

# 分级特征反馈融合的深度图像超分辨率重建

张帅勇<sup>1,2</sup> 刘美琴<sup>1,2</sup> 姚超<sup>3</sup> 林春雨<sup>1,2</sup> 赵耀<sup>1,2</sup>

**摘 要** 受采集装置的限制,采集的深度图像存在分辨率较低、易受噪声干扰等问题.本文构建了分级特征反馈融合网络(Hierarchical feature feedback network, HFFN),以实现深度图像的超分辨率重建.该网络利用金字塔结构挖掘深度-纹理特征在不同尺度下的分层特征,构建深度-纹理的分层特征表示.为了有效利用不同尺度下的结构信息,本文设计了一种分级特征的反馈式融合策略,综合深度-纹理的边缘特征,生成重建深度图像的边缘引导信息,完成深度图像的重建过程.与对比方法相比,实验结果表明 HFFN 网络提升了深度图像的主、客观重建质量.

**关键词** 深度图像,超分辨率重建,特征融合,残差学习

**引用格式** 张帅勇,刘美琴,姚超,林春雨,赵耀.分级特征反馈融合的深度图像超分辨率重建.自动化学报,2022,48(4):992-1003

**DOI** 10.16383/j.aas.c200542

## Hierarchical Feature Feedback Network for Depth Super-resolution Reconstruction

ZHANG Shuai-Yong<sup>1,2</sup> LIU Mei-Qin<sup>1,2</sup> YAO Chao<sup>3</sup> LIN Chun-Yu<sup>1,2</sup> ZHAO Yao<sup>1,2</sup>

**Abstract** Due to the limitations of the depth acquisition devices, the depth maps are easily distorted by noise and with low resolution. In this paper, we propose a hierarchical feature feedback network (HFFN) to reconstruct depth map with high-resolution. The HFFN uses a pyramid structure to mine the features of depth-texture at different scales, and constructs the hierarchical feature representations of depth-texture. In order to effectively utilize the structural information at different scales, a feedback fusion strategy based on hierarchical features is designed, which integrates the edge features of depth-texture to generate edge guidance information to assist the reconstruction process of depth map. Compared with the comparison methods, the experimental results show that the HFFN can achieve better subjective and objective quality.

**Key words** Depth map, super-resolution reconstruction, feature fusion, residual learning

**Citation** Zhang Shuai-Yong, Liu Mei-Qin, Yao Chao, Lin Chun-Yu, Zhao Yao. Hierarchical feature feedback network for depth super-resolution reconstruction. *Acta Automatica Sinica*, 2022, 48(4): 992-1003

近年来,深度信息在计算机视觉领域的应用越来越广.在远程医疗、无人驾驶以及安防监控等领域,深度信息可以显著地提升产品的性能.然而,受采集设备的限制,采集的深度图像分辨率比较低,如 Mesa Swiss Ranger 4000 采集的深度图像的分辨率仅为  $176 \times 144$ ,微软 Kinect V2 采集的深度图

像的分辨率为  $512 \times 424$ ,均难以满足实际需求.因此,如何由低分辨率深度图像重建高分辨率的深度图像已经成为当前多媒体领域的研究热点之一.

与深度图像相比,同一场景下的高分辨率纹理图像更容易获取,且二者之间存在着高度关联的结构相似性.因此,许多学者利用同一场景下高分辨率纹理图像所包含的先验信息,引导完成深度图像的超分辨率重建过程.如:在早期阶段,一般采用低通滤波器<sup>[1-8]</sup>,这类方法实现简单,却会过滤掉大部分的高频信息,导致重建的深度图像出现边缘模糊现象.基于优化表示的方法<sup>[9-16]</sup>设计了相应的优化函数和正则项,约束边缘信息,充分考虑了深度-纹理的边缘相似性,但是手工设计的优化函数并不具有普适性,且求解最优化问题耗时巨大,难以满足实际应用的需求.

近年来,基于深度学习的方法在深度图像超分辨率重建领域的应用越来越广,按照是否利用纹理图像的引导信息大致可分为两类:单深度图像超分辨率算法和纹理图像引导的深度图像超分辨率算

收稿日期 2020-07-13 录用日期 2020-11-04

Manuscript received July 13, 2020; accepted November 4, 2020  
国家自然科学基金(61972028, 61902022, U1936212),国家重点研发计划(2018AAA0102100),中央高校基本科研业务费(2019JBM018, FRF-TP-19-015A1)资助

Supported by National Natural Science Foundation of China (61972028, 61902022, U1936212), National Key Research and Development of China (2018AAA0102100), and the Fundamental Research Funds for the Central Universities (2019JBM018, FRF-TP-19-015A1)

本文责任编辑 桑农

Recommended by Associate Editor SANG Nong

1. 北京交通大学信息科学研究所 北京 100044 2. 现代信息科学与网络技术北京市重点实验室 北京 100044 3. 北京科技大学计算机与通信工程学院 北京 100083

1. Institute of Information Science, Beijing Jiaotong University, Beijing 100044 2. Beijing Key Laboratory of Modern Information Science and Network Technology, Beijing 100044 3. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083

法. 与图像超分辨率算法<sup>[17-21]</sup>类似, 单深度图像超分辨率算法<sup>[22-23]</sup>通过构建卷积神经网络, 学习从低分辨率深度图像到高分辨率深度图像的非线性映射关系, 实现深度图像的超分辨率重建; 基于纹理图像引导的深度图像超分辨率算法<sup>[24-29]</sup>通常采用上采样或者下采样的方式, 获取丰富的层次特征, 利用对应尺度下纹理图像提供的引导信息, 完成深度图像的重建过程, 如: Hui 等<sup>[25]</sup>构建了金字塔结构, 输入的低分辨率深度图像经逐级采样后得到多尺度的特征表示, 再融合对应尺度的纹理特征; Guo 等<sup>[29]</sup>采用 U-Net 结构对插值后的深度图像进行编码, 在解码过程中融合对应尺度的纹理特征. 这两种特征融合结构以串联方式从低尺度到高尺度地逐步重建高分辨率的深度图像, 均可以提取丰富的层次特征信息, 却未能有效利用多尺度特征间的关联性进行特征融合.

针对以上问题, 本文充分考虑了不同尺度间深度图像与纹理图像的结构相似性, 提出了一种基于分级特征反馈融合策略的深度图像超分辨率重建算法, 并相应地设计了一种端到端的分级特征反馈融合网络 (Hierarchical feature feedback network, HFFN). HFFN 网络采用了并行结构, 结合单深度图像重建网络和深度-纹理融合特征增强网络分别提取深度图像和纹理图像的深层特征, 获取深度图像与纹理图像隐含的结构信息; 改进的金字塔结构, 可以构建不同尺度下深度-纹理的分层特征表示; 设计了一种分级特征的反馈式融合策略, 实现并行金字塔不同层特征的反馈融合, 搭建起不同尺度间深度-纹理特征信息传递的桥梁, 充分挖掘不同尺度间深度-纹理的结构相似性; 结合深度-纹理边缘特征, 获取深度图像的边缘引导信息, 生成边缘引导图像; 采用残差学习的方式, 将边缘引导图像与单深度图像重建网络的重建结果相加, 以进一步提升深度图像的重建质量. 实验结果表明, 与对比方法相比, 本文算法获得了高质量的深度图像重建结果.

本文的组织结构如下: 第 1 节简要回顾了深度图像的超分辨率重建算法; 第 2 节具体介绍了 HFFN 网络的整体框架; 第 3 节详细给出了对比实验及分析结果; 第 4 节总结全文.

## 1 相关工作

深度图像超分辨率重建算法主要包括基于滤波、优化表示和深度学习三类.

### 1.1 基于滤波的深度图像超分辨率算法

基于滤波的深度图像超分辨率算法大多利用低

通滤波器, 在上采样过程中采用与深度图像点-点相关的纹理图像的高频信息, 恢复深度图像的边缘信息. 如, Lei 等<sup>[3]</sup>利用深度-纹理间的相似性, 构建了一种三边上采样滤波器; Yang 等<sup>[4]</sup>利用同场景纹理图像确定了像素上采样过程中的权值, 提出了一种自适应权值滤波的深度图像超分辨率算法; Lu 等<sup>[5]</sup>在稀疏深度样本的指导下分割纹理图像, 恢复每个子图像中的深度值, 利用了纹理边缘与深度边缘之间的依赖关系; He 等<sup>[6]</sup>通过建立引导图像与输出之间的线性关系对目标图像进行引导滤波, 相比双边滤波器保留了更多的边缘细节; Barron 等<sup>[7]</sup>提出了一种新的边缘感知平滑算法双边求解器, 实现纹理图像引导下的深度图像快速超分辨率重建; Yang 等<sup>[8]</sup>面向三维视点合成, 利用深度-纹理相似性构建了像素级的置信度计算方法, 构造了基于置信度的联合引导滤波器, 不仅考虑了深度像素间的平滑性, 而且融入了深度-纹理的相似性, 提高了深度图像上采样的性能及合成视点的质量.

