

DOI: 10.12086/oe.2020.190260

彩色图像多尺度引导的深度图像超分辨率重建

于淑侠*, 胡良梅, 张旭东, 付绪文

合肥工业大学计算机与信息学院, 安徽 合肥 230601

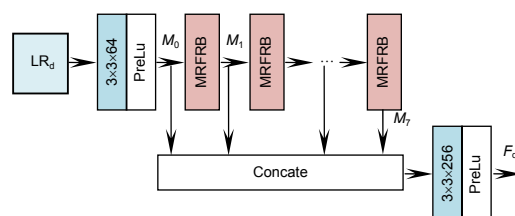
摘要: 为获得更优的深度图像超分辨率重建结果, 本文构建了彩色图像多尺度引导深度图像超分辨率重建卷积神经网络。该网络使用多尺度融合方法实现高分辨率(HR)彩色图像特征对低分辨率(LR)深度图像特征的引导, 有益于恢复图像细节信息。在对 LR 深度图像提取特征的过程中, 构建了多感受野残差块(MRFRB)提取并融合不同感受野下的特征, 然后将每一个 MRFRB 输出的特征连接、融合, 得到全局融合特征。最后, 通过亚像素卷积层和全局融合特征, 得到 HR 深度图像。实验结果表明, 该算法得到的超分辨率图像缓解了边缘失真和伪影问题, 有较好的视觉效果。

关键词: 深度图像; 超分辨率重建; 卷积神经网络; 多尺度引导; 多感受野特征

中图分类号: TP391.41; TP183

文献标志码: A

引用格式: 于淑侠, 胡良梅, 张旭东, 等. 彩色图像多尺度引导的深度图像超分辨率重建[J]. 光电工程, 2020, 47(4): 190260



Color image multi-scale guided depth image super-resolution reconstruction

Yu Shuxia*, Hu Liangmei, Zhang Xudong, Fu Xuwen

School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230601, China

Abstract: In order to obtain better super-resolution reconstruction results of depth images, this paper constructs a multi-scale color image guidance depth image super-resolution reconstruction convolutional neural network. In this paper, the multi-scale fusion method is used to realize the guidance of high resolution (HR) color image features to low resolution (LR) depth image features, which is beneficial to the restoration of image details. In the process of extracting features from LR depth images, a multiple receptive field residual block (MRFRB) is constructed to extract and fuse the features of different receptive fields, and then connect and fuse the features of each MRFRB output to obtain global fusion features. Finally, the HR depth image is obtained through sub-pixel convolution layer and global fusion features. The experimental results show that the super-resolution image obtained by this method alleviates the edge distortion and artifact problems, and has better visual effects.

Keywords: depth image; super-resolution; convolutional neural network; multi-scale guidance; multiple receptive field characteristics

Citation: Yu S X, Hu L M, Zhang X D, *et al.* Color image multi-scale guided depth image super-resolution reconstruction[J]. *Opto-Electronic Engineering*, 2020, 47(4): 190260

收稿日期: 2019-05-17; 收到修改稿日期: 2019-10-21

基金项目: 国家自然科学基金资助项目(61876057)

作者简介: 于淑侠(1991-), 女, 硕士研究生, 主要从事智能信息处理的研究。E-mail: ysx123@mail.hfut.edu.cn

版权所有©2020 中国科学院光电技术研究所

1 引言

近年来,随着计算机视觉领域对场景深度信息的需求不断扩大,高分辨率深度图像的获取变得至关重要。然而,由于传感器等硬件条件的限制,由深度相机得到的深度图像分辨率普遍不高,很难满足实际应用需求^[1-3]。例如 PMD Camcube 相机分辨率只有 200×200 ^[4],微软的 Kinect 相机分辨率也只有 640×480 ^[5]。如果通过提高硬件设施来提高深度图像的分辨率,成本将会很高,并且也存在一些很难克服的技术难题,因此通常通过软件处理的方法提高深度图像分辨率。

目前深度图像超分辨率重建问题,根据是否加入高分辨率(high resolution, HR)彩色图像作为引导,可以分为:1) 仅用低分辨率(low resolution, LR)深度图像超分辨率重建,2) HR 彩色图像引导 LR 深度图像超分辨率重建。许多学者对仅用 LR 深度图像超分辨率重建的问题展开了研究^[6-10]。Mandal 等^[6]提出了一种可以基于图像噪声级别自适应的鲁棒超分辨重建算法。Aodha 等^[7]使用基于图像块的多感受块(multiple receptive block, MRB)模型对 LR 深度图进行超分辨率重建。Li 等^[8]通过添加自相似结构的几何约束来扩展文献[7]中的模型框架。Xie 等^[9]提出了一种新的框架,首先重建深度图像的边缘图,然后重建 HR 深度图。Chen 等^[10]提出了一种利用卷积神经网络实现单幅深度图像超分辨率重建的方法,首先利用卷积神经网络获得 HR 深度图像边缘图,然后用 HR 深度图像边缘图引导 LR 深度图像超分辨率重建。在该类方法中,由于输入网络的 LR 深度图像包含高频信息很少且没有同场景 HR 彩色图像作为引导,所以由此类方法得到的重建图像通常在边缘部分不是很理想。

同场景的 HR 彩色图像与 LR 深度图像包含相对应的关系,使用同场景 HR 彩色图像引导 LR 深度图像超分辨率重建有利于 LR 深度图像恢复高频信息,因此在对 LR 深度图像进行超分辨率重建时,通常会使用同场景彩色图像引导^[11-16]。Liu 等^[11]提出了一种基于 HR 彩色图像引导的 LR 深度图像超分辨率重建鲁棒优化框架。Kiechle 等^[12]提出了一种双模共稀疏模型重建 LR 深度图像。Li 等^[13]将基本的引导插值问题转化为一个分层的全局优化框架,然后对 LR 深度图像进行重建。Park 等^[14]在马尔科夫模型中加入了非局部平均项,使重建图像更加可靠。Li 等^[15]提出了一种深层次的端到端深度图像超分辨率重建卷积神经网络,

