CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)

光电工程
Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

基于子光场遮挡融合的无监督光场深度估计

李豪宇，陈晔曜，蒋志迪，蒋刚毅，郁梅

引用本文:

李豪宇，陈晔曜，蒋志迪，等. 基于子光场遮挡融合的无监督光场深度估计[J]. 光电工程，2024， **51**(10): 240166.

Li H Y, Chen Y Y, Jiang Z D, et al. Unsupervised light field depth estimation based on sub-light field occlusion fusion[J]. *Opto-Electron Eng*, 2024, **51**(10): 240166.

https://doi.org/10.12086/oee.2024.240166

收稿日期: 2024-07-15; 修改日期: 2024-09-06; 录用日期: 2024-09-10

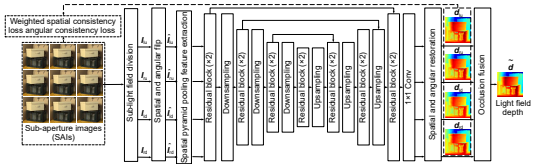
相关论文

Opto-Electronic Engineering
光 电 工 程

Article
2024 年, 第 51 卷, 第 10 期

DOI: 10.12086/oee.2024.240166 CSTR: 32245.14.oee.2024.240166

基于子光场遮挡融合的无监督光场深度估计



李豪宇[1]，陈晔曜[1]，蒋志迪[2]，蒋刚毅[1]，郁　梅[1*]

[1]宁波大学信息科学与工程学院，浙江 宁波 315211；
[2]宁波大学科学技术学院，浙江 宁波 315300

摘要：光场深度估计是光场处理和应用领域的重要科学问题。然而，现有研究忽略了光场视图间的几何遮挡关系。本文通过对不同视图间遮挡的分析，提出了一种基于子光场遮挡融合的无监督光场深度估计方法。该方法首先采用一种有效的子光场划分机制来考虑不同角度位置处的深度关系，具体是将光场子孔径阵列的主副对角线上的视图按左上、右上、左下、右下分为四个子光场。然后，利用空间金字塔池化特征提取和 U-Net 网络来估计子光场深度。最后，设计了一种遮挡融合策略来融合所有子光场深度以得到最终深度，该策略对在遮挡区域具有更高精度的子光场深度图赋予更大的权重，从而减小遮挡影响。此外，引入了加权空间一致性损失和角度一致性损失以约束网络训练并增强鲁棒性。实验结果表明，所提出方法在定量指标和定性比较上展现出了良好的性能。

关键词：光场；深度估计；无监督；子光场划分；遮挡融合

中图分类号：TP391.4　　**文献标志码：**A

李豪宇，陈晔曜，蒋志迪，等. 基于子光场遮挡融合的无监督光场深度估计 [J]. 光电工程，2024，**51**(10): 240166
Li H Y, Chen Y Y, Jiang Z D, et al. Unsupervised light field depth estimation based on sub-light field occlusion fusion[J]. *Opto-Electron Eng*, 2024, **51**(10): 240166

Unsupervised light field depth estimation based on sub-light field occlusion fusion

Li Haoyu[1], Chen Yeyao[1], Jiang Zhidi[2], Jiang Gangyi[1], Yu Mei[1*]

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China;
[2]College Science & Technology, Ningbo University, Ningbo, Zhejiang 315300, China

Abstract: Light field depth estimation is an important scientific problem of light field processing and applications. However, the existing studies ignore the geometric occlusion relationship among views in the light field. By analyzing the occlusion among different views, an unsupervised light field depth estimation method based on sub-light field occlusion fusion is proposed. The proposed method first adopts an effective sub-light field division mechanism to consider the depth relationship at different angular positions. Specifically, the views on the primary and secondary diagonals of the light field sub-aperture arrays are divided into four sub-light fields, i.e., top-left, top-right, bottom-left, and bottom-right. Then, a spatial pyramid pooling feature extraction and a U-Net network are leveraged to estimate the depths of the sub-light fields. Finally, an occlusion fusion strategy is designed to fuse all sub-light field depths to obtain the final depth. This strategy assigns greater weights to the sub-light field depth with

收稿日期：2024-07-15；修回日期：2024-09-06；录用日期：2024-09-10
基金项目：国家自然科学基金资助项目 (62271276, 62071266)；浙江省自然科学基金资助项目 (LQ24F010002)
*通信作者：郁梅，yumei@nbu.edu.cn。

higher accuracy in the occlusion region, thus reducing the occlusion effect. In addition, a weighted spatial and an angular consistency loss are employed to constrain network training and enhance robustness. Experimental results demonstrate that the proposed method exhibits favorable performance in both quantitative metrics and qualitative comparisons.

Keywords: light field; depth estimation; unsupervised; sub-light field division; occlusion fusion

1 引　言

光场描述了三维空间中光线的强度和方向信息，为计算机视觉和图像处理提供了更加全面的基础数据。光场深度估计是计算机视觉中一个经典的研究课题，也是理解场景的基本任务，在新视点合成[1]、三维重建[2]、多曝光融合[3]、超分辨率[4]等各种视觉应用中具有重要意义。

现有光场深度估计方法可大致分为传统优化方法、有监督学习方法和无监督学习方法三种。这其中，传统优化方法[5-8]一般是采用手工特征提取以构造成本代价函数来进行初始深度估计，而后通过深度优化以得到最终深度。该类方法在具有清晰纹理和结构的场景中具有较好的深度估计性能，但难以有效应对弱纹理场景，并且存在计算时间长以及易受噪声和遮挡影响的问题。有监督学习方法[9-11]是在以真实深度图为监督信号的前提下，通过网络来学习光场图像与深度图之间的非线性映射关系，其包括三部分，即对光场空间、角度、极平面图特征提取、代价体积构造，以及深度回归。相比于传统优化方法而言，有监督学习方法在节省计算时间的同时显著提高了深度估计的精度，但其在缺乏真值深度的真实场景中泛化能力有限。此外，对于现实任务而言，真值深度通常是难以获取的，这进而限制了有监督学习方法的实际价值。无监督学习方法是指在无需真值深度的前提下，通过探索光场内在纹理和几何信息以预测场景深度。通常地，该类方法是基于光度一致性假设，即通过预测深度来计算光场各子视图之间的绘制误差，进而约束网络训练。Srinivasan 等[12]使用卷积神经网络从单幅彩色图像合成四维光场，并采用无监督学习进行深度推理。Peng 等[13]通过预定义深度将子孔径图像绘制到中心视图，再用网络学习均值和方差特征来估计深度。然而，由于遮挡和无纹理问题没有明确解决，其估计的深度图表现出相对有限的精度。Zhou 等[14]研究了光场的多方位极线几何特性，并引入光度损失、散焦损失和对称损失来联合约束网络训练以提高深度估计性能。Jin 等[15]将光场进行分割来分别估计各子部分的深度图，之后根据遮挡关系进行融合以得到最终深度图，但其融合策略会产生噪点，进而影响精度。Zhang 等[16]提出了一种基于估计误差的视差融合策略，并设计一种由粗到细的网络架构来估计深度。综上而言，无监督学习方法具有更高的泛化性，但由于缺少真值深度，估计精度尚待提高。此外，现有无监督光场深度估计方法仍存在以下挑战，一是未充分考虑光场视图间的几何关系，二是难以准确估计遮挡区域深度。

针对上述问题，本文提出了一种基于子光场遮挡融合的无监督光场深度估计方法，其通过探索光场各角度位置处的遮挡关系以准确预测场景深度。实验结果表明，所提出方法在合成场景和真实场景都取得较好性能，并与现有方法相比具有更好的稳健性和泛化能力。本文的主要贡献如下：1) 提出了一种子光场划分机制，其反映了光场视图在不同角度位置处的遮挡和深度关系；2) 设计了一种遮挡融合策略，旨在对遮挡区域具有更高精度的子光场深度图赋予更大融合权重，从而提高深度估计精度；3) 构建了加权空间一致性损失和角度一致性损失以有效约束网络训练，并提高网络估计的鲁棒性。

