

基于深监督跨尺度注意力网络的深度图像超分辨率重建

李 滔¹,董秀成¹,林宏伟²

(1. 西华大学电气与电子信息学院,四川成都 610039;2. 西北民族大学电气工程学院,甘肃兰州 730000)

摘 要: 消费级深度相机拍摄的深度图像具有分辨率较低的问题,深度图像超分辨率重建是解决该问题的有效方法. 为了提高重建性能,提出一种基于深监督跨尺度注意力网络的深度图像超分辨率重建算法. 网络逐级放大,在损失函数中对每一级的输出都进行约束,实现深监督的目的. 采用高阶跨尺度注意力模块,将多尺度特征尺度内及跨尺度相关性与注意力机制结合起来,实现多尺度特征的自适应调整. 采用内层为宽激活残差、外层为基本残差的双层残差块作为网络基本构成元素,以提高网络对复杂非线性关系的学习能力. 实验结果表明,本文算法在主观视觉效果和客观质量评价指标方面都优于当前主流的深度图像超分辨率重建算法.

关键词: 深度图像超分辨率;深度学习;深监督;多尺度特征表示;残差块

基金项目: 国家自然科学基金(No.61901392, No.62041109);四川省科技计划(No.2021YJ0109, No.2021ZYD0034)

中图分类号: TP751.1;TP183 **文献标识码:** A **文章编号:** 0372-2112(2023)01-0128-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210659

Depth Map Super-Resolution Reconstruction Based on Deeply Supervised Cross-Scale Attention Network

LI Tao¹, DONG Xiu-cheng¹, LIN Hong-wei²

(1. School of Electrical Engineering and Electronic Information, Xihua University, Chengdu, Sichuan 610039, China;

2. College of Electrical Engineering, Northwest Minzu University, Lanzhou, Gansu 730000, China)

Abstract: Depth maps captured by consumer depth cameras usually suffer from low spatial resolution. Depth map super-resolution (SR) is an effective method to solve this problem. To improve the reconstruction performance, this paper proposes a depth map super-resolution reconstruction algorithm based on deeply supervised cross-scale attention network. A multi-stage up-sampling strategy is introduced. The loss function of the network contains the constraint on the output of each stage for a deep supervision. A high-order cross-scale attention block is proposed to adaptively adjust multi-scale features by integrating the in-scale and cross-scale correlations of multi-scale features with the attention mechanism. A bilayer residual block, which contains inner wide-activated residual learning and outer basic residual learning, is used as the basic component of network for more powerful ability of complex non-linear relationship learning. Experimental results demonstrate the superiority of the proposed algorithm over several state-of-the-art depth map SR methods in terms of visual comparison and quantitative evaluation.

Key words: depth map super-resolution; deep learning; deep supervision; multi-scale feature representation; residual block

Foundation Item(s): National Natural Science Foundation of China (No.61901392, No.62041109); Department of Science and Technology of Sichuan Province (No.2021YJ0109, No.2021ZYD0034)

1 引言

深度图像超分辨率重建旨在由一幅或多幅低分辨率深度图像恢复出一幅高分辨率深度图像,其一直是图像处理领域的一个重要研究课题. 目前深度图像超分辨率重建算法主要分为三类:基于滤波的方法,基于

优化的方法和基于学习的方法.

基于滤波的方法是根据局部或非局部邻域关系来逐个估计高分辨率深度值. Hornacek 等^[1]提出了基于刚体变化三维块自相似性的滤波方法. Lei 等^[2]设计出一种综合考虑了深度平滑性、纹理相似性和视图合成质量的上采样滤波器. 虽然基于滤波的方法运算复杂

度较低,但其重建的深度图像边缘比较模糊。

基于优化的方法通过求一个全局能量最小化问题来实现重建,并采用各种先验知识来约束最小化过程。其中,Aodha等^[3]将深度图像超分辨率重建模型化表示为马尔科夫随机场,Xie等^[4]首先通过马尔可夫随机场重建深度边缘图,然后用深度边缘图引导深度重建。此类方法的性能严重依赖于所用的先验知识;此外,最小化问题的迭代求解也会导致较高的运算复杂度。

传统的基于学习的方法主要借助稀疏编码策略来学习高低分辨率空间的映射关系。Xie等^[5]提出了一种鲁棒的耦合字典学习方法。Mandal等^[6]构建例子字典,并施加了边缘保持约束来进行深度重建。近年来,基于深度学习的深度图像超分辨率重建方法已取得了较大的成功。其中,Rieger等^[7]提出的ATGV-Net在深度卷积神经网络末端增加了总广义变分约束。Song等^[8]将重建任务分解为一组视图合成子任务,每个子任务旨在合成某个固定相机位置下的深度图像,同时还采用了多级融合和强监督的网络结构。Huang等^[9]使用密集残差网络搭建金字塔结构,逐步完成重建任务。Song等^[10]提出使用迭代残差学习由粗至精地学习深度图像的高频成分。此外,方法^[11-17]通过研究同场景彩色图像与深度图像间的结构共生关系,在网络中利用彩色信息来引导深度图像重建。如Hui等^[11]提出的多尺度引导卷积网络(MSG-Net),在强度分支的特征图引导下,逐级实现深度分支的上采样。Guo等^[13]在残差U-Net结构中采用了分层次特征驱动残差学习。Li等搭建了多尺度对称网络^[16],利用彩色与深度的相关性来控制彩色引导特征的传递,从而提高引导的正确性;随后,又提出了多输入单输出的RYNet^[17]来完成彩色引导下的深度重建。

虽然基于深度学习的方法已经取得了显著进展,但如何有效地恢复图像高频信息仍具有挑战性。研究者通过研究人类视觉特性发现,人眼在观察事物时会选择性地关注重要目标区域,而忽略其它无关紧要的区域。基于此发现,研究者提出了通道注意力机制^[18-20]和空间注意力机制^[20-22],通过对特征通道相关性或特征空间相关性的学习,来重点关注重要的特征通道或空间区域,以提高网络表述能力。但目前这些关于注意力机制的研究都局限在了单一的尺度空间内。

深度图像的纹理特征较少,主要包含了图像目标的结构信息。因此深度图像重建需要重点关注网络对结构特征的提取能力。而目标结构尺度差异性的存在表明,对深度图像采用多尺度特征表示是非常必要的,它能够更准确地获取图像的结构和上下文

信息;同时,通过研究多尺度特征间的关系来自适应地强调重要的多尺度特征通道,抑制不重要的多尺度特征通道,能够提高网络对多尺度特征表示的辨析能力。

目前多尺度特征表示已在多种计算机视觉任务^[17, 23-25]中发挥作用,但是关于多尺度特征间关系的研究仍然比较匮乏。鉴于此,本文在有效结合多尺度特征表示和注意力机制的基础上,提出了一种新的深度图像超分辨率重建方法。主要工作有:设计了一种高阶跨尺度注意力模块,在多尺度特征表示的基础上,将尺度内特征相关性及跨尺度特征相关性与注意力机制结合起来,实现多尺度特征权重的自适应调整;将一种双层残差块作为网络的基本元素,外层使用基本残差块学习一般残差特征,内层使用宽激活残差块学习非线性更高的残差特征,以提高网络对高低分辨率空间映射关系的学习能力。提出了一种新颖的跨尺度注意力网络,来完成深度图像超分辨率重建任务。网络采用了逐级放大模式,并在每一级末端都增加了深度图像输出分支,以便于为该级引入相应的目标函数,从而实现对该级放大的强监督约束。

2 相关技术

2.1 注意力机制

注意力机制能通过学习对有用特征投入更多的注意力资源,对无用特征加以抑制,是各种深度学习任务中被广泛使用的技术之一。其中,Zhang等^[18]在残差块中引入通道注意力,构成了残差通道注意力网络。Hu等^[20]在残差块中整合了通道和空间注意力,以全局和局部的方式自适应地调整特征表示。此外,文献^[20-22]在网络中引入非局部空间注意力,来学习长距离特征依赖关系。但这些方法都仅研究了同一尺度下的特征相关性,忽略了跨尺度的特征相关性。

