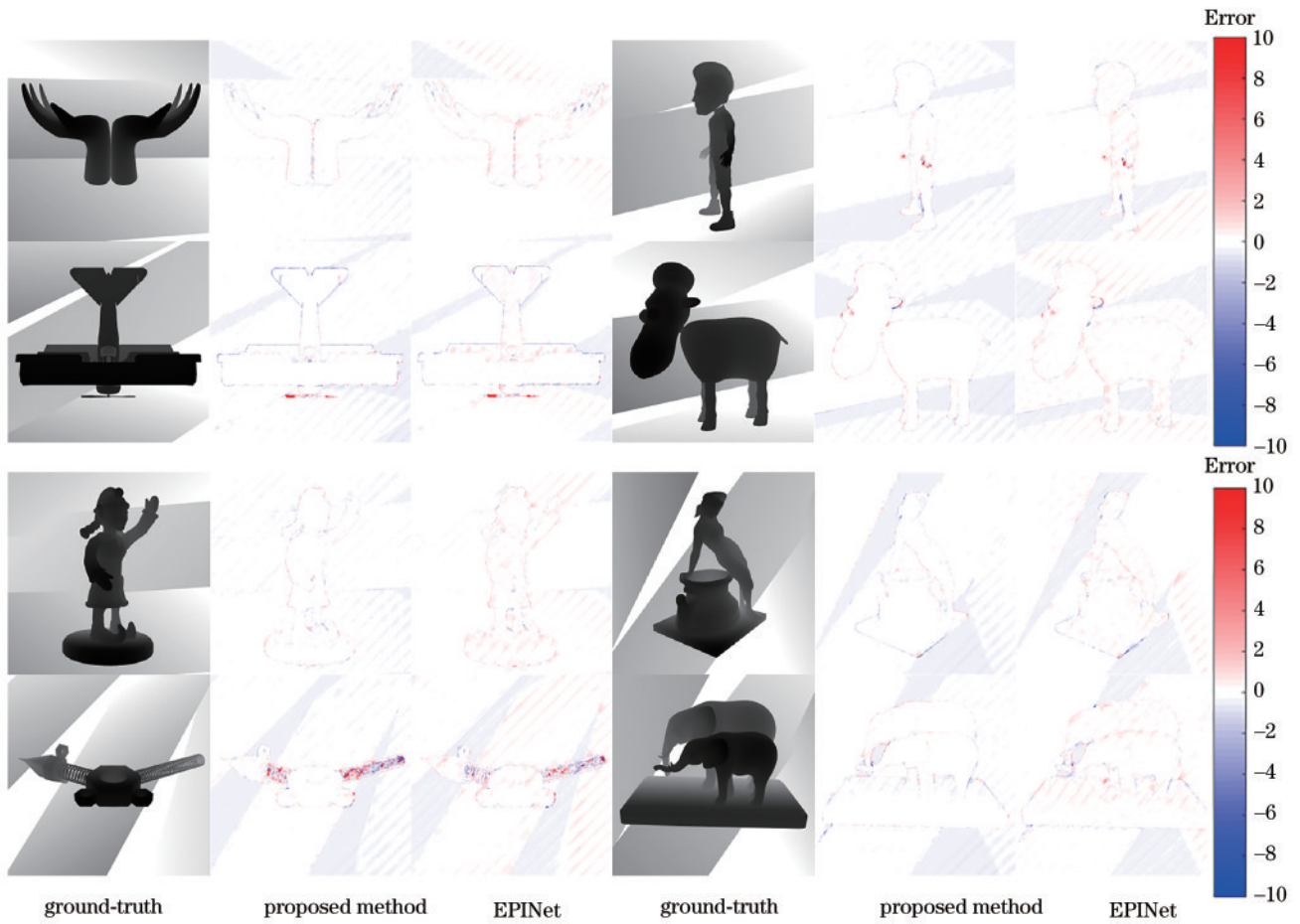


图7 所提方法与EPINet的误差图对比

Fig. 7 Comparison of error maps of the proposed method and EPINet

表3 不同算法的效率对比

Table 3 Efficiency comparison of different algorithms

Method	Average time/s	Number of parameters
CAE	806.5161	
OCC	19.3286	
SPO	283.0965	
REFOCUS	166.4036	
EPINet	0.5533	5,124,281
Proposed method	0.2282	1,402,065