2 基于子光场遮挡融合的无监督光场深度估计方法

基于上述分析，本文提出了一种基于子光场遮挡融合的无监督光场深度估计方法，其通过融合子光场深度以减小遮挡影响并提高估计精度。本节内容首先对遮挡现象进行分析，而后提出子光场划分机制；其次，阐述所提出方法中涉及的网络结构以及设计的遮挡融合策略；最后描述用于约束网络训练的损失函数，包括加权空间一致性损失和角度一致性损失。

2.1 遮挡分析和子光场划分机制

光场记录了光线的强度和方向信息，可表示为四维结构，表示为 $\boldsymbol{I}_{\boldsymbol{u}}(\boldsymbol{x})=\boldsymbol{I}(u,v,x,y)\in\mathbb{R}^{M\times N\times H\times W}$，其中，

$\boldsymbol{u}=(u,v)$ 表示角度坐标，$\boldsymbol{x}=(x,y)$ 表示空间坐标，$M\times N$ 表示角度分辨率，$H\times W$ 表示空间分辨率。需要指出，本文中角度分辨率取 $M=N$。子孔径图像阵列是四维光场的有效可视化方式，其中心视图和周围视图的几何关系可表示如下：

$$\hat{\boldsymbol{I}}_{\boldsymbol{u}\to\boldsymbol{u}_c}(\boldsymbol{x})=\boldsymbol{I}_{\boldsymbol{u}}(\boldsymbol{x}+(\boldsymbol{u}_c-\boldsymbol{u})\boldsymbol{d}(\boldsymbol{x}))\,, \tag{1}$$

式中：$\boldsymbol{u}_c=(u_c,v_c)$ 表示中心视图的角度坐标，$\boldsymbol{u}=(u,v)$ 表示其余视图的角度坐标，$\boldsymbol{d}(\boldsymbol{x})$ 表示中心视图在 $\boldsymbol{x}=(x,y)$ 处的视差 (由于视差与深度成反比关系，本文对两者不作区分)。

基于此，可以通过最小化光度一致性损失[17]来以无监督方式训练深度估计网络，如下所示：

$$L_{\mathrm{rec}}(\tilde{\boldsymbol{d}})=\sum_{\boldsymbol{u}}\sum_{\boldsymbol{x}}\left\|\hat{\boldsymbol{I}}_{\boldsymbol{u}\to\boldsymbol{u}_c}\left(\boldsymbol{x};\tilde{\boldsymbol{d}}\right)-\boldsymbol{I}_{\boldsymbol{u}_c}(\boldsymbol{x})\right\|_1\,, \tag{2}$$

式中：$\hat{\boldsymbol{I}}_{\boldsymbol{u}\to\boldsymbol{u}_c}\left(\boldsymbol{x};\tilde{\boldsymbol{d}}\right)$表示利用预测的深度图$\tilde{\boldsymbol{d}}$将周围视图绘制到中心视点后得到的图像，$\boldsymbol{I}_{\boldsymbol{u}_c}(\boldsymbol{x})$表示中心视图。

在遮挡区域，光度一致性损失，即式 (2) 将失去深度估计精度，但光场的角度信息为遮挡检测和补偿提供了有利条件。图 1 展示了利用真实深度将左右视图绘制到中心视点所得图像与真实中心视点图像之间的误差图，可观察到高亮区域 (即大误差区域) 显示了遮挡。特别地，该区域通常位于对象边界，并且在左右视图中处于相反位置，即左视图中遮挡出现在对象右边界，右视图中遮挡出现在对象左边界，这表明光场中左右视图至少有一个候选者可用来进行准确的深度估计。在光场上下视图绘制误差中也有类似现象，即上视图中遮挡出现在对象下边界，下视图中遮挡出现在对象上边界。因此，降低用左视图预测的深度在对象右边界的权重，以及降低用右视图预测的深度在对象左边界的权重，再进行融合可以消减遮挡对深度估计的影响。

基于上述遮挡分析，本文将光场子孔径阵列分成左上、右上、左下、右下四部分以分别估计深度。这里，以左上部分为例进行阐述，其反映的遮挡存在于物体右下边界。为减小视图信息冗余，可进一步选用对角线上的视图作为输入，其他部分同理。

图 2 展示了划分的子光场，具体地，将子孔径阵列主副对角线上的光场视图按左上、右上、左下、右下分成四个子光场记为$\boldsymbol{I}_{\mathrm{lu}}=\boldsymbol{I}_{\boldsymbol{u}\in\boldsymbol{U}_{\mathrm{lu}}}(\boldsymbol{x})$，$\boldsymbol{I}_{\mathrm{ru}}=\boldsymbol{I}_{\boldsymbol{u}\in\boldsymbol{U}_{\mathrm{ru}}}(\boldsymbol{x})$，$\boldsymbol{I}_{\mathrm{ld}}=\boldsymbol{I}_{\boldsymbol{u}\in\boldsymbol{U}_{\mathrm{ld}}}(\boldsymbol{x})$，$\boldsymbol{I}_{\mathrm{rd}}=\boldsymbol{I}_{\boldsymbol{u}\in\boldsymbol{U}_{\mathrm{rd}}}(\boldsymbol{x})$，式中，$\boldsymbol{U}_{\mathrm{lu}}=\{\boldsymbol{u}=(u,v)|\,1\leqslant u\leqslant M_c,\ v=u\}$，$\boldsymbol{U}_{\mathrm{ru}}=\{\boldsymbol{u}=(u,v)|\,1\leqslant u\leqslant M_c,\ v=M-u+1\}$，$\boldsymbol{U}_{\mathrm{ld}}=\{\boldsymbol{u}=(u,v)|\,M_c\leqslant u\leqslant M,\ v=M-u+1\}$，$\boldsymbol{U}_{\mathrm{rd}}=\{\boldsymbol{u}=(u,v)|\,M_c\leqslant u\leqslant M,\ v=u\}$，$M_c=(M+1)/2$ 表示中心视图的角度位置。

为减小网络参数，将上述构造的四个子光场输入到权重共享的深度估计网络中进行深度推断。考虑到不同子光场的位移方向不同，利用角度和空间翻转以保证不同子光场中各边界视图与中心视图的位移方向一致，进而统一优化网络训练。简单而言，1) 左上角子光场保持不变；2) 右上角子光场在空间和角度上进行水平翻转；3) 左下角子光场在空间和角度上进行垂直翻转；4) 右下角子光场在空间和角度上进行水平翻转和垂直翻转，具体翻转过程表示如下：

$$\hat{\boldsymbol{I}}_{\mathrm{lu}}(u,v,x,y)=\boldsymbol{I}_{\mathrm{lu}}(u,v,x,y)\,,$$

$$\hat{\boldsymbol{I}}_{\mathrm{ru}}(u,v,x,y)=\boldsymbol{I}_{\mathrm{ru}}(u,N-v,x,W-y)\,,$$

$$\hat{\boldsymbol{I}}_{\mathrm{ld}}(u,v,x,y)=I_{\mathrm{ld}}\left(M-u,v,H-x,y\right)\,,$$

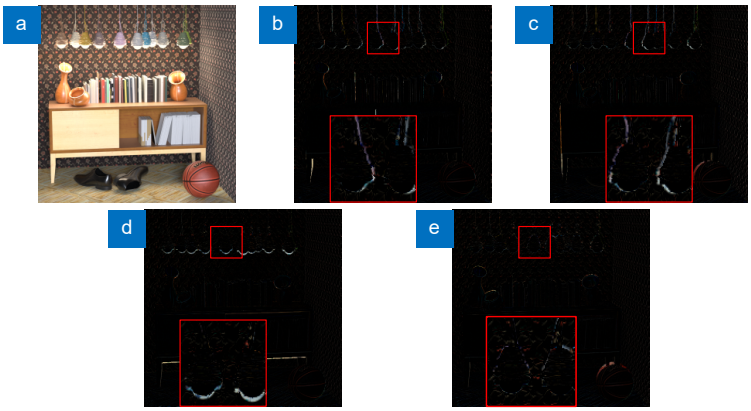


图 1 中心视图与左、右、上、下视图绘制误差示例。

(a) 中心视图；(b) 左视图绘制误差；(c) 右视图绘制误差；(d) 上视图绘制误差；(e) 下视图绘制误差

Fig. 1 Illustrations of center view and warping errors of left, right, top, and bottom views. (a) Center view; (b) Warping error of left view; (c) Warping error of right view; (d) Warping error of top view; (e) Warping error of bottom view