2.2 多尺度特征表示

常规神经网络在固定网络深度所提取特征的尺度是固定的。而多尺度特征表示技术能够改变固定网络深度处的感知域,获取不同尺度的特征并加以融合,从而更好地学习图像的结构特性,已被广泛用于面部分析^[23]、边缘检测^[24]、图像语义分割^[25]、目标检测^[26]等领域。空间金字塔池化^[27,28]和并行空洞卷积^[25]是获取多尺度特征的常用方法。

为了提高网络对多尺度特征表示的辨析能力,进而增强深度图像重建性能,本文将多尺度特征表示与注意力机制有机结合起来,通过学习尺度内特征相关性和跨尺度特征相关性,来了解多尺度特征通道的重要程度并进行自适应调整,以便给予不同的关注。

3 本文方法

3.1 深监督跨尺度注意力网络

本文构建了一个如图1所示的深监督跨尺度注意力网络. 为了更好地处理高倍率下的超分辨率重建, 使用了逐级放大的模式, 令 $U = 2^u$ 为重建倍率, 则网络总共有 u 级, 每级实现2倍放大. 深监督跨尺度注意力网络主要包括: 浅层特征提取、多个单级子网络、以及位于每个单级子网络后的上采样模块和重建模块.

令 D_{LR} 和 D_{SR} 分别表示输入的低分辨率深度图像和重建得到的高分辨率深度图像. 首先使用一个卷积层 (CONV) 从 D_{LR} 中提取浅层特征 F_0 , 即

$$F_0 = f_{SF}(D_{LR}) \quad (1)$$

其中 f_{SF} 表示浅层特征提取函数.

接着, 将 F_0 传入单级子网络以提取深层特征, 设第 i ($i=1, \dots, u$) 个单级子网络所输出的深层特征为 $F_{D,i}$. $F_{D,i}$ 被随后的上采样层^[29]进行分辨率增强得到 $F_{D,i}^*$,

则有:

$$F_{D,i}^* = f_{UP,i}(f_{SUB,i}(F_{D,i-1}^*)) \quad (2)$$

$f_{UP,i}$ 和 $f_{SUB,i}$ 分别表示第 i 个上采样函数和第 i 个单级子网络的合成函数. $F_{D,0}^*$ 即为 F_0 .

第 i 个单级子网络的结构如图1上半部分所示. 单级子网络包括 G_i 个串联的多尺度残差组和末端两个 CONV 层. 单级子网络采用全局残差学习, 可表示为

$$F_{D,i} = f_{GS,i}(F_{D,i-1}^*) + F_{D,i-1}^* \quad (3)$$

$f_{GS,i}$ 表示 G_i 个多尺度残差组以及末端两个 CONV 层的合成函数. 多尺度残差组间采用密集跳线连接, 以促进网络内部的信息流通. 第 g ($g=1, \dots, G_i$) 个多尺度残差组首先使用 1×1 CONV 层, 融合前 $g-1$ 个多尺度残差组通过密集跳线传递来的特征, 然后用 B 个双层残差块提取特征, 最后嵌入一个高阶跨尺度注意力模块来挖掘多尺度特征相关性.

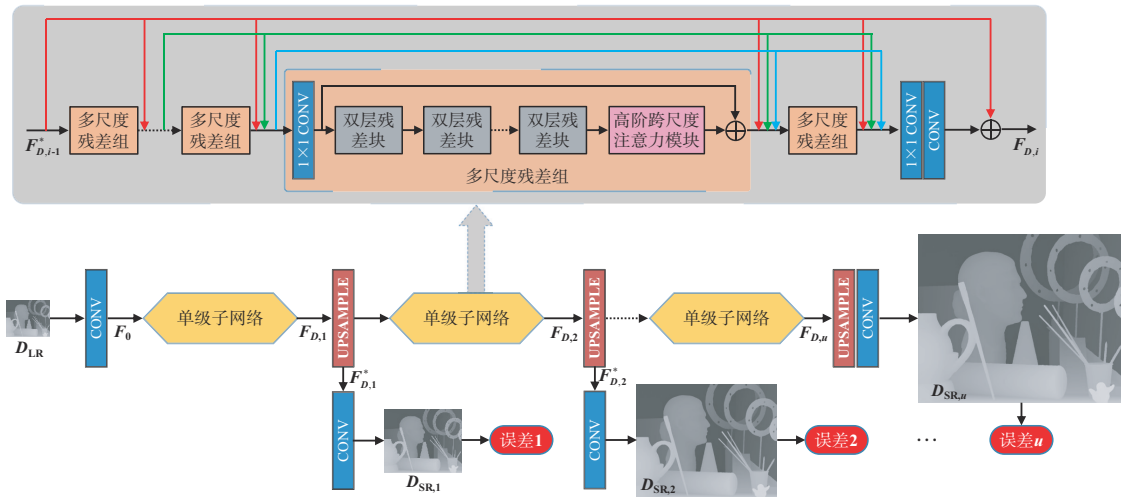


图1 深监督跨尺度注意力网络结构图

每一级最后都使用一个 CONV 层完成该级的重建, 即

$$D_{SR,i} = f_{REC,i}(F_{D,i}^*) \quad (4)$$

其中 $f_{REC,i}$ 为第 i 级的重建函数. $D_{SR,i}$ 表示第 i 级重建的深度图像, $D_{SR,u}$ 对应了网络的最终输出 D_{SR} .

为了确保每一级子网络都能有效地恢复高频信息, 本文采用了深监督的方法, 在训练过程中, 除了用真实高分辨率深度图像 D_{HR} 来监督网络的最终输出 D_{SR} , 还用了 D_{HR} 的下采样图像 $D_{HR,i}$ 来监督各级子网络的输出 $D_{SR,i}$, 其中 D_{HR} 的下采样图像可表示为

$$D_{HR,i} = \downarrow_2 D_{HR,i} \quad (5)$$

符号 \downarrow_2 表示2倍下采样, 这里使用常见的 Bicubic 插值作为下采样策略. $D_{HR,u}$ 对应了 D_{HR} . 训练集 $\{D_{LR}^j, D_{HR,1}^j, D_{HR,2}^j, \dots, D_{HR,u}^j\}$ ($j=1, \dots, N$) 包含了 N 个低分

率深度图像和真实高分辨率深度图像各级下采样图像, 网络损失函数同时考虑了各级子网络输出 $D_{SR,i}$ 与 $D_{HR,i}$ 间的 $L2$ 误差, 即

$$L(\Theta) = \frac{1}{N} \sum_{j=1}^N \left(\sum_{i=1}^u \|D_{SR,i}^j - D_{HR,i}^j\|^2 \right) \quad (6)$$

3.2 双层残差块

基本残差块^[30]有效解决了深度网络难以训练收敛的问题, 而宽激活残差块^[31]作为其变体, 通过“通道扩展-ReLU-通道收缩”过程增强了残差块的非线性学习能力. 本文在融合基本残差块和宽激活残差块的基础上, 提出了双层残差块, 该模块外层使用基本残差块学习一般的残差特征, 内层使用宽激活残差块学习非线性更高的残差特征, 然后将二者结合起来, 确保更广泛的残差特征学习, 提高网络对复杂映射关系的学习能力.

双层残差块的结构如图 2 所示,令 x 和 z 分别表示双层残差块的输入特征和输出特征,基本残差块输出的残差特征为 $F_B=f_B(x)$,宽激活残差块学习得到的残差特征表示为 $F_{WA}=f_{WA}(F_B)$,其中 f_B 和 f_{WA} 分别表示基本残差块和宽激活残差块的学习函数,则双层残差块对应的数学描述为

$$z=F_B+F_{WA}+x=f_B(x)+f_{WA}(f_B(x))+x \quad (7)$$

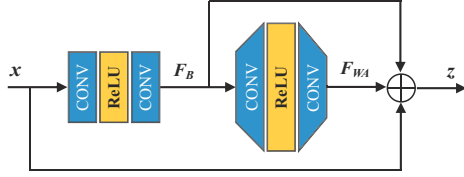


图 2 双层残差块结构图

3.3 高阶跨尺度注意力模块

由于深度图像各结构间具有明显的尺度差异,因此在重建时采用多尺度特征表示是非常必要的. 尽管已有相关工作探讨了网络的多尺度特征表示方法,但大多忽略了对多尺度特征相关性的研究. 为了解决此问题,本文设计了一个高阶跨尺度注意力模块,学习多尺度特征在尺度内和跨尺度上的相关性,并将它们与注意力机制结合起来,实现多尺度特征的自适应调整,以提高网络的辨析能力和学习能力.

