基于混合多模注意力融合矫正的深度图超分辨率重建

侯兵[1)], 董秀成[1,2)*], 杨陈成[1)], 雍萧[1)], 睢雅玲[1)]

[1)] (西华大学电气与电子信息学院 成都 610039)
[2)] (四川大学锦江学院 眉山 620860)
(dxc@mail.xhu.edu.cn)

摘 要: 针对低成本深度相机捕获的深度图中通常存在模糊和分辨率低等多种退化问题, 联合同场景高分辨率彩色图像, 提出一种混合多模注意力融合矫正网络. 为了有效地利用彩色图像的结构信息, 使用注意力机制迭代对齐及融合深度特征和彩色结构特征, 并通过彩色结构特征和感知边缘图对融合特征进行矫正. 首先在深度边缘感知模块, 联合低分辨率深度图和彩色边缘图提取高分辨率深度边缘图; 然后使用侧窗边缘保留模块提取引导彩色图像中有用的结构信息, 减少多余纹理信息的影响; 最后在多模注意力融合矫正模块中, 利用注意力机制提取深度特征和彩色结构特征的结构注意力分布, 使 2 种模态特征实现对齐及融合, 并利用彩色结构特征和高分辨率深度边缘图的空间注意力权重增强对深度边缘的监督, 减少模糊和伪影问题. 在 NYU v2 数据集和 Middlebury 数据集(2005)上的实验结果表明, 所提方法均方根误差平均测试结果分别为 2.96 和 1.45, 相对于次优方法分别降低了 0.10 和 0.08; 该网络能重建出边缘更加清晰和伪影更少的高分辨率深度图.

关键词: 深度图超分辨率; 注意力机制; 侧窗卷积; 可变卷积
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024-00105

Depth Map Super-Resolution Reconstruction Based on Hybrid Multimodal Attention Fusion Correction

Hou Bing[1)], Dong Xiucheng[1,2)*], Yang Chencheng[1)], Yong Xiao[1)], and Ju Yaling[1)]

[1)] (*School of Electrical Engineering and Electronic Information, Xihua University, Chengdu* 610039)
[2)] (*College of Jinjiang, Sichuan University, Meishan* 620860)

Abstract: To solve the problems of blurring and low resolution in depth maps captured by low-cost depth cameras, a hybrid multimode attention fusion correction network is proposed by combining high-resolution color images in the same scene. To effectively leverage the structure information of color images, the attention mechanism is employed to iteratively align and integrate depth features with color structural features, and the fused features are corrected through the incorporation of color structure features and edge aware maps. Initially, the deep edge aware block combines low-resolution depth maps with color edge maps to extract high-resolution depth edge maps. Then, the side-window edge preserve block is utilized to extract useful structural information from the color image, reducing the impact of extraneous texture information. Finally, multimodal attention fusion correction block utilizes attention mechanisms to extract the structural attention distribution of both depth features and color structural features, aligning and fusing these modalities.

收稿日期: 2024-02-28; 修回日期: 2024-06-20. **基金项目**: 国家自然科学基金(11872069); 四川省中央引导地方科技发展专项(2021ZYD0034); 四威高科-西华大学产学研联合实验室项目(2016-YF04-00044-JH). **侯兵**(2000—), 男, 硕士研究生, 主要研究方向为计算机视觉、深度学习; **董秀成**(1963—), 男, 硕士, 教授, 硕士生导师, 论文通信作者, 主要研究方向为智能信息处理、计算机视觉、智能控制; **杨陈成**(1998—), 女, 硕士研究生, 主要研究方向为图像处理、深度学习; **雍萧**(2001—), 女, 硕士研究生, 主要研究方向为图像处理、深度学习; **睢雅玲**(2000—), 女, 硕士研究生, 主要研究方向为图像处理、深度学习.

The spatial attention weights of the color structural features and high-resolution depth edge maps are used to enhance the supervision of the depth edges, reducing blurring and artifacts. Experimental results on the NYU v2 dataset and Middlebury dataset (2005) show that the average root mean square error test results are 2.96 and 1.45, respectively, which are reduced by 0.10 and 0.08 compared to suboptimal methods. The network can reconstruct high-resolution depth maps with clearer edges and fewer artifacts.

Key words: depth map super-resolution; attention mechanism; side-window convolution; deformable convolution

深度图包含当前场景的空间距离信息，广泛应用于计算机视觉领域中需要高质量深度信息的场景识别、自动驾驶和三维重建等任务. 然而，由于消费级深度相机的性能限制和场景光影干扰，直接拍摄的深度图的分辨率较低并且带有明显的噪声退化. 为了解决这个问题，目前已有大量的深度图超分辨率(depth map super-resolution, DSR)方法，它们可以分为基于滤波的方法、基于优化的方法和基于学习的方法 3 类.

基于滤波的方法利用局部信息和全局信息的先验假设手工设计滤波核实现 DSR 重建. He 等[1]以引导图像的信息作为参考，基于局部线性模型恢复目标图像; Lei 等[2]综合考虑图像的平滑性和纹理相似性，提出一种基于置信度的多视图合成上采样滤波器. 虽然基于滤波的方法设计难度较低，但难以适应结构复杂的深度图，重建时通常因语义信息变化和缺失导致恢复效果较差.

基于优化的方法将 DSR 转化为一个优化约束问题，采用多种先验假设迭代优化，缩小输出结果与目标结果之间的差异. Ferstl 等[3]将 DSR 视为凸优化问题，使用引导图像的各向异性扩散张量引导深度图上采样; Liu 等[4]基于 k 奇异值分解 (k singular value decomposition, kSVD)抑制深度图的锯齿状边缘，并使用改进的联合双边滤波器对深度图进行插值，在保持边缘的锐利度的同时抑制了噪声. 但是，此类方法依赖复杂的计算和特定的先验假设，如果深度图场景发生变化，重建效果将受到较大影响.

近年来，基于学习的方法在 DSR 重建中取得了更优异的成果. Hui 等[5]提出一种多尺度引导 DSR 方法，通过逐层地提取并利用引导纹理特征消除模糊; Li 等[6]设计一种相关控制彩色引导模块，并基于高层对称结构分析 2 个子网络之间的相关性和传递信息，有效地恢复了深度细节; Zhong 等[7]通过多模态注意力机制提取并融合深度特征和彩色特征，还使用双向层次特征协作模块，充分地利用多尺度特征之间的低层空间信息和高层空间信息，解决了结构不一致难以融合的问题. 然而，这些 DSR 方法存在 2 个问题: (1) 从引导图像中提取的结构特征不准确且带有大量噪声，导致多模态特征融合困难; (2) 通过简单的叠加或注意力融合多模特征难以实现特征的准确对齐，得到的融合图像存在伪影和模糊.

为了解决上述问题，本文提出一种混合多模注意力融合矫正网络(hybrid multimodal attention fusion correction network, HMAFCN)，以迭代的方式重建深度图. 本文联合彩色边缘图和高分辨率(high-resolution, HR)插值深度图，设计一种深度边缘感知模块(deep edge aware block, DEAB)，其中，彩色边缘图为深度图定位深度边缘，深度图能够帮助彩色边缘图像消除多余结构信息，更精准地提取深度边缘; 针对彩色图像中多余纹理噪声影响重建效果的问题，提出一种基于侧窗卷积的侧窗边缘保留模块(side-window edge preserve block, SEPB)，能够自适应地提取并保留多种边缘形状的彩色结构信息来匹配深度特征，以提升融合效率和重建效果; 综合 3 种模态特征的特性，提出一种多模注意力融合矫正模块(multimodal attention fusion correction block, MAFCB)，其中，针对不同特征使用不同的注意力处理方式，有效地实现了特征对齐融合，并且在融合后，使用可变卷积和 HR 边缘图注意力权重监督融合特征中的高频信息，减少边缘模糊和伪影.