提升了重建性能,但是该网络提取的特征单一。Xiao 等^[16]通过两个金字塔子网络提取 LR 深度图像和 HR 彩色图像特征。以上提到的关于 HR 彩色图像引导 LR 深度图像超分辨率重建的卷积神经网络,虽然重建效果相对较优,但是仍然存在一些问题,在使用彩色图像引导时,使用了单尺度引导,不利于对图像精细结构的恢复;在网络的每一层只提取了单一感受野特征,没有充分提取图像特征。

针对以上问题,本文构建了彩色图像多尺度引导深度图像超分辨率重建卷积神经网络。在不同的尺度上,该网络可以在 HR 彩色图像分支中学习到丰富的 HR 特征,将其作为相对应尺度上深度图像分支中特征的补充,这种多尺度引导结构有益于恢复图像细节信息。在 LR 深度图像分支,首先构建了多感受野残差块(multiple receptive field residual block, MRFRB),该结构能提取并融合不同感受野下的特征,然后将每一个 MRFRB 输出的特征连接、融合,从而得到全局融合特征。最后,通过亚像素卷积层和融合后的特征,重建 HR 深度图像。综上,本文采用彩色图像特征多尺度引导深度图像重建有益于对细节信息的恢复;本文构建的 MRFRB 可以提取不同感受野下的特征,有益于图像结构信息重建。

2 理论推导

为有效提高深度图像分辨率,本文构建了彩色图像多尺度引导深度图像超分辨率重建卷积神经网络。该网络包含三个分支:彩色图像分支、深度图像分支、图像重建分支。由于 LR 深度图像包含较少高频信息,同场景的 HR 彩色图像包含较多与其相关的高频信息,因此可以通过同场景 HR 彩色图像引导 LR 深度图像超分辨率重建以获得更优的重建结果。由于图像中不同的结构信息具有不同的尺度,因此在彩色图像分支,本文提取了不同尺度下的彩色图像特征,用以引导相对应尺度的深度图像重建。对于深度图像超分辨率重建问题来说,输入的 LR 深度图像与输出的 HR 深度图像是高度相关的,如果能充分提取 LR 深度图像的特征,将会得到更优的重建结果。因此在深度图像分支中,本文构建了 MRFRB,该结构可以提取并融合不同感受野下的特征,然后将每一个 MRFRB 输出的特征连接、融合,从而得到全局融合特征。在图像重建分支中,通过亚像素卷积和全局融合特征重建 HR 深度图像。

2.1 网络概述

本文构建的整体网络架构如图 1 所示。假设超分辨率重建倍数为 2^s ，那么彩色图像分支总共 $4s-1$ 层；在深度图像分支，除去深度图像特征提取(deep image feature extract, DIFE)部分总共有 $4s$ 层，包含 s 个亚像素卷积层、 s 个连接层、 $2s$ 个卷积层。在本文框架中设超分辨率重建倍数为 4。

在彩色图像分支中，所有的卷积层卷积核大小均为 3×3 ，步长为 1，输出特征图个数为 64，卷积过程中不改变特征图大小。HR 彩色图像用 HR_c 表示，假设 HR_c 的大小为 $4m \times 4n$ 。输入的 HR_c 首先经过三个卷积层，得到 F_{c1} ， F_{c1} 大小为 $4m \times 4n$ ，包含 64 张特征图。 F_{c1} 经过池化核大小为 2×2 ，步长为 2 的池化层，得到 F_c 。具体过程如下：

$$F_c = \text{Maxpooling}(F_{c1}) , \quad (1)$$

式中： Maxpooling 表示最大池化， F_c 的大小 $2m \times 2n$ ， F_c 经过三个卷积层，得到 F_{c2} 。 F_{c2} 大小为 $2m \times 2n$ ，包含 64 张特征图。至此，该网络在彩色图像分支提取了尺度不同的特征 F_{c1} 、 F_{c2} ，用于在不同尺度下引导深度图像超分辨率重建。

在深度图像分支中，LR 深度图像用 LR_d 表示，其大小 $m \times n$ 。 LR_d 经过深度图像特征提取块(deep image feature extract block, DIFE)得到 F_{d0} ， F_{d0} 的大小与 LR_d 相同，包含 256 张特征图。经过亚像素卷积层得到 F_{d1} 。具体过程如下：

$$F_{d1} = \text{SF}(F_{d0}) , \quad (2)$$

式中： SF 表示亚像素卷积操作，超分辨率重建倍数为 2， F_{d1} 的大小为 $2m \times 2n$ ，因此 F_{d1} 包含 64 张特征图，进而将由深度图像提取的特征 F_{d1} 与彩色图像提取的特征 F_{c2} 进行连接操作，得到 F_{d2} 。具体过程如下：

$$F_{d2} = [F_{d1}, F_{c2}] , \quad (3)$$

式中： $[F_{d1}, F_{c2}]$ 表示对 F_{d1} 、 F_{c2} 的连接操作， F_{d2} 包含 128 张特征图。 F_{d2} 首先经过一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 64 的卷积层，再经过一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 256 的卷积层，得到 F_{d3} 与 F_{c1} 融合后的特征 F_{d3} 。 F_{d3} 大小为 $2m \times 2n$ ，从而实现了彩色图像分支提取的特征 F_{c1} 在尺度 $2m \times 2n$ 上对深度图像特征 F_{d1} 的引导。 F_{d3} 经过亚像素卷积层得到 F_{d4} 。具体过程如下：

$$F_{d4} = \text{SF}(F_{d3}) , \quad (4)$$

式中： SF 表示亚像素卷积操作，超分辨率重建倍数为 2， F_{d4} 的大小为 $4m \times 4n$ ，因此 F_{d4} 包含 64 张特征图，进而将 F_{d4} 与 F_{c1} 进行连接操作，得到 F_{d5} 。具体过程如下：

$$F_{d5} = [F_{d4}, F_{c1}] , \quad (5)$$

式中： $[F_{d4}, F_{c1}]$ 表示对 F_{d4} 、 F_{c1} 的连接操作， F_{d5} 包含 128 张特征图。 F_{d5} 经过一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 64 的卷积层，再经过一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 1 的卷积层，从而得到最终融合后的特征图 F_{d6} ，实现了彩色图像分支提取的特征 F_{c1} 在尺度 $4m \times 4n$ 上对深度图像特征 F_{d4} 的引导。