3.3.1 跨尺度通道注意力生成

用 S 个并行的具有不同扩张率的 CONV 层来提取多尺度特征. 令 $F_i \in \mathbb{R}^{C \times H \times W}$ 表示第 i 个尺度的特征,其中 C, H 和 W 分别表示特征 F_i 的通道数,高和宽;用 $(F_i)_{c_i}$ 表示 F_i 的第 c_i 个通道 ($1 \leq c_i \leq C$). 受跨模态相关矩阵^[32]的启发,可将特征跨尺度的通道相关性模型化表示为高阶相关张量 T ,则:

$$\begin{aligned} T(c_1, c_2, \dots, c_S) &= \mathbf{1}_{H \times W} \left[(F_1)_{c_1} \circ (F_2)_{c_2} \circ \dots \circ (F_S)_{c_S} \right] \\ &= \mathbf{1}_{H \times W} \left[\prod_{i=1}^S (F_i)_{c_i} \right] \end{aligned} \quad (8)$$

其中 \circ 表示逐像素相乘,如果相乘的两个张量在某一个维度上尺寸不一致,但其它维度的尺寸完全一致,则先在不一致的维度上将两个张量广播至相同的尺寸,再逐像素相乘. $\mathbf{1}_{H \times W}[\cdot]$ 表示沿高、宽方向进行 2 维求和.

将二维矩阵 $(F_i)_{c_i}$ 视为 F_i 的基本原子,多尺度特征原子间的乘积由式(8)表示,则 T 可用张量乘积算子表示为

$$T = \otimes (F_1, F_2, \dots, F_S) = \bigotimes_{i=1}^S F_i \quad (9)$$

由式(9)可知,张量 T 的阶数为 S .

给定尺度 l 和通道 c_l ,可用 $T(:, \dots, c_l, \dots, :)$ 表示尺

度 l 的特征通道 $(F_l)_{c_l}$ 与其余 $S-1$ 个尺度间的相关性, $T(:, \dots, c_l, \dots, :)$ 简写为 $T(c_l)$, 有:

$$\begin{aligned} T(c_l) &= \otimes \left\{ (F_l)_{c_l} \circ F_1, F_2, \dots, F_{l-1}, F_{l+1}, \dots, F_S \right\} \\ &= \bigotimes_{i=1, i \neq l}^S F_i^* \end{aligned} \quad (10)$$

其中 $F_i^* = (F_l)_{c_l} \circ F_i$, 即 $(F_l)_{c_l}$ 与 F_i 的各通道逐像素相乘;其余的 F_i^* 等于 F_i ($i \neq l$ 且 $i \neq l$). 由式(10)可知, $T(c_l)$ 为 $S-1$ 阶张量.

通过学习跨尺度特征间的相关性,实现多尺度特征通道的自适应调整,调整的权重系数采用式(11)来得到.

$$(\alpha_l)_{c_l} = \sigma \left(\sum [W_l \circ T(c_l)] \right) \quad (11)$$

其中 $(\alpha_l)_{c_l}$ 为通道 $(F_l)_{c_l}$ 对应的权重系数; W_l 为 $S-1$ 阶权重张量; Σ 为高阶张量求和算子; $\sigma(\cdot)$ 为 Sigmoid 激活函数.

但直接按式(9)和式(10)计算 T 需要较大的运算量,因此可使用低秩近似的方式来降低运算复杂度. 首先将 W_l 低秩分解为(设张量 W_l 的秩为 k):

$$W_l = \sum_{j=1}^k \bigotimes_{i=1, i \neq l}^S w_i^j \quad (12)$$

其中 $w_i^j \in \mathbb{R}^{C \times 1 \times 1}$.

接着,将式(10)和式(12)代入式(11),得到:

$$(\alpha_l)_{c_l} = \sigma \left(\sum_{j=1}^k \left[\sum_{i=1, i \neq l}^S \left[w_i^j \circ \bigotimes_{i=1, i \neq l}^S F_i^* \right] \right] \right) \quad (13)$$

由式(13)可得到式(14)

$$(\alpha_l)_{c_l} = \sigma \left(\mathbf{1}_{H \times W} \left[(F_l)_{c_l} \circ \left(\sum_{j=1}^k \prod_{i=1, i \neq l}^S \mathbf{1}_c [w_i^j \circ F_i] \right) \right] \right) \quad (14)$$

其中 $\mathbf{1}_c[\cdot]$ 表示沿通道方向求和.

式(14)表明,跨尺度通道注意力权重的计算可用一个常规的卷积神经网络模块来实现. 以尺度 l 的注意力权重 α_l 为例,其实现过程为:采用数据驱动的方式获取 w_i^j , 则 $\mathbf{1}_c[w_i^j \circ F_i]$ 可用输出通道数为 1 的 1×1 CONV 层(简称为 CONV-CC)实现, w_i^j 对应了 CONV-CC 层的权重参数. j 表示秩的索引,它的值域 ($j=1, \dots, k$) 表明每个尺度特征 F_i 需要并行使用 k 个 CONV-CC 层. 对于第 j 个秩, $\prod_{i=1, i \neq l}^S \mathbf{1}_c [w_i^j \circ F_i]$ 表示需将除尺度 l 外其余 $S-1$ 个尺度对应秩 (j) 位置的 CONV-CC 层输出进行逐像素相乘. 将所有秩索引 ($j=1, \dots, k$) 上得到的乘积矩阵相加并乘以 F_l . 随后的 2 维求和 $\mathbf{1}_{H \times W}[\cdot]$ 可用全局平均池化层实现, $\sigma(\cdot)$ 运算用 Sigmoid 层实现. 由于空间有限,图 3 中只给出了 α_1 的详细生成结构,其余 α_l ($l=2, \dots, S$) 以此类推.

3.3.2 尺度内通道注意力生成

除了跨尺度通道注意力,本文还考虑了尺度内通道注意力^[18],其网络结构如图3右下角所示.给定尺度 l ,先对 F_l 进行全局平均池化,得到 F_l 的通道统计特征 z_l ,即

$$z_l = \frac{1}{H \times W} \mathbf{1}_{H \times W} [F_l] \quad (15)$$

按照式(16)所示的通道压缩-扩展方式从 z_l 中学习尺度内特征通道的依赖关系:

$$\beta_l = \sigma(W_l^E \delta(W_l^S z_l)) \quad (16)$$

$\delta(\cdot)$ 为ReLU激活函数, W_l^S 表示通道压缩 CONV

层的权重参数, W_l^E 表示通道扩展 CONV 层的权重参数.

3.3.3 跨尺度和尺度内通道注意力的结合

将跨尺度通道注意力和尺度内通道注意力结合起来,实现多尺度特征的自适应调整,以强调更重要的多尺度信息,即

$$\hat{F}_l = (\alpha_l \circ \beta_l) \circ F_l \quad (17)$$

其中 α_l, β_l 表示注意力的结合.紧接着,直接采用Concat层和CONV层来融合调整后的多尺度特征,并将融合特征与高阶跨尺度注意力模块的输入特征相加以实现局部残差学习.