1 相关工作

1.1 单 DSR 重建

基于学习的 DSR 利用神经网络学习低分辨率(low-resolution, LR)深度图到 HR 深度图之间的非线性映射关系，直接从 LR 深度图重建缺失的结构细节. 得益于卷积神经网络、残差网络和多尺度网络的不断进步，基于学习的单 DSR 重建取得了远优于传统方法的性能. 范佩佩等[8]利用图像的自相似性对亚像素卷积网络上采样得到的深度图施加

非局部均值约束，重建出细节清晰的深度图；Chen 等[9]结合深度学习和全变分模型提升了重建深度图的边缘锐利度；Song 等[10]结合多尺度融合及深监督网络结构将 DSR 分解为多个子任务，有效地降低了边缘模糊和伪影；李滔等[11]结合多尺度网络和注意力机制自适应选择多尺度特征，并对网络进行深监督约束，有效地提升了重建性能．虽然上述方法提升了重建效果，但深度结构细节的缺失导致重建深度图的边缘模糊．

1.2　引导 DSR 重建

通常，彩色图像及其相关的深度图是同一场景的光度和几何表示，二者具有很强的结构相似性．通过滤波和聚合等方式获得 2 种图像之间的非线性映射关系，可以有效地提升深度重建的效果．现有方法大多采用多尺度提取、彩色图像边缘引导和注意力融合等方法设计网络结构．

多尺度提取可以建立彩色图像与深度图多层级特征之间的内在联系，有效地重建退化的深度图．Zuo 等[12-14]提出一种基于多尺度结构的网络，并进行了一系列改进．首先提出一种多尺度残差网络，由粗到细地逐步整合多尺度特征[12]；为了减少深度图和彩色图像之间的分布差异造成的伪影和边缘模糊，提出一种深度引导的仿射变换，用于过滤和细化特征[13]；还基于全局和局部残差策略，通过多尺度强度图像引导将深度图从粗到细地重建[14]．Zhong 等[15]通过构建可学习滤波核，从不同尺度的深度图和彩色图像中自适应地选择有用信息，取得了更好的重建性能．虽然多尺度结构取得了良好的性能，但在不同尺度特征的提取过程中会出现信息错位的情况，导致难以建模．

深度图由低频纹理和高频边缘组成，高频边缘决定了深度信息的准确度．Ye 等[16]利用深度卷积神经网络构建一个深度边缘推理分支，并提出了一种深度填充方法来推理深度值；Wang 等[17]采用与文献[16]类似的结构，通过边缘感知卷积网络预测深度边缘，然后利用深度填充和级联恢复网络重建深度图；Zhao 等[18]提出离散余弦变换重建网络，利用边缘引导注意力模块提取彩色图像边缘特征；Yuan 等[19]使用多个不同尺寸的卷积核提取不同平滑度的引导图像特征，并在相邻平滑度特征之间进行逐元素减法，以获得高频信息．然而，上述方法大多通过简单的卷积提取单模态边缘，难以联合利用 2 种模态的互补信息，提取的高频信息并不能准确地补充缺失的深度边缘．

彩色图像包含大量的结构特征和丰富的低频细节，与深度图存在较大的模态差异．有效的特征融合模块能够准确地匹配有用的特征信息，有利于减少重建边缘模糊和伪影．Li 等[20]采用空间注意力(spatial attention, SA)机制和通道注意力(channel attention, CA)机制，从并行多尺度特征提取分支中提取不同尺度的特征进行融合；Ye 等[21]将通道叠加的 2 个模态特征输入融合模块，利用 CA 机制自适应地选择重要特征；Zhang 等[22]通过串联分支注意模块和边缘感知空间注意模块，将有用的引导信息和边缘特征融合为深度特征；Ariav 等[23]通过交叉注意 Transformer 建立彩色图像和深度图的全局联系，有效地提高了图像重建能力．然而，上述方法仅考虑了深度特征和彩色特征在非关联注意力机制下的联系，难以有效地对齐多模特征．

2　HMAFCN

2.1　网络总体结构

彩色图像引导的 DSR 网络的目的在于探寻 HR 彩色图像 I_{HR} 和 LR 深度图 D_{LR} 之间的非线性映射，引导 DSR 重建．HMAFCN 的结构如图 1 所示，由深度边缘提取分支、彩色图像特征提取分支和多模特征融合重建分支 3 部分组成；其中，卷积层框内 (k,c) 分别表示卷积核尺寸和输出通道数．在 DSR 重建时，在深度边缘提取分支中，利用一个简化的 U-Net 从彩色边缘图和插值后的深度图中保留深度细节和彩色结构，得到准确的 HR 深度边缘图；彩色图像特征提取分支和多模特征融合重建分支迭代地提取各级彩色结构特征和融合多模特征，其中，彩色图像特征提取分支利用多级 SEPB 为深度图补充有用的结构信息；多模特征融合重建分支则在级联融合模块(cascade fusion block, CFB)中使多模特征的空间分辨率匹配后，在 MAFCB 中根据彩色结构信息的特性与深度图进行注意力融合和矫正，并利用感知边缘图的边缘注意力特征进一步监督高频边缘．HMAFCN 通过重建模块(reconstruction block, RB)将融合矫正后的深度特征上采样至目标分辨率．网络加入全局残差结构，有利于使重建网络专注于高频残差信息，防止梯度消失，能够更有效地重建深度图．在全局残差中，将 D_{LR} 经过双三次插值至目标分辨率与上采样后的重建特征进行残差连接，得到 HR 深度图 D_{SR}，网络处理过程可表示为

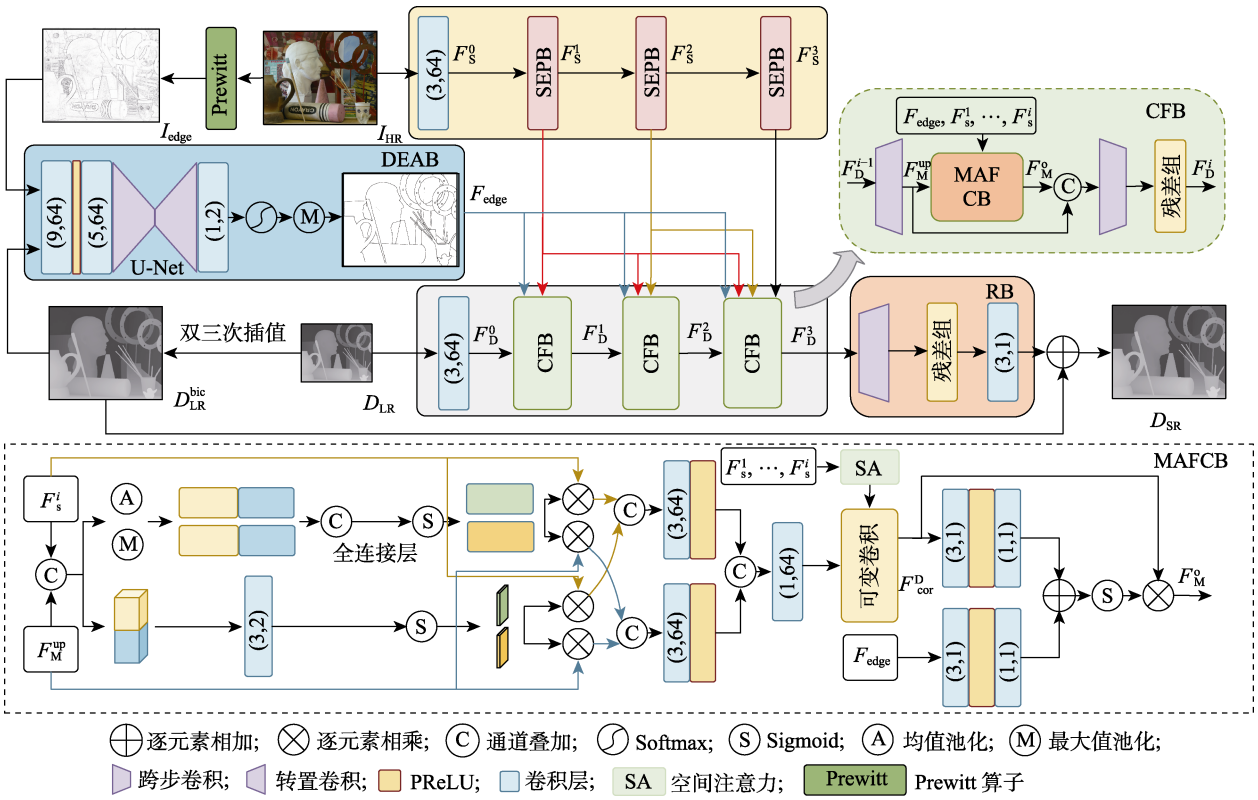


图 1 HMAFCN 结构

$$D_{SR} = f_{HMAFCN}^{p}\left(D_{LR}, I_{HR}; \theta\right) + D_{LR}^{bic}.$$

其中，f_{HMAFCN}^{p} 表示上采样倍率为 p 的 HMAFCN，$p = 2^{u}, u \in (2,3,4)$；θ 表示 HMAFCN 的参数；$D_{LR}^{bic} = f_{bic}\left(D_{LR}\right)$，$f_{bic}$ 表示双三次插值函数.