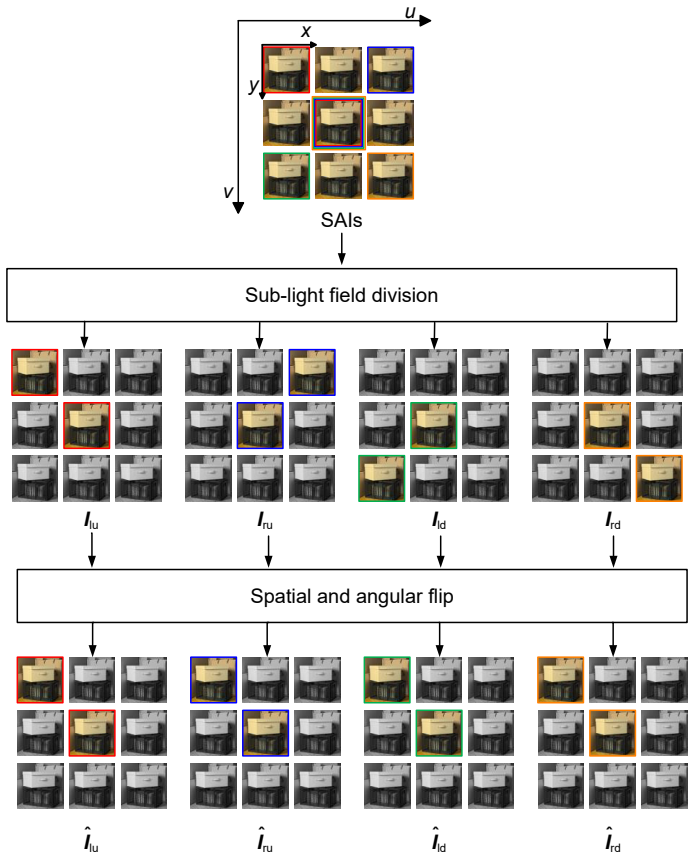


图 2 子光场划分和空间角度翻转，以 3×3 为示例

Fig. 2 Sub-light field division and spatial and angular flip (3×3 as an example)

$$\hat{\boldsymbol{I}}_{\mathrm{rd}}(u,v,x,y)=\boldsymbol{I}_{\mathrm{rd}}(M-u,N-v,H-x,W-y), \quad (3)$$

式中：$\boldsymbol{I}_{\mathrm{lu}}(u,v,x,y)$，$\boldsymbol{I}_{\mathrm{ru}}(u,v,x,y)$，$\boldsymbol{I}_{\mathrm{ld}}(u,v,x,y)$，$\boldsymbol{I}_{\mathrm{rd}}(u,v,x,y)$ 分别表示翻转前左上、右上、左下、右下的子光场，$M\times N$ 表示角度分辨率，$H\times W$ 表示空间分辨率。

2.2 基于子光场遮挡融合的无监督光场深度估计网络

为有效从子光场中估计深度，本文设计了一个基于 U-Net 架构的深度估计网络，如图 3 所示。首先将子孔径图像阵列按主副对角线分成四个子光场并进行空间角度翻转以得到 $\hat{\boldsymbol{I}}_{\mathrm{lu}}$, $\hat{\boldsymbol{I}}_{\mathrm{ru}}$, $\hat{\boldsymbol{I}}_{\mathrm{ld}}$, $\hat{\boldsymbol{I}}_{\mathrm{rd}}$(见式 (3))。之后，利用空间金字塔池化 (spatial pyramid pooling, SPP)[9] 进行初始特征提取以将输入子光场映射到高维特征空间。如图 4 所示，SPP 模块由多个二维卷积、残差块和四层平均池化块组成，其中平均池化块中的核尺寸依次为 2×2、4×4、8×8 和 16×16。通过四种池化可得到多尺度上下文信息，而后通过双线性插值来将这些多尺度特征上采样到相同空间尺寸并进行级联以作为初始特征。在挑战性区域，如无纹理区域和反射区域，这些特征可互为补充，进而促进后续深度估计。

其后，将初始特征输入权重共享的 U-Net 网络中进行深度回归。在 U-Net 编码阶段的每个尺度中，采用两个残差块[18] 和一层基于最大池化的下采样层进行特征提取。在解码阶段，采用对称的结构，即联合残差块和上采样层来将编码特征进行空间尺寸恢复，其中，上采样由转置卷积实现。此外，为增强编码和解码之间特征信息流，在每个尺度中引入跳转连接，其通过特征级联实现。最后，使解码特征通过一个残差块和一个 1×1 卷积层以输出深度图；同时还采用额外的一个 1×1 卷积层来估计置信度图，用于损失计算。考虑到光场中的遮挡问题，对每个子光场所估计的深度图进行遮挡融合 (详见 2.3 节所述) 以生成最终深度图，见图 3 右侧所示。

2.3 遮挡融合策略

由于不同角度位置的视图中遮挡出现在对象的不同边界，导致各子光场在不同边界处的深度不准确。为此，采用遮挡融合策略以消减边界处遮挡对深度估计的影响。具体地，首先对输入子光场深度 $\boldsymbol{d}_{\mathrm{lu}}$, $\boldsymbol{d}_{\mathrm{ru}}$,

$\boldsymbol{d}_{\rm ld}$, $\boldsymbol{d}_{\rm rd}$ 中的对象边界处的遮挡区域进行检测，过程如下：

$$\boldsymbol{o}_k = \text{clip}\left\{\left|\text{gray}\left[\hat{\boldsymbol{I}}_{\boldsymbol{u}_k\to\boldsymbol{u}_c}(\boldsymbol{x};\boldsymbol{d}_k) - \boldsymbol{I}_{\boldsymbol{u}_c}(\boldsymbol{x})\right]\right|\right\}, \quad (4)$$

式中：$\boldsymbol{o}_k$ 表示检测得到的遮挡掩膜，k = lu, ld, ru, rd (lu 表示左上，ld 表示左下，ru 表示右上，rd 表示右下)，$\boldsymbol{u}_{\rm lu}$=(1,1)，$\boldsymbol{u}_{\rm ld}$=(M,1)，$\boldsymbol{u}_{\rm ru}$=(1,M)，$\boldsymbol{u}_{\rm rd}$=(M,M)，gray(·) 表示灰度化，clip(·) 表示将数值限制在到 [0,1] 范围内。

$\boldsymbol{o}_k$ 中数值越大表示遮挡越严重，因而为降低子光场深度中遮挡区域的负面影响，可以用$\|1\text{-}\boldsymbol{o}_k\|_2$ 来表示子光场深度权重，并用 softmax 函数以获取子光场深度的权重相对重要性，如下所示：

$$\hat{\boldsymbol{o}}_{\rm lu}, \hat{\boldsymbol{o}}_{\rm ru}, \hat{\boldsymbol{o}}_{\rm ld}, \hat{\boldsymbol{o}}_{\rm rd} = softmax(\|1-\boldsymbol{o}_{\rm lu}\|_2, \cdots \|1-\boldsymbol{o}_{\rm ru}\|_2, \|1-\boldsymbol{o}_{\rm ld}\|_2, \|1-\boldsymbol{o}_{\rm rd}\|_2). \quad (5)$$