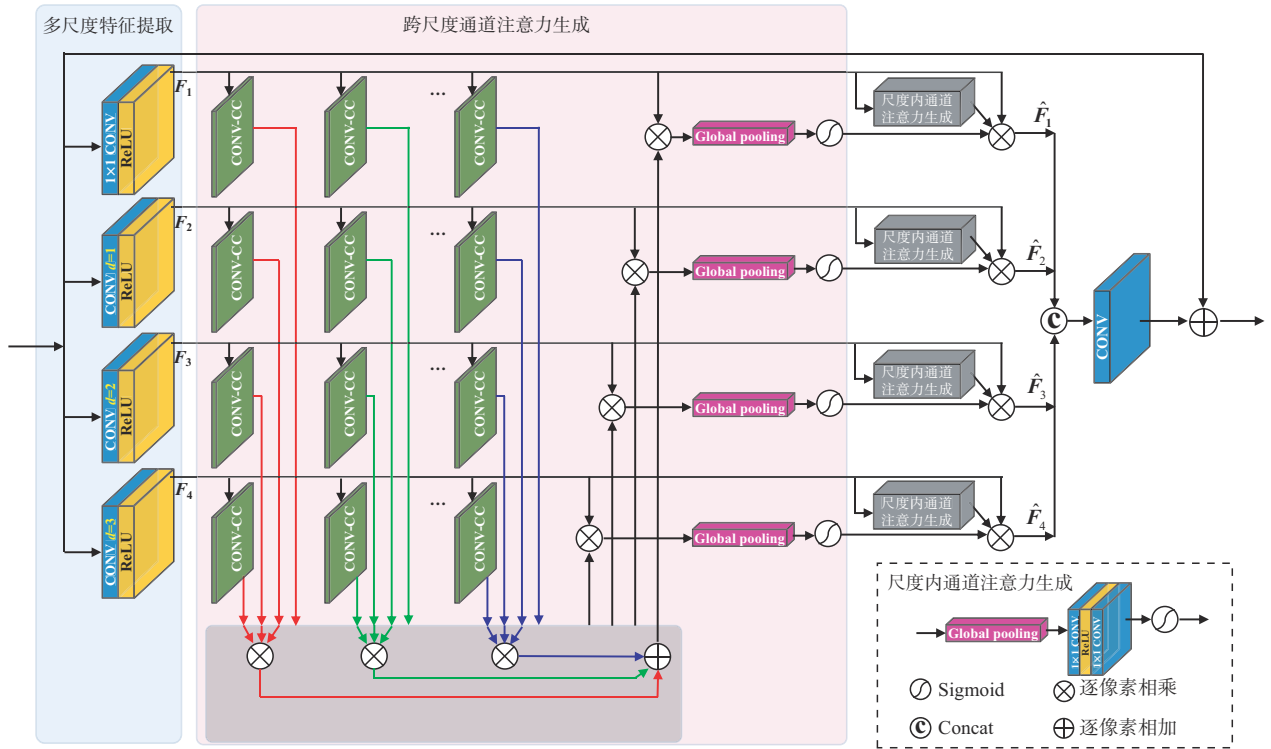


图3 高阶跨尺度注意力模块结构图

4 实验结果与分析

4.1 实验设置

4.1.1 数据集

训练集:使用MSG-Net公开提供的深度图像训练集^[11],其包含82幅RGB-D图像对.实验中将训练深度图像进行旋转和翻转数据增强操作,同时在增强数据集上将高分辨率图像裁剪为大小为 $P \times P$ 的子图,裁剪步幅为24;无噪实验中2倍和4倍率下的 P 值均为96,8倍率下的 P 为128;有噪实验中 P 取192.将高分辨率子图双三次下采样可得到相应倍率下的低分辨率子图及各下采样子图.所有的训练样本都归一化至 $[0,1]$.

测试集:与多数深度图像重建算法^[11,13,14,17]一样,

选取Middlebury^[33]中的10幅图像进行有噪和无噪实验.有噪实验中,对低分辨率深度图像叠加了方差为25的高斯噪声.此外在标准测试集Middlebury,ToFMark和NYU_Depth v2上进行了实验比较,给出了统计分析.数值实验采用峰值信噪比(Peak Signal to Noise Ratio, PSNR)和误像素百分比(Percentage of error pixels, PE)作为评价指标.误像素百分比衡量了重建深度误差大于给定阈值(T_e)的像素数占图像总像素数的百分比,无噪和有噪实验中 T_e 分别为5和10.

4.1.2 参数设置

G_i 的取值规则为:最后一个单级子网络的 G_u 设为10,其余单级子网络的 $G_i (1 \leq i < u)$ 设为4.每个多尺度残差组中双层残差块数 B 为10.高阶跨尺度注意力模

块中,多尺度数为4,低秩分解的秩 k 为3.除非特别说明,CONV层的滤波器大小为 3×3 ,输出通道数为64.网络模型采用Pytorch框架实现,使用Adam算法($\beta_1=0.9$, $\beta_2=0.999$)完成优化.学习率初始化为 10^{-4} ,每隔40 epochs减半,总共训练80 epochs.为了便于比较,每组实验中的最优客观评价指标都加粗显示,第二优的加下划线显示.

4.2 实验结果

4.2.1 无噪声深度图像超分辨率重建实验

本小节主要通过无噪声深度图像超分辨率重建实验来验证方法的有效性.除了Bicubic方法外,还选取了单幅深度图像重建方法RCAN^[18],SAN^[19]和ATGV-Net^[7],以及彩色引导下的深度图像重建方法DJFR^[15],JID^[12],MSG-Net^[11],DepthSR-Net^[13]和RYNet^[17]作为比较算法.表1对各个算法在2倍,4倍和8倍率下的重建PSNR和PE值进行了比较.可以看出,2倍率下单幅深度图像重建方法RCAN,SAN和ATGV-Net能够取得与MSG-Net,DepthSR-Net和RYNet相当甚至更高的PSNR值,但随着倍率的增大,前三者的性能就明显低于后三者了,这表明彩色引导信息能够消除特征不确定性,从而提高重建性能,在大倍率重建时发挥的作用非常明显.尽管本文方法没有彩色引导,但由于具有对高低分

辨率空间映射关系的良好学习能力和较强的特征辨析能力,取得的性能仍然优于其它算法.具体地,与第二优具有彩色引导的RYNet相比,本文方法的平均PSNR在2倍率时提高了0.23 dB,在4倍率时提高了0.72 dB,在8倍率时提高了0.40 dB.平均PE在2倍率时比RYNet低了0.01,4倍率时虽略高于DepthSR-Net,但仍然比RYNet低了0.03,8倍率时比DepthSR-Net降低0.23.

图4和图5分别给出了深度图像Cones 4倍率和对Venus 8倍率重建结果的视觉比较,同时以热图形式展示了重建深度图与真实高分辨率深度图之间的绝对误差.可以看出,ATGV-Net,DJFR和JID具有明显的重建失真,MSG-Net结果图的部分区域比较模糊,RCAN,SAN,DepthSR-Net和RYNet虽然能够恢复锐化的深度边缘,但仍存在明显的重建误差(如放大区域所示).相比之下,本文方法在恢复边缘清晰度和减少重建误差方面都具有较好的性能,重建的图像具有比其它算法更好的视觉效果,且更逼近真实高分辨率深度图像.