在图像重建分支中，LR 深度图像首先经过一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 16 的卷积层，特征图的大小为 $m \times n$ (该层卷积层输出特征图的个数与超分辨率重建倍数相关，假设超分辨率重建倍数为 r ，则该层输出特征图个数为 r^2)。然后经过亚像素卷积层得到 F_d ，其大小为 $4m \times 4n$ 。具体过程如下：

$$F_d = \text{SF}(\varphi(W_d * LR_d + b_d)) , \quad (6)$$

式中： φ 表示激活函数， W_d 和 b_d 分别表示卷积核和偏

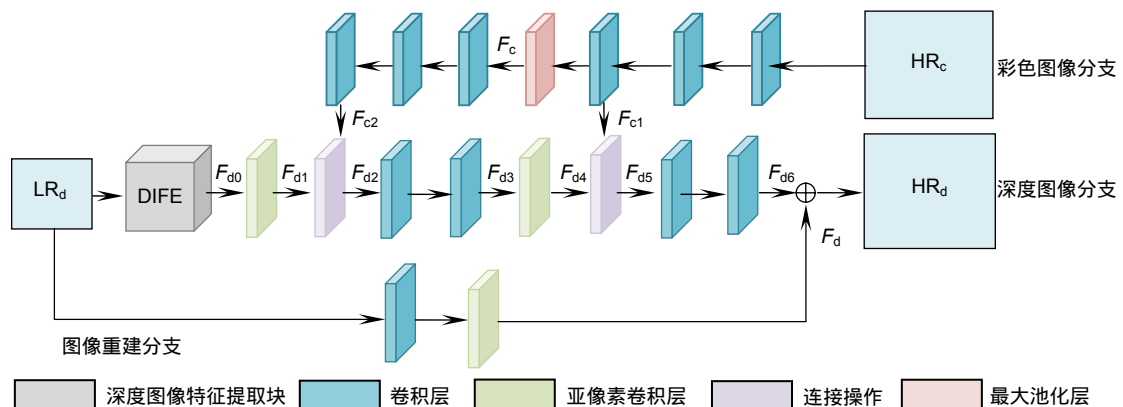


图 1 当超分辨率重建倍数 r 为 4 时的网络结构

Fig. 1 Network structure when super-resolution reconstruction multiple r is 4

置，“*”表示卷积操作， SF 表示亚像素卷积，这里的超分辨率重建倍数即为期望的超分辨率重建倍数 r ，因为该网络是以超分辨率重建倍数 4 为例，所以这里取 r 为 4。

假设最终得到的 HR 深度图像用 HR_d 表示，具体过程如下：

$$HR_d = F_d + F_{d6}, \quad (7)$$

式中：“+”表示元素级相加操作，即两幅图片对应元素相加。

2.2 深度图像特征提取块

在对深度图像提取特征时，本文直接将 LR 深度图像 LR_d 输入网络，这样使得后续的 DIFE 过程都是在 LR 空间下进行，降低了计算复杂度，同时使网络可以提取图像的原始特征信息。该部分网络结构如图 2 所示，包含两部分：多层次特征融合和 MRFRB。对于深度图像超分辨率重建问题来说，输入网络的 LR 深度图像所包含的信息是最可靠的，因此为了充分挖掘 LR 深度图像的特征，本文构建了 MRFRB，该结构可以提取不同感受野下的特征，并将其融合。然后使用多层次特征融合，将各个 MRFRB 块所提取的特征进行融合，得到全局融合特征。

2.2.1 多层次特征融合

为了能充分利用每个 MRFRB 块所提取的特征，本文将各个 MRFRB 块所提取的特征进行连接操作。具体过程如下：

$$M = [M_0, M_1, \dots, M_7], \quad (8)$$

式中： $[M_0, M_1, \dots, M_7]$ 表示对 M_0, M_1, \dots, M_7 特征的连接操作， M 表示连接后的特征。因为 M_0, M_1, \dots, M_7 各包含 64 个特征图，因此 M 包含 512 张特征图。这 512 张特征图会包含一定的冗余信息，为了去除这些冗余信息、降低计算复杂度，将 M 输入一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 256 的卷积层，得到全局融合特征 F_{d0} 。具体过程如下：

$$F_{d0} = \varphi(W_{d0} * M + b_{d0}), \quad (9)$$

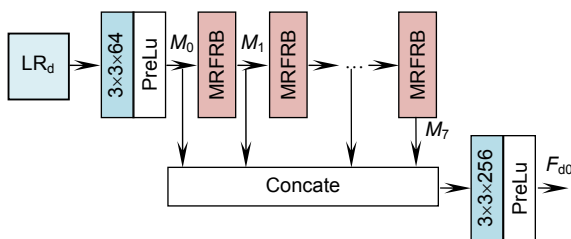


图 2 深度图特征提取块

Fig. 2 Feature extraction block of depth map

式中： φ 表示激活函数， W_{d0} 和 b_{d0} 分别表示卷积核和偏置，“*”表示卷积操作。

2.2.2 多层次特征融合多感受野残差块(MRFRB)