### 1.2 基于优化表示的深度图像超分辨率算法

基于优化表示的深度图像超分辨率算法是构建一种深度-纹理的联合表示, 以字典或稀疏表示的方式学习深度-纹理间的映射. 如, Diebel 等<sup>[9]</sup>用马尔科夫随机场 (Markov random field, MRF) 将低分辨率深度图像与高分辨率纹理图像集成后进行上采样操作; 除应用 MRF 以外, An 等<sup>[10]</sup>通过正则项局部误差均值自适应地计算平衡正则项和保真项的正则化参数矩阵, 较好地保持了边缘信息; Park 等<sup>[11]</sup>采用非局部方法对深度图像进行正则化, 改善了深度图像的细节; Yan 等<sup>[12]</sup>采用非局部均值算法先获取初始上采样深度图像, 再利用边缘检测算法对其优化, 提高了深度图像的重建质量; Yang 等<sup>[13]</sup>采用自适应回归模型, 通过纹理图像的引导自动生成超分辨率深度图像; Ferstle 等<sup>[14]</sup>提出了一个深度图像超分辨率重建的凸优化问题, 利用纹理图像计算出各向同性扩散张量, 指导深度图像的上采样过程; Li 等<sup>[15]</sup>基于加权最小二乘法和快速全局平滑技术, 提出了一种分层、多遍的全局优化框架, 避免了不必要的纹理复制伪影; Kiechle 等<sup>[16]</sup>构建了一种双模共稀疏分析模型, 能够捕捉到纹理图像与深度图像之间的相互依赖关系, 实现深度图像的超分辨率重建.

### 1.3 基于深度学习的深度图像超分辨率算法

近年来, 卷积神经网络在图像超分辨率重建领域取得了显著的成就. Dong 等<sup>[17]</sup>首次提出了端到端的 SRCNN (Super resolution convolution neural

network) 网络, 直接学习从低分辨率到高分辨率的映射关系, 解决了基于深度学习的图像超分辨率重建问题; Shi 等<sup>[18]</sup> 和 Lim 等<sup>[19]</sup> 改进了 SRCNN, 加深了网络的深度; Zhou 等<sup>[20]</sup> 和 Zhang 等<sup>[21]</sup> 利用了递归残差网络, 在提高超分重建质量的同时限制了模型参数, 有效降低了计算复杂度。

不同于直接堆叠卷积层, Huang 等<sup>[22]</sup> 采用基于深度密集残差网络实现金字塔结构的深度图像超分辨率重建网络, 利用不同层次特征逐步重建高分辨率的深度图像; Song 等<sup>[23]</sup> 利用通道注意力机制, 实现了深度图像超分辨率重建网络, 从粗到细地恢复深度图像的高频分量; Zhao 等<sup>[24]</sup> 提出了纹理-深度条件生成对抗网络, 学习纹理图像和低分辨率深度图像之间的结构相似性; Hui 等<sup>[25]</sup> 提出了多尺度引导卷积网络 MSG (Multi-scale guided convolutional network), 利用丰富的分层纹理特征, 消除了深度图像重建后存在的模糊现象; Zuo 等提出了纹理引导增强的残差稠密网络 RDN-GDE<sup>[26]</sup> (Residual dense network for guided depth enhancement) 和多尺度融合残差网络 MFR-SR<sup>[27]</sup> (Multi-scale fusion residual network for depth map super resolution), 探究如何利用纹理图像提供的多尺度引导信息, 逐步指导深度图像的上采样重建; Ye 等<sup>[28]</sup> 提出了渐进的多分支聚合网络 PMBA (Progressive multi-branch aggregation network), 利用多分支融合方式逐步恢复退化的深度图像; Guo 等<sup>[29]</sup> 提出了深度图像超分辨率重建网络 DepthSR (Hierarchical features driven residual learning for depth map super resolution), 利用 U-Net 残差深层网络, 获取多尺度的结构信息, 完成深度图像的

超分辨率重建。

## 2 分级特征反馈融合网络

### 2.1 网络结构

本文提出的分级特征反馈融合网络 HFFN (如图 1 所示) 采用并行结构, 由单深度图像重建和深度-纹理融合特征增强两个子网络构成。首先, 两个子网络分别提取低分辨率深度图像和高分辨率纹理图像的深层特征, 构建深度-纹理的深层特征表示。其次, 利用并行金字塔结构, 设计深度-纹理反馈式融合模块 (Depth-Color feedback fusion module, DCF), 提取深度-纹理在不同尺度下的结构信息, 构建深度-纹理的多尺度分层特征表示; 应用反馈式融合策略, 挖掘不同尺度下深度-纹理隐含的关联性, 生成包含多级感受野的边缘引导信息。最后, 由边缘引导信息生成的边缘引导图像, 与单深度图像重建网络的重建结果相加, 实现深度-纹理的引导重建过程, 获得纹理图像引导的高分辨率深度图像的重建结果。

### 2.2 深度-纹理深层特征表示

由于深度图像和纹理图像表征的内容不同, 采用同一网络将难以区分所提取的纹理图像和深度图像的信息。因此, HFFN 网络利用单深度图像重建网络和深度-纹理融合特征增强网络, 提取低分辨率深度图像  $D_L$  和高分辨率纹理图像  $Y_C$  的深层特征  $F_{DL}$  和  $F_{DC}$ 。子网络的特征提取模块均由一个卷积层和  $M$  个残差块<sup>[19]</sup> 组成, 每个残差块包含一个线性整流激活函数 (Rectified linear unit, ReLU)

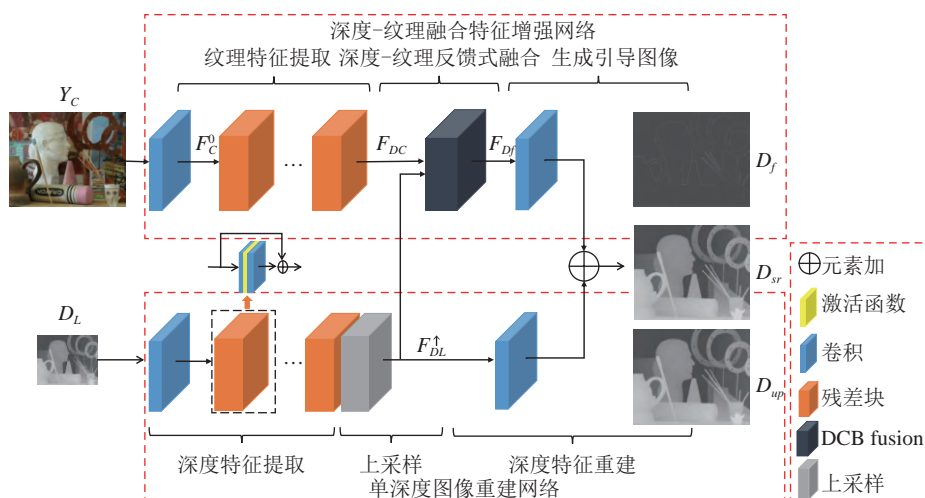


图 1 分级特征反馈融合网络

Fig. 1 Hierarchical feature feedback network

和两个卷积层。

为了降低运算复杂度, 特征提取阶段的输入为低分辨率的深度图像, 在特征融合阶段, 深度图像和纹理图像的深层特征会出现空间不匹配的问题。因此, 深度图像的深层特征  $F_{DL}$  需经上采样操作后才能与纹理特征  $F_{DC}$  融合, 以便学习二者的相关性。单深度图像重建网络采用亚像素卷积层<sup>[18]</sup>  $H_{\uparrow}$  上采样  $F_{DL}$ , 生成与纹理图像空间一致的深层特征  $F_{DL}^{\uparrow}$ , 构建相应的深度-纹理深层特征表示, 如式 (1) 所示:

$$F_{DL}^{\uparrow} = H_{\uparrow}(F_{DL}) \quad (1)$$

此处,  $F_{DL}^{\uparrow}$  有两个作用: 一是与纹理图像的深层特征  $F_{DC}$  进行融合, 生成深度图像的边缘引导信息; 二是作为单深度图像重建网络的重建特征, 生成高分辨率的深度图像。