4.2.2 有噪声深度图像超分辨率重建实验

选取Bicubic,RCAN^[18],SAN^[19],ATGV-Net^[7],MSG-Net^[11],RYNet^[17]和DepthSR-Net^[13]作为有噪实验的比较算法.表2给出了噪声方差为25时各个算法的8倍

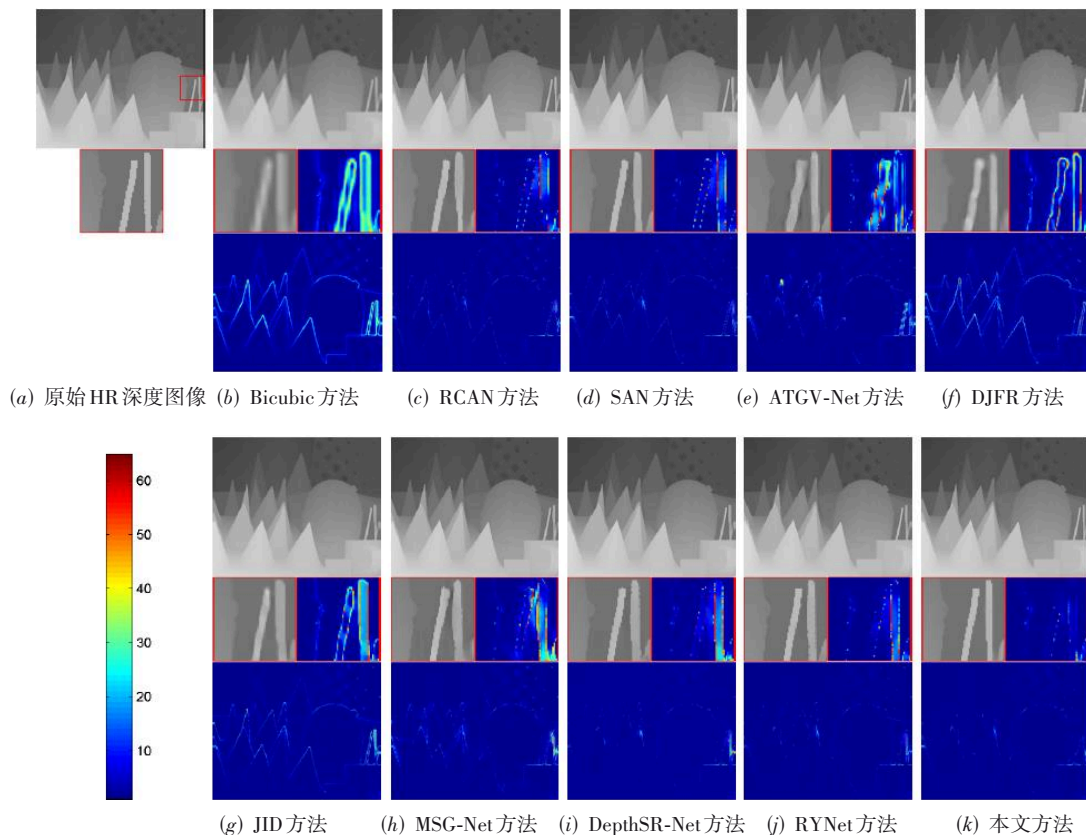


图4 无噪声深度图像Cones 4倍超分辨率重建结果的视觉比较

表 1 无噪声深度图像超分辨率重建结果的 PSNR(dB)和 PE(%)比较

方法	Art		Book		Dolls		Laundry		Moebius		Reindeer		Cones		Teddy		Tsukuba		Venus		平均	
	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE
2倍																						
Bicubic	39.74	2.750	47.77	0.687	48.96	0.709	44.03	1.305	49.32	0.754	42.45	1.309	40.14	3.649	42.40	2.613	32.82	6.884	45.84	1.044	43.35	2.170
RCAN	55.35	0.084	57.11	0.023	55.01	0.070	56.10	0.058	55.99	0.073	55.94	0.048	51.31	0.271	51.25	0.284	<u>51.58</u>	0.269	55.92	0.010	54.56	0.119
SAN	55.10	0.094	57.19	0.023	55.00	0.073	56.07	0.061	55.89	0.074	55.90	0.050	51.19	0.254	51.09	0.299	53.66	<u>0.253</u>	55.32	0.014	54.64	0.120
ATGV-Net	54.11	0.108	60.29	0.023	57.04	0.054	56.10	0.073	58.46	0.055	55.59	0.061	46.77	0.643	49.97	0.482	39.65	1.488	46.41	2.336	52.44	0.532
JID	46.06	0.685	51.62	0.262	50.58	0.355	50.36	0.398	52.30	0.301	47.55	0.411	44.21	1.500	46.56	1.165	37.22	2.386	52.19	0.249	47.87	0.771
MSG-Net	53.16	0.148	59.94	0.025	56.70	0.043	56.75	0.061	58.17	0.059	55.59	0.069	49.00	0.611	51.07	0.384	42.80	1.247	65.08	<u>0.006</u>	54.83	0.265
DepthSR-Net	53.64	0.050	55.54	<u>0.011</u>	54.27	0.024	55.23	0.032	55.29	<u>0.029</u>	53.86	0.033	50.15	<u>0.171</u>	49.78	0.276	45.46	0.336	54.06	0.001	52.73	0.096
RYNet	<u>59.88</u>	<u>0.040</u>	63.09	0.008	59.45	0.026	61.23	<u>0.027</u>	<u>60.66</u>	0.028	<u>60.12</u>	<u>0.026</u>	54.10	0.175	<u>52.84</u>	<u>0.269</u>	47.49	0.312	65.58	0.001	<u>58.44</u>	<u>0.091</u>
Proposed	60.33	0.037	63.05	0.008	<u>58.41</u>	<u>0.025</u>	<u>61.02</u>	0.026	60.73	<u>0.029</u>	60.14	0.025	<u>54.02</u>	0.167	52.91	0.255	51.16	0.233	<u>64.94</u>	0.001	58.67	0.081
4倍																						
Bicubic	36.38	5.507	44.05	1.465	45.80	1.488	40.52	2.755	45.63	1.658	39.19	2.546	36.45	7.207	39.04	5.276	29.47	13.052	42.52	2.200	39.91	4.32
RCAN	45.19	0.470	52.80	0.125	49.35	0.315	49.45	0.308	50.93	0.295	48.00	0.265	41.61	1.690	44.09	2.204	37.15	2.031	51.28	0.138	46.99	0.78
SAN	45.06	0.565	52.90	0.128	49.70	0.310	49.47	0.317	51.25	0.288	47.91	0.291	41.57	1.804	44.38	1.939	36.79	2.236	51.06	0.175	47.01	0.81
ATGV-Net	44.52	0.542	54.63	0.084	50.30	0.238	48.61	0.405	52.26	0.211	47.56	0.244	38.84	2.670	44.70	<u>1.761</u>	31.65	4.697	58.05	0.041	47.11	1.09
DJFR	36.73	3.109	44.27	0.779	45.38	1.019	40.99	1.601	46.07	0.978	39.91	1.392	37.11	4.570	39.22	3.540	29.68	8.252	43.32	0.993	40.27	2.62
JID	42.29	1.238	49.73	0.417	48.56	0.610	45.85	0.867	49.51	0.576	44.93	0.690	39.03	3.157	43.18	2.508	32.66	4.699	49.38	0.427	44.51	1.52
MSG-Net	45.23	0.550	54.94	0.089	50.85	0.206	50.21	0.256	52.86	0.192	48.27	0.260	39.85	2.277	44.69	1.581	35.48	4.350	57.34	0.046	47.97	0.98
DepthSR-Net	46.42	<u>0.224</u>	52.35	0.056	49.97	<u>0.152</u>	50.22	<u>0.161</u>	51.54	0.157	48.40	<u>0.137</u>	40.77	1.249	45.37	1.129	<u>37.78</u>	1.171	52.01	0.014	47.48	0.45
RYNet	<u>48.29</u>	0.236	<u>57.11</u>	<u>0.049</u>	52.66	0.160	<u>52.06</u>	<u>0.161</u>	<u>54.18</u>	<u>0.152</u>	<u>50.70</u>	0.138	<u>42.46</u>	<u>1.193</u>	<u>45.48</u>	1.819	37.73	1.286	60.28	0.019	<u>50.10</u>	0.52
Proposed	48.88	0.219	<u>57.13</u>	0.047	<u>52.37</u>	0.150	52.96	0.138	54.30	0.146	51.63	0.136	45.17	0.935	45.72	1.966	40.01	1.115	<u>60.03</u>	0.014	50.82	<u>0.49</u>
8倍																						
Bicubic	33.39	10.534	40.79	3.098	42.75	3.075	37.38	5.600	42.23	3.610	36.14	4.967	33.29	12.932	35.99	10.268	26.31	20.858	39.35	4.454	36.76	7.94
RCAN	38.67	2.476	46.59	0.681	45.86	0.955	42.98	1.366	46.88	0.912	41.40	1.189	35.38	5.823	38.77	5.798	30.30	8.041	44.61	0.941	41.14	2.82
SAN	39.21	1.637	48.26	0.425	45.71	0.841	44.23	1.008	47.08	0.752	42.36	0.784	35.58	5.010	39.01	5.377	30.59	6.432	45.31	0.574	41.73	2.28
ATGV-Net	38.15	2.829	47.50	0.630	45.65	0.990	42.71	2.139	47.18	0.850	41.86	1.092	33.90	7.339	39.01	5.520	26.87	12.184	47.27	0.406	41.01	3.40
DJFR	34.43	5.463	41.48	1.544	43.27	1.976	38.45	2.928	42.54	2.132	37.12	2.559	32.39	7.674	37.27	6.077	25.11	14.445	41.22	1.843	37.33	4.66
JID	39.27	1.794	47.81	0.602	46.16	0.992	42.74	1.483	46.61	1.003	41.16	1.185	35.04	5.403	39.53	4.984	28.42	8.064	46.35	0.622	41.31	2.61
MSG-Net	40.45	1.411	49.24	0.356	47.34	0.609	44.53	1.044	48.70	0.508	43.24	0.744	35.61	5.625	39.31	4.550	29.62	10.059	47.79	0.294	42.58	2.52
DepthSR-Net	41.02	0.892	49.03	<u>0.212</u>	47.21	0.469	45.81	0.648	48.47	0.417	44.23	0.420	<u>37.36</u>	<u>3.797</u>	<u>40.16</u>	<u>4.105</u>	<u>31.30</u>	5.990	47.26	0.194	43.19	<u>1.71</u>
RYNet	<u>41.94</u>	<u>0.795</u>	<u>50.88</u>	0.215	48.40	0.441	<u>46.47</u>	<u>0.581</u>	<u>49.96</u>	<u>0.388</u>	<u>45.13</u>	<u>0.354</u>	37.68	3.226	40.97	3.665	31.44	<u>4.899</u>	<u>51.26</u>	<u>0.142</u>	<u>44.41</u>	1.48
Proposed	42.58	0.696	52.09	0.166	<u>48.11</u>	<u>0.455</u>	<u>47.55</u>	<u>0.495</u>	50.27	0.358	45.30	0.351	37.09	<u>3.729</u>	40.03	4.606	32.22	3.846	52.82	0.113	44.81	1.48

