



计算机科学

COMPUTER SCIENCE

基于彩色图像高频信息引导的深度图超分辨率重建算法研究

李嘉莹, 梁宇栋, 李少吉, 张昆鹏, 张超

引用本文

李嘉莹, 梁宇栋, 李少吉, 张昆鹏, 张超. 基于彩色图像高频信息引导的深度图超分辨率重建算法研究[J]. 计算机科学, 2024, 51(7): 197-205.

LI Jiaying, LIANG Yudong, LI Shaoji, ZHANG Kunpeng, ZHANG Chao. [Study on Algorithm of Depth Image Super-resolution Guided by High-frequency Information of Color Images](#) [J]. Computer Science, 2024, 51(7): 197-205.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[自编码器端到端通信系统后门攻击方法](#)

Backdoor Attack Method in Autoencoder End-to-End Communication System

计算机科学, 2024, 51(7): 413-421. <https://doi.org/10.11896/jsjcx.230400113>

[针对系统调用的基于语义特征的多方面信息融合的主机异常检测框架](#)

Host Anomaly Detection Framework Based on Multifaceted Information Fusion of Semantic Features for System Calls

计算机科学, 2024, 51(7): 380-388. <https://doi.org/10.11896/jsjcx.230400023>

[Deep-Init:基于深度学习的视觉惯性里程计非联合初始化方法](#)

Deep-Init: Non Joint Initialization Method for Visual Inertial Odometry Based on Deep Learning

计算机科学, 2024, 51(7): 327-336. <https://doi.org/10.11896/jsjcx.230500036>

[基于外部先验和自先验注意力的图像描述生成方法](#)

Image Captioning Generation Method Based on External Prior and Self-prior Attention

计算机科学, 2024, 51(7): 214-220. <https://doi.org/10.11896/jsjcx.230600167>

[基于特征重要性的深度学习自动调度优化研究](#)

Study on Deep Learning Automatic Scheduling Optimization Based on Feature Importance

计算机科学, 2024, 51(7): 22-28. <https://doi.org/10.11896/jsjcx.230500220>

基于彩色图像高频信息引导的深度图超分辨率重建算法研究

李嘉莹¹ 梁宇栋^{1,2} 李少吉¹ 张昆鹏¹ 张超^{1,2}

1 山西大学计算机与信息技术学院 太原 030006

2 山西大学计算智能与中文信息处理教育部重点实验室 太原 030006

(202122407023@email.sxu.edu.cn)

摘要 深度图像信息是三维场景信息的重要组成部分,然而,由于采集设备的局限性和成像环境的多样性,深度传感器获取的深度图像往往分辨率较低、高频信息较少,限制了其在各种计算机视觉任务中的进一步应用。深度图超分辨率试图提高深度图的分辨率,是一项实用而有价值的任务。同一场景下的 RGB 图像分辨率高,纹理信息丰富,部分深度图超分辨率算法通过引入来自同一场景下的 RGB 图像提供指导信息,实现了算法性能的显著提升。然而,由于 RGB 图像和深度图之间的模态不一致,如何充分、有效地利用 RGB 信息辅助深度图像进行图像超分辨率重建仍然极具挑战。为此,提出了一种基于彩色图像高频信息引导的深度图超分辨率重建算法。具体地,设计了一个高频特征提取模块来自适应地学习彩色图像中的高频信息,以指导深度图边缘的重建。另外,设计了一个特征自注意力模块来获取特征之间的全局依赖,同时提取更深层次的特征,以帮助深度图细节信息的恢复。经过跨模态融合,重组深度图像特征和彩色图像引导特征,并使用多尺度特征融合模块融合不同尺度特征之间的空间结构信息,获取包含多级感受野的重建信息。最后,通过深度重建模块,恢复相应的高分辨率深度图。公开数据集上的实验结果表明所提方法在定量和定性两方面均优于对比方法,验证了所提方法的有效性。

关键词: 深度图超重建;深度学习;跨模态特征融合;高频信息;自注意力机制

中图分类号 TP391

Study on Algorithm of Depth Image Super-resolution Guided by High-frequency Information of Color Images

LI Jiaying¹, LIANG Yudong^{1,2}, LI Shaoji¹, ZHANG Kunpeng¹ and ZHANG Chao^{1,2}

1 School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

2 Key Laboratory of Ministry of Education for Computation Intelligence and Chinese Information Processing, Shanxi University, Taiyuan 030006, China

Abstract Depth image information is an important part of 3D scene information. However, due to the limitations of acquisition equipment and the diversity of imaging environments, the depth images acquired by depth sensors often have low resolution and less high-frequency information, which limits their further applications in various computer vision tasks. Depth image super-resolution attempts to improve the resolution of depth images and is a practical and valuable task. The RGB image in the same scene has high resolution and rich texture information, and some depth image super-resolution algorithms achieve significant improvement in algorithm performance by introducing RGB images from the same scene to provide guidance information. However, due to the structural inconsistency between RGB images and depth maps, how to utilize RGB information fully and effectively is still extremely challenging. To this end, this paper proposes a depth image super-resolution guided by high-frequency information of color images. Specifically, a high-frequency feature extraction module is designed to adaptively learn high-frequency information of color images to guide the reconstruction of depth map edges. In addition, a feature self-attention module is designed to capture the global dependencies between features, extract deeper features to help recover details in the depth image. After cross-modal fusion,

到稿日期:2023-04-16 返修日期:2023-09-21

基金项目:国家自然科学基金(61802237, 62272284);山西省基础研究计划项目(202203021221002, 202203021211291);山西省自然科学基金(201901D211176, 202103021223464);山西省高等学校科技创新项目(2019L0066);山西省科技重大专项计划(202101020101019);山西省重点研发计划(202102070301019);山西省科技创新青年人才团队项目(202204051001015)

This work was supported by the National Natural Science Foundation of China(61802237, 62272284), Fundamental Research Program of Shanxi Province(202203021221002, 202203021211291), Natural Science Foundation of Shanxi Province, China(201901D211176, 202103021223464), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi(2019L0066), Science and Technology Major Project of Shanxi Province, China(202101020101019), Key R & D Program of Shanxi Province(202102070301019) and Special Fund for Science and Technology Innovation Teams of Shanxi(202204051001015).

通信作者:梁宇栋(liangyudong@sxu.edu.cn)

the depth image features and color image-guided features are reconstructed, and the proposed multi-scale feature fusion module is used to fuse the spatial structure information between different scale features to obtain reconstruction information including multi-level receptive fields. Finally, through the depth reconstruction module, the corresponding high-resolution depth map is recovered. Comprehensive qualitative and quantitative experimental results on public datasets have demonstrated that the proposed method outperforms comparative methods, which verifies its effectiveness.