### 2.3 深度-纹理反馈式融合模块

由文献 [25-29] 可知, 多级别的感受也可以获取丰富的层次特征信息。因此, 本文构建深度-纹理反馈式融合模块 DCF, 提取深度-纹理的分层特征表示, 并利用分级特征反馈融合策略, 充分融合深度-纹理在不同尺度下的特征, 生成包含多级感受野的边缘引导信息, 如式 (2) 所示:

$$F_{DF} = H_{DCB}(F_{DL}^{\uparrow}, F_{DC}) \quad (2)$$

其中,  $H_{DCB}$  表示深度-纹理反馈式融合函数。

该模块包含多尺度分层特征构建子模块和特征

融合反馈子模块, 如图 2 所示, 图中的数字表示通道的数量。其中, 多尺度分层特征构建子模块提取  $F_{DL}^{\uparrow}$  和  $F_{DC}$  在不同尺度下的结构特征信息, 构建深度-纹理的分层特征表示; 融合特征反馈子模块对不同尺度下的结构特征进行融合反馈, 丰富融合特征包含的结构信息。

#### 2.3.1 多尺度分层特征构建子模块

多尺度分层特征构建子模块通过并行金字塔结构获取深度-纹理图像的多尺度分层特征, 以并行的方式构建深度-纹理的分层特征表示。为了避免因层数过高引起深度图像边缘结构的退化现象, 多尺度分层特征构建子模块选用三层金字塔 ( $1 \leq j \leq 3$ ) 结构。具体地, 第  $j-1$  层的分层特征  $F_{PL}^{j-1}$  和  $F_{PC}^{j-1}$  通过卷积提取其边缘结构信息, 并利用最大池化层构建深度图像和纹理图像第  $j$  层的分层特征  $F_{PL}^j$  和  $F_{PC}^j$ , 如式 (3) 和 (4) 所示:

$$F_{PL}^j = H_{\downarrow} \left( \sigma \left( W_{PL}^j * F_{PL}^{j-1} \right) + b_{PL}^j \right) \quad (3)$$

$$F_{PC}^j = H_{\downarrow} \left( \sigma \left( W_{PC}^j * F_{PC}^{j-1} \right) + b_{PC}^j \right) \quad (4)$$

其中,  $F_{PL}^0 = F_{DL}^{\uparrow}$ ,  $F_{PC}^0 = F_{DC}$ , “\*” 表示卷积操作, “ $\sigma$ ” 表示 ReLU 函数,  $H_{\downarrow}$  表示  $2 \times$  下采样操作,  $W_{PL}^j$  和  $W_{PC}^j$  表示卷积的权重,  $b_{PL}^j$  和  $b_{PC}^j$  表示卷积的偏置。

多尺度分层特征构建子模块将并行金字塔每一层构建的分层特征传递到金字塔的下一层, 以提供不同感受野下深度-纹理的层次特征信息。

#### 2.3.2 特征融合反馈子模块

如何利用上文构建的多尺度分层特征获取深度图像的边缘引导信息, 是重建高质量深度图像的关键。大多数算法<sup>[25-27]</sup> 采用级联某一尺度下的深度特征和纹理特征的方式生成融合特征, 显然无法有效利用深度-纹理图像间的结构相似性。因此, 本文构建的特征融合反馈子模块基于分级特征反馈融合策略, 在不同层间传递结构信息, 并利用卷积自适应地分配不同层特征融合时的权重系数, 用于提高不同尺度间特征的相容性。

特征融合反馈子模块融合金字塔第三层的深度特征  $F_{PL}^3$  和纹理特征  $F_{PC}^3$  后, 通过反卷积  $H_D^{\uparrow}$  进行上采样操作, 生成金字塔第三层的深度-纹理融合特征  $R^3$ , 将  $R^3$  和金字塔第二层的特征  $F_{PL}^2$ 、 $F_{PC}^2$  融合后生成第二层的深度-纹理融合特征  $R^2$ , 第一层的深度-纹理融合特征  $R^1$  与  $R^2$  采取相同的融合方式获得, 如式 (5) 所示:

$$R^j = H_D^{\uparrow} \left( \sigma \left( W_r^j * \left[ F_{PL}^j, F_{PC}^j, R^{j+1} \right] + b_r^j \right) \right) \quad (5)$$

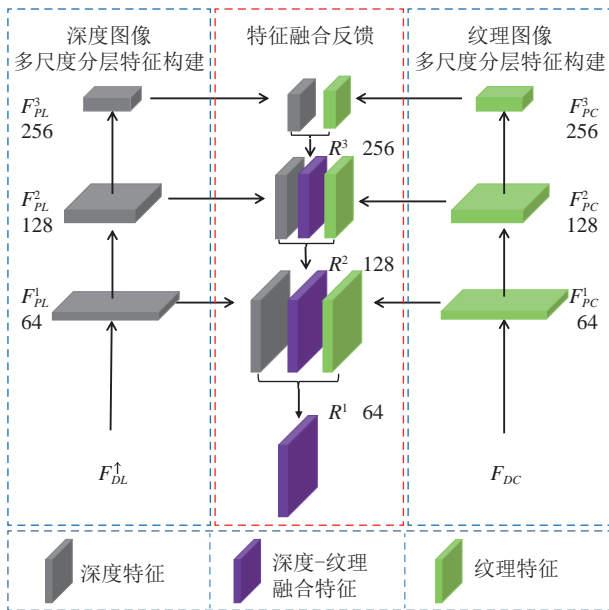


图 2 深度-纹理反馈式融合模块

Fig.2 Depth-Color feedback fusion module



其中,  $R^4$  为空集,  $W_r^j$  和  $b_r^j$  分别表示  $1 \times 1$  卷积的权重和偏置,  $[\cdot, \cdot]$  表示级联. 至此, 纹理图像对深度图像的边缘引导信息  $F_{DF}$  可由式 (6) 获得:

$$F_{DF} = \sigma(W_r * [F_{DL}^\dagger, F_{DC}, R^1] + b_r) \quad (6)$$

其中,  $W_r$  和  $b_r$  分别表示生成边缘引导信息  $F_{DF}$  时卷积的权重和偏置.

应用反馈式融合策略, 特征融合反馈子模块可以挖掘不同尺度下深度特征与纹理特征潜在的关联性, 生成包含多级感受野的边缘引导信息.

## 2.4 深度-纹理的引导重建

在构建单深度图像重建网络的基础上, 本文算法通过深度-纹理融合特征增强网络生成深度图像的边缘引导图像, 引导深度图像的重建过程. 单深度图像重建网络的重建结果  $D_{UP}$  如图 3(a) 所示, 由提取的深度特征生成; 深度-纹理融合特征增强网络的重建结果  $D_F$  如图 3(b) 所示, 包含丰富的边缘信息.

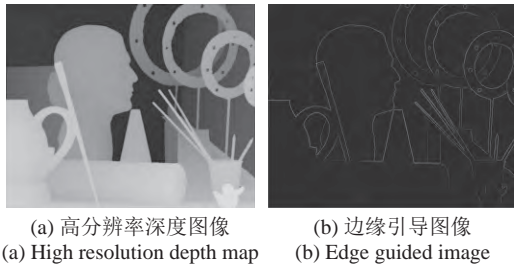


图 3 特征重建结果

Fig. 3 Feature reconstruction results

HFFN 网络采用残差学习的方式将  $D_{UP}$  与  $D_F$  相加, 得到深度图像的超分辨率重建结果  $D_{SR}$ , 利用边缘引导图像进一步增强深度图像的边缘细节信息, 如式 (7) 所示:

$$D_{SR} = H_{RL}(F_{DL}^\dagger) + H_{RC}(F_{DF}) = D_{UP} + D_F \quad (7)$$

其中,  $H_{RL}(\cdot)$  表示高分辨率深度图像重建函数,  $H_{RC}(\cdot)$  表示边缘引导图像生成函数.