表 2 有噪声深度图像 8 倍超分辨率重建结果的 PSNR(dB)和 PE(%)比较

方法	Art		Book		Dolls		Laundry		Moebius		Reindeer		Cones		Teddy		Tsukuba		Venus		平均	
	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE
Bicubic	31.42	9.075	34.57	3.995	34.96	3.813	33.52	5.093	34.92	3.896	32.93	5.264	31.36	9.493	32.97	7.455	25.85	15.353	34.01	5.336	32.65	6.877
RCAN	35.75	2.229	42.54	0.540	40.55	0.954	39.12	1.424	41.55	0.906	38.71	1.015	32.40	5.565	35.94	3.718	27.99	8.457	41.12	0.803	37.57	2.561
SAN	35.67	2.067	<u>42.60</u>	0.542	40.90	0.912	38.94	1.329	<u>41.67</u>	0.875	38.38	0.944	31.94	5.642	35.91	3.514	28.15	7.978	41.59	0.742	37.58	2.455
ATGV-Net	35.01	2.876	40.70	0.714	40.19	0.875	37.96	1.199	40.35	0.960	37.84	1.145	32.78	5.540	35.84	3.810	26.70	9.325	39.57	0.979	36.69	2.742
MSG-Net	35.46	2.393	40.89	0.584	40.24	0.819	38.29	1.091	39.84	1.000	37.89	1.281	32.88	5.468	36.02	3.389	27.15	9.580	40.02	0.750	36.87	2.636
DepthSR-Net	36.17	<u>1.611</u>	41.77	<u>0.420</u>	<u>41.07</u>	<u>0.623</u>	<u>39.33</u>	<u>0.722</u>	41.35	<u>0.687</u>	<u>38.72</u>	<u>0.733</u>	<u>33.53</u>	<u>4.407</u>	<u>36.43</u>	<u>2.847</u>	27.50	7.502	<u>41.58</u>	<u>0.366</u>	<u>37.75</u>	<u>1.992</u>
RYNet	37.33	1.197	42.66	0.351	42.23	0.465	40.29	0.633	42.67	0.531	39.55	0.620	34.61	3.455	37.37	2.403	<u>28.34</u>	6.264	41.37	0.330	38.64	1.625
Proposed	<u>36.22</u>	1.998	42.21	0.535	40.83	0.812	39.20	0.984	41.46	0.847	38.68	0.916	32.99	4.997	36.27	3.261	28.80	<u>7.489</u>	40.58	0.810	37.72	2.265

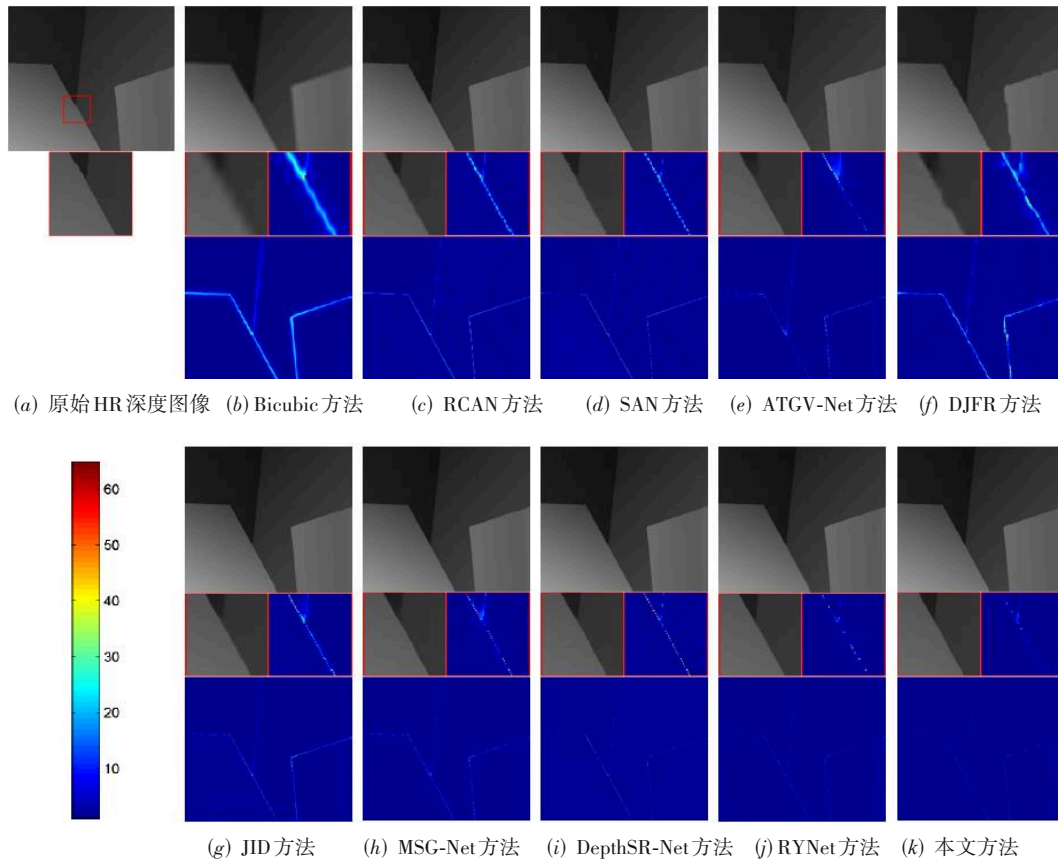


图5 无噪声深度图像Venus 8倍超分辨率重建结果的视觉比较

率重建结果. 具有彩色引导的DepthSR-Net和RYNet方法取得的性能明显优于单幅深度图像重建方法RCAN, SAN和ATGV-Net, 表明彩色信息的引入有助于抑制深度图像的噪声. 本文方法虽然没有引入彩色信息, 但仍然取得了比RCAN, SAN和ATGV-Net更优的PSNR和PE值, 平均PSNR值甚至比具有彩色引导的MSG-Net方法提高了0.85 dB, 平均PE值比MSG-Net降低了0.371. 图6给出了各个算法对噪声深度图像Tsukuba 8倍率的重建结果图. 具有彩色引导的DepthSR-Net和RYNet方法重建图中出现明显的纹理过度转移失真, 而本文方法不仅有效抑制了噪声, 还恢复出清晰锐化的边缘. 通过比较客观指标和视觉质量, 证实了算法对噪声的鲁棒性.