通过将深度与对应权重逐元素相乘并相加，可消减对象边界处遮挡的影响，进而实现有效的子光场深度融合，如下所示：

$$\tilde{\boldsymbol{d}} = \hat{\boldsymbol{o}}_{\rm lu}\odot\boldsymbol{d}_{\rm lu} + \hat{\boldsymbol{o}}_{\rm ld}\odot\boldsymbol{d}_{\rm ld} + \hat{\boldsymbol{o}}_{\rm ru}\odot\boldsymbol{d}_{\rm ru} + \hat{\boldsymbol{o}}_{\rm rd}\odot\boldsymbol{d}_{\rm rd}, \quad (6)$$

式中：⊙表示逐元素相乘。

根据 2.1 小节的分析可知，由左上角光场视图估计出来的深度 $\boldsymbol{d}_{\rm lu}$ 在对象右下边界处的估计是不准确的。因此，计算出来的对应权重$\hat{\boldsymbol{o}}_{\rm lu}$消减了对象右下边界处遮挡对深度估计的影响。同理，右上深度权重$\hat{\boldsymbol{o}}_{\rm ru}$降低了对象左下边界处遮挡对深度估计的影响；左下深度权重$\hat{\boldsymbol{o}}_{\rm ld}$中降低了对象右上边界处遮挡的影响；右下深度权重$\hat{\boldsymbol{o}}_{\rm rd}$中降低了对象左上边界处遮挡的影响。通过上述遮挡融合可消减对象边界处遮挡的影响，进而提高所估计深度图的精度。

2.4 损失函数

尽管式 (2) 在遮挡处会失去估计精度，但根据划分的子光场而估计得到的子光场深度 $\boldsymbol{d}_{\rm lu}$, $\boldsymbol{d}_{\rm ru}$, $\boldsymbol{d}_{\rm ld}$, $\boldsymbol{d}_{\rm rd}$ 中总有一个深度是可靠的。因此，采用加权空间一致性损失 ($L_{\rm spa}$) 来优化式 (2)，即基于绘制误差的可靠性来对子光场深度进行加权，表示如下：

$$L_{\rm spa}(\boldsymbol{d}) = \sum_{i=1}^{4}\sum_{\boldsymbol{u}\in U_i}\sum_{\boldsymbol{x}} \boldsymbol{c}_i \left|\hat{\boldsymbol{I}}_{\boldsymbol{u}\to\boldsymbol{u}_c}(\boldsymbol{x};\boldsymbol{d}_i) - \boldsymbol{I}_{\boldsymbol{u}_c}(\boldsymbol{x})\right|, \quad (7)$$

式中：为方便起见，将 $\boldsymbol{d}_{\rm lu}$, $\boldsymbol{d}_{\rm ru}$, $\boldsymbol{d}_{\rm ld}$, $\boldsymbol{d}_{\rm rd}$ 分别表示为 $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, $\boldsymbol{d}_3$, $\boldsymbol{d}_4$，$\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$, $\boldsymbol{c}_4$ 表示对应的置信度图，由 U-Net 网络最后一层 1×1 卷积预测得到，且$\sum_{i=1}^{4}\boldsymbol{c}_i = 1$。注意，$\boldsymbol{U}_1$, $\boldsymbol{U}_2$, $\boldsymbol{U}_3$, $\boldsymbol{U}_4$ 与划分子光场的 $\boldsymbol{U}_{\rm lu}$, $\boldsymbol{U}_{\rm ru}$, $\boldsymbol{U}_{\rm ld}$, $\boldsymbol{U}_{\rm rd}$ 不同，其不仅包含对角线上的视图，还增加了对角线周围的视图，即是以中心视图为顶点的左上、右上、左下、右下的所有视图。具体而言，$\boldsymbol{U}_1$={$\boldsymbol{u}$=(u,v)| 1≤ u≤

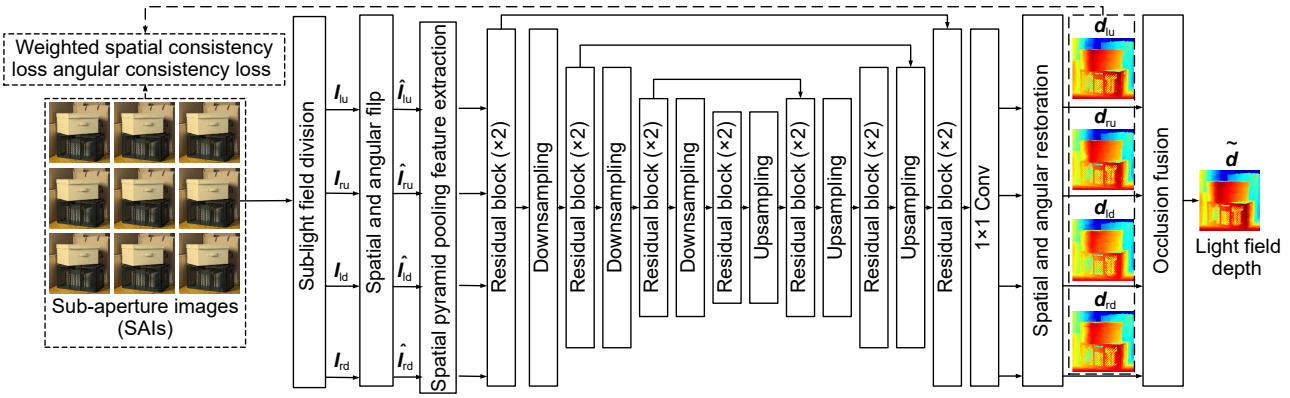


图 3 所提出方法的整体网络框架

Fig. 3 Overall network framework of the proposed method

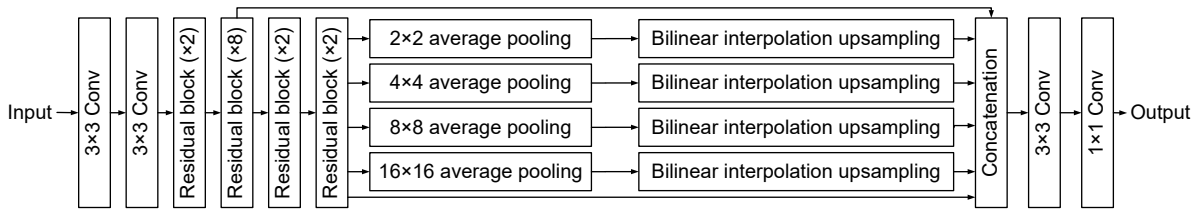


图 4 空间金字塔池化特征提取模块

Fig. 4 Spatial pyramid polling feature extraction model

$M_c,1\leqslant v\leqslant M_c\}$，$\boldsymbol{U}_2=\{\boldsymbol{u}=(u,v)|1\leqslant u\leqslant M_c,M_c\leqslant v\leqslant M\}$，$\boldsymbol{U}_3=\{\boldsymbol{u}=(u,v)|\ M_c\leqslant u\leqslant M,1\leqslant v\leqslant M_c\}$，$\boldsymbol{U}_4=\{\boldsymbol{u}=(u,v)|\ M_c\leqslant u\leqslant M,M_c\leqslant v\leqslant M\}$。

加权空间一致性损失只考虑了单幅视图按深度绘制后与中心视图的误差。显然，当深度估计准确时，每个视图按深度绘制后应与中心视图一致，且它们之间也应一致，因而进一步提出了角度一致性损失 (L_{ang}) 来正则网络训练，计算过程如下所示：

$$L_{\text{ang}}(\boldsymbol{d})=\frac{1}{4}\sum_{i=1}^{4}\text{std}\left(\hat{\boldsymbol{I}}_{(\boldsymbol{u}\in\boldsymbol{U}_i)\to\boldsymbol{u}_c}(\boldsymbol{x};\boldsymbol{d}_i),\boldsymbol{I}_{\boldsymbol{u}_c}(\boldsymbol{x})\right), \tag{8}$$