## 2.5 实现细节

除标记参数的卷积层以外, HFFN 网络其他部分卷积核的大小为  $k = 3$ , 通道数为  $C = 64$ , 采用经典的均方误差 (Mean square error, MSE) 作为训练模型的损失函数, 如式 (8) 所示:

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^N \|H(D_C^i, Y_C^i, \Theta) - D_H^i\|^2 \quad (8)$$

其中,  $\Theta$  表示网络学习到的参数,  $H$  表示重建模型,  $N$  表示每批的训练样本数,  $N = 16$ , 初始的学习率  $lr = 1 \times 10^{-4}$ , 且在第 60 个和第 120 个周期时  $lr$  减半, 共训练 200 个周期. HFFN 网络采用 ADAM 算法<sup>[30]</sup> 进行优化, 参数设为  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\varepsilon = 1 \times 10^{-8}$ . 本文实验在 PyTorch 框架下使用一块 NVIDIA 2080Ti 的 GPU 卡完成.

## 3 实验

### 3.1 实验说明

**训练集:** 本文采用与文献 [22, 25, 29] 相同的数据集, 包含 34 对来自于 “Middlebury 2001 Datasets”<sup>[31]</sup>、 “Middlebury 2003 Datasets”<sup>[32]</sup> 和 “Middlebury 2014 Datasets”<sup>[33]</sup> (分辨率最小为  $432 \times 368$ 、最大为  $2300 \times 1992$ ) 以及 58 对来自于 “MPI Sintel depth” 数据集<sup>[34]</sup> (分辨率为  $1024 \times 436$ ) 的深度-纹理图像. 其中, 82 对深度-纹理图像作为训练集, 10 对深度-纹理图像作为验证集.

**测试集:** 为了评估不同算法的性能, 本文从 “Middlebury 2003 Datasets”<sup>[35]</sup> 和 “Middlebury 2005 Datasets”<sup>[32, 36]</sup> 上选取 10 对深度-纹理图像作为测试集. 参考文献 [25–27], 测试集分为 A (Art, Books, Moebius)、B (Dolls, Laundry, Reindeer)、C (Tsukuba, Venus, Teddy, Cones) 三个部分.

**数据处理:** 本文将原始深度图像  $D_H$  经双三次插值算法下采样得到  $D_L$ . 在训练阶段, 选取比例因子分别为  $2 \times$ 、 $3 \times$ 、 $4 \times$ 、 $8 \times$  时, 原始深度图像  $D_H$  和纹理图像  $Y_C$  被分割成大小为 128、144、128、128 的块, 对应的低分辨率深度图像  $D_L$  被分割成大小为 64、48、32、16 的块, 并利用随机旋转和翻转操作扩大数据集. 在训练阶段, HFFN 网络的输入为纹理图像和深度图像的 Y 通道, 输出为单通道的高分辨率深度图像. 在测试阶段, 采用均方根误差 (Root mean square error, RMSE) 和峰值信噪比 (Peak signal to noise ratio, PSNR) 评价网络的超分辨率重建性能.

### 3.2 消融分析

#### 3.2.1 参数分析

当比例因子为  $4 \times$  时, 本文评估特征提取模块残差块的数目、金字塔层数对 HFFN 网络性能的影响分别如表 1、表 2 所示.

在表 1 中, 网络模型 H\_R3、H\_R5、H\_R7、H\_R10 分别表示 Basic 网络中有 3、5、7、10 个残差块, 且金字塔层数为 3. 由表 1 可知, 随着残差块数

表 1 残差块数目对 HFFN 网络性能的影响

Table 1 Influence of the number of residual blocks on the performance of HFFN

网络模型	H_R3	H_R5	H_R7	H_R10
训练时间(h)	3	3.8	4.4	5.4
模型参数(M)	2.67	2.96	3.26	3.70
重建结果(dB)	48.14	48.23	48.45	48.54

表 2 金字塔层数对 HFFN 网络性能的影响

Table 2 Influence of pyramid layers on the performance of HFFN

网络模型	H_P2	H_P3	H_P4
训练时间(h)	3.6	3.8	8.6
模型参数(M)	1.62	2.96	8.34
重建结果(dB)	48.04	48.24	48.32

目的增加, HFFN 模型重建结果的 PSNR 值也随之增加. 同时, 残差块数目的增加也带来了网络模型训练时间和参数数量的增长, 如: 模型 H\_R7 和 H\_R10 比 H\_R5 的训练时间提高了 0.6 h 和 1.6 h、参数量提高了 0.30 M 和 0.74 M. 因此, 设置残差块超参不仅需要考量重建结果的 PSNR 值, 也需考虑模型的训练时间和参数量.

在表 2 中, 金字塔层数为 2、3、4 对应的网络模型分别为 H\_P2、H\_P3、H\_P4, 残差块数目为 5. 可以看出, 模型 H\_P4 比 H\_P3 重建的 PSNR 值提高了 0.08 dB、参数量增加了 5.38 M、训练时间增加了 4.8 h; 而模型 H\_P3 比 H\_P2 重建的 PSNR 值提高了 0.20 dB、参数量增加了 1.34 M、训练时间增加了 0.2 h.

因此, 在模型性能的验证阶段, 本文均衡考虑了 HFFN 模型的重建质量、计算复杂度及训练时间等因素, 设置残差块数目为 7、金字塔层数为 3, 不仅可以保证深度图像高分辨率重建质量, 还可以控制网络模型的参数规模和训练时间.

### 3.2.2 模型分析

为了验证 HFFN 各部分的性能, 本文将模型分为纹理、分层、深层和融合四部分. 其中纹理表示引入纹理图像的先验信息; 分层表示利用金字塔结构获取多尺度分层特征; 深层表示采用深层网络提取深度-纹理图像的特征; 融合表示加入自适应特征融合算法. 基准网络 Basic 包含五层残差块的单深度图像超分辨率重建网络, 结构与文献 [19] 类似, 相应构建的 H\_Color、H\_CP 和 H\_Res 网络的配置及实验结果如表 3 所示.

由表 3 可知, 对比 Basic 网络、H\_Color 网络、H\_CP 网络和 HFFN-100 网络的重建结果, HFFN-

表 3 消融分析结果

Table 3 Results of ablation study

网络	纹理	分层	深层	融合	结果(RMSE/PSNR)
Basic			✓		3.2238/40.071
H_Color	✓		✓		2.8239/41.438
H_CP	✓	✓	✓		2.8544/41.578
H_Res	✓	✓		✓	2.9352/41.285
HFFN-100	✓	✓	✓	✓	2.7483/41.671

100 网络的性能最优, 说明纹理图像、并行金字塔结构以及反馈式融合策略在深度图像超分辨率重建中的有效性; 此外, H\_Res 网络与 HFFN-100 网络的实验结果表明, 通过多层网络提取深度-纹理图像的深层特征, 有利于挖掘深度图像与纹理图像隐含的关联性.

本文对比 Basic 网络和 HFFN-100 网络在 “Art” 上的重建结果, 以验证纹理图像在深度图像超分辨率重建中的作用, 如图 4 所示. 由于缺失纹理图像提供的引导信息, Basic 网络的重建结果出现了伪影和噪点, 见图 4(b); 图 4(c) 为 HFFN-100 网络的重建结果, 保留了更多的边缘细节信息.

### 3.3 实验结果

实验对比算法为: 传统算法 (Bicubic、GF<sup>[6]</sup>、TGV<sup>[14]</sup>、JID<sup>[16]</sup>) 和基于深度学习的超分辨率算法 (纹理图像超分辨率算法 SRCNN<sup>[17]</sup>、单深度图像超分辨率算法 Huang<sup>[21]</sup>、纹理图像引导下的深度图像超分辨率算法 MSG<sup>[25]</sup>、DepthSR<sup>[29]</sup>、MFR-SR<sup>[27]</sup>、RDN-GDE<sup>[26]</sup>、PMBA<sup>[28]</sup>), HFFN+表示在 HFFN 算法中引入 Self-Ensemble<sup>[37]</sup> 方法. 当比例因子为 2×、3×、4×、8× 时, 深度图像超分辨率重建的客观质量对比结果如表 4 到表 6 所示. 实验结果的最优值用粗体表示、次优值用下划线表示.