对于深度图像超分辨率重建问题来说，输入的 LR 深度图像与输出的 HR 深度图像是高度相关的，因此如果能充分挖掘 LR 深度图像的特征信息，将会得到更优的重建结果。本文构建了 MRFRB，如图 3 所示，该结构能提取不同感受野下的特征并将其融合，MRFRB 使用了双支路，在支路一使用两个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 64 的卷积层，因此在上面的支路输出是感受野为 5×5 的特征，支路二使用一个卷积核大小为 3×3 ，步长为 1，输出特征图个数为 64 的卷积层，因此下面的支路输出是感受野为 3×3 的特征。具体过程如下：

$$u_0 = \varphi(W_{u0} * M_{n-1} + b_{u0}), \quad (10)$$

$$u = \varphi(W_u * u_0 + b_u), \quad (11)$$

$$v = \varphi(W_v * M_{n-1} + b_v), \quad (12)$$

式中： φ 表示激活函数， W 和 b 分别表示卷积核和偏置，“*”表示卷积操作， u, v 分别为感受野为 $5 \times 5, 3 \times 3$ 的特征。

对两个支路得到的特征 u, v 进行连接操作， u, v 连接之后具有 128 个特征图，为了保证 MRFRB 块输入输出特征图数目相同，将 u, v 连接之后的特征图输入卷积核大小为 1×1 ，步长为 1，输出特征图个数为 64 的卷积层。 s 即为不同感受野特征融合之后的特征。具体过程如下：

$$s = \varphi(W_s * [u, v] + b_s), \quad (13)$$

式中： φ 为激活函数， $[u, v]$ 表示对 u, v 的连接操作， W_s 和 b_s 分别表示卷积核和偏置，“*”表示卷积操作。

为了提高网络性能，本文在该部分使用了残差连接，具体过程如下：

$$M_n = s + M_{n-1}, \quad (14)$$

式中： M_{n-1} 和 M_n 分别表示 MRFRB 块的输入和输出， s 表示所提取的特征。

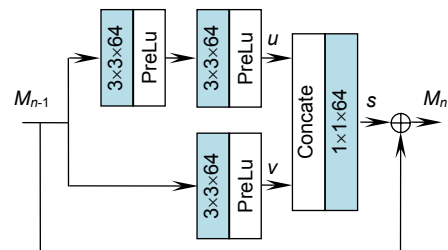


图 3 多感受野残差块

Fig. 3 Multiple receptive field residual block

2.3 网络训练

本文基于反向传播的随机梯度下降法^[17]训练网络,使用均方误差(mean-square error, MSE)作为损失函数,则:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - X_i\|^2, \quad (15)$$

式中: n 为每次训练样本的个数, θ 表示需要学习的参数, $F(Y_i; \theta)$ 表示重建得到的深度图像, X_i 表示参考图像。

3 实验结果与分析

3.1 实验设置

本文实验的PC配置为 NVIDIA Titan X、RAM 16 G、Linux 64 位操作系统,训练测试通过 Tensorflow 实现。

本文使用的训练数据集包含 4 部分,分别是 Middlebury^[18]、RGBZ^[19]和 RGBD^[20]、ICL-NUIM^[21]。Middlebury 数据集包含丰富的 HR 深度信息,图像的边缘清晰,是目前验证超分辨率重建算法优劣常用的数据集,从 Middlebury 数据集随机选取了 52 个场景。RGBZ 和 RGBD 数据集是 TOF 拍摄的真实场景深度图像,从 RGBD 和 RGBZ 数据集随机选取了 10 个场景。ICL-NUIM 数据集包含丰富的 HR 深度信息,从中随机选取了 65 个场景。为了充分利用这些深度图像,本文对图像进行了旋转 90°, 180°, 270°之后再对所有的图像进行水平镜像,最终得到 1016 个场景的图片。在每个训练批次中,当超分辨率重建倍数为 2 或 4 时,从 1016 个场景中随机剪裁大小为 64×64 的 4 个场景的图像块进行训练。每个 epoch 迭代 100 次。当超分辨率重建倍数为 8 时,考虑到图像本身的大小,调整图像形块的大小为 32×32。本文通过双三次下采样来获得 LR 的训练图像。

网络参数:本文使用 He 等^[22]的方法初始化卷积

层,所有的卷积层后使用激活函数 PreLu。网络的初始学习率被设置为 5×10^{-4} ,每 15 个 epoch 学习率下降 2 倍,直到学习率小于 5×10^{-9} 。

评价指标:本文用均方根误差(root mean squared error, RMSE)和结构相似度指数(structural similarity Index, SSIM)作为评价指标。RMSE 是 MSE 的算术平方根,数值越小,表明重构图像质量越好。SSIM 是指两幅图像之间的结构信息相似程度,SSIM 值越接近 1,表明重建深度图与原始图像越接近,重建效果越好。

$$E_{\text{RMSE}} = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (Y(i, j) - X(i, j))^2}, \quad (16)$$

$$I_{\text{SSIM}} = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (17)$$

式中: $M \times N$ 是重建的 HR 图像 Y 的大小, X 代表参考图像, μ_x 和 σ_x 分别表示参考图像的灰度平均值和方差, μ_y 和 σ_y 分别表示重建的 HR 图像的灰度平均值和方差, σ_{xy} 是参考图像与重建 HR 图像的协方差, C_1 和 C_2 是防止分母为 0 的常数。

3.2 彩色图像对重建结果的影响

本小节对有无彩色图像引导对实验结果的影响进行了实验分析,构建了无彩色图像引导卷积神经网络(去除了本章所构建网络中的彩色图像分支,网络的其他结构均无变化)如图 4 所示,在下文中用 without color 表示此结构。

图 5 为有无彩色图像引导在 Middlebury 数据集上的实验结果的定性比较,放大倍数为 4。图 5(a)为真值图像;图 5(b)为文献[6]的重建结果,由局部放大图可知其存在边缘模糊的问题;图 5(c)为无彩色引导的重建结果,由局部放大图可知其在边缘存在伪影问题;图 5(d)为本文方法重建结果,由局部放大图可知,本文方法在边缘部分重建效果较好,验证了彩色图像的引导有益于重建性能的提升。