Keywords Depth image super-resolution reconstruction, Deep learning, Cross-modal fusion, High-frequency information, Self-attention mechanism

1 引言

随着深度传感器技术的不断发展,场景深度信息的获取更加方便,深度图像作为三维场景信息的普遍表达方式,在社会生活的各个领域都得到了广泛的应用,如自动驾驶辅助、虚拟现实、三维重建、航空航天以及医学成像等。然而,由于采集设备的局限性和成像环境的多样性,深度图像往往易受噪声干扰,存在分辨率低、边缘失真、高频信息较少等问题,限制了深度图像的广泛应用。因此,如何提高深度图像的分辨率成为计算机视觉领域的重要研究问题。

深度图像超分辨率算法研究的早期阶段主要是采用基于滤波或者插值^[1-2]的方法,利用图像的局部信息,来实现深度图像的超分辨率重建。但是,由于低分辨率深度图像本身所含结构信息较少,这些局部算法的重建效果往往不尽如人意。当放大倍数较大时,深度图像的边缘、细节等结构信息损失较多,导致重建结果出现边缘模糊、锐度消失等问题。

与深度图像相比,同一场景下的彩色图像更容易采集,且具有更高的分辨率、更细致的纹理特征以及更清晰的边缘信息。深度图像和彩色图像是同一场景下不同模态的表达,在边缘结构上具有一定的相似性。因此,研究者们引入高分辨率的彩色图像作为参考图像,提出了基于彩色图像引导的深度图像超分辨率算法,利用彩色图像包含的边缘信息来帮助实现深度图像的超分辨率重建,极大地提升了深度图像超分辨率算法的性能。起初,研究人员使用人工设计的滤波器^[3-4],对深度图与彩色图中对应像素值进行加权平均以得到高分辨率深度图的像素值,这类方法实现简单,但往往会过滤掉大部分高频信息,导致深度图的重建结果出现边缘模糊、纹理伪影等问题。之后,基于优化的深度图像超分辨率算法^[5-6]利用彩色图像与深度图像之间的结构相似性来设计模型的优化函数与正则项,从而约束深度图像边缘的结构信息,进而帮助深度图像的超分辨率重建。另一方面,基于深度学习的超分辨率算法^[7-9]通过多次上-下采样来获取多尺度的彩色图像与深度图像特征,在此基础上,对不同尺度的特征进行逐级融合,以获取彩色图像中的纹理信息,从而引导深度图像的超分辨率重建。近年来,随着 Transformer 在计算机视觉领域的广泛应用,结合 Transformer 的超分辨率算法不断涌现, Yang 等^[10]提出了一种新颖的纹理迁移网络,引入 Self-Attention 的思想,将 LR 图像作为查询向量 Q ,降采样的参考图像为键向量 K ,高分辨率参考图像为值向量 V ,从清晰的参考图像中迁移纹理信息到低分辨率图像中,从而实现低分辨率图像的纹理重建。虽然这些方法有效地提升了深度图像的重建效果,但彩色图像和深度图像在结构上仍然存在很大的不同。

例如,彩色图像包含更多的纹理信息,这就容易导致在引导深度图像超分辨率重建的过程中出现纹理信息的过度迁移,引入不必要的噪声。现有基于彩色图像引导的深度图像超分辨率方法仍面临以下问题:(1)上采样因子较大时,重建结果存在结构与细节信息丢失严重的问题;(2)由彩色图像与深度图像的结构不一致性引起的特征融合问题。因此,需要进一步探索如何更好地提取并融合彩色图像特征来引导深度图像的超分辨率重建过程。

为解决上述问题,本文提出了一种基于彩色图像高频信息引导的深度图超分辨率重建算法,通过学习彩色图像中的结构信息,来弥补低分辨率深度图像中缺失的信息,进而帮助高分辨率深度图的有效重建。为了解决重建深度图边缘缺失、细节模糊等问题,本文构建了高频特征提取模块(High-frequency Feature Extraction Module, HFEM),自适应地突出彩色图像中的高频信息并抑制低频信息,通过高频特征来引导深度图边缘信息的恢复,发掘彩色图像与深度图像之间的结构一致性,同时,抑制低频分量可以降低模型计算量,减少训练时间。为了进一步提升重建效果并恢复丰富的细节信息,本文采用了特征自注意力模块(Feature Self-attention Module, FSM)和多尺度特征融合模块(Multi-scale Feature Fusion Module, MFFM)来增强模型的特征表达能力。特征自注意力模块(FSM)帮助模型建模任意两点像素之间的关系,在深度图像重建过程中引入全局信息,增强了特征表达能力,提升了重建效果。同时,多尺度特征融合模块(MFFM)能够实现跨尺度特征之间的信息交流,并充分融合不同尺度之间的特征,通过分层特征学习更为丰富的信息,使模型更好地适应大比例因子的超分辨率重建任务。

本文的主要贡献如下:

(1)提出了一个基于彩色图像高频信息引导、多尺度跨模态融合的深度图像超分辨率重建算法。文中提出了一种新颖的彩色图像和深度图像间的跨模态融合计算方式,通过特征自注意力计算,捕获高频特征之间的全局依赖关系,并根据彩色图像和深度图像的对应关系,采用彩色图像高频信息计算的注意力权重加权深度图像与彩色图像的融合特征,进一步实现特征之间的跨模态融合。在融合过程中,彩色图像高频信息能够有效引导深度图特征的超分辨率增强,更好地帮助深度图像中细节信息的恢复。

(2)基于 Octave 卷积^[11]构建了一个高频引导分支来有效提取彩色图像的高频信息,引导深度图超分辨率的映射过程,有效促进了高分辨率深度图的高频信息重建,提升了模型计算效率。

(3)在算法的多尺度特征融合模块中,3个尺度的跨模态

融合特征可以进一步充分融合,以获取更丰富的层次结构信息,帮助模型在大比例放大因子的超分辨率重建过程中更好地恢复细节信息。

实验表明,本文算法能够显著提升深度图像的超重建效果,同时更有效地利用彩色图像的结构信息实现对深度图像超分辨率重建任务的指导。

2 相关工作

目前深度图像超分辨率重建方法大致分为两类:基于单幅图像的深度图超分辨率重建和基于彩色图像引导的深度图超分辨率重建。

2.1 基于单幅图像的深度图超分辨率重建

目前基于单幅图像的深度图超分辨率重建方法主要分为两类:基于传统方法的超分辨率重建方法和基于学习的超分辨率重建方法。

基于传统的超分辨率重建方法借助传统的图像处理算法对低分辨率深度图像进行重建。Xie等^[12]提出了一种单幅深度图像超分辨率模型,通过基于 patch 合成的马尔可夫随机优化模型从低分辨率深度图像的边缘构建高分辨率边缘引导图,在高分辨率边缘图的引导下,通过改进的联合双边滤波器对低分辨率深度图像进行上采样。该模型可以有效地避免引入纹理伪影并保留尖锐的边缘。