由表 4 可知, 当比例因子为 2×、3×、4×、8× 时, 所提算法在 “Books” 和 “Moebius” 上的重建结果均优于对比算法. 其中, HFFN 算法重建 “Books” 的 RMSE 值比 Bicubic 算法降低了 0.8 (2×)、1.02 (3×)、1.14 (4×)、1.49 (8×); HFFN 算法重建 “Books” 的 RMSE 值比 GF<sup>[6]</sup> 算法降低了 1.21 (2×)、1.22 (3×)、1.24 (4×)、1.48 (8×); 与 Huang<sup>[21]</sup> 提出的算法相比, HFFN 算法重建 “Books” 的 RMSE 值降低了 0.26 (2×)、0.34 (4×)、0.32 (8×); 与 PMBA<sup>[28]</sup> 算法相比, HFFN 算法重建 “Books” 的 RMSE 值降低了 0.14 (2×)、0.45 (4×)、0.87 (8×). 随着放大比例因子的增加, 本文提出的 HFFN 算法明显优于传统算法 GF<sup>[6]</sup>、单深度图像算法 Huang<sup>[21]</sup> 和纹理图像引导的算法 PMBA<sup>[28]</sup>, RMSE 值之差也

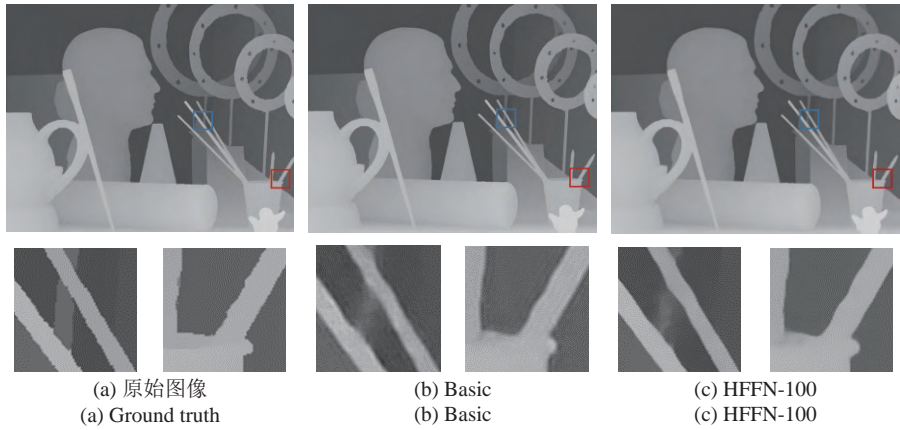


图 4 “Art”的视觉质量对比结果  
Fig.4 Visual quality comparison results of “Art”

表 4 测试数据集 A 的客观对比结果 (RMSE)  
Table 4 Objective comparison results (RMSE) on test dataset A

对比算法	Art				Books				Moebius			
	2×	3×	4×	8×	2×	3×	4×	8×	2×	3×	4×	8×
Bicubic	2.66	3.34	3.90	5.50	1.08	1.39	1.63	2.36	0.85	1.08	1.29	1.89
GF <sup>[6]</sup>	3.63	3.84	4.14	5.49	1.49	1.59	1.73	2.35	1.25	1.32	1.42	1.91
TGV <sup>[14]</sup>	3.03	3.31	3.78	4.79	1.29	1.41	1.60	1.99	1.13	1.25	1.46	1.91
JID <sup>[16]</sup>	1.24	1.63	2.01	3.23	0.65	0.76	0.92	1.27	0.64	0.71	0.89	1.27
SRCNN <sup>[17]</sup>	2.48	3.05	3.71	5.28	1.03	1.26	1.58	2.30	0.81	1.03	1.23	1.84
Huang <sup>[21]</sup>	0.66	/	1.59	2.71	0.54	/	0.83	1.19	0.52	/	0.86	1.21
MSG <sup>[25]</sup>	0.66	/	1.47	2.46	0.37	/	0.67	1.03	0.36	/	0.66	1.02
RDN-GDE <sup>[26]</sup>	0.56	/	1.47	2.60	0.36	/	0.62	1.00	0.38	/	0.69	1.05
MFR-SR <sup>[27]</sup>	0.71	/	1.54	2.71	0.42	/	0.63	1.05	0.42	/	0.72	1.10
PMBA <sup>[28]</sup>	0.61	/	2.04	3.63	0.41	/	0.92	1.68	0.39	/	0.84	1.41
DepthSR <sup>[29]</sup>	0.53	0.89	<b>1.20</b>	<u>2.22</u>	0.42	0.56	0.60	0.89	/	/	/	/
HFFN	<u>0.41</u>	<u>0.84</u>	1.28	2.29	<u>0.28</u>	<u>0.37</u>	<u>0.49</u>	<u>0.87</u>	<u>0.31</u>	<u>0.45</u>	<u>0.57</u>	<u>0.89</u>
HFFN+	<b>0.38</b>	<b>0.81</b>	<u>1.24</u>	<b>2.19</b>	<b>0.27</b>	<b>0.36</b>	<b>0.47</b>	<b>0.84</b>	<b>0.30</b>	<b>0.44</b>	<b>0.55</b>	<b>0.85</b>

越来越大, 验证了 HFFN 算法融合不同尺度下深度-纹理结构相似性的优势. 当比例因子为 4× 时, 各算法在 “Books” 上的 RMSE 值与运行时间如图 5 所示. 由图 5 可知, 综合考虑运行时间和重建质量, HFFN 算法取得了最优的结果.

由表 5 可知, 与 MFR-SR<sup>[27]</sup> 算法相比, HFFN 算法重建 “Laundry” 的 RMSE 值减少了 0.29 (2×)、0.38 (4×)、0.42 (8×), 与 PMBA<sup>[28]</sup> 算法相比, HFFN 算法重建 “Laundry” 的 RMSE 值降低了 0.08 (2×)、0.44 (4×)、0.93 (8×), 与 DepthSR<sup>[29]</sup> 算法相比, HFFN 算法重建 “Laundry” 的 RMSE 值减少了 0.12 (2×)、0.10 (3×)、0.05 (4×); 与 Huang<sup>[21]</sup> 提出的算法相比, HFFN 算法重建 “Reindeer” 的 RMSE 值减少了 0.25 (2×)、0.15 (4×)、1.16 (8×), 与 RDN-

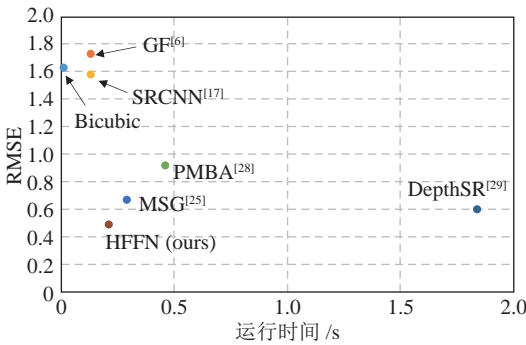


图 5 各算法重建 “Books” 的 RMSE 值走势图  
Fig.5 The trend of the RMSE for “Books”

GDE<sup>[26]</sup> 算法相比, HFFN 算法重建 “Reindeer” 的 RMSE 值降低了 0.17 (2×)、0.25 (4×)、0.47 (8×).

RDN-GDE<sup>[26]</sup>算法、MFR-SR<sup>[27]</sup>算法以及 DepthSR<sup>[29]</sup>算法都是利用金字塔结构获取丰富的层次特征信息, 而 HFFN算法改进了金字塔结构、实现深度-纹理的反馈式融合、有效利用了多尺度特征间的关联性, 其重建结果优于以上对比算法. 同时, 考虑到模型参数以及运行时间的影响, HFFN网络只采用了三层金字塔结构, 而 DepthSR<sup>[29]</sup>算法在 U-Net 结构的基础上进行了四次下采样, 并且与对应尺度下输入金字塔提取的特征进行级联, 相比 HFFN网络获取了更丰富的层次特征, 更有利于大比例因子的重建. 因此, 当比例因子为  $8\times$  时, HFFN网络重建 “Laundry” 和 “Reindeer” 的 RMSE 值相比 DepthSR<sup>[29]</sup>算法增加了 0.02 和 0.07. 由上述实验结果可知, HFFN算法的重建结果总体上优于对比算法, 验证了 HFFN算法在深度图像超分辨率重

建任务中的有效性.