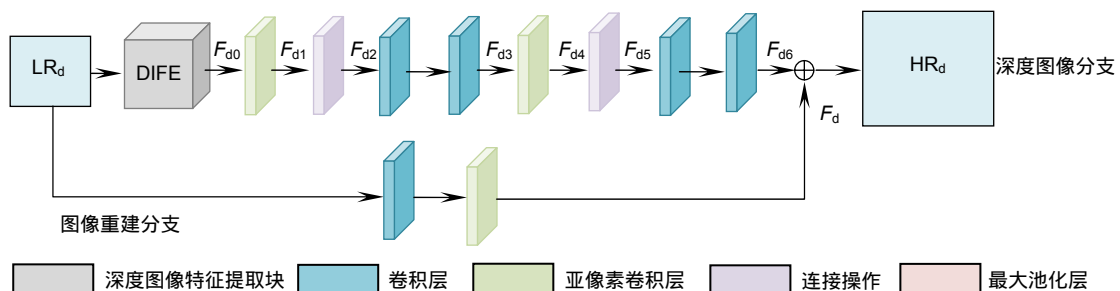


图 4 无彩色图像引导网络结构

Fig. 4 Network structure without color image guidance

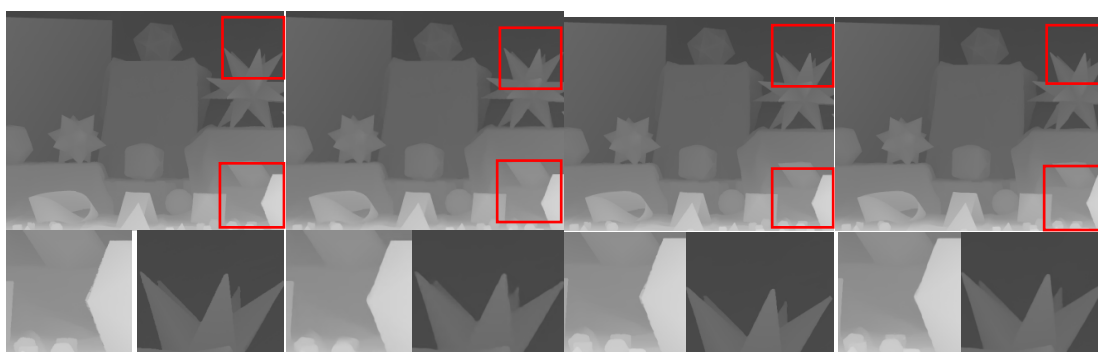


图 5 有无彩色图像引导在 Middlebury 数据集上的实验结果的定性比较。

(a) 真值图像; (b) 文献[6]方法; (c) 无彩色图引导; (d) 本文方法

Fig. 5 Qualitative comparison of experimental results on the data set at Middlebury with and without color image guidance.

(a) Truth image; (b) Ref. [6] method; (c) No color image guide; (d) This article method

表 1 给出了有无彩色图像引导在 Middlebury 数据集上的实验结果的定量比较,放大倍数为 4。从实验结果的定量比较分析可知,加入彩色图像的引导有助于重建结果的提升。与未加入彩色图像引导结果做比较,加入彩色图像引导的结果 RMSE 平均降低约 0.02, SSIM 平均增加约 0.0001。

综上,从定性和定量两个方面分析结果可知,加入彩色图像作为引导有助于重建性能的提升。

3.3 与其他方法的比较

Middlebury 数据集为标准的深度图像数据集,也是目前验证深度图像超分辨率重建算法优劣最常用的数据集,在 Middlebury 数据集上评估了本文提出的方法。由于篇幅限制,本文将实验的图像分为了 A、B 两组。对比的方法包括代表性的算法:双三次插值, Mandal^[6]、Chen^[10]、Liu^[11]、Kiechle^[12]、Li^[13]、Park^[14]、Dong^[23]、Hui^[24]、Kim^[25]、Lai^[26]。其中 Chen^[10]的定量结果引自文献[10], Dong^[23]和 Hui^[24]的定量结果引自文献[24]。

在 Middlebury 数据集上采用 RMSE、SSIM 指标分别评估了本文的方法,实验结果如表 2~5 所示。表 2、3 分别是不同的方法在 Middlebury 数据集 A、B 重建结果的 RMSE 定量分析,由表 2、3 可知,本文算法均优于传统算法(Mandal^[6]、Liu^[11]、Kiechle^[12]、Li^[13]、Park^[14]),和基于深度学习的算法(Chen^[10]、Dong^[23]、Hui^[24]、Kim^[25]、Lai^[26])相比,取得了 11 个最优值,4 个次优值。表 4 和表 5 分别是不同的方法在 Middlebury 数据集 A、B 重建结果的 SSIM 定量分析,由表 4 和表 5 可知,本文算法取得了 14 个最优值,1 个次优值。和次优结果相比,本文的 RMSE 指数平均降低了约 0.083,降低比例约 5.87%。SSIM 指数平均增加了约 0.0011。

图 6 是当超分辨率重建倍数为 8 时本文方法和其他方法在 Middlebury 数据集上的重构结果图像对比,红色矩形区域对应局部放大区域。从图 6(b)、6(c)可以看出, Liu^[11]和 Li^[13]两种方法在边缘处出现模糊的问题。Kim^[25]需要用双三次插值进行预处理,并且在网络每一层只提取了单感受野特征,从图 6(d)可以看出,

表 1 有无彩色图像引导在 Middlebury 数据集上的实验结果的定量比较

Table 1 Quantitative comparison of experimental results on the data set at Middlebury with and without color image guidance

	RMSE			SSIM		
	Art	Books	Moebius	Art	Books	Moebius
Bicubic	3.8564	1.5789	1.4005	0.9683	0.9899	0.9888
Mandal ^[6]	2.6869	1.2520	1.1255	0.9854	0.9934	0.9925
Kim ^[25]	1.7986	0.7560	0.7657	0.9934	<u>0.9966</u>	0.9962
Without color	<u>1.6267</u>	<u>0.7586</u>	<u>0.7448</u>	<u>0.9949</u>	0.9969	<u>0.9966</u>
Ours	1.6000	0.7484	0.7244	0.9951	0.9969	0.9967