基于学习的超分辨率重建方法通过大量的数据来学习低分辨率图像与高分辨率图像之间的映射关系,构建非线性模型来对低分辨率图像进行重建。Ferstl等^[18]通过从高分辨率和低分辨率图像对中学习边缘先验字典,利用稀疏编码估计边缘先验知识,进而实现对低分辨率深度图像的重建。但该方法需要对数据集进行预处理,且图像之间的映射关系难以准确建立,导致重建结果出现模糊和伪影等问题。随着深度学习的发展,基于深度学习的超分辨率算法获得了广泛的关注,面向深度图像的超分辨率算法也得到了较大的发展。Dong等^[14]提出的 SRCNN 通过三层卷积神经网络结构实现端到端的网络训练,获取低分辨率与高分辨率图像之间映射关系,用于超分辨率重建。随后,Dong等^[15]又提出了 FSRCNN,对 SRCNN 进行了改进,增加了卷积神经网络的层数,同时减小了卷积核尺寸。受 ResNet^[16]的启发,Kim等^[17]设计了一个更深的网络 VDSR,同时采用残差结构,学习高分辨率图像和低分辨率图像之间的高频部分。此外,Kim等^[18]还提出了一种基于深度递归卷积网络的超分辨率模型 DRCN,加深网络层数,增大感受野,采用跨层连接降低训练难度。2016年,Riegler等^[19]通过展开原始-对偶算法的优化步骤,将具有各向异性 TGV 正则化的变分模型集成到深度学习网络中,并基于合成数据端到端地训练整个模型,从而获得单幅深度图像超分辨率结果。2019年,Song等^[20]提出将深度图超分辨率任务表示为一系列视图合成子任务,训练阶段每个子任务可以并行执行,并通过端到端的学习方式来求解,最后采用多尺度融合策略获得深度图像的超分辨率重建结果。随后,为了解决低分辨率深度图在现实世界中的退化问题,Song等^[21]提出将真实世界 LR 深度图像的噪声进行分类,并提出了一种新的迭代残差学习模型来解决不同

深度传感器的真实退化问题。

由于深度图像本身所含信息较少,且基于单幅图像的超分辨率算法多针对彩色图像的超分辨率重建任务进行设计,因此直接将单图像超分模型应用于深度图像可能达不到很好的效果。相比之下,彩色图像引导的深度图像超分辨率方法可以挖掘彩色图像和低分辨率深度图之间一致的结构特征,从而有效地帮助深度图像中几何结构信息的重建。

2.2 基于彩色图像引导的深度图超分辨率重建

基于彩色图像引导的深度图超分辨率重建方法主要分为3类:基于滤波、基于优化和基于深度学习的超分辨率重建算法。

基于滤波的深度图超分辨率重建算法主要是在重建过程中利用滤波器来融合彩色图像中与深度图像对应区域相关的结构信息,实现彩色图像信息到深度图像的迁移,从而帮助深度图的超分辨率重建。Kopf等^[3]基于联合双边滤波器,提出了联合双边上采样方法,引入高清彩色图像作为先验知识,将给定的引导图权值与低分辨率深度图像的结构信息相结合,得到高分辨率重建图像。但当上采样因子较大时,超分辨率重建图像会出现边缘模糊现象。Yang等^[4]利用双边滤波对低分辨率图像上采样得到的视差图进行代价体优化,再进行亚像素精度提升,从而实现深度图像的超分辨率重建。He等^[2]提出了一个能够保持边缘的引导滤波算法,将彩色图像作为引导图,引导滤波可以帮助算法重建保持边缘的深度图像。然而这类方法在上采样过程中可能会滤掉深度图像中大量的高频信息,导致重建结果出现边缘模糊或纹理复制等问题。

基于优化的深度图超重建算法将超重建任务看作凸优化问题,获取彩色图像与深度图像之间的相关结构信息,借助优化函数与正则化项来约束深度图的上采样过程。2005年,Diebel等^[22]首次在多模态数据集中应用马尔可夫随机场 MRF 公式来填补分辨率之间的差距,但这种方法的重建结果过于平滑。Park等^[5]基于高分辨率彩色图像的若干图像特征,结合多个加权因子和非局部均值滤波来保持低分辨率深度图像重建结果中深度边界的清晰。Ferstl等^[6]使用更高阶正则化约束对深度图像上采样构造凸优化问题,使用从高分辨率彩色图像计算出的各向同性扩散张量来指导深度图的上采样过程。Gu等^[23]提出了一种加权分析表示模型,并利用任务驱动的训练策略来学习任务中的参数,从而引导高质量深度图像的重建。基于优化的方法充分考虑了深度图与彩色图像之间的相似性,但由于其算法复杂度较高,手工设计的函数不具有普适性,对重建的结果影响较大。

随着卷积神经网络的发展,基于深度学习的超分辨率重建算法取得了显著的成就。Li等^[24]通过学习的方法来构造基于卷积神经网络的联合滤波器,设计了两个子网络,分别提取深度图像和彩色图像的信息特征,然后这些特征被选择性地从彩色图像转移到深度图像,同时抑制两个图像中不一致的结构信息,最后得到超分辨率重建图像。而 DJFR^[25]在其基础上进行了改进,直接学习彩色图像与深度图像两者之间的高频残差。Su等^[26]提出了一种像素自适应卷积(PAC)操作,采用双边滤波器的思想来扩展标准卷积操作,其中滤波器

权重与可学习的局部像素特征的空间变化内核相乘,可以有效地学习和利用彩色图像引导信息,从而帮助深度图的恢复。

Hui 等^[27]提出了一个新的框架,通过多尺度引导卷积网络(MSGNet),LR 深度图在彩色图像的引导下逐级上采样,缓解了深度图在上采样中模糊的问题。Guo 等^[9]基于 U-Net 网络结构,通过分层特征驱动的残差学习从低分辨率图像推理出高分辨率深度图。Lutio 等^[28]提供了另外一种解释,将引导超分辨率任务视为彩色引导图像到深度图图像域之间像素到像素的映射。Kim 等^[29]设计了一个新的图像滤波器,对于每一个像素点,学习一个 3×3 大小的权重 *weight* 和偏置 *offset*,根据 *offset* 采样,然后与 *weight* 相乘,得到每个点的预测值。He 等^[30]首先构建了一个“RGB-D”数据集,此外,还提供了一种轻量深度图超分辨率(FDSR)模型,从彩色图像中自适应分解高频分量来指导深度图的重建。Sun 等^[31]首次尝试在训练阶段从彩色图像与深度图像中学习跨模态知识,但测试仅在深度图像数据集上进行,通过一个跨模态交互模块,实现模态之间的知识迁移。Deng 等^[32]提出了 CU-Net,使用两个模块通过多模态卷积稀疏编码分离公共、唯一特征,并详细说明了模型的可解释性。Zhao 等^[33]提出了一种新的基于彩色图像引导的深度图像超分辨率模型 DCT-Net,该模型基于离散余弦变换、半耦合特征提取和边缘注意力,以缓解彩色图像纹理过度转移、无效跨模态特征提取等问题。Mallick 等^[34]通过结合 CNN 与 Transformer 编码器的优势,以高效的局部图像注意力来解决基于引导的深度图像超分辨率重建任务。Dong 等^[35]提出了一种基于空间变异线性表示模型(SVLRM)的联合滤波方法,利用 CNN 来估计空间变异线性表示系数,对引导图像和目标图像的结构信息进行建模,