由表 6 可知, HFFN 算法重建 “Cones” 的 RMSE 值均优于比较算法. 与 DepthSR<sup>[29]</sup> 算法相比, HFFN 算法重建 “Teddy” 的 RMSE 值减少了 0.27 ( $2\times$ )、0.22 ( $3\times$ )、0.16 ( $4\times$ ); 与 MSG<sup>[25]</sup> 算法相比, HFFN 算法重建 “Teddy” 的 RMSE 值减少了 0.1 ( $2\times$ )、0.28 ( $4\times$ )、0.49 ( $8\times$ ), 上述结果表明了 HFFN 算法采用反馈式融合策略, 充分融合了不同尺度下的深度特征和纹理特征, 其重建深度图像 “Cones” 和 “Teddy” 上的 RMSE 值优于对比算法. 由表 4 到表 6 可知, HFFN+算法在测试集的重建结果均优于 HFFN 算法. 因此, 采用 Self-Ensemble<sup>[37]</sup>方法可以进一步提升 HFFN 算法的重建质量.

当比例因子为  $4\times$ 、 $8\times$  时, HFFN 与 Bicubic、GF<sup>[6]</sup>、SRCNN<sup>[17]</sup>、DepthSR<sup>[29]</sup>、PMBA<sup>[28]</sup> 算法在测试

表 5 测试数据集 B 的客观对比结果 (RMSE)  
Table 5 Objective comparison results (RMSE) on test dataset B

对比算法	Dolls				Laundry				Reindeer			
	$2\times$	$3\times$	$4\times$	$8\times$	$2\times$	$3\times$	$4\times$	$8\times$	$2\times$	$3\times$	$4\times$	$8\times$
Bicubic	0.94	1.15	1.33	1.87	1.61	2.05	2.39	3.43	1.97	2.46	2.86	4.05
GF <sup>[6]</sup>	1.25	1.31	1.41	1.86	2.21	2.36	2.54	3.42	2.68	2.84	3.05	4.06
TGV <sup>[14]</sup>	1.12	1.21	1.36	1.86	1.99	2.22	2.51	3.76	2.40	2.56	2.71	3.79
JID <sup>[16]</sup>	0.70	0.79	0.92	1.26	0.75	0.94	1.21	2.08	0.92	1.21	1.56	2.58
SRCNN <sup>[17]</sup>	0.90	1.01	1.28	1.82	1.52	1.74	2.31	3.32	1.84	2.17	2.73	3.92
Huang <sup>[21]</sup>	0.58	/	0.91	1.31	0.52	/	0.92	1.52	0.59	/	1.11	1.80
MSG <sup>[25]</sup>	<u>0.35</u>	/	<b>0.69</b>	<b>1.05</b>	0.37	/	0.79	1.51	0.42	/	0.98	1.76
RDN-GDE <sup>[26]</sup>	0.56	/	0.88	1.21	0.48	/	0.96	1.63	0.51	/	1.17	2.05
MFR-SR <sup>[27]</sup>	0.60	/	0.89	1.22	0.61	/	1.11	1.75	0.65	/	1.23	2.06
PMBA <sup>[28]</sup>	0.36	/	0.95	1.47	0.38	/	1.14	2.19	0.40	/	1.39	2.74
DepthSR <sup>[29]</sup>	/	/	/	/	0.44	0.62	0.78	<u>1.31</u>	0.51	0.77	<u>0.96</u>	<b>1.57</b>
HFFN	0.36	<u>0.59</u>	0.75	1.11	<u>0.32</u>	<u>0.52</u>	<u>0.73</u>	1.33	<u>0.35</u>	<u>0.66</u>	<u>0.96</u>	1.64
HFFN+	<b>0.34</b>	<b>0.57</b>	<u>0.74</u>	<u>1.09</u>	<b>0.30</b>	<b>0.51</b>	<b>0.70</b>	<b>1.26</b>	<b>0.34</b>	<b>0.63</b>	<b>0.92</b>	<u>1.58</u>

表 6 测试数据集 C 的客观对比结果 (RMSE)  
Table 6 Objective comparison results (RMSE) on test dataset C

对比算法	Tsukuba				Venus				Teddy				Cones			
	$2\times$	$3\times$	$4\times$	$8\times$	$2\times$	$3\times$	$4\times$	$8\times$	$2\times$	$3\times$	$4\times$	$8\times$	$2\times$	$3\times$	$4\times$	$8\times$
Bicubic	5.81	7.17	8.56	12.3	1.32	1.64	1.91	2.76	1.99	2.48	2.90	4.07	2.45	3.06	3.60	5.30
GF <sup>[6]</sup>	8.12	8.63	9.40	12.5	1.63	1.75	1.93	2.69	2.49	2.67	2.93	3.98	3.33	3.55	3.87	5.29
TGV <sup>[14]</sup>	7.20	7.78	10.3	17.5	2.15	2.34	2.52	4.04	2.71	2.99	3.3	5.39	3.51	3.97	4.45	7.14
JID <sup>[16]</sup>	3.48	4.91	5.95	10.9	0.8	<u>0.91</u>	1.17	1.76	1.28	1.53	2.94	2.76	1.69	2.42	4.17	5.11
SRCNN <sup>[17]</sup>	5.47	6.32	8.11	11.8	1.27	1.43	1.85	2.67	1.88	2.25	2.77	3.95	2.34	2.52	3.43	5.15
Huang <sup>[21]</sup>	1.41	/	3.73	7.79	0.56	/	0.72	1.09	0.85	/	1.58	2.88	0.88	/	2.38	4.66
MSG <sup>[25]</sup>	1.85	/	4.29	8.42	<b>0.14</b>	/	<b>0.35</b>	1.04	0.71	/	1.49	2.76	0.90	/	2.60	4.23
DepthSR <sup>[29]</sup>	1.33	<u>2.25</u>	<u>3.26</u>	<b>6.89</b>	/	/	/	/	0.83	1.15	1.37	<b>1.85</b>	/	/	/	/
HFFN	<u>1.37</u>	2.49	3.53	7.67	0.21	<b>0.28</b>	0.42	<u>0.84</u>	<u>0.61</u>	<u>0.93</u>	<u>1.21</u>	2.27	<u>0.65</u>	<u>1.24</u>	<u>1.71</u>	<u>3.91</u>
HFFN+	<b>1.14</b>	<b>2.22</b>	<b>3.21</b>	<u>7.60</u>	<u>0.20</u>	<b>0.28</b>	<u>0.40</u>	<b>0.78</b>	<b>0.56</b>	<b>0.86</b>	<b>1.13</b>	<u>2.12</u>	<b>0.61</b>	<b>1.14</b>	<b>1.59</b>	<b>3.66</b>



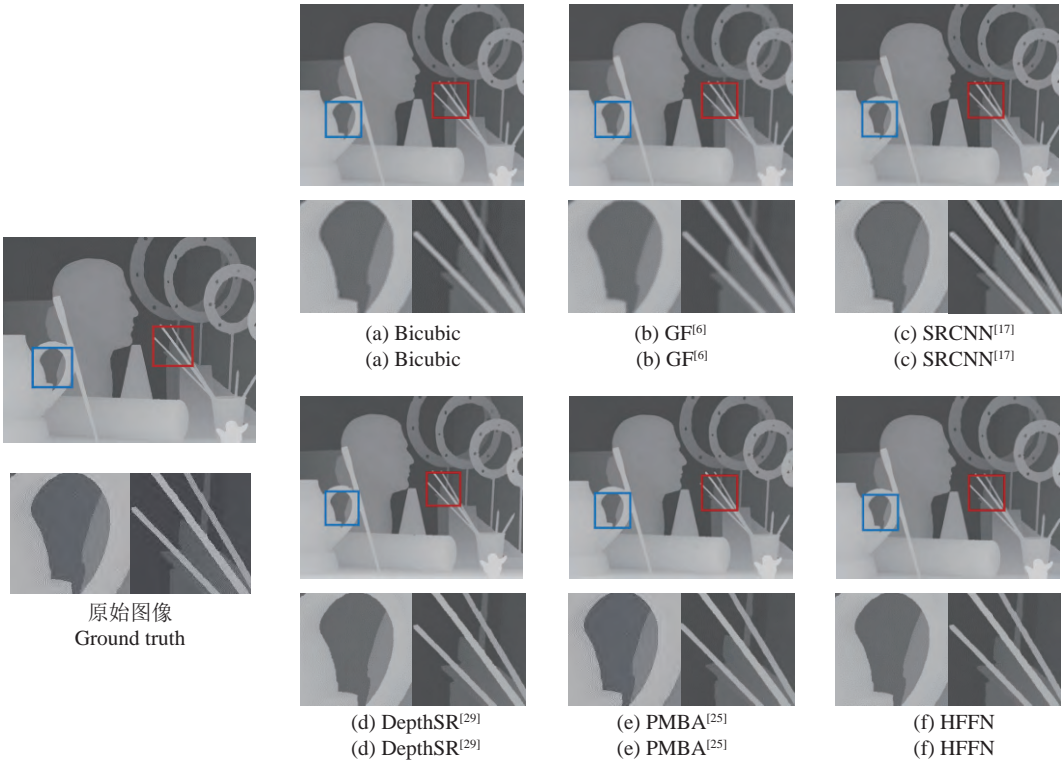


图 6 4×尺度下测试图片“Art”的视觉质量对比结果  
Fig.6 Visual quality comparison results of “Art” at scale 4×