式中：std(·) 表示标准差算子，该值越小说明图像间误差越小。

为提高深度的平滑度，同时保持对象的边界结构，采用边界感知的平滑度损失 (L_{sm})，计算如下：

$$\begin{aligned}L_{\text{sm}}=&\frac{1}{4}\sum_{i=1}^{4}\frac{1}{2}\sum_{x}\exp\left[-\gamma\left|\frac{\partial\boldsymbol{I}_{\boldsymbol{u}_c}}{\partial x}(\boldsymbol{x})\right|\right]\left|\frac{\partial\boldsymbol{d}_i}{\partial x}(\boldsymbol{x})\right|\\&+\exp\left[-\gamma\left|\frac{\partial\boldsymbol{I}_{\boldsymbol{u}_c}}{\partial y}(\boldsymbol{x})\right|\right]\left|\frac{\partial\boldsymbol{d}_i}{\partial y}(\boldsymbol{x})\right|,\end{aligned} \tag{9}$$

式中：exp(·) 表示以实数 e 为底的指数函数，γ 表示边界权重，根据文献 [15,17]，其值设置为 150 以平衡深度平滑度与边界深度不连续性。

上述三个损失共同组成最终损失，表示如下：

$$Loss=L_{\text{spa}}+L_{\text{ang}}+0.1L_{\text{sm}}. \tag{10}$$

3 实验结果与分析

在本小节中，首先介绍实验采用的数据集以及实验环境，然后将所提出方法与现有方法进行多维度比较，最后通过消融实验来验证所提出方法中核心部分的有效性。

3.1 数据集与实验环境

本文采用网上公开的合成光场数据集和真实光场数据集进行实验。对于前者，使用 HCI new 数据集[19] 中的 Additional 集 (包含 16 个场景) 作为训练集，另选取 HCI new 数据集中额外 4 个场景以及 DLF 数据集[20] 中 4 个场景作为测试集，其空间分辨率为 512×512，角度分辨率为 7×7。上述合成数据集是通过开源软件 Blender[21] 进行渲染而建立的。对于后者，从 EPFL 数据集[22] 和 Stanford Lytro 数据集[23] 中选取 92 个场景作为训练集，并使用 Stanford Lytro 数据集中额外 4 个场景作为测试集，其空间分辨率为 368×528，角度分辨率为 7×7。上述真实数据集是通过 Lytro Illum 光场相机而建立的。实验中所使用的训练集和测试集没有重叠。

实验环境配置为 Intel(R) Xeon(R) Gold 6230 CPU，64.0 GB 内存，NVIDIA RTX 2080Ti (11 GB)。本文方法采用 PyTorch 深度学习框架进行实现。所构建的网络采用 Adam 优化器进行优化，其中 β_1=0.9，β_2=0.999。批量大小设置为 3，初始学习率设置为 1×10^{-4}，其在每 250 个 epochs 后以 0.5 为比例因子进行衰减。本文方法共训练 1000 个 epochs 可达到收敛。在训练阶段，采用随机水平翻转、垂直翻转和 90°旋转来执行数据增强。在本实验环境下，对于一幅空间分辨率为 512×512、角度分辨率为 7×7 的光场图像，本文算法在 GPU 上运行时间约为 0.01 s，在 CPU 上运行时间约为 12.97 s。

在下文定量比较中，采用均方误差 (mean square error, MSE) 和坏像素率 (bad pixel ratio, BPR) 作为客观质量评价指标来度量不同光场深度估计方法的性能。上述两个指标计算如下：

$$MSE=\frac{1}{H\times W}\sum_{x}|(\boldsymbol{d}_{\text{gt}}(\boldsymbol{x})-\tilde{\boldsymbol{d}}(\boldsymbol{x}))|^2, \tag{11}$$

$$BPR(>t)=\frac{1}{H\times W}\sum_{x}|(\boldsymbol{d}_{\text{gt}}(\boldsymbol{x})-\tilde{\boldsymbol{d}}(\boldsymbol{x}))|>t, \tag{12}$$

式中：H 和 W 表示深度图的高和宽，$\boldsymbol{d}_{\text{gt}}$ 表示真实深度图，$\tilde{\boldsymbol{d}}$表示估计的深度图，t 表示定义坏像素的阈值，根据文献 [8,15]，其值设置为 0.07。

3.2 实验分析

为了评估所提出方法的有效性，将其与现有光场深度估计方法进行定量和定性比较，包括三种传统优化方法，即 OCC [6]、SPO [7] 和 OAVC [8]，以及两种无监督学习方法，即 Unsup [13] 和 OccUnNet [15]。需要指出的是，由于 Unsup 方法未提供 PyTorch 实现版本，本文根据原文对网络结构和超参数设置的描述进行复现。为了保证比较公平，所有无监督学习方法均在相同的场景上进行重新训练并测试。

1) 定量比较：表 1 和表 2 给出了所提出方法与现有方法在合成数据集上的 MSE 及 BPR 的定量比较结果，其值越小表示深度图质量越高。MSE 主要揭示深度图的全局精度，而 BPR 则反映深度图的局部精度。由表 1 可以看出，所提出方法在 4 个测试场景中的 MSE 指标排名第一，在 1 个测试场景中排名第二；特别地，在最终平均 (即 Ave.)MSE 指标上，所提出

方法实现了最低的分数，这表明所提出方法估计的深度的全局精度优于其他方法。由表 2 可以看出，所提出方法的 BPR 指标在 8 个测试场景中均排名第一或第二，并且在平均 BPR 指标上取得了可竞争的结果，这表明所提出方法具有相对较好的局部估计精度。进一步地，联合表 1 和表 2 的对比结果，可以发现所提出方法估计的深度图在整体和局部精度两方面均取得了较为优异的结果。此外，表 3 给出了不同方法的计算效率 (即运行时间) 比较。可以看到，所提出方法在运行时间上远优于所有对比方法，这是由于子光场输入方式减少了光场冗余信息和网络计算量，进而提高运行速率。

2) 定性比较：图 5 至图 8 展示了不同方法对四个合成场景所估计的深度图以及对应坏像素图 (BPR 的可视化结果，红色表示误差大于 0.07 像素值的坏像素点，绿色则表示好像素点) 的视觉结果。从图 5 中黑色箭头所指的弱纹理区域以及网格区域中可以观察到，所提出方法得到的深度图比对比方法所得更接近真值深度图。图 6 是带有分级噪声的场景，通过观察黑色箭头所指区域，可以观察到所提出方法估计的深度图具有更清晰的结构边界，且坏像素更少，这表明提出方法具有一定的抗噪声能力。观察图 7 中上方箭

表 1　所提出方法与其他光场深度估计方法在 MSE(×100) 指标上的定量比较 (加粗表示第一，下划线表示第二)

Table 1　Quantitative comparison of different light field depth estimation methods in terms of MSE (×100) (bolded indicates first, underlined indicates second)

Type	Methods	Boxes	Cotton	Pyramids	Sideboard	Antiques	Pinenuts	Smiling	Toys	**Ave.**
Tradi-tional	OCC [6]	12.22	9.47	1.81	18.76	38.76	46.95	237.26	10.49	46.97
	SPO [7]	9.57	1.99	0.20	**1.34**	**3.00**	1.32	6.53	0.90	3.11
	OAVC [8]	7.46	**1.47**	0.08	1.55	5.38	1.62	**4.77**	1.02	2.92
Unsup-ervised	Unsup [13]	12.21	7.37	0.43	3.82	11.87	30.69	17.68	2.38	10.81
	OccUnNet [15]	6.94	1.68	0.17	7.54	3.23	19.04	4.93	0.70	5.53
	Proposed	**6.61**	2.22	**0.04**	2.07	4.10	**0.70**	4.80	**0.67**	**2.65**

表 2　所提出方法与其他光场深度估计方法在 BPR(>0.07) 指标上的定量比较 (加粗表示第一，下划线表示第二)

Table 2　Quantitative comparison of different light field depth estimation methods in terms of BPR (>0.07) (bolded indicates first, underlined indicates second)