从而实现将有意义的结构信息转移到目标图像中。Zhong 等^[36]提出了一种深度注意力引导的图像滤波框架,通过学习引导图像和目标图像生成的两组滤波器,并建模两者之间的像素依赖关系进而整合互补信息,同时采用多尺度滤波和多尺度融合策略由粗到细地生成重建结果。Yang 等^[37]提出了一种基于跨域注意力的方法(CODON)来解决深度图超分问题,使用递归多尺度卷积模块和跨域注意力模块,在彩色图像与深度图像中高效地提取局部结构并通过建模二者之间的共有特征来引导模型注意力,继而逐步实现跨域特征的融合,用于重构高分辨率深度图像。Zhou 等^[38]提出了一种新颖的循环逐尺度特征融合网络,从不同模态图像本身特征出发来融合多模态特征。

尽管基于深度学习的方法已经取得了突破性的成果,但不恰当地引入彩色图像特征会导致纹理伪影等问题。因此,如何有效地融合彩色图像与深度图像特征仍是一个值得探索的问题。

3 本文工作

本文提出的基于彩色图像高频信息引导的深度图超分辨率模型如图 1 所示,包括 3 部分,分别是高频特征提取模块(High-frequency Feature Extraction Module, HFEM)、特征自注意力模块(Feature Self-attention Module, FSM),以及多尺度特征融合模块(Multi-scale Feature Fusion Module, MFFM)。给定输入图像即低分辨率深度图 LR,以及对应的高分辨率彩色参考图像 Ref,将参考图像 Ref 输入高频特征提取模块,获取对应的高频分量,用作深度图超分辨率重建的早期特征指导。

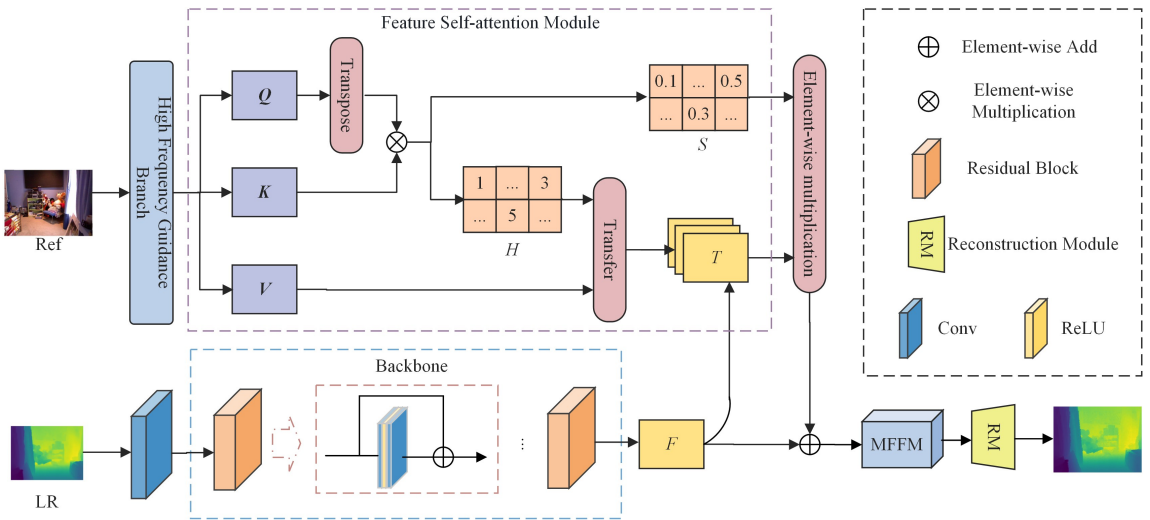


图 1 基于彩色图像高频信息引导的深度图超分辨率重建模型

Fig. 1 Depth super-resolution guided by high-frequency information of color images

在此基础上,将获取的高频特征输入特征自注意力模块,学习特征之间的相关性,使相似特征相互关联,从而捕获全局特征,恢复更多的细节信息,并将其作为深度图像重建的引导特征。根据彩色图像和深度图像的对应关系,算法采用彩色图像提取的引导特征计算注意力权重,对低分辨率深度图像 LR 的特征与引导特征融合后的特征进行加权,作为深度图像重建的特征信息。此外,算法利用多尺度特征融合模块对

多个尺度的特征信息进行集成,以在不同尺度的特征之间交换信息,将不同尺度的特征融合到重建图像中,更好地恢复图像边缘信息。最后,经过超分辨率重建模块得到高分辨率深度图像的重建结果。接下来将介绍各个模块以及损失函数。

3.1 高频特征提取模块

高分辨率的彩色图像相比深度图更容易采集,且包含更丰富的图像信息。二者作为同一场景下不同模态的表示,

理论上具有一定的结构相似性。例如,深度图中的边缘在彩色图像中也会存在,故可以利用彩色图像的清晰边缘来帮助深度图像的边缘信息重建。但彩色图像包含更多的纹理信息,而深度图中并不包含这些复杂的纹理结构,不恰当的引导方法会导致纹理伪影等问题。

为了更好地帮助深度图恢复边缘信息,本文构建了高频特征提取模块(HFEM),其中包含3个高频提取层(High-frequency layer, HFL),图1中的高频引导分支(High-Frequency Guidance Branch)则是由卷积层与HFEM模块组成。如图2所示,高频引导分支分别对彩色图像的3个尺度进行高频特征提取。本文利用Octave Convolution^[11]将彩色图像特征分解为高频和低频分量,以自适应地突出高频分量并抑制低频分量,高频分量被有效地用于引导深度图的超分辨率重建。模型提取有用的高频细节信息来提高性能,同时由于其未使用低频分量,因此能够降低模型复杂度,减小模型计算量。如图3所示,上一阶段的高频和低频分量经过高频提取层(HFL)以生成当前阶段的高低频分量,经过3次高频特征提取,将最后阶段的高频分量作为特征自注意力模块(FSM)的输入,公式如下:

$$Y_{i+1}^H = f(Y_i^H; W_i^{H \rightarrow H}) + up(f(Y_i^L; W_i^{L \rightarrow H}), s) \quad (1)$$

$$Y_{i+1}^L = f(Y_i^L; W_i^{L \rightarrow L}) + f(down(Y_i^H; s); W_i^{H \rightarrow L}) \quad (2)$$