注:粗体字表示最优值,下划线标识次优值。

表 2 不同的方法在 Middlebury 数据集 A 上重建结果的定量分析(RMSE)

Table 2 Quantitative analysis of reconstruction results on dataset A of Middlebury by different methods (RMSE)

	Art			Books			Moebius		
	X2	X4	X8	X2	X4	X8	X2	X4	X8
Bicubic	2.5837	3.8564	5.5279	1.0319	1.5789	2.2682	0.9303	1.4005	2.0578
Mandal ^[6]	1.7502	2.6869	6.0125	0.6152	1.2520	2.6268	0.7046	1.1255	2.2823
Liu ^[11]	2.3406	3.4221	4.6056	1.1873	1.6225	2.2459	1.0183	1.5159	2.1717
Kiechle ^[12]	1.1414	1.8789	2.8130	0.5541	0.8137	1.2378	0.6037	0.8497	1.5094
Li ^[13]	2.9834	4.0422	5.3825	1.2582	1.8400	2.4619	1.2400	1.7089	2.2352
Park ^[14]	2.9923	4.2771	6.4019	1.2339	1.8707	2.6495	1.1248	1.6981	2.5554
Dong ^[23]	1.1330	2.0170	3.8290	0.5230	0.9350	1.7260	0.5370	0.9130	1.5790
Hui ^[24]	0.8130	<u>1.6270</u>	<u>2.7690</u>	0.4170	0.7240	1.0720	<u>0.4130</u>	<u>0.7410</u>	<u>1.1380</u>
Kim ^[25]	1.0208	1.7986	3.0044	0.4553	0.7560	1.2075	0.4769	0.7657	1.1854
Lai ^[26]	<u>0.7568</u>	2.0198	3.3320	0.4470	0.8653	1.3639	0.4367	0.8843	1.3811
Ours	0.6306	1.6000	2.6063	<u>0.4197</u>	<u>0.7484</u>	<u>1.0917</u>	0.3905	0.7244	1.0864

注：粗体字表示最优值，下划线标识次优值。

表 3 不同的方法在 Middlebury 数据集 B 上重建结果的定量分析(RMSE)

Table 3 Quantitative analysis of reconstruction results on dataset B of Middlebury by different methods (RMSE)

	Cones		Venus		Tsukuba	
	X2	X4	X2	X4	X2	X4
Bicubic	2.5418	3.8659	1.3370	1.9364	5.8185	8.5637
Mandal ^[6]	1.9663	3.3027	0.7919	1.3592	4.4127	7.0468
Chen ^[10]	—	3.1742	—	0.9955	—	6.3472
Liu ^[11]	2.6181	4.5441	1.1619	1.5082	5.6176	7.9656
Kiechle ^[12]	1.7212	3.2005	0.6102	0.8758	3.6626	6.1629
Li ^[13]	3.6702	5.5222	1.2277	1.7814	6.3763	9.0654
Park ^[14]	4.1455	6.3944	1.2662	1.8346	6.9523	9.8982
Dong ^[23]	1.4840	3.5850	0.4560	0.7890	3.2750	7.9390
Hui ^[24]	1.1000	2.7700	<u>0.2590</u>	0.4220	2.4720	<u>4.9960</u>
Kim ^[25]	1.3261	<u>2.6033</u>	0.5517	0.8808	3.0955	5.0385
Lai ^[26]	<u>0.9976</u>	2.7728	0.3855	1.1368	<u>2.0191</u>	5.3103
Ours	0.8688	2.5339	0.2516	<u>0.7055</u>	1.8346	4.5424

注：粗体字表示最优值，下划线标识次优值。

该方法得到的超分辨率重建结果在边缘有伪影问题。从图 6(e)可以看出，本文重建的深度图具有较好的视觉效果，接近真实值，有效地缓解了边缘失真和伪影的问题，因为本文的网络使用多尺度引导的方法重建图像，提取多感受野特征。

4 结 论

针对深度图像分辨率低的问题，本文构建了一种

彩色图像多尺度引导深度图像超分辨率重建卷积神经网络。在彩色图像分支，该网络可以提取不同尺度下的彩色图像特征，用以引导相对应尺度的深度图像重建。在深度图像分支，构建的 MRFRB 块可以提取并融合不同感受野下的特征。实验结果表明，本文方法得到的重建图像，从定性和定量两个方面均取得了较优的结果。接下来的工作考虑通过构建更优的网络结构，从而得到更优的重建结果。

表 4 不同的方法在 Middlebury 数据集 A 上重建结果的定量分析(SSIM)

Table 4 Quantitative analysis of reconstruction results on data set A of Middlebury by different methods (SSIM)

	Art			Books			Moebius		
	X2	X4	X8	X2	X4	X8	X2	X4	X8
Bicubic	0.9870	0.9683	0.9438	0.9956	0.9899	0.9833	0.9950	0.9888	0.9812
Mandal ^[6]	0.9943	0.9854	0.9407	0.9980	0.9934	0.9830	0.9969	0.9925	0.9788
Liu ^[11]	0.9902	0.9810	0.9706	0.9939	0.9910	0.9868	0.9942	0.9897	0.9834
Kiechle ^[12]	0.9971	0.9931	<u>0.9863</u>	0.9980	0.9965	<u>0.9938</u>	0.9975	0.9955	0.9920
Li ^[13]	0.9842	0.9742	0.9653	0.9936	0.9897	0.9870	0.9924	0.9884	0.9851
Park ^[14]	0.9847	0.9704	0.9546	0.9944	0.9895	0.9857	0.9937	0.9884	0.9836
Kim ^[25]	0.9974	<u>0.9934</u>	0.9833	0.9984	<u>0.9966</u>	0.9931	0.9983	<u>0.9962</u>	<u>0.9921</u>
Lai ^[26]	<u>0.9984</u>	0.9918	0.9795	0.9986	0.9962	0.9926	<u>0.9986</u>	0.9955	0.9909
Ours	0.9988	0.9951	0.9881	<u>0.9985</u>	0.9969	0.9944	0.9988	0.9967	0.9938