4.2.3 深度图像集统计实验

为了评估算法的统计性能, 在三个深度图像集: Middlebury, ToFMark和NYU_Depth上开展了4倍率下的重建实验. 如表3所示, 具有彩色引导的深度图像重建方法, 如MSG-Net、DepthSR-Net和RYNet, 在多数情况下重建性能优于RCAN, SAN和ATGV-Net. 本文方法虽然没有彩色引导, 在Middlebury和ToFMark深度图像集上仍取得了最高PSNR值和最低PE值, 在NYU_

Depth深度图像集上也仅次于RYNet, 明显优于其余方法.

4.2.4 消融实验

本小节通过消融实验来分析主要网络组件的性能贡献, 本文网络主要组件包括: 高阶跨尺度注意力模块、双层残差块和多级深监督. 这三种网络组件可以组合成不同的网络配置, 表4中列出了5种具有代表性的网络配置(C1~C5). C1为基本残差块组成的比较基准网络; C2在C1基础上增加了高阶跨尺度注意力模块; C3则将残差块替换成了双层残差块; C4同时使用了高阶跨尺度注意力模块和双层残差块; C5是本文网络的完整结构.

如表4所示, 与C1相比, C2和C3的平均PSNR值和平均PE值都有较大的改善, 证实了高阶跨尺度注意力模块和双层残差块各自的有效性, 而C4平均PSNR值和平均PE值的进一步改善验证了高阶跨尺度注意力模块和双层残差块的组合有效性. C5取得了最高的PSNR值和最低的PE值, 反应出深监督的多级强约束模式能够有效降低重建误差, 提高重建质量. 消融实验充分验证了本文网络各组件的有效性和互补性.

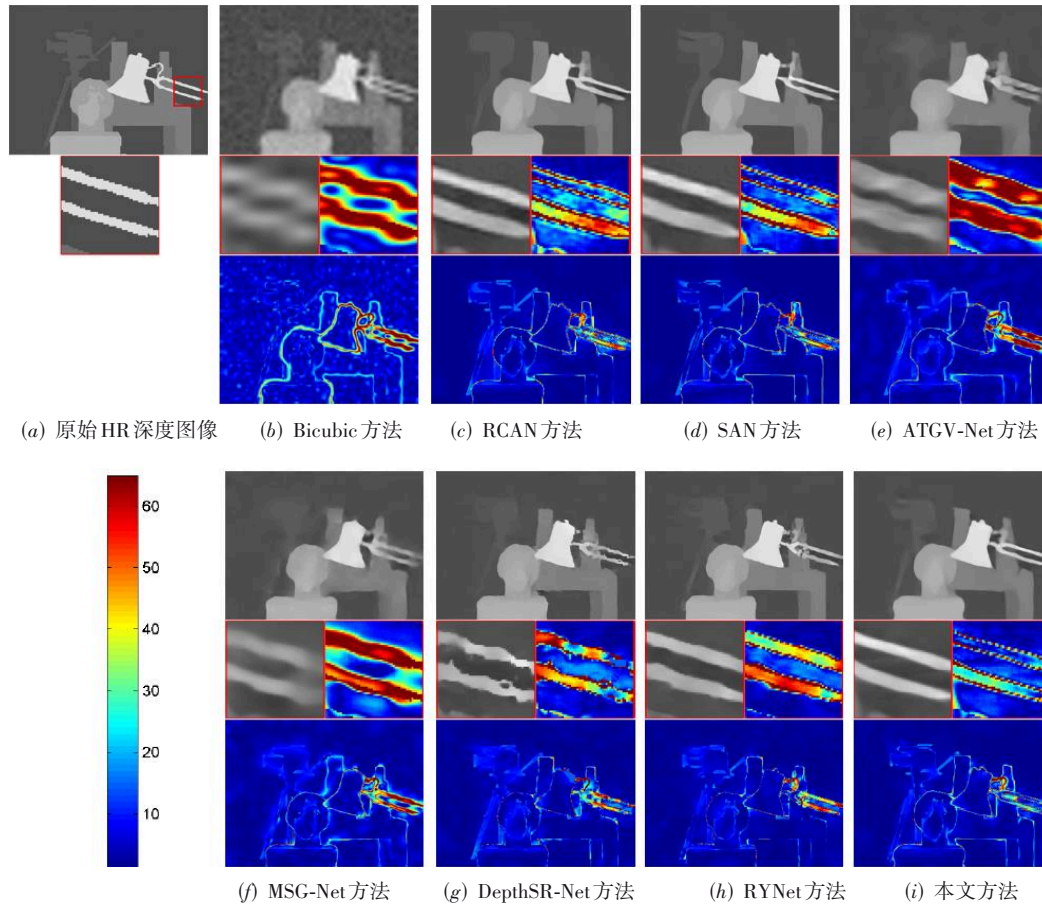


图6 有噪声深度图像Tsukuba 8倍超分辨率重建结果的视觉比较

表3 三个深度图像集重建结果的PSNR(dB)和PE(%)比较(4倍)

方法	Middlebury		ToFMark		NYU_Depth		平均	
	PSNR	PE	PSNR	PE	PSNR	PE	PSNR	PE
Bicubic	37.27	3.516	40.00	3.204	41.44	2.643	41.27	2.68
RCAN	47.34	0.575	45.43	0.656	45.84	1.023	45.90	1.00
SAN	47.39	0.606	45.58	0.694	46.10	0.986	46.15	0.97
ATGV-Net	47.34	0.787	45.23	0.677	46.31	1.050	46.35	1.04
MSG-Net	48.07	0.691	45.47	0.735	46.18	1.018	46.25	1.00
DepthSR-Net	47.84	0.369	45.59	0.545	45.88	0.830	45.96	0.81
RYNet	50.34	0.354	46.75	0.554	48.02	0.694	48.10	0.68
Proposed	50.89	0.342	47.15	0.525	47.73	0.772	47.85	0.75

表4 消融研究(8倍)

结构	C1	C2	C3	C4	C5
高阶跨尺度注意力模块	×	√	×	√	√
双层残差块	×	×	√	√	√
深监督	×	×	×	×	√
平均PSNR	41.14	43.15	43.37	43.73	44.81
平均PE	2.82	2.06	1.99	1.80	1.48

4.2.5 训练收敛性、运行时间及参数量分析

根据不同训练代(epoch)时重建结果的PSNR值

来绘制PSNR变化曲线,以验证训练收敛性.不同重建倍率下具有相似的结论,因此图7中只提供了2倍率下的PSNR变化曲线.从图7可以看出,随着训练代的增加,虽然PSNR曲线在局部区域有轻微的波动,但总体呈现上升的趋势,且在约50~60 epochs之后达到稳定收敛的状态.因此,最大训练代设置为80是合理的.