式(1)表示当前阶段的高频特征 Y_{i+1}^H 是由上一阶段的高频特征 Y_i^H 及低频特征 Y_i^L 分别经过卷积操作及上采样操作的结果相加得到;式(2)表示当前阶段的低频特征 Y_{i+1}^L 是由上一阶段的低频特征 Y_i^L 经过卷积操作的结果与上一阶段的高频特征 Y_i^H 经过下采样操作的结果相加得到。其中, W 代表卷积核, $f(Y; W)$ 为卷积操作, $up(Y; s)$ 为最近邻插值的上采样操作,而 $down(Y; s)$ 是使用平均池化的下采样操作, s 代表放大或缩小因子。

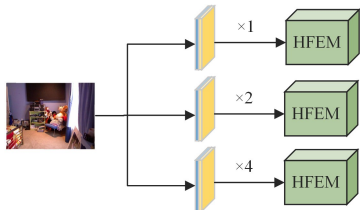


图2 高频引导分支

Fig. 2 High-frequency guidance branch

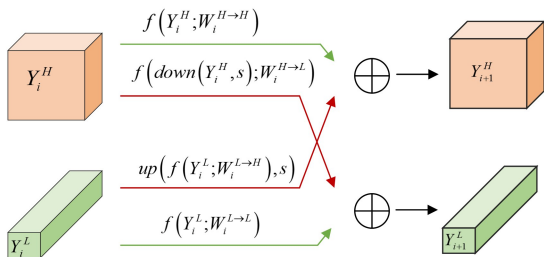


图3 高频提取层结构图

Fig. 3 Structure of high-frequency extraction layer

3.2 特征自注意力模块

受文献^[10]的启发,本文在网络中采用了特征自注意力模块(FSM)来捕获特征之间的全局依赖关系,提取更深层次

的特征,增强特征表达能力,同时更好地帮助深度图像中细节信息的恢复。

该模块的输入为HFEM的输出特征图,如图1所示。首先,利用unfold操作将高频特征展开为patches,得到查询特征 Q 、键特征 K 和值特征 V ,然后将经过转置的 Q 与 K 做自注意力计算,建模彩色图像特征中任意两个像素之间的关系,关联度计算式如式(3)所示。自注意力计算的查询特征 Q 和键特征 K 分别表示为 $q_i (i \in [1, H_x \times W_x])$ 和 $k_j (j \in [1, H_x \times W_x])$ 。对于 Q 中的每个 q_i 和 K 中的每个 k_j ,通过归一化内积来计算两者之间的相关性 $r_{i,j}$, $r_{i,j}$ 用于进一步获取相关位置引导图 H 以及权值引导图 S 。

$$r_{i,j} = \left\langle \frac{q_i}{\|q_i\|}, \frac{k_j}{\|k_j\|} \right\rangle \quad (3)$$

具体地,首先计算相关位置引导图 H 对值特征 V 进行选择, H 中的每一个 $h_i (i \in [1, H_x \times W_x])$ 的计算式如下:

$$h_i = \arg \max_j r_{i,j} \quad (4)$$

其中, h_i 可以被看作是一个索引,表示对于 Q 中的第 i 个位置的特征,在 K 中与其相关性最大的特征所在位置。为了从彩色图像中获取迁移的高频引导特征 T ,本文使用相关位置引导图 H 作为索引,对自注意力计算中的值特征 V ,表示为 $v_m (m \in [1, H_x \times W_x])$,进行特征选择操作:

$$t_i = v_{h_i} \quad (5)$$

其中, t_i 表示 T 的第 i 个位置,该值从 V 中的第 h_i 个位置进行选取。最后,获得引导特征表示 T ,它将在后续的跨模态特征融合中使用。

其次,计算权值引导图 S ,用于增强跨模态相关特征。权值引导图 S 表示 T 中每个位置特征的权重值,对于权值引导图 S 中的每一个 $s_i (i \in [1, H_x \times W_x])$:

$$s_i = \max_j r_{i,j} \quad (6)$$

本文并不是直接将权值引导图 S 应用于迁移的高频引导特征 T ,而是首先对 T 与低分辨率深度图特征 F 进行跨模态特征融合,以利用LR深度图中的更多信息。之后,跨模态融合特征与权重引导图 S 进一步做逐像素相乘,并与 F 相加以获得跨模态融合的最终输出。该过程可以表示为:

$$F_{out} = F + Conv(Concat(F, T)) \cdot S \quad (7)$$

其中, F_{out} 表示跨模态融合输出特征, $Conv$ 和 $Concat$ 分别表示卷积层和级联操作,“ \cdot ”表示特征图之间的逐像素相乘。

3.3 多尺度特征融合模块

在特征融合阶段,大多数算法只将当前尺度的彩色图像信息与深度信息进行融合,这将导致部分空间结构信息丢失,模态之间信息融合不彻底等。针对这一问题,本文提出了多尺度特征融合模块(MFFM),在不同尺度的特征之间交换信息,充分融合多个尺度的特征,以获取更丰富的层次结构信息。

多尺度特征融合结构如图4所示,3个分支代表3个不同尺度的跨模态融合特征。经过跨模态融合得到3个分辨率尺度(对应不同放大因子 $1 \times, 2 \times$ 和 $4 \times$)的融合特征 F_{out} ,实现了不同尺度的高频特征与LR特征图之间的融合。当LR特征图上采样到下一个尺度时,都会使用MFFM模块。

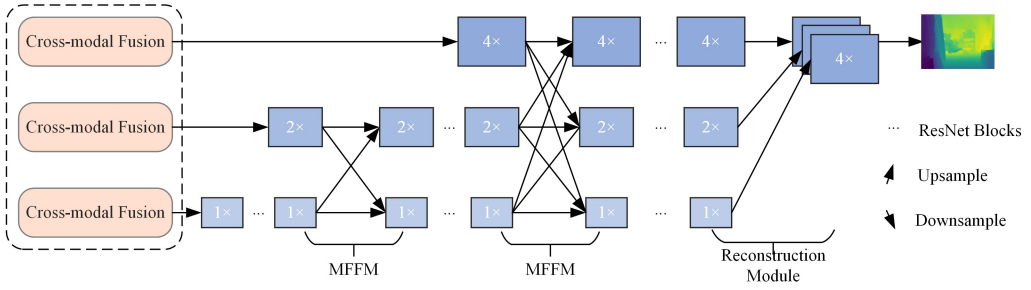


图4 多尺度特征融合结构图

Fig. 4 Architecture of multi-scale feature fusion module(MFFM)