注：粗体字表示最优值，下划线标识次优值。

表 5 不同的方法在 Middlebury 数据集 B 上重建结果的定量分析(SSIM)

Table 5 Quantitative analysis of reconstruction results on dataset B of Middlebury by different methods (SSIM)

	Cones		Venus		Tsukuba	
	X2	X4	X2	X4	X2	X4
Bicubic	0.9814	0.9585	0.9935	0.9861	0.9700	0.9295
Mandal ^[6]	0.9877	0.9713	0.9979	0.9934	0.9829	0.9553
Chen ^[10]	—	0.9711	—	0.9897	—	0.9599
Liu ^[11]	0.9761	0.9473	0.9941	0.9919	0.9665	0.9461
Kiechle ^[12]	0.9904	0.9757	0.9983	<u>0.9970</u>	0.9904	0.9733
Li ^[13]	0.9633	0.9331	0.9951	0.9919	0.9618	0.9341
Park ^[14]	0.9633	0.9300	0.9943	0.9893	0.9589	0.9177
Kim ^[25]	0.9935	<u>0.9826</u>	0.9980	0.9963	0.9931	<u>0.9781</u>
Lai ^[26]	<u>0.9952</u>	0.9803	<u>0.9988</u>	0.9947	<u>0.9956</u>	0.9749
Ours	0.9960	0.9834	0.9993	0.9973	0.9970	0.9834

注：粗体字表示最优值，下划线标识次优值。

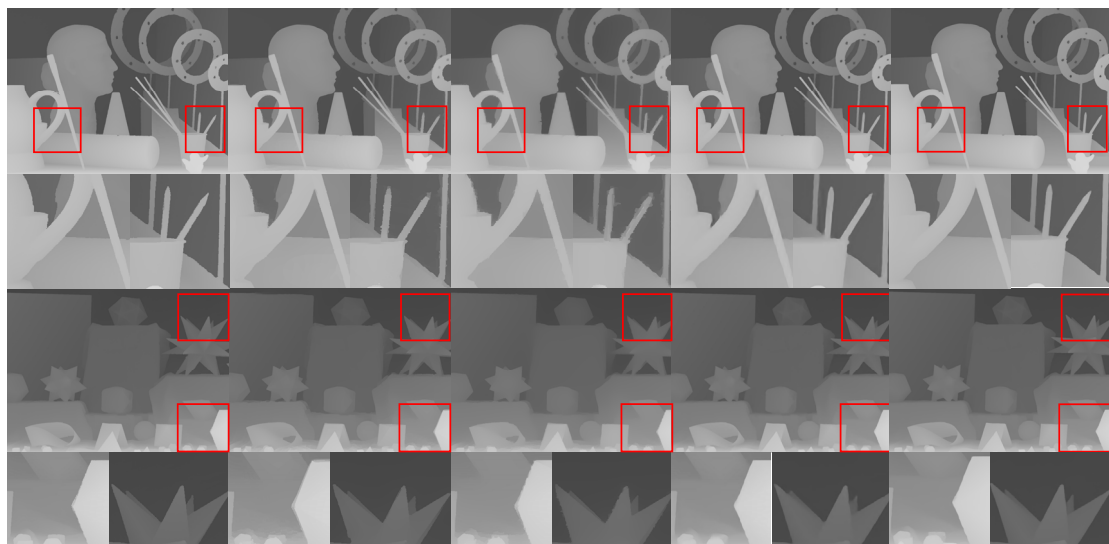


图 6 不同方法在 Middlebury 数据集上的超分辨率重建结果。

(a) 真值图像; (b) 文献[11]方法; (c) 文献[13]方法; (d) 文献[25]方法; (e) 本文方法

Fig. 6 Super-resolution reconstruction results on the Middlebury dataset by different methods.
(a) Truth image; (b) Ref. [11] method; (c) Ref. [13] method; (d) Ref. [25] method; (e) This article method