使用视点数量为EPINet的一半,网络参数量只有EPINet的27.4%,简化的视点输入和轻量的网络模型使得所提方法推理深度图的时间仅仅为EPINet的41.2%。综合来看,在得到准确度相近的深度图的情况下,所提方法的效率远远高于所有对比方法,实现了高精度且高效率的三维感知。

表4 所提网络有无中心视图支路的平均客观指标对比

Table 4 Comparison of average objective metrics of the proposed network with/without center view

Method	MAE /10 ²	BP7 /%	BP5 /%	BP3 /%
Proposed method (without center view)	41.4286	0.0894	0.1702	0.3657
Proposed method (with center view)	19.7986	0.0885	0.1591	0.3551

为了验证网络中心视图支路对深度值估计的有效性,对该模块进行了消融实验并与所提三支路网络进行对比分析。综合所有场景的平均客观指标如表4所示。

从表4可以看出,去掉中心视图支路的网络的平均绝对误差是有中心视图支路网络的2.09倍,各阈值的坏点率比有中心视图支路的网络有所上升,即中心视图支路的加入能有效减小网络预测深度值的整体误差,并对深度细节有较好的还原能力。两个网络的主观结果比较如图8和图9所示。

从图8和图9可以看出,去掉中心视图支路后,深度结果的整体误差变大,在场景里除主体外的其他部分中,误差随着真实深度的增加而变大,而带有中心视图支路的网络在场景的渐变背景部分也具有较好的预测能力。即由中心视图支路提取特征并结合注意力机制后,网络能很好地区别前景和背景并进行针对性处理,引导准确的深度结果生成。