深度边缘提取分支的输入为 D_{LR}^{bic} 和 I_{edge}，通过多尺度处理和真实深度边缘图的监督得到感知边缘图 $F_{edge} = f_{DEAB}\left(D_{LR}^{bic}, I_{edge}\right)$；其中，$f_{DEAB}$ 表示 DEAB 操作，I_{edge} 由 Prewitt 算子提取.

彩色图像含有丰富的纹理信息和结构信息，利用高频结构特征能帮助深度图定位深度边缘. 为了提取并保留有用的彩色边缘特征，彩色图像特征提取分支中首先使用 1 层卷积提取浅层特征；然后迭代地使用 n 个 SEPB，该模块中用多种形状的侧窗卷积，通过密集链接的方式传递不同形状之间的特征，能够自适应地匹配各种图形边缘并有效地保留有用高频信息. 该分支将 I_{HR} 作为输入，通过 1 层卷积得到 F_{S}^{0}，然后通过各级 SEPB 得到彩色结构特征 $F_{S}^{i}, i \in [1,n]$，过程可表示为

$$F_{S}^{0} = f_{3}\left(I_{HR}\right),\quad F_{S}^{i} = f_{SEPB}^{i}\left(F_{S}^{i-1}\right).$$

其中，$f_{k}, k \in (1,3,5,9)$ 表示 $k \times k$ 卷积；f_{SEPB}^{i} 表示第 i 个 SEPB.

多模特征融合重建分支中，利用多种注意力将已经得到的边缘特征和各级彩色结构特征与各级输入深度特征实现特征对齐，有效地提升了边缘清晰度并消除了伪影. 该分支首先通过 1 层卷积提取浅层特征得到 $F_{D}^{0} = f_{3}\left(D_{LR}\right)$；然后将 F_{edge} 和 $\left(F_{S}^{1}, F_{S}^{2}, \cdots, F_{S}^{i}\right)$ 输入 n 级 CFB 中，得到各级级联输出 F_{D}^{i}. CFB 的结构如图 1 所示. 由于不同模态特征的空间分辨率不同，首先在 CFB 中使用转置卷积将 F_{D}^{i-1} 上采样 p 倍得到 F_{M}^{up}；然后将多模特征共同输入到 MAFCB 中进行特征对齐和矫正，得到 F_{M}^{o}，为了复用特征，将融合矫正特征 F_{M}^{o} 与 F_{M}^{up} 叠加后输入跨步卷积，下采样至输入空间分辨率；最后通过残差组(residual group, RG)[24]进行特征增强，每个 RG 包含 m 个残差通道注意力块(residual channel attention block, RCAB)[24]. 上述过程可表示为

$$F_{M}^{o} = f_{MAFCB}^{i}\left(\uparrow\left(F_{D}^{i-1}\right), F_{S}^{1}, \cdots, F_{S}^{i}, F_{edge}\right),$$

$$F_{D}^{i} = f_{RG}^{i}\left(\downarrow\left[F_{M}^{up}, F_{M}^{o}\right]\right).$$

其中，f_{MAFCB}^{i} 和 f_{RG}^{i} 分别表示第 i 个 MAFCB 和 RG 函数；$\uparrow$ 和 $\downarrow$ 分别表示转置卷积和跨步卷积操作；$[,]$ 表示通道叠加操作.

F_{D}^{n} 经注意力融合矫正后，通过 RB 实现采样

倍数为 p 的超分辨率重建. 为了在上采样后进一步增强特征和减少噪声，在 RB 中使用转置卷积上采样，RG 增强特征和通道调整卷积输出重建深度图. 最后，$D_{\mathrm{LR}}^{\mathrm{bic}}$ 通过全局残差与重建输出进行逐元素相加，使网络能专注于重建高频信息，网络重建过程可表示为 $D_{\mathrm{SR}}=f_3\left(f_{\mathrm{RG}}\left(\uparrow F_{\mathrm{D}}^{n}\right)\right)+D_{\mathrm{LR}}^{\mathrm{bic}}$.

2.2　DEAB

深度图由界定物体边界的高频边缘和物体表面的低频信息组成. 通常，现有的方法通过假定相同场景彩色图像和深度图的光度和结构信息之间有很强的相关性，并从彩色图像中转移有用纹理结构信息来辅助重建深度图[17]. 然而，彩色图像的提取特征中包含大量无用的纹理信息，将所有信息传递到深度图会引入伪影. 因此，本文提出 DEAB，从 $D_{\mathrm{LR}}^{\mathrm{bic}}$ 和 I_{HR} 中提取感知边缘图 F_{edge}. 由于彩色图像中的低频平滑区域影响对深度边缘的推断准确度，并且浪费计算资源，因此使用 Prewitt 算子预先处理 I_{HR} 得到 I_{edge}；然后使用 9×9 和 5×5 的卷积和 PReLU 激活函数提取彩色边缘图和深度图的浅层特征. 为了能有效地使用彩色边缘特征，利用 U-Net 多层级特征提取的特点，将叠加的 2 种特征作为输入，得到浅层深度边缘特征

$$F'_{\mathrm{edge}}=W_{\mathrm{U}}*\left(\left[f_5\left(\sigma\left(f_9\left(I_{\mathrm{edge}}\right)\right)\right),f_5\left(\sigma\left(f_9\left(D_{\mathrm{LR}}^{\mathrm{bic}}\right)\right)\right)\right]\right).$$

其中，W_{U} 表示 U-Net 的权重参数；σ 表示 PReLU 激活函数；$*$ 表示卷积操作.

由于深度图中高频信息的占比极小，在训练时会造成分类不平衡，导致边缘恢复不能得到有效监督，因此本文将深度图中的边缘和非边缘预测视为一个分类问题[17]，0 值代表非边缘，1 值代表边缘. 为了减少直接生成边缘图的计算成本，使用 1×1 的卷积和 Softmax 函数进行分类，得到分类特征 $X^j, j\in(0,1)$. 然而，直接得到的边缘图通常不能覆盖全部的边缘，导致边缘图并不能起到有效的监督作用. 因此，本文使用最大值池化函数处理 X^j，以提高边缘覆盖范围. 最终，DEAB 取 $j=0$ 作为深度感知边缘特征，即 $F_{\mathrm{edge}}=f_{\mathrm{mpool}}\left(X^0\right)$；其中，$f_{\mathrm{mpool}}$ 表示最大值池化操作.

2.3　SEPB

侧窗滤波[25]由多种形状不同的边缘卷积核组成，边缘卷积核将权重参数放在滤波核的边缘，能准确地识别并保留高频信息. 为了增加对各种图像特征边缘的自适应提取能力，本文选取 8 种形状不同的滤波核，并按照形状分为 2 组，其中，U(up)，D(down)，L(left)和 R(right)为一组；NW(northwest)，NE(northeast)，SW(southwest)和 SE(southeast)为一组. 图 2 所示为 SEPB 的结构，为了使卷积核能自适应地组合以匹配不同形状的边缘，2 组卷积核在组间分离 3/4 通道的输出特征传递给后续的每一层进行叠加，并将剩余的 1/4 通道的输出特征传递给同一层的另一组特征进行通道叠加，最后将每种形状的侧窗卷积的输出特征取 1/4 叠加，组合所有形状的边缘特征. 非完全通道的特征传递既可以与其他形状特征建立联系，又避免大幅影响其他侧窗卷积核提取特征，能够自适应地使用各种特征信息，以提取更加准确的边缘特征.