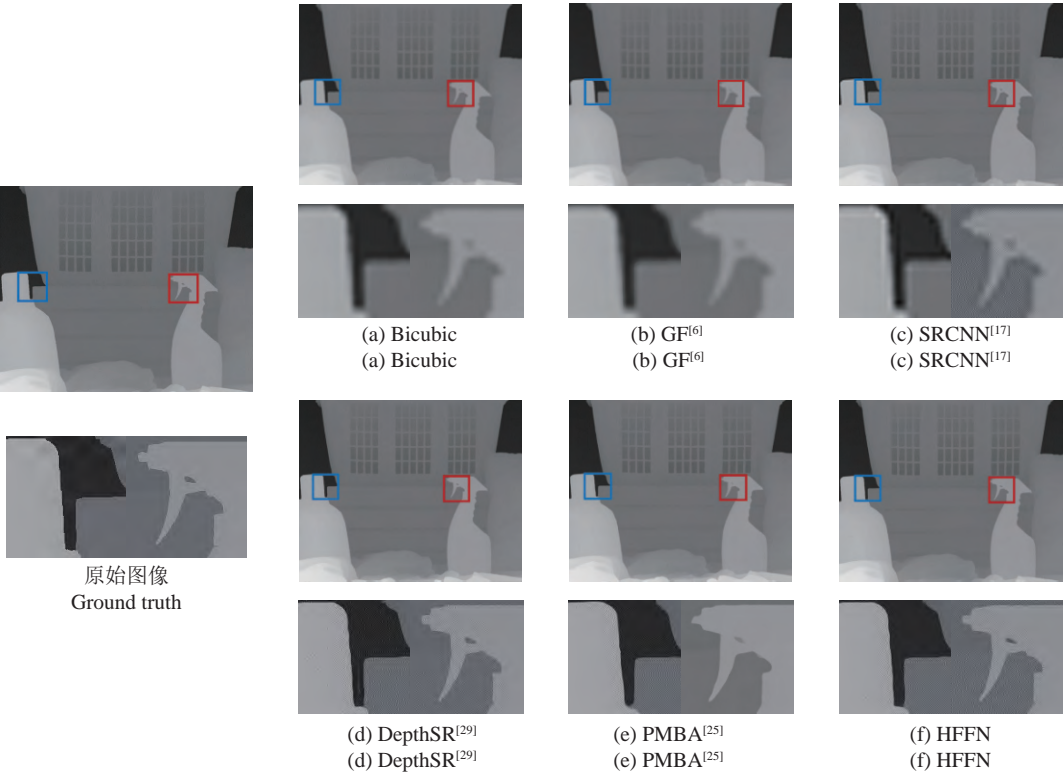


图 7 8×尺度下测试图片“Laundry”的视觉质量对比结果  
Fig.7 Visual quality comparison results of “Laundry” at scale 8×

图像 “Art” 和 “Laundry” 上超分辨率重建的对比结果分别如图 6 和图 7 所示。

由图 6 可知, 与基于滤波的 Bicubic 算法、GF<sup>[6]</sup> 算法以及基于深度学习的单图像超分辨率算法 SRCNN<sup>[17]</sup> 相比, 基于纹理图像引导的 DepthSR<sup>[29]</sup> 算法、PMBA<sup>[28]</sup> 算法和 HFFN 算法在重建过程中融合了纹理图像提供的引导信息, 指引深度图像的重建过程, 其重建结果有着清晰的边缘。

由图 7(d) 框中区域可知, DepthSR<sup>[29]</sup> 算法在深度图像 “Laundry” 上的重建结果出现了原始深度图像上不存在的白色条纹, 可能是由于 DepthSR<sup>[29]</sup> 算法在解码过程中只融合浅层的纹理特征, 纹理图像中包含的纹理信息干扰了深度图像的重建过程; 如图 9(f) 蓝框区域所示, HFFN 算法利用深层网络充分挖掘了纹理图像的结构信息, 并利用反馈式融合策略获得了包含多级感受野的边缘引导信息, 重建结果未出现白色条纹, 避免了纹理图像内部纹理信息的干扰。

### 3.4 计算复杂度分析

本文对比 HFFN 与 Bicubic、GF<sup>[6]</sup>、TGV<sup>[14]</sup>、SRCNN<sup>[17]</sup>、MSG<sup>[25]</sup>、PMBA<sup>[28]</sup>、DepthSR<sup>[29]</sup> 的计算复杂度。在 Intel(R)E5-1620v4, 3.5 GHz, 32 GB RAM、NVIDIA GeForce GTX 2080Ti 12 GB 计算平台上, 各算法将低分辨率深度图像 (264 × 324) 重建为高分辨率深度图像 (1 056 × 1 296) 的平均运行时间, 如表 7 所示。

表 7 平均运行时间  
Table 7 Average running time

对比算法	时间 (s)	对比算法	时间 (s)
Bicubic	0.01	MSG <sup>[25]</sup>	0.29
GF <sup>[6]</sup>	0.13	PMBA <sup>[28]</sup>	0.46
TGV <sup>[14]</sup>	894.43	DepthSR <sup>[29]</sup>	1.84
SRCNN <sup>[17]</sup>	0.13	HFFN	0.21

由表 7 可知, 基于滤波的 Bicubic 算法、GF<sup>[6]</sup> 算法, 运行时间分别为 0.01 s、0.13 s, 相比其他算法运行时间最快、计算复杂度最低; 基于优化表示的 TGV<sup>[14]</sup> 算法在 CPU 上的运行时间达到了 893 s, 时间消耗最大; 得益于 GPU 计算资源的优势, 基于深度学习的 SRCNN<sup>[17]</sup> 算法、MSG<sup>[25]</sup> 算法、DepthSR<sup>[29]</sup> 算法、PMBA<sup>[28]</sup> 算法和 HFFN 算法的运行时间分别为: 0.13 s、0.29 s、1.84 s、0.46 s。在这些网络结构中, SRCNN<sup>[17]</sup> 只使用了 3 层卷积层, 计算复杂度最低; DepthSR<sup>[29]</sup> 利用了 4 层输入金字塔提取不同尺度下的结构特征, 再与 U-Net 结构中对尺度

的特征进行级联, 提取各层的特征信息, 增加了模型参数量、降低了运行速度; HFFN 网络采用了 3 层并行金字塔结构、有效利用深度-纹理多尺度间的相关性, 在提升深度图像重建质量的同时, 控制了模型的参数量和计算复杂度。

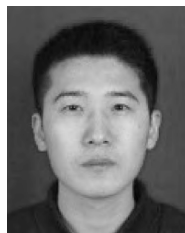
## 4 结论

本文提出了一种基于分级特征反馈融合策略的深度图像超分辨率重建算法, 并构建了端到端的分级特征反馈融合网络。在提取深度-纹理图像的深层特征的基础上, 该算法利用并行金字塔结构构建深度-纹理在不同尺度下的分层特征表示, 提取深度-纹理在不同尺度下的结构信息; 进而, 应用反馈式融合策略, 融合不同尺度下的结构特征, 生成包含多级感受野的边缘引导信息, 引导深度图像的超分辨率重建过程。与其他算法相比, 实验结果验证了所提算法的有效性。