Type	Methods	Boxes	Cotton	Pyramids	Sideboard	Antiques	Pinenuts	Smiling	Toys	**Ave.**
Tradi-tional	OCC [6]	40.05	46.39	12.17	48.81	72.53	62.81	71.03	83.33	54.64
	SPO [7]	42.58	29.79	14.87	32.32	31.63	40.07	18.93	41.51	31.46
	OAVC [8]	**18.59**	**5.19**	2.91	19.59	**6.71**	**11.40**	23.13	**10.17**	**12.21**
Unsup-ervised	Unsup [13]	43.75	23.97	14.75	26.37	33.43	48.94	42.45	29.50	32.90
	OccUnNet [15]	27.39	7.03	6.41	17.50	17.57	43.58	21.77	16.52	19.72
	Proposed	24.70	6.39	**0.78**	**16.26**	9.32	29.95	**15.63**	10.64	14.21

表 3　所提出方法与其他光场深度估计方法在运行时间 (s) 指标上的定量比较，传统方法在 CPU 上运行，无监督方法在 GPU 上运行 (加粗表示第一，下划线表示第二)

Table 3　Quantitative comparison of different light field depth estimation methods in terms of runtime (s), where traditional methods run on CPU and unsupervised methods run on GPU (bolded indicates first, underlined indicates second)

Type	Methods	Boxes	Cotton	Pyramids	Sideboard	Antiques	Pinenuts	Smiling	Toys	**Ave.**
Tradi-tional	OCC [6]	192.05	210.73	319.46	222.10	205.32	172.03	242.53	166.85	216.38
	SPO [7]	831.91	820.72	790.02	814.70	794.63	807.51	803.15	807.74	808.80
	OAVC [8]	16.33	16.60	16.50	16.47	16.62	16.74	16.62	16.72	16.58
Unsup-ervised	Unsup [13]	39.38	39.14	39.55	39.49	39.72	39.47	39.21	38.38	39.29
	OccUnNet [15]	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
	Proposed	**0.01**	**0.01**	**0.01**	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**

头指示区域的分层纹理以及下方箭头指示的松果纹理，所提出方法估计的深度图比对比方法估计的深度图更加清晰。从图 8 中箭头指示的灯区域以及小车区域可以观察到，所提出方法预测的深度图比对比方法预测的深度图更为平滑。综上，所提出方法估计的深度图在整体视觉感知上更接近真值深度图，且在弱纹理和细节纹理区域估计的精度均优于对比方法所得结果。

进一步地，图 9 展示了在真实数据集上的深度估计结果。值得注意的是，真实场景是由 Lytro 相机采集得到，其窄基线特性导致采集的光场图像具有相对较小的视差范围。因此，这里选择具有明显前背景区别的场景进行测试。通过对比可以看到，所提出方法相对其他方法更能将前景与背景深度做出区分，并且在对象边缘以及弱纹理区域估计得更为准确。特别地，在后两个场景中对位于前景处的网格的纹理和边缘估计更为精细，并且具有良好连续性。上述视觉结果表明所提出方法在真实数据集上具有较好的准确性和泛化性。

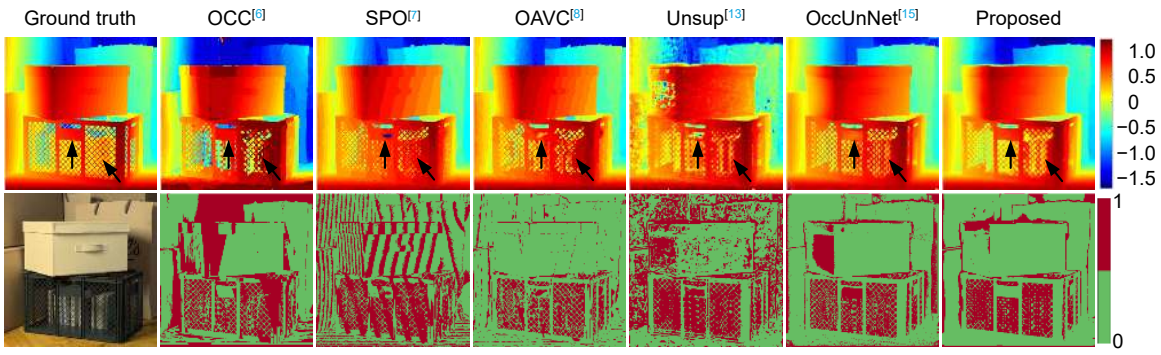


图 5 不同方法在 HCI new 数据集[19] 中的场景 Boxes 上估计的深度图和坏像素图比较

Fig. 5 Comparison of depths and bad pixel maps estimated by different methods on Boxes from the HCI new dataset [19]

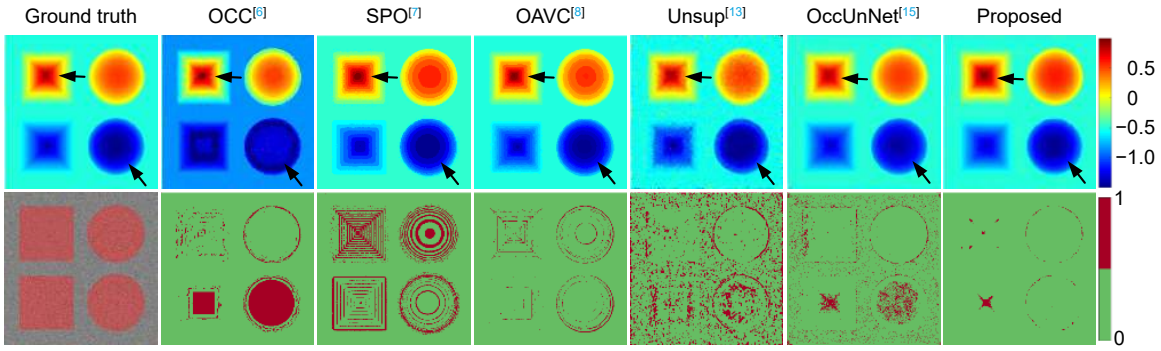


图 6 不同方法在 HCI new 数据集[19] 中的场景 Pyramids 上估计的深度图和坏像素图比较

Fig. 6 Comparison of depths and bad pixel maps estimated by different methods on Pyramids from the HCI new dataset [19]

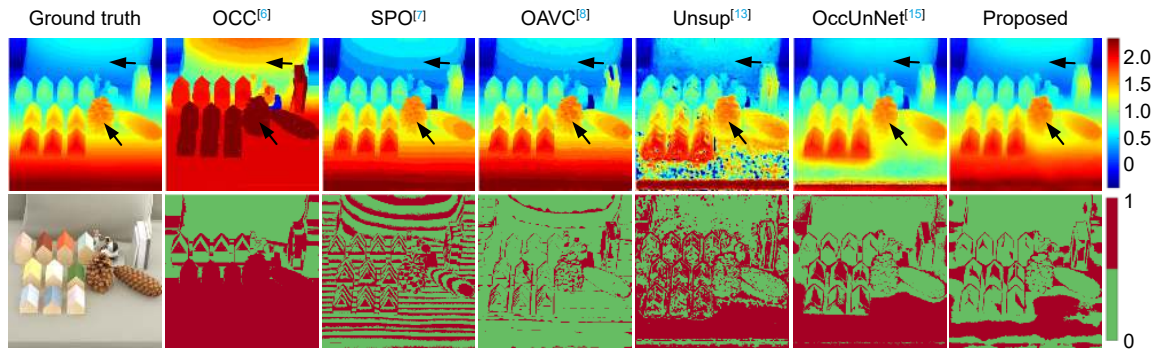


图 7 不同方法在 DLF 数据集[20] 中的场景 Pinenuts 上估计的深度图和坏像素图比较

Fig. 7 Comparison of depths and bad pixel maps estimated by different methods on Pinenuts from the DLF dataset [20]