基于运行时间来比较重建算法的效率,在配备了

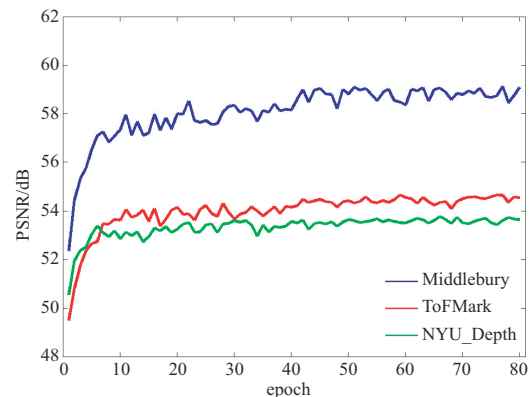


图7 三个深度图像集随训练代的重建结果PSNR曲线

NVIDIA RTX 2080Ti GPU 和 Intel Core i9 3.5 GHz CPU 的台式机上,对 Middlebury 中 1080×1320 深度图像进行测试,实验结果如表 5 所示. 本文方法的运行时间开销明显低于方法 SAN, DJFR, JID 和 DepthSR-Net; 与 RCAN,

ATGV-Net, MSG-Net 和 RYNet 相比,本文方法虽然增加了稍许时间复杂度,却获得了性能上的提升. 此外,表 5 中还列举了几种比较算法的参数量显存开销. 本文方法的显存开销明显低于 DepthSR-Net 和 RYNet.

表 5 深度图像重建的平均运行时间(秒)和显存开销(MB)比较

方法	RCAN	SAN	ATGV-Net	DJFR	JID
时间	0.45	4.99	0.39	5.62	869.60
显存开销	16	15.7	—	0.5	—
方法	MSG-Net	DepthSR-Net	RYNet	Proposed	—
时间	0.37	1.87	0.71	1.17	—
显存开销	0.8	43.3	62.7	20.5	—

5 结论与探讨

本文提出了一种基于深监督跨尺度注意力网络的深度图像超分辨率重建算法,该算法构建了高阶跨尺度注意力模块,通过学习多尺度特征跨尺度和尺度内的相关性来分配注意力权重,实现多尺度特征的自适应调整,从而提高网络的辨析能力;算法还构造了双层残差块,通过双层残差学习结构提高网络的非线性学习能力;此外,整体网络采用了逐级放大的模式,将每一级的输出都引入损失函数,实现深监督的目的. 实验表明,本文方法取得了较好的 PSNR 和 PE 客观评价值,并呈现出较好的主观视觉效果. 这也为后续应用提供了高质量的深度图像数据,因此算法具有较高的实用性.

参考文献

- [1] HORNACEK M, RHEMANN C, GELAUTZ M, et al. Depth super resolution by rigid body self-similarity in 3d [C]//IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 1123-1130.
- [2] LEI J, LI L, YUE H, et al. Depth map super-resolution considering view synthesis quality[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1732-1745.
- [3] MAC A O, CAMPBELL N D F, NAIR A, et al. Patch based synthesis for single depth image super-resolution [C]//European Conference on Computer Vision. Berlin: Springer, 2012: 71-84.
- [4] XIE J, FERIS R S, SUN M T. Edge-guided single depth image super resolution[J]. IEEE Transactions on Image Processing, 2015, 25(1): 428-438.
- [5] XIE J, FERIS R S, YU S S, et al. Joint super resolution and denoising from a single depth image[J]. IEEE Transactions on Multimedia, 2015, 17(9): 1525-1537.
- [6] MANDAL S, BHAVSAR A, SAO A K. Depth map restoration from undersampled data[J]. IEEE Transactions on Image Processing, 2016, 26(1): 119-134.
- [7] RIEGLER G, et al. Atgv-net: Accurate depth super-resolution[C]//European Conference on Computer Vision. Cham: Springer, 2016: 268-284.
- [8] SONG X, DAI Y, QIN X. Deeply supervised depth map super-resolution as novel view synthesis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(8): 2323-2336.
- [9] HUANG L, ZHANG J, ZUO Y, et al. Pyramid-structured depth MAP super-resolution based on deep dense-residual network[J]. IEEE Signal Processing Letters, 2019, 26(12): 1723-1727.
- [10] SONG X, DAI Y, ZHOU D, et al. Channel attention based iterative residual learning for depth map super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 5631-5640.
- [11] HUI T W, LOY C C, TANG X. Depth map super-resolution by deep multi-scale guidance[C]//European Conference on Computer Vision. Cham: Springer, 2016: 353-369.
- [12] KIECHLE M, HAWES S, KLEINSTEUBER M. A joint intensity and depth cosparse analysis model for depth map super-resolution[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013: 1545-1552.
- [13] GUO C, LIC, GUO J, et al. Hierarchical features driven residual learning for depth map super-resolution[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2545-2557.
- [14] ZUO Y, WU Q, FANG Y, 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, 2019, 30(2): 297-306.
- [15] LI Y, HUANG J B, AHUJA N, YANG M H. Joint image filtering with deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1909-1923.
- [16] LI T, LIN H, DONG X, et al. Depth image super-resolu-

- tion using correlation-controlled color guidance and multi-scale symmetric network[J]. Pattern Recognition, 2020, 107: 107513.
- [17] LI T, DONG X, LIN H. Guided depth map super-resolution using recumbent Y network[J]. IEEE Access, 2020, 8: 122695-122708.
- [18] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//European Conference on Computer Vision. Munich: Springer, 2018: 286-301.
- [19] DAI T, CAI J, ZHANG Y, et al. Second-order attention network for single image super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 11065-11074.
- [20] HU Y, LI J, HUANG Y, et al. Channel-wise and spatial feature modulation network for single image super-resolution[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(11): 3911-3927.
- [21] ZHANG Y, LI K, LI K, et al. Residual non-local attention networks for image restoration[C]//International Conference on Learning Representations. New Orleans: Brown Walker Press, 2019.
- [22] MEI Y, FAN Y, ZHOU Y, et al. Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 5690-5699.
- [23] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. SSH: Single stage headless face detector[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 4875-4884.
- [24] LIU Y, CHENG M M, HU X, et al. Richer convolutional features for edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(8): 1939-1946.
- [25] CHEN L C, ZHU Y, PAPANDEOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//European Conference on Computer Vision. Munich: Springer, 2018: 801-818.
- [26] 李雅倩, 盖成远, 肖存军, 等. 基于细化多尺度深度特征的目标检测网络[J]. 电子学报, 2020, 48(12): 2360-2366. LI Y, GAI C, XIAO C, et al. Object detection networks based on refined multi-scale depth feature [J]. Acta Electronica Sinica, 2020, 48(12): 2360-2366. (in Chinese)
- [27] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2881-2890.
- [28] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [29] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1874-1883.
- [30] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [31] FAN Y, YU J, HUANG T S. Wide-activated deep residual networks based restoration for BPG-compressed images [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2621-2624.
- [32] JIN T, HUANG S, LI Y, et al. Low-rank HOCA: Efficient high-order cross-modal attention for video captioning[C]//2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: ACL, 2019: 2001-2011.
- [33] Middlebury. The Middlebury Stereo Datasets[DB/OL]. (2015). <http://vision.middlebury.edu/stereo/data/>.

作者简介



李 滔 女, 1983 年 8 月 出生, 四川 资 阳 人. 分 别 于 2005 年、2008 年 和 2017 年 在 西 华 大 学 获 得 学 士、硕 士 和 博 士 学 位. 现 为 西 华 大 学 副 教 授、硕 士 生 导 师. 主 要 从 事 数 字 图 像 处 理 及 计 算 机 视 觉 方 面 的 研 究 工 作.
E-mail: lucia634@163.com



董 秀 成 男, 1963 年 4 月 出生, 陕 西 咸 阳 人. 分 别 于 1985 年 和 1990 年 在 重 庆 大 学 获 得 学 士 和 硕 士 学 位. 现 为 西 华 大 学 教 授、硕 士 生 导 师. 主 要 从 事 现 代 控 制 理 论 及 机 器 人 方 面 的 研 究 工 作.

林 宏 伟 男, 1983 年 2 月 出生. 于 2019 年 在 西 华 大 学 获 得 博 士 学 位. 现 为 西 北 民 族 大 学 副 教 授、硕 士 生 导 师. 主 要 从 事 数 字 图 像 处 理、视 频 压 缩 及 通 信 方 面 的 研 究 工 作.