## References

- Kopf J, Cohen M F, Lischinski D, Uyttendaele M. Joint bilateral upsampling. *ACM Transactions on Graphics*, 2007, **26**(3): 96.1–96.4
- Liu M Y, Tuzel O, Taguchi Y. Joint geodesic upsampling of depth images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA: IEEE, 2013: 169–176
- Lei J, Li L, Yue H, Wu F, Ling N, Hou C. Depth map super-resolution considering view synthesis quality. *IEEE Transactions on Image Processing*, 2017, **26**(4): 1732–1745
- Yang Yu-Xiang, Zeng Yu, He Zhi-Wei, Gao Ming-Yu. Depth map super-resolution via adaptive weighting filter. *Journal of Image and Graphics*, 2014, **19**(8): 1210–1218  
(杨宇翔, 曾毓, 何志伟, 高明煜. 基于自适应权重滤波的深度图像超分辨率重建. *中国图象图形学报*, 2014, **19**(8): 1210–1218)
- Lu J, Forsyth D. Sparse depth super resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA: IEEE, 2015: 2245–2253
- He K, Sun J, Tang X. Guided image filtering. In: *Proceedings of the European Conference on Computer Vision*. Heraklion, Greece: Springer, 2010: 1–14
- Barron J T, Poole B. The fast bilateral solver. In: *Proceedings of the European Conference on Computer Vision*. Amsterdam, The Netherlands: Springer, 2016: 617–632
- Yang Y, Lee H S, Oh B T. Depth map upsampling with a confidence-based joint guided filter. *Signal Processing: Image Communication*, 2019, **77**: 40–48
- Diebel J, Thrun S. An application of Markov random fields to range sensing. In: *Proceedings of the Advances in Neural Information Processing Systems*. Vancouver, Canada: NIPS, 2006: 291–298
- An Yao-Zu, Lu Yao, Zhao Hong. An adaptive-regularized image super-resolution. *Acta Automatica Sinica*, 2012, **38**(4): 601–608  
(安耀祖, 陆耀, 赵红. 一种自适应正则化的图像超分辨率算法. *自动化学报*, 2012, **38**(4): 601–608)

- 11 Park J, Kim H, Tai Y W, Brown M S, Kweon I. High quality depth map upsampling for 3D-TOF cameras. In: Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 1623–1630
- 12 Yan Xu-Le, An Ping, Zheng Shuai, Zuo Yi-Fan, Shen Li-Quan. Super-resolution reconstruction for depth map based on edge enhancement. *Journal of Optoelectronics-Laser*, 2016, **27**(4): 437–447  
(严徐乐, 安平, 郑帅, 左一帆, 沈礼权. 基于边缘增强的深度图超分辨率重建. 光电子·激光, 2016, **27**(4): 437–447)
- 13 Yang J, Ye X, Li K, Hou C, Wang Y. Color-guided depth recovery from RGB-D data using an adaptive autoregressive model. *IEEE Transactions on Image Processing*, 2014, **23**(8): 3443–3458
- 14 Ferstl D, Reinbacher C, Ranftl R, Ruether M, Bichof H. Image guided depth upsampling using anisotropic total generalized variation. In: Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 993–1000
- 15 Li Y, Min D, Do M N, Lu J. Fast guided global interpolation for depth and motion. In: Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 717–733
- 16 Kiechle M, Hawe S, Kleinstaub M. A joint intensity and depth co-sparse analysis model for depth map super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 1545–1552
- 17 Dong C, Loy C C, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: Proceedings of the European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 184–199
- 18 Shi W, Caballero J, Huszár F, Totz J, Aitken A P, Bishop R et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 1874–1883
- 19 Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE, 2017: 136–144
- 20 Zhou Deng-Wen, Zhao Li-Juan, Duan Ran, Chai Xiao-Liang. Image super-resolution based on recursive residual networks. *Acta Automatica sinica*, 2019, **45**(6): 1157–1165  
(周登文, 赵丽娟, 段然, 柴晓亮. 基于递归残差网络的图像超分辨率重建. 自动化学报, 2019, **45**(6): 1157–1165)
- 21 Zhang Yi-Feng, Liu Yuan, Jiang Cheng, Cheng Xu. A curriculum learning approach for single image super resolution. *Acta Automatica Sinica*, 2020, **46**(2): 274–282  
(张毅锋, 刘袁, 蒋程, 程旭. 用于超分辨率重建的深度网络递进学习方法. 自动化学报, 2020, **46**(2): 274–282)
- 22 Huang L, Zhang J, Zuo Y, Wu Q. Pyramid-structured depth map super-resolution based on deep dense-residual network. *IEEE Signal Processing Letters*, 2019, **26**(12): 1723–1727
- 23 Song X, Dai Y, Zhou D, Liu L, Li W, Li H et al. Channel attention based iterative residual learning for depth map super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 5631–5640
- 24 Zhao L, Bai H, Liang J, Zeng B, Wang A, Zhao Y. Simultaneous color-depth super-resolution with conditional generative adversarial networks. *Pattern Recognition*, 2019, **88**: 356–369
- 25 Hui T W, Loy C C, Tang X. Depth map super-resolution by deep multi-scale guidance. In: Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 353–369
- 26 Zuo Y, Fang Y, Yang Y, Shang X, Wang B. Residual dense network for intensity-guided depth map enhancement. *Information Sciences*, 2019, **495**: 52–64
- 27 Zuo Y, Wu Q, Fang Y, An P, Huang L, Chen Z. Multi-scale frequency reconstruction for guided depth map super-resolution via deep residual network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, **30**(2): 297–306
- 28 Ye X, Sun B, Wang Z, Yang J, Xu R, Li H, Li B. PMBANet: Progressive multi-branch aggregation network for scene depth super-resolution. *IEEE Transactions on Image Processing*, 2020, **29**: 7427–7442
- 29 Guo C, Li C, Guo J, Cong R, Fu H, Han P. Hierarchical features driven residual learning for depth map super-resolution. *IEEE Transactions on Image Processing*, 2018, **28**(5): 2545–2557
- 30 Kingma D P, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA, 2015: 1–15
- 31 Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 2002, **47**(1-3): 7–42
- 32 Scharstein D, Pal C. Learning conditional random fields for stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA: IEEE, 2007: 1–8
- 33 Scharstein D, Hirschmüller H, Kitajima Y, Krathwohl G, Nesić N, Wang X, et al. High-resolution stereo datasets with subpixel-accurate ground truth. In: Proceedings of the German Conference on Pattern Recognition. Münster, Germany: Springer, 2014: 31–42
- 34 Butler D J, Wulff J, Stanley G B, Black M J. A naturalistic open source movie for optical flow evaluation. In: Proceedings of the European Conference on Computer Vision. Florence, Italy: Springer, 2012: 611–625
- 35 Scharstein D, Szeliski R. High-accuracy stereo depth maps using structured light. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, USA: IEEE, 2003, 1: I–I
- 36 Hirschmüller H, Scharstein D. Evaluation of cost functions for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007: 1–8
- 37 Timofte R, Rothe R, Van Gool L. Seven ways to improve example-based single image super resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 1865–1873

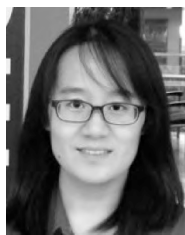


张帅勇 北京交通大学信息科学研究所硕士研究生。主要研究方向为图像/视频超分辨率重建。

E-mail: 19125150@bjtu.edu.cn

(ZHANG Shuai-Yong Master student at Institute of Information Science, Beijing Jiaotong University.

His research interest covers image and video super resolution.)



**刘美琴** 北京交通大学信息科学研究所副教授. 主要研究方向为图像/视频编码, 三维视频处理. 本文通信作者. E-mail: mqliu@bjtu.edu.cn

**(LIU Mei-Qin** Assistant professor at Institute of Information Science, Beijing Jiaotong University. Her re-

search interest covers image and video compression, three-dimension video processing. Corresponding author of this paper.)



**姚超** 北京科技大学计算机与通信工程学院助理教授. 主要研究方向为图像和视频处理, 计算机视觉.

E-mail: yaochao@ustb.edu.cn

**(YAO Chao** Assistant professor at School of Computer and Communication Engineering, University of

Science and Technology Beijing. His research interest covers image and video processing, computer vision.)



**林春雨** 北京交通大学信息科学研究所教授. 主要研究方向为图像视频编码, 三维视觉处理.

E-mail: cylin@bjtu.edu.cn

**(LIN Chun-Yu** Professor at Institute of Information Science, Beijing Jiaotong University. His research

interest covers image and video compression, three-dimension vision processing.)



**赵耀** 北京交通大学信息科学研究所教授. 主要研究方向为图像编码, 数字水印, 多媒体信息处理.

E-mail: yzhao@bjtu.edu.cn

**(ZHAO Yao** Professor at Institute of Information Science, Beijing Jiaotong University. His research in-

terest covers image compression, digital watermarking, and multimedia information processing.)