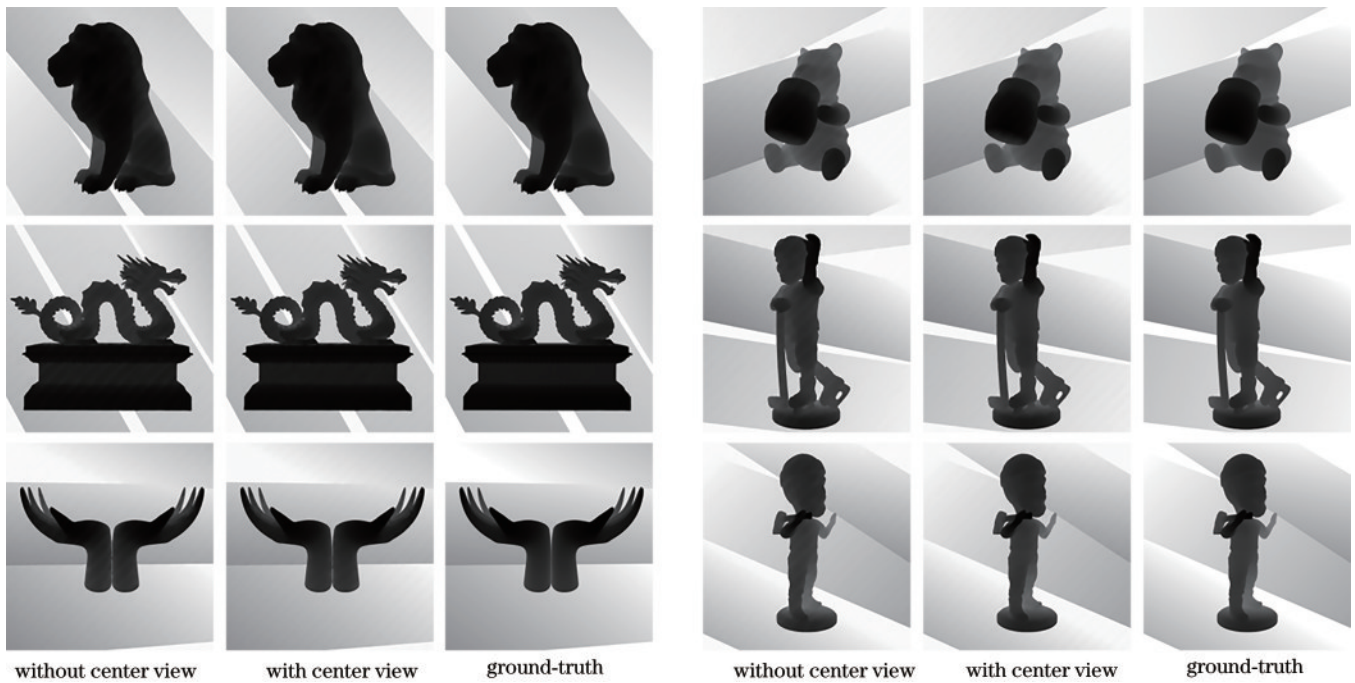


图 8 所提网络有无中心视图支路的深度图结果对比

Fig. 8 Comparison of depth map results of the proposed network with/without center view

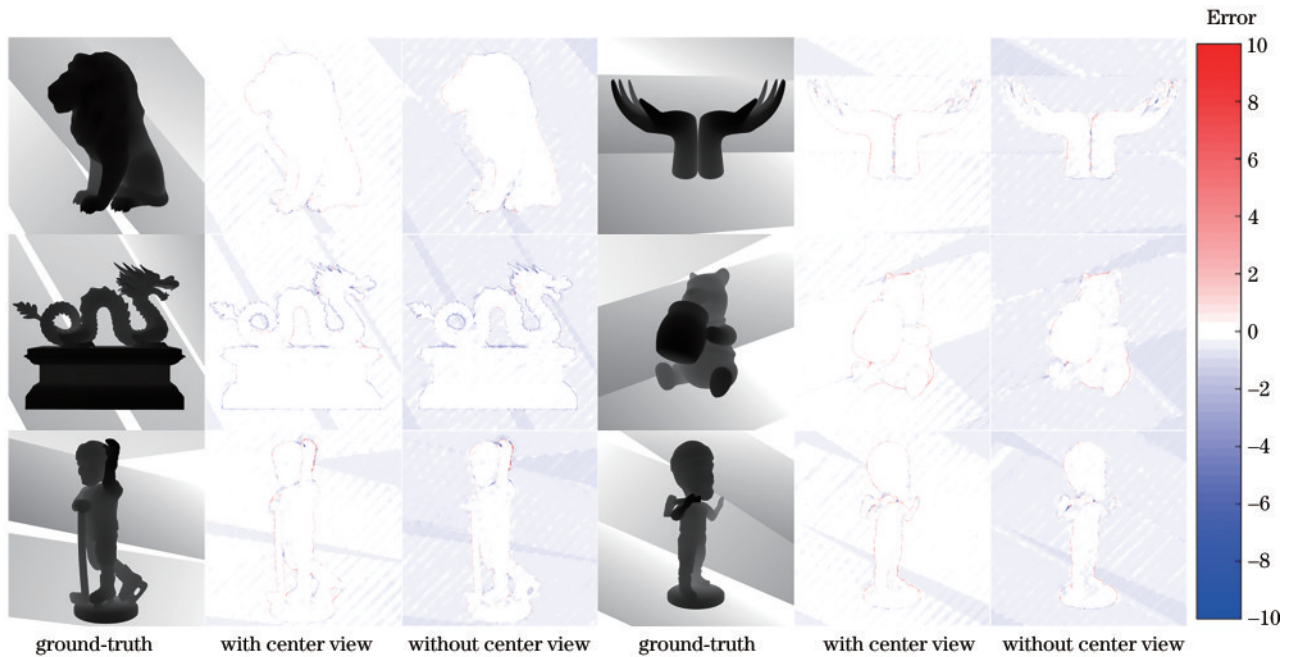


图 9 所提网络有无中心视图支路的误差图对比

Fig. 9 Comparison of error map results of the proposed network with/without center view