给定输入 $F_{\mathrm{S}}^{i-1}\in\mathbb{R}^{C\times H\times W}$，其中，$C$ 表示通道数量，$H\times W$ 表示空间尺度. 首先使用 2 层卷积和激活函数初步处理彩色图像特征中的噪声，得到侧窗卷积组的输入 $F_{\mathrm{S}}^{\mathrm{in}}=f_3\left(\sigma\left(f_3\left(F_{\mathrm{S}}^{i-1}\right)\right)\right)$. 为了避免较大的输入特征尺寸影响所提取结构信息的准确性，SEPB 对特征 $F_{\mathrm{S}}^{\mathrm{in}}$ 进行下采样，以增加感受野；并在侧窗滤波完成后上采样，以保持尺度不变. 生

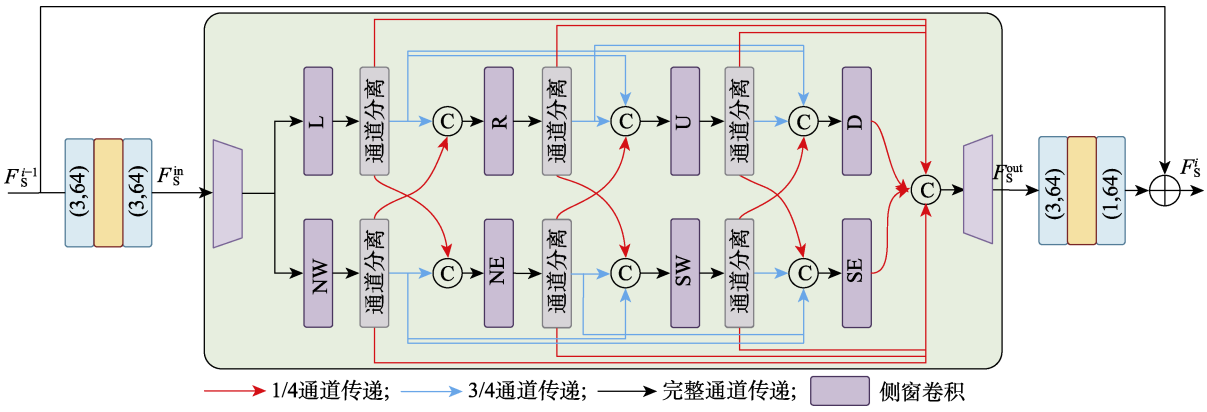


图 2　SEPB 结构

成彩色结构特征 $F_{\mathrm{S}}^{\mathrm{out}}$ 的过程为

$$F_{\mathrm{S}}^{\mathrm{out}}=\uparrow\left[F_{\mathrm{C}_{1/4}}^{\mathrm{U}},F_{\mathrm{C}_{1/4}}^{\mathrm{D}},F_{\mathrm{C}_{1/4}}^{\mathrm{L}},F_{\mathrm{C}_{1/4}}^{\mathrm{R}},F_{\mathrm{C}_{1/4}}^{\mathrm{NW}},F_{\mathrm{C}_{1/4}}^{\mathrm{NE}},F_{\mathrm{C}_{1/4}}^{\mathrm{SW}},F_{\mathrm{C}_{1/4}}^{\mathrm{SE}}\right].$$

其中，$F_{\mathrm{C}_{1/4}}^{\mathrm{U}}$，$F_{\mathrm{C}_{1/4}}^{\mathrm{D}}$，$F_{\mathrm{C}_{1/4}}^{\mathrm{L}}$，$F_{\mathrm{C}_{1/4}}^{\mathrm{R}}$，$F_{\mathrm{C}_{1/4}}^{\mathrm{NW}}$，$F_{\mathrm{C}_{1/4}}^{\mathrm{NE}}$，$F_{\mathrm{C}_{1/4}}^{\mathrm{SW}}$和 $F_{\mathrm{C}_{1/4}}^{\mathrm{SE}}$分别表示形状为 U, D, L, R, NW, NE, SW 和 SE 的侧窗卷积的 $1/4$ 通道输出特征.

为了调整结构特征的通道相关性，在模块尾部使用 3×3 和 1×1 的卷积和激活函数. 同时，为了避免梯度消失，并让模块注重学习结构特征的提取，在 SEPB 中加入全局残差. 最后, SEPB 的输出为 $F_{\mathrm{S}}^{i}=f_1\left(\sigma\left(f_3\left(F_{\mathrm{S}}^{\mathrm{out}}\right)\right)\right)+F_{\mathrm{S}}^{i-1}$.

2.4 MAFCB

现有的引导 DSR 重建方法中，通常将深度图特征和彩色特征通过简单的叠加、相加或注意力处理进行融合，但这些简单的处理不能有效地界定深度边缘和消除多余的彩色纹理，导致边缘模糊并引入伪影. MAFCB 的结构如图 1 所示，通过混合深度特征、彩色结构特征和引导边缘图的注意力权重，融合深度特征和彩色结构并矫正多余纹理，实现准确的特征对齐及融合. 融合过程利用深度特征和彩色结构特征的 CA 权重和 SA 权重，实现 2 种模态的特征对齐与融合；矫正过程通过各级彩色结构特征作为指导，结合可变卷积对融合特征中的边缘模糊和特征错位进行矫正，然后利用引导边缘图的空间边缘感知注意力权重作为监督，进一步去除噪声.

MAFCB 将已经得到的各级彩色结构特征 F_{S}^{i} 和上采样过的深度特征 $F_{\mathrm{M}}^{\mathrm{up}}$ 进行叠加，其中，$F_{\mathrm{M}}^{\mathrm{up}}=\uparrow\left(F_{\mathrm{D}}^{i-1}\right)$. 叠加后的特征经过 CA 分支和 SA 分支实现特征对齐. CA 分支使用均值池化和最大值池化提取通道维度中的显著特征，然后将 2 组特征叠加并通过全连接层充分混合，得到 CA 对齐特征 $F_{\mathrm{CA}}\in\mathbb{R}^{4C\times1\times1}$. SA 分支则使用卷积将叠加特征沿通道维度聚合，输出 SA 对齐特征 $F_{\mathrm{SA}}\in\mathbb{R}^{2\times H\times W}$. MAFCB 使用 Sigmoid 函数处理对齐注意力特征，得到 CA 特征和 SA 特征的对齐注意力权重分别为

$$W_{\mathrm{CA}}^{g}=\mathrm{Sigmoid}\left(F_{\mathrm{CA}}\right),\quad W_{\mathrm{SA}}^{h}=\mathrm{Sigmoid}\left(F_{\mathrm{SA}}\right).$$

其中，$W_{\mathrm{CA}}^{g},g\in(0,1)$；$W_{\mathrm{SA}}^{h},h\in(0,1)$.

注意力权重中包含深度特征和彩色结构特征的特征分布，能界定特征结构并实现对齐. 为了避免不同模态之间的干扰，深度特征和彩色结构特征首先分组进行对应模态特征的提取和融合；然后叠加 2 组特征并使用 1×1 卷积进行融合，即

$$F_{\mathrm{fuse}}^{\mathrm{D}}=\sigma\left(f_3\left[W_{\mathrm{CA}}^{0}\times F_{\mathrm{S}}^{i},W_{\mathrm{SA}}^{0}\times F_{\mathrm{S}}^{i}\right]\right),$$

$$F_{\mathrm{fuse}}^{\mathrm{I}}=\sigma\left(f_3\left[W_{\mathrm{CA}}^{1}\times F_{\mathrm{M}}^{\mathrm{up}},W_{\mathrm{SA}}^{1}\times F_{\mathrm{M}}^{\mathrm{up}}\right]\right),$$

$$F_{\mathrm{M,fuse}}^{\mathrm{D}}=f_1\left(\left[F_{\mathrm{fuse}}^{\mathrm{D}},F_{\mathrm{fuse}}^{\mathrm{I}}\right]\right).$$

其中，$F_{\mathrm{fuse}}^{\mathrm{D}}$ 和 $F_{\mathrm{fuse}}^{\mathrm{I}}$ 分别表示深度对齐特征和彩色对齐特征；$F_{\mathrm{M,fuse}}^{\mathrm{D}}$ 表示初步融合深度特征.