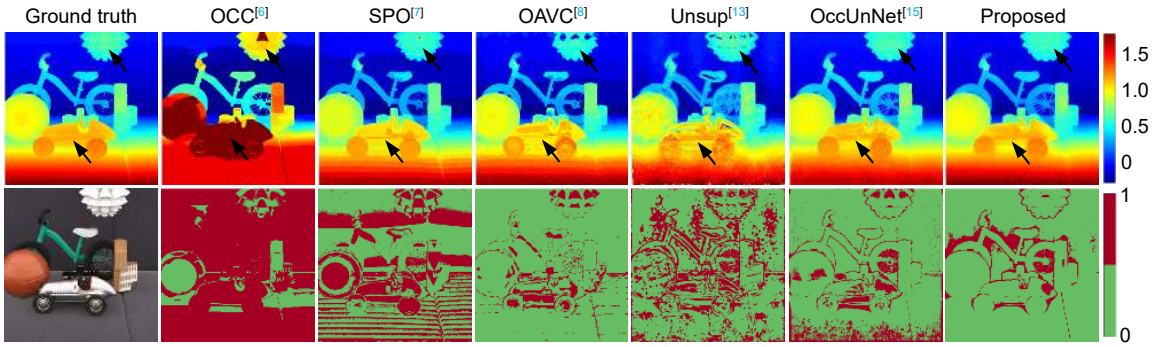


图 8 不同方法在 DLF 数据集[20] 中的场景 Toys 上估计的深度图和坏像素图比较

Fig. 8 Comparison of depths and bad pixel maps estimated by different methods on Toys from the DLF dataset [20]

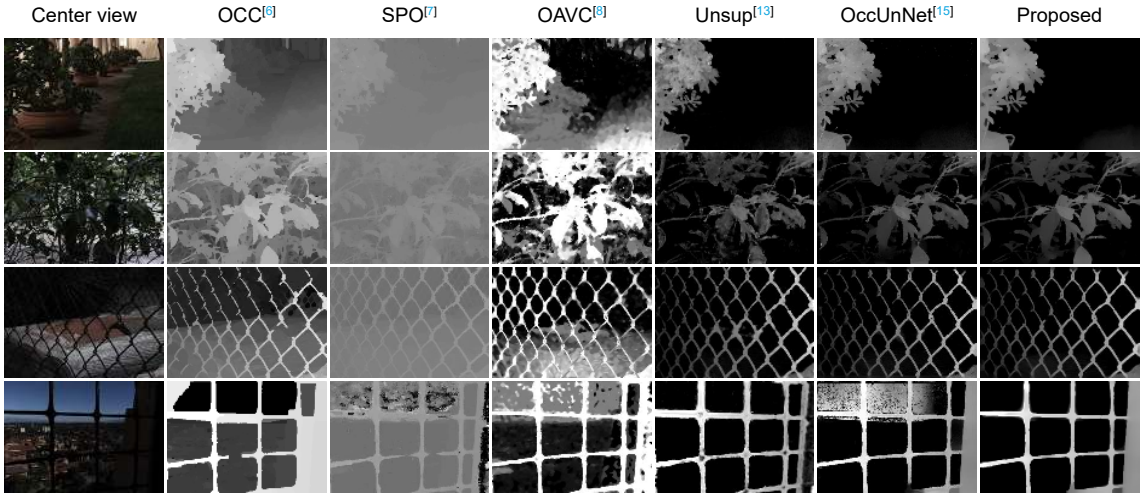


图 9 不同方法在 Stanford Lytro 真实数据集[23] 中估计的深度图比较

Fig. 9 Comparison of depths estimated by different methods on real-world data from the Stanford Lytro dataset[23]

3.3 消融实验

在本小节中，将构建多个消融方案来验证所提出方法中核心部分的有效性。这里，以合成数据集进行实验，并报告平均客观指标 (MSE 和 BPR) 来评估不同消融方案的性能。表 4 给出了消融实验结果。

表 4 消融实验结果

Table 4 Results of ablation experiment

Schemes	MSE(×100)	BPR(>0.07)
Scheme 1	5.85	15.23
Scheme 2	2.74	17.34
Scheme 3	3.04	20.52
Scheme 4	3.54	20.88
Scheme 5	3.74	18.16
Scheme 6	4.02	18.98
Scheme 7	2.89	21.60
Scheme 8	2.75	16.24
Scheme 9	2.89	18.32
Proposed	**2.65**	**14.21**

1) 子光场划分和训练策略：为了验证所提出的子光场划分和训练策略的有效性，本文构建了消融方案 1 和 2。其中，消融方案 1 是将全部视图输入求得一个深度图，并且也以全部视图进行绘制来计算损失函数；消融方案 2 则用本文同样的子光场划分作为输入，但在训练阶段采用子光场划分的视图来计算损失，其相比所提出方法缺少一些视图。从表 4 列出的定量结果中可看到，所提出方法在 MSE 和 BPR 上均优于所构建的两个消融方案，这验证了所提出的子光场划分和训练策略的有效性。

2) 遮挡融合策略：为了验证所设计的遮挡融合策略的有效性，本文构建了消融方案 3 至 7。其中，消融方案 3 直接采用融合前的左上子光场深度计算指标；消融方案 4 采用融合前的右上子光场深度计算指标；消融方案 5 采用融合前的左下子光场深度计算指标；消融方案 6 采用融合前的右下子光场深度计算指标；消融方案 7 不使用式 (5) 中的 softmax 函数来获取遮

挡掩膜权重。通过表 4 中数值比较可以发现，所提出方法在 MSE 和 BPR 上均优于所构建的五个消融方案，这验证了所提出的遮挡融合策略的有效性。

3) 损失函数：为了验证所提出的损失函数的有效性，本文构建了消融方案 8 和 9。其中，消融方案 8 对空间一致性损失去除加权策略，即使用式 (2) 直接计算空间一致性损失；消融方案 9 去除角度一致性损失。从表 4 中可看到，所提出方法在 MSE 和 BPR 上均优于所构建的两个消融方案，这验证了损失函数的有效性。

4 结　论

针对光场视图间遮挡对深度估计的影响，本文提出了一种基于子光场遮挡融合的无监督光场深度估计方法。该方法通过对划分子光场分别进行深度估计并利用遮挡融合策略以得到最终深度，可有效减小遮挡对深度估计的影响，从而提高估计精度。此外，本文还引入了新的加权空间一致性损失和角度一致性损失，增强了网络估计的鲁棒性。实验结果表明，所提出方法在合成和真实数据集都获得了良好性能，所估计的深度在对象边缘更加清晰，且在无纹理区域更加平滑。在未来工作中，将进一步关注无纹理区域中深度精度的提高以及非朗伯区域的深度估计，并将探索利用先进的对比学习策略等来提升无监督学习精度。

参考文献

[1] Rabia S, Allain G, Tremblay R, et al. Orthoscopic elemental image synthesis for 3D light field display using lens design software and real-world captured neural radiance field[J]. *Opt Express*, 2024, **32**(5): 7800−7815.

[2] Charatan D, Li S L, Tagliasacchi A, et al. pixelSplat: 3D gaussian splats from image pairs for scalable generalizable 3d reconstruction[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Piscataway, 2024: 19457–19467.

[3] Li Y L, Chen Y Y, Cui Y L, et al. LF-UMTI: unsupervised multi-exposure light field image fusion based on multi-scale spatial-angular interaction[J]. *Opto-Electron Eng*, 2024, **51**(6): 240093.
李玉龙, 陈晔曜, 崔跃利, 等. LF-UMTI: 基于多尺度空角交互的无监督多曝光光场图像融合[J]. 光电工程, 2024, **51**(6): 240093.

[4] Lv T Q, Wu Y C, Zhao X L. Light field image super-resolution network based on angular difference enhancement[J]. *Opto-Electron Eng*, 2023, **50**(2): 220185.
吕天琪, 武迎春, 赵贤凌. 角度差异强化的光场图像超分网络[J]. 光电工程, 2023, **50**(2): 220185.

[5] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015: 1547–1555. https://doi.org/10.1109/CVPR.2015.7298762.

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

[7] Zhang S, Sheng H, Li C, et al. Robust depth estimation for light field via spinning parallelogram operator[J]. *Comput Vis Image Underst*, 2016, **145**: 148−159.

[8] Han K, Xiang W, Wang E, et al. A novel occlusion-aware vote cost for light field depth estimation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2022, **44**(11): 8022−8035.