4 结 论

提出了一种高精度、高效率的相位光场数据深度值估计方法。该方法采用编码光场数据,结合了结构光的像素级编码和 4D 光场高维数据的优点,解决了普通光场中特征不唯一导致的深度值错误问题。该方法中的三支路卷积神经网络分别提取并利用水平视图

流、垂直视图流、中心视点,进一步进行特征提取和融合,最终预测得到深度值。实验结果表明,所提方法实现了准确的深度值估计,其精度远高于传统深度值估计方法,与光场深度值估计网络 EPINet 表现相近。与此同时,所提方法的计算效率优于所有对比方法,网络参数量和推理耗时分别只有 EPINet 的 27.4% 和 41.2%,实现了高精度且高效率的深度值估计。

参 考 文 献

- [1] Levoy M, Hanrahan P. Light field rendering[C]// Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, August 4-9, 1996, New Orleans, LA, USA. New York: ACM Press, 1996: 31-42.
- [2] Sitzmann V, Rezkikov S, Freeman W T, et al. Light field networks: neural scene representations with single-evaluation rendering[EB/OL]. (2021-06-04) [2022-02-04]. <https://arxiv.org/abs/2106.02634>.
- [3] Wu G C, Liu Y B, Fang L, et al. Revisiting light field rendering with deep anti-aliasing neural network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9): 5430-5444.
- [4] Liang Z Y, Wang Y Q, Wang L G, et al. Light field image super-resolution with transformers[J]. IEEE Signal Processing Letters, 2022, 29: 563-567.
- [5] Ko K, Koh Y J, Chang S, et al. Light field super-resolution via adaptive feature remixing[J]. IEEE Transactions on Image Processing, 2021, 30: 4114-4128.
- [6] Alain M, Smolic A. A spatio-angular filter for high quality sparse light field refocusing[C]//2021 IEEE International Conference on Multimedia & Expo Workshops, July 5-9, 2021, Shenzhen, China. New York: IEEE Press, 2021.
- [7] Jayaweera S S, Edussooriya C U S, Wijenayake C, et al. Multi-volumetric refocusing of light fields[J]. IEEE Signal Processing Letters, 2021, 28: 31-35.
- [8] Wang T C, Efros A A, Ramamoorthi R. Depth estimation with occlusion modeling using light-field cameras[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2170-2181.
- [9] Feng W, Gao J H, Qu T, et al. Three-dimensional reconstruction of light field based on phase similarity[J]. Sensors, 2021, 21(22): 7734.
- [10] 殷永凯, 于锴, 于春展, 等. 几何光场三维成像综述[J]. 中国激光, 2021, 48(12): 1209001.
Yin Y K, Yu K, Yu C Z, et al. 3D imaging using geometric light field: a review[J]. Chinese Journal of Lasers, 2021, 48(12): 1209001.
- [11] Wanner S, Goldluecke B. Globally consistent depth labeling of 4D light fields[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 41-48.
- [12] Lü H J, Gu K Y, Zhang Y B, et al. Light field depth estimation exploiting linear structure in EPI[C]//2015 IEEE International Conference on Multimedia & Expo Workshops, June 29-July 3, 2015, Turin. New York: IEEE Press, 2015.
- [13] Zhang Y B, Lü H J, Liu Y B, et al. Light-field depth estimation via epipolar plane image analysis and locally linear embedding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(4): 739-747.
- [14] Tao M W, Hadap S, Malik J, et al. Depth from combining defocus and correspondence using light-field cameras[C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 673-680.
- [15] Tao T Y, Chen Q, Feng S J, et al. Active depth estimation from defocus using a camera array[J]. Applied Optics, 2018, 57(18): 4960-4967.
- [16] Williem, Park I K, Lee K M. Robust light field depth estimation using occlusion-noise aware data costs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(10): 2484-2497.
- [17] Cai Z W, Liu X L, Peng X, et al. Structured light field 3D imaging[J]. Optics Express, 2016, 24(18): 20324-20334.
- [18] Cai Z W, Liu X L, Peng X, et al. Ray calibration and phase mapping for structured-light-field 3D reconstruction [J]. Optics Express, 2018, 26(6): 7598-7613.
- [19] Zhang X J, Cai Z W, Liu X L, et al. Improved 3D imaging and measurement with fringe projection structured light field[J]. Proceedings of SPIE, 2020, 11438: 114380X.
- [20] Zhou P, Zhang Y T, Yu Y L, et al. 3D reconstruction from structured light field by Fourier transformation profilometry[J]. Proceedings of SPIE, 2019, 11338: 113381K.
- [21] Zhou P, Zhang Y T, Yu Y L, et al. 3D shape measurement based on structured light field imaging[J]. Mathematical Biosciences and Engineering, 2019, 17(1): 654-668.
- [22] Wang Z W, Yang Y, Liu X L, et al. Light-field-assisted phase unwrapping of fringe projection profilometry[J]. IEEE Access, 2021, 9: 49890-49900.
- [23] Cai Z W, Liu X L, Pedrini G, et al. Accurate depth estimation in structured light fields[J]. Optics Express, 2019, 27(9): 13532-13546.
- [24] Cai Z W, Liu X L, Pedrini G, et al. Structured-light-field 3D imaging without phase unwrapping[J]. Optics and Lasers in Engineering, 2020, 129: 106047.
- [25] Liu L, Xiang S, Deng H P, et al. Fast geometry estimation for phase-coding structured light field[C]//2020 IEEE International Conference on Visual Communications and Image Processing, December 1-4, 2020, Macao, China. New York: IEEE Press, 2020: 124-127.
- [26] Shin C, Jeon H G, Yoon Y, et al. EPINET: a fully-convolutional neural network using epipolar geometry for depth from light field images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4748-4757.
- [27] Tsai Y J, Liu Y L, Ouhyoung M, et al. Attention-based view selection networks for light-field disparity estimation [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12095-12103.
- [28] Huang Z C, Hu X M, Xue Z, et al. Fast light-field disparity estimation with multi-disparity-scale cost aggregation[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 6300-6309.
- [29] 罗姚翔. 基于卷积神经网络的光场图像深度估计技术研究