融合后的特征仍存在少量不确定的伪影和噪声. 可变卷积[26]可以通过计算引导图像的偏移信息和掩码信息，自适应地提取各种形状的结构信息，能够在不引入噪声的情况下增加融合特征的边缘清晰度. 为了在不引入噪声的情况下矫正干扰并充分地利用各层彩色结构特征的不同结构信息，首先使用 SA 机制提取 $(1,2,\cdots,i)$ 级彩色结构特征的 SA 特征作为偏置和掩码权重，然后结合可变卷积[26]处理得到初步矫正输出特征

$$F_{\mathrm{cor}}^{\mathrm{D}}=f_{\mathrm{defconv}}\left(f_{\mathrm{SA}}\left(F_{\mathrm{S}}^{1},F_{\mathrm{S}}^{2},\cdots,F_{\mathrm{S}}^{i}\right),F_{\mathrm{M,fuse}}^{\mathrm{D}}\right).$$

其中，f_{defconv} 和 f_{SA} 分别表示可变卷积操作和 SA 权重提取操作.

DEAB 输出的边缘引导图是内容为 0 和 1 的二值图，可以监督 $F_{\mathrm{cor}}^{\mathrm{D}}$ 的高频信息. 为了避免边缘引导图滤除深度低频纹理，在参考边缘信息时使用 2 组卷积和激活函数处理 2 组特征，使得 $F_{\mathrm{cor}}^{\mathrm{D}}$ 能自适应地选择是否参考边缘引导特征. 为了区分 2 组特征的重要程度，卷积处理后，将 2 组特征相加并通过 Sigmoid 函数得到自适应选择权重

$$W_{\mathrm{SEL}}=\mathrm{Sigmoid}\left(f_1\left(\sigma\left(f_3\left(F_{\mathrm{edge}}\right)\right)\right)+f_1\left(\sigma\left(f_3\left(F_{\mathrm{cor}}^{\mathrm{D}}\right)\right)\right)\right).$$

其中，1×1 卷积层的初始化权重为 1. 最后，将 $F_{\mathrm{cor}}^{\mathrm{D}}$ 与 W_{SEL} 做逐元素相乘，得到 MAFCB 的完全矫正输出为 $F_{\mathrm{M}}^{\mathrm{o}}=F_{\mathrm{cor}}^{\mathrm{D}}\times W_{\mathrm{SEL}}$.

2.5 损失函数

为了提高输出深度图 D_{SR} 与真实深度图 D_{GT} 的相似度，本文采用 3 种损失函数来优化生成网络：(1) 平均绝对误差损失函数 L_1. 计算 D_{SR} 与 D_{GT} 之间的平均绝对误差，反映它们的整体差异. (2) 快速傅里叶变换(fast Fourier transform, FFT)损失函数 L_{FFT}. 捕捉 D_{SR} 和 D_{GT} 之间的高频信息差异，增强输出深度图的结构细节. (3) 边缘感知损失函数 L_{edge}. 基于交叉熵损失函数，能够平衡深度平滑区域的损失权重，增加对边缘区域的监督. 3 种损失函数的定义分别为

$$L_1=\frac{1}{N}\sum_{t=1}^{N}\left\|D_{\mathrm{SR}}^{t}-D_{\mathrm{GT}}^{t}\right\|^{1},$$

$$L_{\mathrm{FFT}}=\frac{1}{N}\sum_{t=1}^{N}\left\|\mathrm{FFT}\left(D_{\mathrm{SR}}^{t}\right)-\mathrm{FFT}\left(D_{\mathrm{GT}}^{t}\right)\right\|^{1},$$

$$L_{\mathrm{edge}}=-\frac{1}{N}\sum_{t=1}^{N}F_{\mathrm{GT,edge}}^{t}\ln\left(F_{\mathrm{edge}}^{t}\right).$$

其中，N 表示像素数量；$F_{\mathrm{GT,edge}}$ 表示由 Prewitt 算子提取的真实深度边缘图.

HMAFCN 的损失函数定义为

$$L(\theta)=L_1+\gamma_1 L_{\mathrm{FFT}}+\gamma_2 L_{\mathrm{edge}}.$$

其中，γ_1 和 γ_2 表示超参数.

3　实验结果与分析

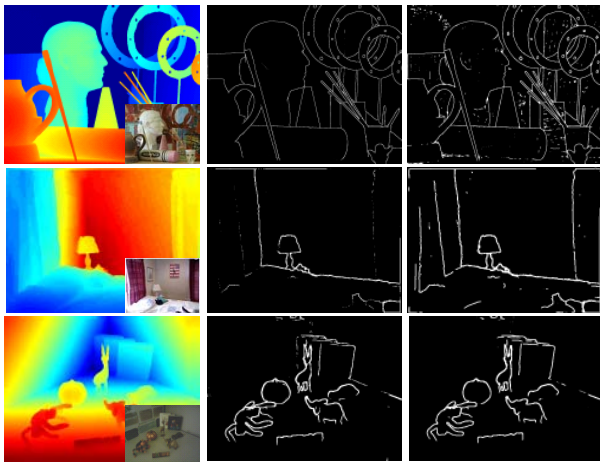
3.1　实验设置

在训练阶段，使用真实数据集和合成数据集验证 HMAFCN 的有效性. 在真实数据集中，使用来自 NYU v2 数据集[27]中的前 1 000 对 RGB-D 图像对作为训练数据集，训练图像切裁剪为 256×256 像素的图像块，剩余的 449 对 RGB-D 图像对作为测试集. 为了验证真实数据集训练的泛化性，使用 Zhao 等[18]提供的 Lu RGB-D 数据集(6 对)和 Middlebury RGB-D 数据集(30 对)进行泛化性验证. 在合成数据集中，从 Middlebury 数据集[28](34 对)和 MPI Sintel 数据集[29](58 对)中选取 82 对 RGB-D 图像对作为训练集. 测试集选用 Middlebury 数据集(2005)[30]中的 6 对 RGB-D 图像对. 为了增加数据量和减少计算，将 RGB-D 图像对裁剪为 256×256 像素的图像块，裁剪步幅为 64，该处理应用于所有采样倍数. 真实深度边缘图像由 Prewitt 算子提取，LR 深度图由真实深度图通过双三次插值下采样获得. 训练过程中，所有的图像对使用 90°旋转、180°旋转、水平翻转和垂直翻转来增强数据.

设置 HMAFCN 的迭代次数 n=3，RG 中 RCAB 的数量 m=4，$\gamma_1=10^{-2}$，$\gamma_2=2\times10^{-4}$，初始学习率为 10^{-4}，特征通道数为 64. 除非特别说明，网络中卷积核的核心尺寸均为 3×3. HMAFCN 基于 PyTorch 框架实现，在 RTX 3090 GPU 上进行训练；网络由 Adam 算法优化，优化器参数设置为 $\beta_1=0.900$，$\beta_2=0.999$. 在真实数据集上共训练 300 个周期，每 100 个周期学习率除以 5，批处理大小设为 1，训练持续约 21 h；合成数据集上共训练 80 个周期，每 30 个周期学习率除以 5，批处理大小设为 2，训练持续约 33 h；所有的训练数据均归一化至 $[0,1]$.

3.2　HMAFCN 边缘提取效果

在 Middlebury 数据集[30]，NYU v2 数据集[27]和 Lu 数据集[18]上，HMAFCN 在重建倍数为 4 时的边缘提取效果如图 3 所示. 边缘提取的准确程度取决于 2 个方面：一是边缘是否清晰锐利；二是在深度图界定参考模糊的情况下，能否准确地利用彩色图像进行边缘重建且不引入多余纹理. 可以看出，与真实边缘深度图相比，HMAFCN 输出的边缘图像准确地恢复了锐利程度相当的深度边缘；利用彩色图像边缘作为参考，该网络能够重建深度图中由于深度相近而模糊的边缘部分，如 NYU v2 数据集中的窗帘部分；针对纹理去除的性能，所选的 3 幅深度图的同场景彩色图像中均存在细节丰富的背景海报，然而在 HMAFCN 所提取的感知深度边缘图中的背景海报、书本封面以及铅笔表面都未出现额外纹理；虽然去除了多余结构信息，但图 3c 中的边缘感知图存在少量噪点，原因是 DEAB 同时提取到了插值深度图中不明显的高频信息和彩色图像中的同位置的高频纹理.



a. 真实深度图　b. 真实深度边缘图　c. 本文感知边缘图

图 3　HMAFCN 边缘提取效果

3.3　有效性实验

分别在真实数据集和合成数据集上进行有效性实验，并通过均方根误差(root mean square error, RMSE)进行评估，其值越小越好. 对比方法的结果从作者原文献中获取，如果没有该数据项，使用该方法作者提供的官方代码进行重新训练.