[9] Tsai Y J, Liu Y L, Ouhyoung M, et al. Attention-based view selection networks for light-field disparity estimation[C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, 2020: 12095–12103. https://doi.org/10.1609/aaai.v34i07.6888.

[10] Wang Y Q, Wang L G, Liang Z Y, et al. Occlusion-aware cost constructor for light field depth estimation[C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 2022: 19777–19786. https://doi.org/10.1109/CVPR52688.2022.01919.

[11] Chao W T, Wang X C, Wang Y Q, et al. Learning sub-pixel disparity distribution for light field depth estimation[J]. *IEEE Trans Comput Imaging*, 2023, **9**: 1126−1138.

[12] Srinivasan P P, Wang T Z, Sreelal A, et al. Learning to synthesize a 4D RGBD light field from a single image[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, 2017: 2262–2270. https://doi.org/10.1109/ICCV.2017.246.

[13] Peng J Y, Xiong Z W, Wang Y C, et al. Zero-shot depth estimation from light field using a convolutional neural network[J]. *IEEE Trans Comput Imaging*, 2020, **6**: 682−696.

[14] Zhou W H, Zhou E C, Liu G M, et al. Unsupervised monocular depth estimation from light field image[J]. *IEEE Trans Image Process*, 2020, **29**: 1606−1617.

[15] Jin J, Hou J H. Occlusion-aware unsupervised learning of depth from 4-D light fields[J]. *IEEE Trans Image Process*, 2022, **31**: 2216−2228.

[16] Zhang S S, Meng N, Lam E Y. Unsupervised light field depth estimation via multi-view feature matching with occlusion prediction[J]. *IEEE Trans Circuits Syst Video Technol*, 2024, **34**(4): 2261−2273.

[17] Godard C, Aodha O M, Brostow G J. Unsupervised monocular depth estimation with left-right consistency[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017: 6602–6611. https://doi.org/10.1109/CVPR.2017.699.

[18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016: 770–778. https://doi.org/10.1109/CVPR.2016.90.

[19] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light

fields[C]//*Proceedings of the 13th Asian Conference on Computer Vision*, Taipei, China, 2016: 19–34. https://doi.org/10.1007/978-3-319-54187-7_2.

[20] Shi J L, Jiang X R, Guillemot C. A framework for learning depth from a flexible subset of dense and sparse light field views[J]. *IEEE Trans Image Process*, 2019, **28**(12): 5867−5880.

[21] Blender website[EB/OL]. [2024-09-01]. https://www.blender.org/.

[22] Rerabek M, Ebrahimi T. New light field image dataset[C]//*Proceedings of the 8th International Conference on Quality of Multimedia Experience*, Lisbon, 2016: 1–2.

[23] Raj A S, Lowney M, Shah R, et al. Stanford lytro light field archive[EB/OL]. [2024-07]. http://lightfields.stanford.edu/LF2016.html.

作者简介

李豪宇 (2000-)，男，浙江温州人，硕士研究生，2022 年于宁波大学获得学士学位，现为宁波大学信息科学与工程学院硕士研究生，主要从事光场图像深度估计等方面的研究。

E-mail：1430096977@qq.com

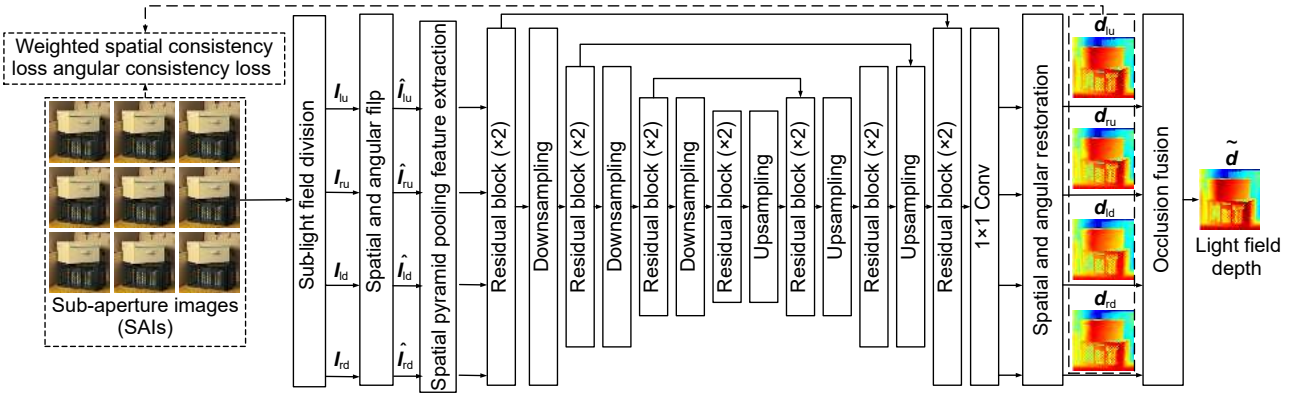
【通信作者】郁梅 (1968-)，女，江苏无锡人，博士，教授，博士生导师，2000 年于韩国 Ajou 大学 (亚洲大学) 获得博士学位，主要从事多媒体信号处理与通信、计算成像、视觉感知与编码、图像与视频质量评价等方面的研究。

E-mail：yumei@nbu.edu.cn

扫描二维码，获取PDF全文

Unsupervised light field depth estimation based on sub-light field occlusion fusion

Li Haoyu[1], Chen Yeyao[1], Jiang Zhidi[2], Jiang Gangyi[1], Yu Mei[1*]



Structure of the proposed unsupervised light field depth estimation network

Overview: Light is an important medium for humans to observe and perceive the real world, while traditional imaging approaches only record limited light information. Light field imaging can simultaneously acquire the intensity and direction information of light rays, thereby enabling a more accurate perception of complex dynamic environments. Currently, it has been applied to many visual tasks such as 3D scene reconstruction, digital refocusing, view synthesis, and occlusion removal. It is regarded as one of the main technologies for immersive media. Light field depth estimation is an important scientific problem of light field processing and applications. In recent years, deep learning has shown strong nonlinear fitting capabilities and achieved favorable results in light field depth estimation, but the generalization capability of supervised methods in real-world scenes is limited. Besides, the existing studies ignore the geometric occlusion relationship among views in the light field. By analyzing the occlusion issue among different views, an unsupervised light field depth estimation method based on sub-light field occlusion fusion is proposed. Firstly, an effective sub-light field division mechanism is employed to consider the depth relationship at different angular positions. Specifically, the view on the primary and secondary diagonals of the light field sub-aperture array are divided into four sub-light fields, i.e., top-left, top-right, bottom-left, and bottom-right. Secondly, a spatial pyramid pooling is leveraged for feature extraction to capture multi-scale context information, along with a U-Net network to estimate the depths of the sub-light fields. Finally, an occlusion fusion strategy is designed to fuse all sub-light field depths to obtain the final depth, which assigns greater weights to the sub-light field depth map with higher accuracy in the occlusion region, so as to reduce the occlusion effect. In addition, a weighted spatial and an angular consistency loss are used to constrain network training and enhance robustness. Extensive experimental results on the benchmark datasets show that the proposed method outperforms the existing methods in both quantitative and qualitative comparison. In particular, the proposed method exhibits favorable performance on real-world datasets established with light field cameras. Moreover, detailed ablation studies validate the effectiveness of sub-light field division, occlusion fusion, and loss functions involved in the proposed method.

Li H Y, Chen Y Y, Jiang Z D, et al. Unsupervised light field depth estimation based on sub-light field occlusion fusion[J]. *Opto-Electron Eng*, 2024, **51**(10): 240166; DOI: 10.12086/oee.2024.240166

Foundation item: Project supported by National Natural Science Foundation of China (62271276, 62071266), and Natural Science Foundation of Zhejiang Province (LQ24F010002)

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China; [2]College Science & Technology, Ningbo University, Ningbo, Zhejiang 315300, China

* E-mail: yumei@nbu.edu.cn