- 究[D]. 杭州: 杭州电子科技大学, 2018.
- Luo Y X. The research of depth estimation for light field based on convolutional neural network[D]. Hangzhou: Hangzhou Dianzi University, 2018.
- [30] 潘志伟. 基于卷积神经网络的光场图像深度估计[D]. 杭州: 浙江工商大学, 2018.
- Pan Z W. Depth estimation on 4D light field based convolutional neural network[D]. Hangzhou: Zhejiang Gongshang University, 2018.
- [31] 马浩鑫. 基于卷积神经网络的光场深度估计方法, 系统及介质: CN109544621A[P]. 2019-03-29.
- Ma H X. Method, system and medium for estimating optical field depth based on convolution neural network: CN109544621A[P]. 2019-03-29.
- [32] Li Y X, Qian J M, Feng S J, et al. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement[J]. Opto-Electronic Advances, 2022(5): 33-48.
- [33] Yin W, Hu Y, Feng S J, et al. Single-shot 3D shape measurement using an end-to-end stereo matching network for speckle projection profilometry[J]. Optics Express, 2021, 29(9): 13388-13407.
- [34] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [35] Chen B, Zhang S. High-quality 3D shape measurement using saturated fringe patterns[J]. Optics and Lasers in Engineering, 2016, 87: 83-89.
- [36] Zhang S, Sheng H, Li C, et al. Robust depth estimation for light field via spinning parallelogram operator[J]. Computer Vision and Image Understanding, 2016, 145: 148-159.