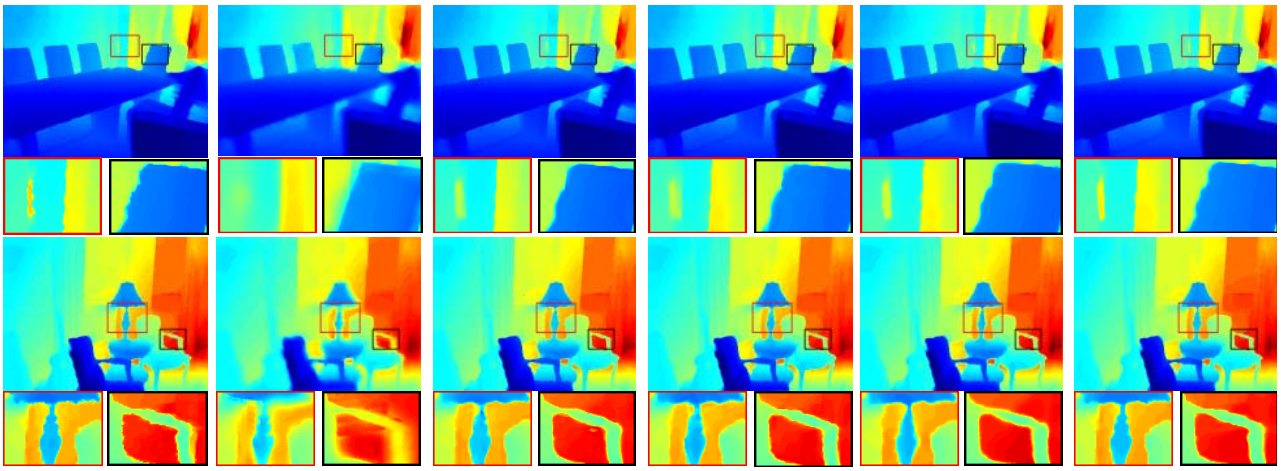
在真实数据集定量实验中，除了基准方法 Bicubic 外，还选用 GF[1]和 TGV[3]这 2 种传统的方法，以及 DKN[31]，DCTNet[18]，AHMF[7]，DAGF[15]，CONet[22]，GeoDSR[32]和 A2G[33]这 7 种基于学习的方法与 HMAFCN 进行实验. 实验中，对 NYU v2，Lu 和 Middlebury 数据集进行上采样倍数为 4，8 和 16 的超分辨率重建测试，结果如表 1 所示；其中，NYU v2 数据集的深度值以厘米为尺度测量 RMSE，其他

数据集缩放至$[0,255]$测量 RMSE(下文同). 可以看出, 由于在全局建模方面存在一些不足, 在 NYU v2 数据集的 4 倍和 8 倍上采样倍数下, HMAFCN 与 A2G[33]和 DAGF[15]结果相当, 但在信息损失更为严重的 16 倍重建实验中, 得益于 SEPB 提供的有用结构信息和感知边缘图的监督, MAFCB 能够针对不同特征的特点为深度图补充结构信息, 使得 HMAFCN 的 RMSE 分别降低 0.30 和 1.12; 在 Lu 数据集 16 倍上采样中, 由于对彩色分支的额外监督不够充分, 与 A2G[33]相比, HMAFCN 存在不足, 但与相同监督策略的 AHMF[7]相比, 该网络仍降低了 0.24. 在 NYU v2, Lu 和 Middlebury 数据集上, HMAFCN 与 GF[1], DKN[31], DCTNet[18]和 DAGF[15]的 8 倍上采样重建结果定性比较分别如图 4~图 6 所示. 从图中放大结果可以看出, GF[1]重建结果模糊, 边缘涂抹严重, DKN[31]存在伪影且边缘扭曲,

表 1　3 个数据集的 RMSE 对比

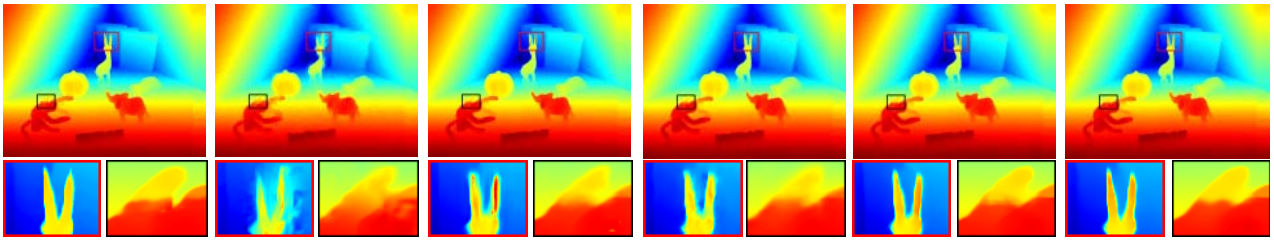
方法	NYU v2			Middlebury			Lu		
	4×	8×	16×	4×	8×	16×	4×	8×	16×
Bicubic	8.16	14.22	22.32	4.44	7.58	11.87	5.07	9.22	14.27
GF[1]	7.32	13.62	22.03	4.01	7.22	11.70	4.87	8.85	14.09
TGV[3]	6.98	11.23	28.13	3.39	5.41	12.03	4.48	7.58	17.46
DKN[31]	1.62	3.26	6.51	1.23	2.12	4.24	0.96	2.16	5.11
DCTNet[18]	1.59	3.16	5.84	1.10	2.05	4.19	0.88	1.85	4.39
AHMF[7]	1.40	2.89	5.64	1.07	1.63	3.14	0.88	1.66	3.71
DAGF[15]	1.36	2.87	6.06	1.15	1.80	3.70	0.83	1.93	4.80
CONet[22]	1.38	2.71	5.60						
GeoDSR[32]	1.48	2.73	5.10	1.04	1.73	3.19	0.82	1.62	4.11
A2G[33]	**1.31**	2.64	5.24	1.02	1.59	3.08	0.76	1.61	**3.41**
HMAFCN	1.33	**2.60**	**4.94**	**0.99**	**1.58**	**2.88**	**0.75**	**1.55**	3.47

注. 粗体表示最优值, 下划线表示次优值.



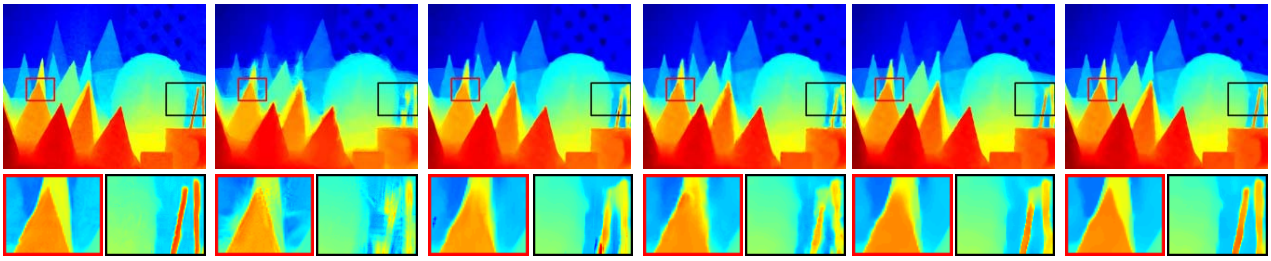
a. 真实深度图　b. GF[1]　c. DKN[31]　d. DCTNet[18]　e. DAGF[15]　f. HMAFCN