首先,将由式(7)跨模态融合得到的融合特征 \mathbf{F}_{out} 输入残差块进行特征变换,用于挖掘融合特征内部的边缘结构信息。其次,通过上采样(双三次插值)、下采样(卷积操作)缩放不同尺度的融合特征。然后,同尺度特征在通道维度进行级联操作,再通过卷积层将特征映射到原始通道数。具体如下(以 $1\times, 2\times$ 为例):

$$\mathbf{F}^{1\times \rightarrow 2\times} = f(\text{up}(\mathbf{F}_{out}^{1\times}, 2); \mathbf{W}^{12}) \quad (8)$$

$$\widetilde{\mathbf{F}}^{2\times} = f(\text{Concat}(\mathbf{F}_{out}^{2\times}, \mathbf{F}^{1\times \rightarrow 2\times}); \mathbf{W}^2) \quad (9)$$

$$\mathbf{F}^{2\times \rightarrow 1\times} = f_{\text{down}}(\mathbf{F}_{out}^{2\times}; \mathbf{W}^{21}) \quad (10)$$

$$\widetilde{\mathbf{F}}^{1\times} = f(\text{Concat}(\mathbf{F}_{out}^{1\times}, \mathbf{F}^{2\times \rightarrow 1\times}); \mathbf{W}^1) \quad (11)$$

其中, $\mathbf{F}_{out}^{1\times}$ 和 $\mathbf{F}_{out}^{2\times}$ 表示跨模态融合得到的 $1\times$ 和 $2\times$ 尺度特征, $\widetilde{\mathbf{F}}^{1\times}$ 和 $\widetilde{\mathbf{F}}^{2\times}$ 表示经过 MFFM 模块融合后的 $1\times$ 和 $2\times$ 尺度特征, Concat 表示级联操作, \mathbf{W} 代表卷积核, $f(\mathbf{F}_{out}; \mathbf{W})$ 为卷积操作, $\text{up}(\mathbf{F}_{out}; 2)$ 为上采样因子为 2 的双三次插值操作, 而 $f_{\text{down}}(\mathbf{F}_{out}; \mathbf{W})$ 是使用卷积进行 2 倍下采样。

式(8)表示跨模态融合特征 $\mathbf{F}_{out}^{1\times}$ 经过上采样及卷积操作得到特征 $\mathbf{F}^{1\times \rightarrow 2\times}$; 式(10)表示跨模态融合特征 $\mathbf{F}_{out}^{2\times}$ 通过卷积进行降采样得到特征 $\mathbf{F}^{2\times \rightarrow 1\times}$; 式(9)表示多尺度融合特征 $\widetilde{\mathbf{F}}^{2\times}$ 由跨模态融合特征 $\mathbf{F}_{out}^{2\times}$ 和特征 $\mathbf{F}^{1\times \rightarrow 2\times}$ 在通道维度上经过级联得到; 式(11)表示 $\widetilde{\mathbf{F}}^{1\times}$ 由跨模态融合特征 $\mathbf{F}_{out}^{1\times}$ 和特征 $\mathbf{F}^{2\times \rightarrow 1\times}$ 在通道维度上经过级联得到。

最后,将 $1\times$ 和 $2\times$ 尺度的融合特征通过双三次上采样到 $4\times$ 尺度,再经过级联和卷积操作对 $4\times$ 尺度的所有特征进行融合,最终得到重建结果 \mathbf{F}_{output} 。如式(12)所示:

$$\mathbf{F}_{output} = f([\widetilde{\mathbf{F}}^{1\times} \uparrow_{4\times}, \widetilde{\mathbf{F}}^{2\times} \uparrow_{2\times}, \widetilde{\mathbf{F}}^{4\times}]; \mathbf{W}) \quad (12)$$

其中,“ $\uparrow_{s\times}$ ”表示缩放因子为 s 的上采样操作, $\widetilde{\mathbf{F}}^{1\times}$, $\widetilde{\mathbf{F}}^{2\times}$ 和 $\widetilde{\mathbf{F}}^{4\times}$ 分别表示经过多尺度特征融合模块得到的 3 个尺度的跨模态融合特征, \mathbf{W} 表示卷积核, f 为卷积操作。

在这种设计中,多尺度特征融合模块可以获得 3 个尺度的跨模态融合特征,这些特征不仅包含与深度图像相对应的所需结构信息,还包括跨不同尺度的特征变换,实现了更强大的特征表示,进一步提高了本文方法的性能。

3.4 损失函数

本文采用了重建损失 L_{rec} 作为模型训练的损失函数。为了更好地恢复出高分辨率深度图像,在像素级别上缩小重建深度图像与 ground-truth 之间的差距,本文使用 L1 损失作为重建损失。重建损失 L_{rec} 可表示为:

$$L_{\text{rec}} = \sum_P \|\mathbf{F}_p^{\text{gt}} - \hat{\mathbf{F}}_p\|_1 \quad (13)$$

其中, \mathbf{F}_p^{gt} 和 $\hat{\mathbf{F}}_p$ 分别代表 ground-truth 和深度图重建结果, P 是所有像素的集合, p 表示图像中的像素。

4 实验结果与分析

4.1 数据集与评价指标

本文实验选取了深度图像超分辨率领域 3 个广泛使用的公开数据集 (NYU v2^[39], Lu^[40], Middlebury^[41-42]) 和一个专门针对彩色图像引导深度图像超分辨率任务的数据集 (RGB-DD 数据集^[30])。

NYU v2 数据集^[39]: 由微软 Kinect 拍摄的 1449 对 RGB-D 图像组成。根据主流深度图像超分辨率算法所使用的训练测试方式,将 NYU v2 数据集中的 1000 对图像用于训练,其余 449 对图像用于测试。

Lu 数据集^[40]: 由 ASUS Xtion Pro 相机采集的 6 对 RGB-D 图像组成。本文使用它来进行模型测试。

Middlebury 数据集^[41-42]: 本文使用 Lu 等^[40] 提供的 2001—2006 年数据集中的 30 个图像对进行测试。

RGBDD 数据集^[30]: CVPR 2021^[30] 提出的新 RGB-D 数据集,包括肖像、模特、植物和灯光 4 个主要类别,其中 RGB 图像和 LR 深度图由华为 P30 Pro 采集,HR 深度图由 LU-CID 视觉实验室生产的 Helios ToF 相机拍摄。在本文的实验中,使用 297 张肖像图、68 株植物图、40 个模特图进行测试。

具体来说,有关数据集的实验设置为: NYU v2、Middlebury 和 Lu 数据集遵循文献[24-26, 29]的使用方式, RGBDD 数据集遵循文献[30]的使用方式。在所有的实验中, LR 深度图像都是通过 HR 深度图的双三次下采样来合成。

为了评估本文深度图像超分方法的性能,本文在测试数据集上采用均方根误差 (RMSE) 进行定量客观评价。RMSE 值越小,预测深度图像质量越好。

4.2 实验设置

本实验基于 Pytorch1.7.0 框架,使用显卡 TITAN Xp 进行实验。训练时,将图像随机裁剪为 160×160 大小作为网络的输入, batchsize 为 8。在训练时网络采用 Adam 优化器,动量衰减指数 β_1 和 β_2 分别为 0.9 和 0.999,初始学习率为 0.0002,使用的学习率衰减策略为 Cosine + warmup, warmup_epochs 设置为 10。