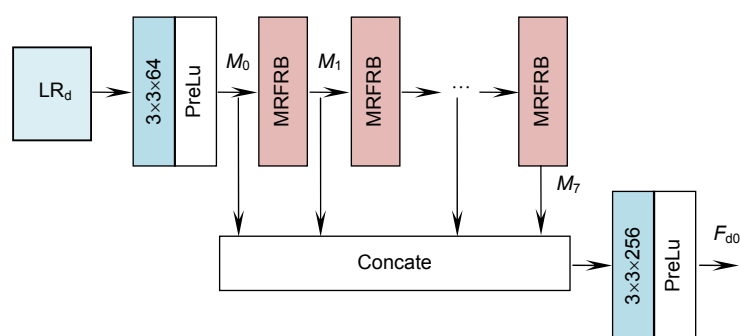
参考文献

- [1] Palacios J M, Sagüés C, Montijano E, *et al.* Human-computer interaction based on hand gestures using RGB-D sensors[J]. *Sensors*, 2013, **13**(9): 11842–11860.
- [2] Nguyen T N, Huynh H H, Meunier J. 3D reconstruction with time-of-flight depth camera and multiple mirrors[J]. *IEEE Access*, 2018, **6**: 38106–38114.
- [3] Yamamoto S. Development of inspection robot for nuclear power plant[C]//*Proceedings of 1992 IEEE International Conference on Robotics and Automation*, Nice, France, 1992: 1559–1566.
- [4] Kolb A, Barth E, Koch R, *et al.* Time-of-flight cameras in computer graphics[J]. *Computer Graphics Forum*, 2010, **29**(1): 141–159.
- [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. Noise adaptive super-resolution from single image via non-local mean and sparse representation[J]. *Signal Processing*, 2017, **132**: 134–149.
- [7] Aodha O M, Campbell N D F, Nair A, *et al.* Patch based synthesis for single depth image super-resolution[C]//*Proceedings of the 12th European Conference on Computer Vision*, Florence, Italy, 2012: 71–84.
- [8] Li J, Lu Z C, Zeng G, *et al.* Similarity-aware patchwork assembly for depth image super-resolution[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014: 3374–3381.
- [9] Xie J, Feris R S, Sun M T. Edge-guided single depth image super resolution[J]. *IEEE Transactions on Image Processing*, 2016, **25**(1): 428–438.
- [10] Chen B L, Jung C. Single depth image super-resolution using convolutional neural networks[C]//*Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, AB, Canada, 2018: 1473–1477.
- [11] Liu W, Chen X G, Yang J, *et al.* Robust color guided depth map restoration[J]. *IEEE Transactions on Image Processing*, 2017, **26**(1): 315–327.
- [12] Kiechle M, Hawe S, Kleinstaub M. A joint intensity and depth co-sparse analysis model for depth map super-resolution[C]//*Proceedings of 2013 International Conference on Computer Vision*, Sydney, NSW, Australia, 2013: 1545–1552.
- [13] Li Y, Min D B, Do M N, *et al.* Fast guided global interpolation for depth and motion[C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, The Netherlands, 2016: 717–733.
- [14] Park J, Kim H, Tai Y W, *et al.* High quality depth map upsampling for 3D-tof cameras[C]//*Proceedings of 2011 IEEE International Conference on Computer Vision*, Barcelona, Spain, 2011: 1623–1630.
- [15] Li W, Zhang X D. Depth image super-resolution reconstruction based on convolution neural network[J]. *Journal of Electronic Measurement and Instrumentation*, 2017, **31**(12): 1918–1928. 李伟, 张旭东. 基于卷积神经网络的深度图像超分辨率重建方法[J]. *电子测量与仪器学报*, 2017, **31**(12): 1918–1928.
- [16] Xiao Y, Cao X, Zhu X Y, *et al.* Joint convolutional neural pyramid for depth map super-resolution[Z]. arXiv:1801.00968, 2018.
- [17] Lecun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324.
- [18] Scharstein D, Szeliski R, Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[C]//*Proceedings of 2001 IEEE Workshop on Stereo and Multi-Baseline Vision*, Kauai, HI, USA, 2001: 131–140.
- [19] Richardt C, Stoll C, Dodgson N A, *et al.* Coherent spatiotemporal filtering, upsampling and rendering of rgbz videos[J]. *Computer Graphics Forum*, 2012, **31**(2): 247–256.
- [20] Lu S, Ren X F, Liu F. Depth enhancement via low-rank matrix completion[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014: 3390–3397.
- [21] Handa A, Whelan T, McDonald J, *et al.* A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM[C]//*Proceedings of 2014 IEEE International Conference on Robotics and Automation*, Hong Kong, China, 2014: 1524–1531.
- [22] He K M, Zhang X Y, Ren S Q, *et al.* Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015: 1026–1034.
- [23] Dong C, Loy C C, He K M, *et al.* Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(2): 295–307.
- [24] Hui T W, Loy C C, Tang X O. Depth map super-resolution by deep multi-scale guidance[C]//*Proceedings of the 14th European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016: 353–369.
- [25] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016: 1646–1654.
- [26] Lai W S, Huang J B, Ahuja N, *et al.* Deep Laplacian pyramid networks for fast and accurate super-resolution[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017: 624–632.

Color image multi-scale guided depth image super-resolution reconstruction

Yu Shuxia*, Hu Liangmei, Zhang Xudong, Fu Xuwen

School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230601, China



Feature extraction block of depth map

Overview: In recent years, as the demand for depth information in the field of computer vision has expanded, the acquisition of high-resolution depth images has become crucial. However, due to the limitations of hardware conditions such as sensors, the depth image resolution obtained by the depth camera is generally not high, and it is difficult to meet the practical application requirements. For example, the PMD Camcube camera has a resolution of only 200×200 , and Microsoft's Kinect camera has a resolution of only 640×480 . If the resolution of the depth image is increased by improving the hardware facilities, the cost will increase, and there are some technical problems that are difficult to overcome, so the depth image resolution is usually improved by a software processing method. In order to obtain better super-resolution reconstruction results of depth images, this paper constructs a multi-scale color image guidance depth image super-resolution reconstruction convolutional neural network. The network consists of three branches: a color image branch, a depth image branch, and an image reconstruction branch. The relationship between high resolution (HR) color image and low resolution (LR) depth image of the same scene is corresponding. Super-resolution reconstruction of LR depth image guided by HR color image of the same scene is conducive to restoring high-frequency information of LR depth image. So the LR depth image super-resolution reconstruction can be guided by the same scene HR color image to obtain more excellent reconstruction results. Because different structural information in the image has different scales, so the multi-scale fusion method is used to realize the guidance of HR color image features to LR depth image features, which is beneficial to the restoration of image details. For the depth image super-resolution reconstruction problem, the input LR depth image is highly correlated with the output HR depth image, so if the features of the LR depth image can be fully extracted, a better reconstruction result will be obtained. Thus in the process of extracting features from LR depth images, this paper constructs a multi-receptive residual block to extract and fuse the features of different receptive fields, and then connect and fuse the features of each multiple receptive field residual block output to obtain global fusion features. Finally, the HR depth image is obtained through sub-pixel convolution layer and global fusion features. The experimental results on different data sets show that the super-resolution image obtained by this algorithm can alleviate the problem of edge distortion and artifacts, and has better visual effect.

Citation: Yu S X, Hu L M, Zhang X D, *et al.* Color image multi-scale guided depth image super-resolution reconstruction[J]. *Opto-Electronic Engineering*, 2020, 47(4): 190260

Supported by National Natural Science Foundation of China (61876057)

* E-mail: ysx123@mail.hfut.edu.cn