图 4　5 种方法在 NYU v2 数据集上的超分辨率重建效果



a. 真实深度图　b. GF[1]　c. DKN[31]　d. DCTNet[18]　e. DAGF[15]　f. HMAFCN

图 5　5 种方法在 Lu 数据集上的超分辨率重建效果



a. 真实深度图　b. GF[1]　c. DKN[31]　d. DCTNet[18]　e. DAGF[15]　f. HMAFCN

图 6　5 种方法在 Middlebury 数据集上的超分辨率重建效果

DCTNet[18]和 DAGF[15]存在边缘平滑问题，与这些方法相比，HMAFCN 利用深度边缘和彩色结构特征的注意力为深度图补充结构信息，有效地修复了深度图的缺陷，使得结果更接近真实深度图．在 NYU v2 数据集上，将 HMAFCN 与 DKN[31]和 A2G[33]的 8 倍重建实验的平均用时和显存占用进行比较，结果如表 2 所示．可以看出，HMAFCN 的重建时间为 48.6 ms，重建深度图误差最小，但由于要处理和保存多种模态的数据，因此占用显存较大．

表 2　不同方法在 NYU v2 数据集上 8 倍上采样指标对比

方法	用时/ms	显存/GB	RMSE
DKN[31]	88.2	2.32	3.26
A2G[33]	167.6	2.41	2.64
HMAFCN	48.6	6.64	2.60

在合成数据集定量实验中，本文选用基准方法 Bicubic，2 种传统的方法 GF[1]和 TGV[3]和 7 种基于学习的方法 MSG[5]，MFR[14]，PMBANet[21]，AHMF[7]，DAGF[15]，CMSFN[34]和 A2G[33]与 HMAFCN 进行实验，合成数据集定量结果如表 3 所示．可以看出，在 18 个案例中，HMAFCN 取得了 15 个最优结果，其余结果也取得了和最优方法相近的结果；由于 DEAB 提供了如图 3 所示的有效的边缘监督和彩色结构信息补充，HMAFCN 在 8 倍和 16 倍上采样结果中依靠更多的结构信息和有针对性的融合矫正，取得了更显著的性能差距；HMAFCN 在 4 倍、8 倍和 16 倍上采样倍率下，RMSE 均值分别为 0.85，1.36 和 2.13，4 倍上采样重建结果与 A2G[33]相当，但在 8 倍和 16 倍上采样结果中分别降低了 0.08 和 0.17．在 16 倍重建倍数下，与 PMBANet[21]，AHMF[7]，DAGF[15]和 A2G[33]的定性比较结果如图 7 所示．从图中的放大区域可以看出，PMBANet[21]存在严重的扭曲和深度缺失，如“Art”图像中的画笔和“Laundry”中的窗户，同样的位置上，AHMF[7]存在扭曲和边缘锯齿，DAGF[15]存在涂抹；A2G[33]虽然取得了良好的效果，但仍存在边缘模糊和边缘锯齿；而 HMAFCN 重建的深度图边缘清晰没有锯齿和模糊，更接近真实深度图．

3.4　消融实验

为了验证本文各模块的有效性，首先将 DEAB，SEPB 和 MAFCB 去除作为基准网络，然后逐一消去各个模块，验证单个模块的有效性．表 4 所示为上采样倍数为 8 时，在 NYU v2 数据集上的消融实验结

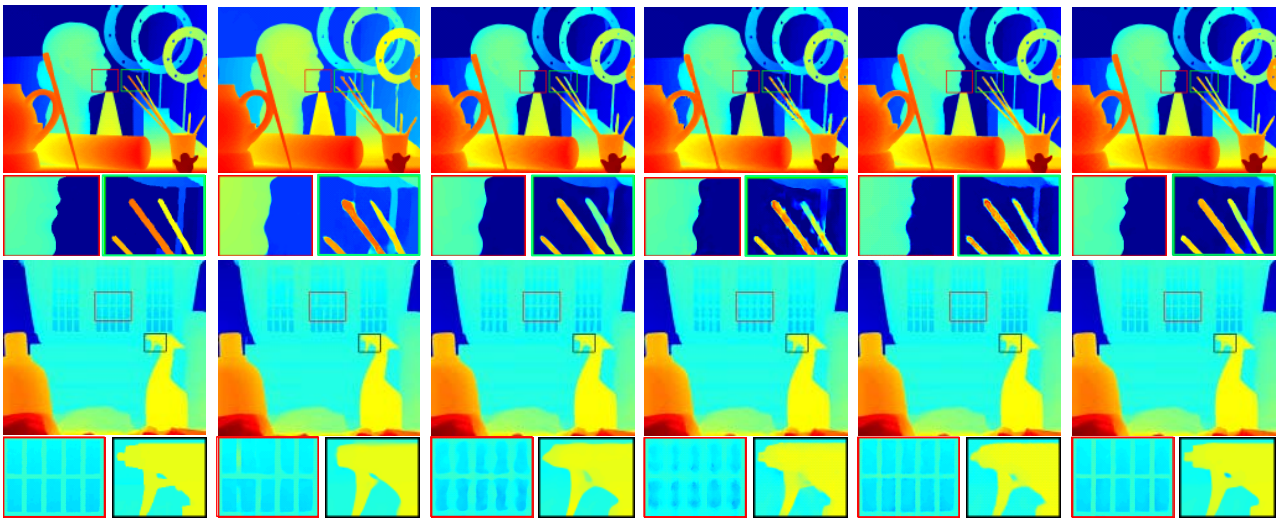
表 3　不同方法在 Middlebury 数据集(2005)上 RMSE 对比

方法	Art			Books			Dolls		
	4×	8×	16×	4×	8×	16×	4×	8×	16×
Bicubic	3.85	5.52	8.37	1.58	2.27	3.36	1.31	1.96	2.63
GF[1]	3.79	4.97	7.88	1.57	2.09	3.18	1.31	1.86	3.62
TGV[3]	3.79	4.79	7.11	1.61	1.99	2.94	1.42	2.05	4.44
MSG[5]	1.65	3.01	5.76	0.69	1.48	2.96	0.92	1.47	3.29
MFR[14]	1.54	2.71	4.35	0.63	1.05	1.78	0.89	1.22	1.74
PMBANet[21]	2.04	3.63	5.38	0.92	1.68	2.55	0.95	1.47	2.03
AHMF[7]	1.39	2.32	4.50	0.54	**0.88**	1.54	0.82	1.21	1.85
DAGF[15]	1.61	2.78	4.67	0.60	1.04	1.84	0.86	1.26	1.77
CMSFN[34]	1.46	2.58	4.13	0.59	0.99	1.65	0.87	1.15	1.65
A2G[33]	**1.31**	2.46	3.96	0.54	0.90	1.57	**0.81**	1.15	1.63
HMAFCN	1.32	**2.27**	**3.78**	**0.53**	**0.88**	**1.38**	0.82	**1.14**	**1.52**

续表

方法	Laundry			Moebius			Reindeer		
	4×	8×	16×	4×	8×	16×	4×	8×	16×
Bicubic	2.41	3.45	5.09	1.40	2.06	2.98	2.81	3.99	5.82
GF[1]	2.54	3.42	4.56	1.44	1.88	2.86	3.05	4.06	5.32
TGV[3]	2.21	3.92	6.75	1.46	1.91	2.63	2.67	4.29	8.80
MSG[5]	1.12	2.09	4.26	0.76	1.44	2.91	1.32	2.43	4.97
MFR[14]	1.11	1.75	3.01	0.72	1.10	1.73	1.23	2.06	3.74
PMBANet[21]	1.14	2.19	3.31	0.84	1.41	2.09	1.39	2.74	4.12
AHMF[7]	0.81	<u>1.35</u>	2.65	**0.64**	1.00	1.69	1.03	<u>1.66</u>	3.04
DAGF[15]	0.87	1.53	2.95	0.68	1.08	1.84	1.15	2.09	3.15
CMSFN[34]	0.93	1.58	2.73	0.67	0.98	1.61	1.11	2.01	3.42
A2G[33]	<u>0.79</u>	1.41	<u>2.29</u>	<u>0.66</u>	<u>0.97</u>	<u>1.52</u>	<u>0.96</u>	1.75	<u>2.85</u>
HMAFCN	**0.78**	**1.32**	**2.16**	0.67	**0.95**	**1.39**	**0.95**	**1.62**	**2.54**

注. 粗体表示最优值, 下划线表示次优值.



a. 真实深度图 b. PMBANet[21] c. AHMF[7] d. DAGF[15] e. A2G[33] f. HMAFCN

图 7 5 种方法在 Middlebury 数据集(2005)上的超分辨率重建效果

表 4 本文模块消融实验结果

DEAB	SEPB	MAFCB	RMSE
×	×	×	2.74
×	√	√	2.69
√	×	√	2.63
√	√	×	2.65
√	√	√	2.60

果. 可以看出, 基准网络的性能较差, 消去不同模块相对于完整网络均有性能下降; 消去 DEAB 时, 由于缺失感知边缘的监督, 性能下降最严重; 缺失 MAFCB 时, 多模态特征无法在通道和空间维度上进行有效的注意力集中, 导致更多的参数被用于模态特征对齐; 消去 SEPB 时, 需要在网络中花费额外的计算来处理彩色图像中的多余纹理, 造成性能浪费和重建效果下降.