4.3 实验结果分析

在本节实验中,我们在 NYU v2, Middlebury, Lu 和 RGB-DD 数据集上进行测试,并将结果与目前性能较优的方法进行定量、定性的比较,相关方法包括 PMBANet^[7], DJF^[24], DJFR^[25], PAC^[26], DKN^[29], FDKN^[29], FDSR^[30], CTKT^[31], CU-Net^[32], DCTNet^[33], LAGJ^[34], SVLRM^[35] 和 DAGF^[36], 以证明本文算法性能的优越性。

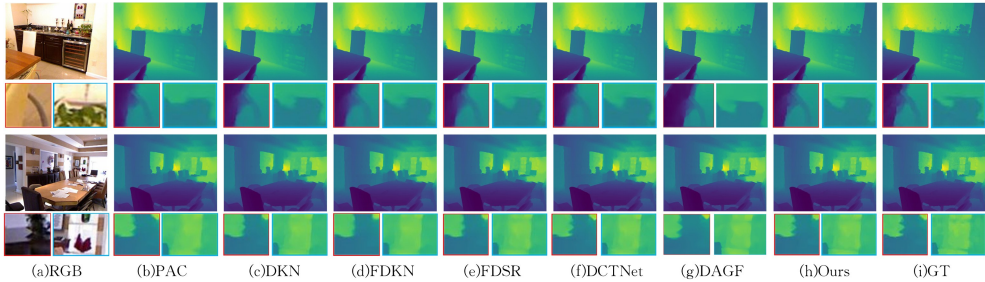


图 5 NYU v2 数据集上的深度图超分对比结果($\times 8$)

Fig. 5 Visual comparison of $\times 8$ depth map SR results in NYU v2 dataset

图 6 为上采样因子为 16 时,本文方法与其他方法在 Middlebury 数据集上的重建结果对比图。对于结构细节,例如第一幅图像 Aloe 中的叶子,以及第二幅图像 Art 中的茶壶柄,

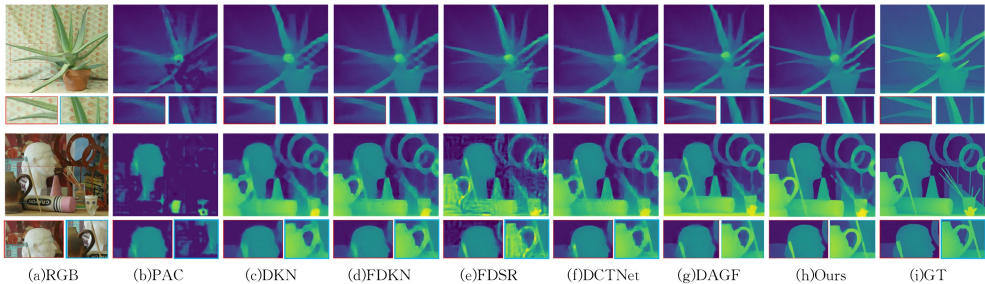


图 6 Middlebury 上深度图像“Aloe, Art”的视觉对比结果($\times 16$)

Fig. 6 Visual comparison of $\times 16$ depth map SR results on Aloe and Art in Middlebury

4.3.1 定性比较

图 5 为上采样因子为 8 时,本文方法与其他方法在 NYU v2 数据集上的重建结果对比图。由重建得到的深度图和放大的局部块可以看出,本文方法获得了最清晰的结果,并保留了正确的物体边缘,同时,其对远处物体边缘的恢复更为清晰,这充分证明了本文方法利用彩色图像中的高频信息作为场景结构引导信息的有效性。

本文方法较其他方法可以更清晰地恢复这些区域。对比结果表明,本文方法可以获得更精确的深度图值,显示了更为精细的边缘和视觉上令人愉快的细节,同时没有额外的噪声。

4.3.2 定量比较

比例因子为 $\times 4$, $\times 8$ 和 $\times 16$ 的超分辨率重建定量对比结果如表 1 所列。与仅在特定数据集或超分辨率因子上表现

良好的现有方法相比,本文方法在多个数据集上和不同的超分辨率尺度实现了最佳或次优性能,有效证明了本文模型相较于以往方法的优势。

表 1 4 个公开数据集上深度图超分结果的定量比较

Table 1 Quantitative comparison with state-of-the-art approaches on four pubic datasets

Methods	Middlebury			NYU V2			Lu			RGBDD		
	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$
DJF ^[24]	1.68	3.24	5.62	2.80	5.33	9.46	1.65	3.96	6.75	3.41	5.57	8.15
DJFR ^[25]	1.32	3.19	5.57	2.38	4.94	9.18	1.15	3.57	6.77	3.35	5.57	7.99
PAC ^[26]	1.32	2.62	4.58	1.89	3.33	6.78	1.20	2.33	5.19	1.25	1.98	3.49
CUNet ^[32]	1.10	2.17	4.33	1.92	3.70	6.78	0.91	2.23	4.99	1.18	1.95	3.45
DKN ^[29]	1.23	2.12	4.24	1.62	3.26	6.51	0.96	2.16	5.11	1.30	1.96	3.42
FDKN ^[29]	1.08	2.17	4.50	1.86	3.58	6.96	<u>0.82</u>	2.10	5.05	1.18	1.91	3.41
FDSR ^[30]	1.13	2.08	4.39	1.61	3.18	5.86	1.29	2.19	5.00	1.16	1.82	3.06
CTKT ^[31]	—	—	—	1.49	2.73	5.11	—	—	—	—	—	—
DCTNet ^[33]	1.10	2.05	4.19	1.59	3.16	5.84	0.88	1.85	4.39	1.08	1.74	3.05
LAGJ ^[34]	<u>1.07</u>	1.86	<u>3.57</u>	1.34	2.71	5.39	0.89	1.73	<u>4.25</u>	—	—	—
SVLRM ^[35]	1.11	2.13	4.34	1.51	3.21	6.98	0.93	2.19	5.44	—	—	—
DAGF ^[36]	1.15	1.80	3.70	1.36	2.87	6.06	0.83	1.93	4.80	1.22	1.97	<u>3.00</u>
PMBANet ^[7]	—	—	—	1.06	2.28	4.98	—	—	—	—	—	—
ours	1.05	<u>1.84</u>	3.42	<u>1.31</u>	<u>2.67</u>	<u>5.05</u>	0.80	<u>1.74</u>	4.05	<u>1.14</u>	<u>1.77</u>	2.81

注:粗体和下划线分别表示最优值和次优值。

4.3.3 消融实验

为了验证本文所提模型框架的有效性,本文在 NYU v2 数据集上进行了 3 项消融实验。消融实验的参数设置与前述实验相同。表 2 的结果表明,3 个模块均可提高模型的性能。高频特征提取模块(HFEM)可以有效提取彩色图像中的高频信息,来帮助深度图像边缘信息的恢复;多尺度特征融合模块(MFFM)能够充分融合不同尺度特征之间的信息,有效提升模型在大比例放大因子的超分辨率重建任务中的细节恢复能力;特征自注意力模块(FSM)能够有效建模特征图全局信息,对算法性能具有较大的影响。

表 2 不同模块对模型性能的影响

Table 2 Effects of different modules on model performance

Methods	NYU v2		
	×4	×8	×16
w/o HFEM	1.52	2.75	5.17
w/o FSM	1.54	2.88	5.22
w/o MFFM	1.41	2.71	5.18
Ours	1.31	2.67	5.05

结束语 本文提出了一种基于彩色图像高频信息引导的深度图像超分辨率算法。在特征提取部分,采用 Octave 卷积构建高频特征提取模块,可以自适应地关注彩色图像的高频分量并抑制低频分量,获取与深度图像相关的边缘信息;设计的高频自注意力模块可以学习特征之间的依赖关系,获取更深层次的特征,有效辅助深度图像的细节恢复;此外,多尺度特征融合模块可以对跨模态融合得到的 3 个尺度的特征进行融合,获取多层次且更丰富的特征表示,进一步提升深度图重建效果。在公开数据集上的实验结果表明,本文方法在定量和定性评估方面均表现出了优越的性能,验证了所提深度图像超分辨率算法的有效性。

参考文献

- [1] RICHARDT C, STOLL C, DODGSON N A, et al. Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos [C]// Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2012: 247-256.
- [2] HE K, SUN J, TANG X. Guided image filtering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(6): 1397-1409.
- [3] KOPF J, COHEN M F, LISCHINSKI D, et al. Joint bilateral up-sampling [J]. ACM Transactions on Graphics (ToG), 2007, 26(3): 96-1-95-5.
- [4] YANG Q, YANG R, DAVIS J, et al. Spatial-depth super resolution for range images[C]// 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.
- [5] PARK J, KIM H, TAI Y W, et al. High quality depth map up-sampling for 3D-TOF cameras[C]// 2011 International Conference on Computer Vision. IEEE, 2011: 1623-1630.
- [6] 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. 2013: 993-1000.
- [7] YE X, SUN B, WANG Z, et al. PMBNet: Progressive multi-branch aggregation network for scene depth super-resolution [J]. IEEE Transactions on Image Processing, 2020, 29: 7427-7442.
- [8] JIANG Z, YUE H, LAI Y K, et al. Deep edge map guided depth super resolution [J]. Signal Processing: Image Communication, 2021, 90: 116040.
- [9] GUO C, LI C, 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.
- [10] YANG F, YANG H, FU J, et al. Learning texture transformer network for image super-resolution [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5791-5800.
- [11] CHEN Y, FAN H, XU B, et al. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3435-3444.
- [12] 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.
- [13] FERSTL D, RUTHER M, BISCHOF H. Variational depth super resolution using example-based edge representations [C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 513-521.
- [14] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution [C]// Computer Vision - ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer International Publishing, 2014: 184-199.
- [15] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network [C]// Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016: 391-407.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [17] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1646-1654.
- [18] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1637-1645.
- [19] RIEGLER G, RÜTHER M, BISCHOF H. Atgv-net: Accurate depth super-resolution [C]// Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer International Publishing, 2016: 268-284.
- [20] 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.

- [21] SONG X, DAI Y, ZHOU D, et al. Channel attention based iterative residual learning for depth map super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020;5631-5640.
- [22] DIEBEL J, THRUN S. An application of markov random fields to range sensing[C]// Proceedings of the 18th International Conference on Neural Information Processing Systems. 2005: 291-298.
- [23] GU S, ZUO W, GUO S, et al. Learning dynamic guidance for depth image enhancement[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3769-3778.
- [24] LI Y, HUANG J B, AHUJA N, et al. Deep joint image filtering [C]// Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 154-169.
- [25] LI Y, HUANG J B, AHUJA N, et al. Joint image filtering with deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1909-1923.
- [26] SU H, JAMPANI V, SUN D, et al. Pixel-adaptive convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11166-11175.
- [27] HUI T W, LOY C C, TANG X. Depth map super-resolution by deep multi-scale guidance[C]// Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer International Publishing, 2016: 353-369.
- [28] LUTIO R, D'ARONCO S, WEGNER J D, et al. Guided super-resolution as pixel-to-pixel transformation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8829-8837.
- [29] KIM B, PONCE J, HAM B. Deformable kernel networks for joint image filtering[J]. International Journal of Computer Vision, 2021, 129(2): 579-600.
- [30] HE L, ZHU H, LI F, et al. Towards fast and accurate real-world depth super-resolution: Benchmark dataset and baseline[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 9229-9238.
- [31] SUN B, YE X, LI B, et al. Learning scene structure guidance via cross-task knowledge transfer for single depth super-resolution [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7792-7801.
- [32] DENG X, DRAGOTTI P L. Deep convolutional neural network for multi-modal image restoration and fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3333-3348.
- [33] ZHAO Z, ZHANG J, 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. 2022: 5697-5707.
- [34] MALLICK A, ENGELHARDT A, BRAUN R, et al. Local Attention Guided Joint Depth Upsampling[C]// Vision, Modeling, and Visualization. The Eurographics Association, 2022: 135-1439.
- [35] DONG J, PAN J, REN J S, et al. Learning spatially variant linear representation models for joint filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 8355-8370.
- [36] ZHONG Z, LIU X, JIANG J, et al. Deep attentional guided image filtering[J]. arXiv:2112.06401, 2023.
- [37] YANG Y, CAO Q, ZHANG J, et al. CODON: on orchestrating cross-domain attentions for depth super-resolution[J]. International Journal of Computer Vision, 2022, 130(2): 267-284.
- [38] ZHOU C, ZHOU Q W, CHEN H M, et al. Recurrent Scale-by-scale Feature Fusion Network for RGBD Salient Object Detection[J]. Journal of Chinese Computer Systems, 2023, 44(10): 2276-2283.
- [39] MARCHAND E, UCHIYAMA H, SPINDLER F. Pose estimation for augmented reality: a hands-on survey[J]. IEEE Transactions on Visualization and Computer Graphics, 2015, 22(12): 2633-2651.
- [40] LU S, REN X, LIU F. Depth enhancement via low-rank matrix completion[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3390-3397.
- [41] HIRSCHMULLER H, SCHARSTEIN D. Evaluation of cost functions for stereo matching[C]// 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.
- [42] SCHARSTEIN D, PAL C. Learning conditional random fields for stereo[C]// 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.



LI Jiaying, born in 1998, master. Her main research interests include computer vision and image processing.



LIANG Yudong, born in 1988, Ph.D., associate professor, is a member of CCF (No. 85977M). His main research interests include computer vision, image processing, and deep learning-based applications.