4 结 语

本文提出一种用于 DSR 重建的 HMAFCN. 首先通过 DEAB 提取 HR 深度边缘; 然后使用 SEPB 去除彩色引导图中多余的纹理信息, 提取彩色结构特征; 最后在 MAFCB 中使用与 HR 深度边缘、彩色结构特征和深度特征相匹配的注意力机制对深度图进行融合矫正, 重建缺失的深度结构信息. 精准的深度边缘监督矫正、彩色结构特征保留和有效的多注意力结构融合, 大幅减少了引导深度图重建存在的伪影和边缘模糊. 在 3 个数据集上的实验结果表明, HMAFCN 能够有效地重建出边缘清

晰和伪影更少的深度图.

为了能更加有效地重建深度图，未来考虑加入额外全局信息和增强对彩色分支的监督，以更有效的方式实现更高精度的 DSR 重建.

参考文献(References):

[1] He K M, Sun J, Tang X O. Guided image filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397-1409

[2] Lei J J, Li L L, Yue H J, *et al*. Depth map super-resolution considering view synthesis quality[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1732-1745

[3] Ferstl D, Reinbacher C, Ranftl R, *et al*. Image guided depth upsampling using anisotropic total generalized variation[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2013: 993-1000

[4] Liu B H, Ling Q. Edge-guided depth image super-resolution based on KSVD[J]. IEEE Access, 2020, 8: 41108-41115

[5] Hui T W, Loy C C, Tang X O. Depth map super-resolution by deep multi-scale guidance[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 353-369

[6] Li T, Lin H W, Dong X C, *et al*. Depth image super-resolution using correlation-controlled color guidance and multi-scale symmetric network[J]. Pattern Recognition, 2020, 107: Article No.107513

[7] Zhong Z W, Liu X M, Jiang J J, *et al*. High-resolution depth maps imaging via attention-based hierarchical multi-modal fusion[J]. IEEE Transactions on Image Processing, 2022, 31: 648-663

[8] Fan Peipei, Dong Xiucheng, Li Tao, *et al*. Super-resolution reconstruction of depth map based on non-local means constraint[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(10): 1671-1678(in Chinese)
(范佩佩, 董秀成, 李滔, 等. 基于非局部均值约束的深度图像超分辨率重建[J]. 计算机辅助设计与图形学学报, 2020, 32(10): 1671-1678)

[9] Chen B L, Jung C. Single depth image super-resolution using convolutional neural networks[C] //Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society Press, 2018: 1473-1477

[10] Song X B, Dai Y C, Qin X Y. Deeply supervised depth map super-resolution as novel view synthesis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(8): 2323-2336

[11] Li Tao, Dong Xiucheng, Lin Hongwei. Depth map super-resolution reconstruction based on deeply supervised cross-scale attention network[J]. Acta Electronica Sinica, 2023, 51(1): 128-138(in Chinese)
(李滔, 董秀成, 林宏伟. 基于深监督跨尺度注意力网络的深度图像超分辨率重建[J]. 电子学报, 2023, 51(1): 128-138)

[12] Zuo Y F, Fang Y M, Yang Y, *et al*. Depth map enhancement by revisiting multi-scale intensity guidance within coarse-to-fine stages[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(12): 4676-4687

[13] Zuo Y F, Fang Y M, An P, *et al*. Frequency-dependent depth map enhancement via iterative depth-guided affine transformation and intensity-guided refinement[J]. IEEE Transactions on Multimedia, 2021, 23: 772-783

[14] Zuo Y F, Wu Q, Fang Y M, *et al*. Multi-scale frequency reconstruction for guided depth map super-resolution via deep residual network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(2): 297-306

[15] Zhong Z W, Liu X M, Jiang J J, *et al*. Deep attentional guided image filtering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 12236-12250

[16] Ye X C, Duan X Y, Li H J. Depth super-resolution with deep edge-inference network and edge-guided depth filling[C] //Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society Press, 2018: 1398-1402

[17] Wang Z H, Ye X C, Sun B L, *et al*. Depth upsampling based on deep edge-aware learning[J]. Pattern Recognition, 2020, 103: Article No.107274

[18] Zhao Z X, Zhang J S, Xu S, *et al*. Discrete cosine transform network for guided depth map super-resolution[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 5687-5697

[19] Yuan J Y, Jiang H B, Li X, *et al*. Recurrent structure attention guidance for depth super-resolution[C] //Proceedings of the 37th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 3331-3339

[20] Li T, Dong X C, Lin H W. Guided depth map super-resolution using recumbent Y network[J]. IEEE Access, 2020, 8: 122695-122708

[21] Ye X C, Sun B L, Wang Z H, *et al*. PMBANet: progressive multi-branch aggregation network for scene depth super-resolution[J]. IEEE Transactions on Image Processing, 2020, 29: 7427-7442

[22] Zhang F, Liu N, Duan F Q. Coarse-to-fine depth super-resolution with adaptive RGB-D feature attention[J]. IEEE Transactions on Multimedia, 2024, 26: 2621-2633

[23] Ariav I, Cohen I. Fully cross-attention Transformer for guided depth super-resolution[J]. Sensors, 2023, 23(5): Article No.2723

[24] Zhang Y L, Li K P, Li K, *et al*. Image super-resolution using very deep residual channel attention networks[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 294-310

[25] Yin H, Gong Y H, Qiu G P. Side window filtering[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 8750-8758

[26] Zhu X Z, Hu H, Lin S, *et al*. Deformable ConvNets v2: more deformable, better results[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 9300-9308

[27] Silberman N, Hoiem D, Kohli P, *et al*. Indoor segmentation and support inference from RGBD images[C] //Proceedings of the 12th European Conference on Computer Vision. Heidelberg: Springer, 2012: 746-760

[28] Hirschmuller H, Scharstein D. Evaluation of cost functions for stereo matching[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2007: 1-8

[29] Butler D J, Wulff J, Stanley G B, *et al*. A naturalistic open source movie for optical flow evaluation[C] //Proceedings of the 12th European Conference on Computer Vision. Heidelberg: Springer, 2012: 611-625

[30] Scharstein D, Pal C. Learning conditional random fields for stereo[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2007: 1-8

[31] Kim B, Ponce J, Ham B. Deformable kernel networks for joint image filtering[J]. International Journal of Computer Vision, 2021, 129(2): 579-600

[32] Wang X H, Chen X H, Ni B B, *et al*. Learning continuous depth representation via geometric spatial aggregator[C] //Proceedings of the 37th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 2698-2706

[33] Zuo Y F, Xu Y P, Zeng Y F, *et al*. A^{2}GSTran: depth map super-resolution via asymmetric attention with guidance selection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(6): 4668-4681

[34] Miao Yongwei, Zhang Xinjie, Ren Hanshi, *et al*. A channel multi-scale fusion network for scene depth map super-resolution[J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(1): 37-47(in Chinese)
(缪永伟, 张新杰, 任瀚实, 等. 基于通道多尺度融合的场景深度图超分辨率网络[J]. 计算机辅助设计与图形学学报, 2023, 35(1): 37-47)